Explaining Cointegration Analysis: Part I

David F. Hendry* and Katarina Juselius
Nuffield College, Oxford — European University Institute, Florence

September 10, 1999

Abstract

'Classical' econometric theory assumes that observed data come from a stationary process, where means and variances are constant over time. Graphs of economic time series, and the historical record of economic forecasting, reveal the invalidity of such an assumption. Consequently, we discuss the importance of stationarity for empirical modeling and inference; describe the effects of incorrectly assuming stationarity; explain the basic concepts of non-stationarity; note some sources of non-stationarity; formulate a class of non-stationary processes (autoregressions with unit roots) that seem empirically relevant for analyzing economic time series; and show when an analysis can be transformed by means of differencing and cointegrating combinations so stationarity becomes a reasonable assumption. We then describe how to test for unit roots and cointegration. Monte Carlo simulations and empirical examples illustrate the analysis.

1 Introduction

Much of 'classical' econometric theory has been predicated on the assumption that the observed data come from a stationary process, meaning a process whose means and variances are constant over time. A glance at graphs of most economic time series, or at the historical track record of economic forecasting, suffices to reveal the invalidity of that assumption: economies evolve, grow, and change over time in both real and nominal terms, sometimes dramatically – and economic forecasts are often badly wrong, although that should occur relatively infrequently in a stationary process.

Figure 1 shows some 'representative' time series to emphasize this point.[1] The first panel (denoted a) reports the time series of broad money in the UK over 1868–1993 on a log scale, together with the corresponding price series (the UK data for 1868–1975 are from Friedman and Schwartz, 1982, extended to 1993 by Attfield, Demery and Duck, 1995). From elementary calculus, since $\partial \log y / \partial y = 1/y$, the log scale shows proportional changes: hence, the apparently small movements between the minor tic marks actually represent approximately 50% changes. Panel b shows real (constant price) money on a log scale, together with real (constant price) output also on a log scale: the wider spacing of the tic marks reveals a much smaller range of variation, but again, the notion of a constant mean seems untenable. Panel c records long-run and short-run interest rates in natural scale, highlighting changes over time in the variability of economic series, as well as in their means, and indeed, of

*Financial support from the U.K. Economic and Social Research Council under grant R000234954, and from the Danish Social Sciences Research Council is gratefully acknowledged. We are pleased to thank Campbell Watkins for helpful comments on, and discussion of, earlier drafts.

[1] Blocks of four graphs are lettered notionally as a, b; c, d in rows from the top left; six graphs are a, b, c; d, e, f; and so on.

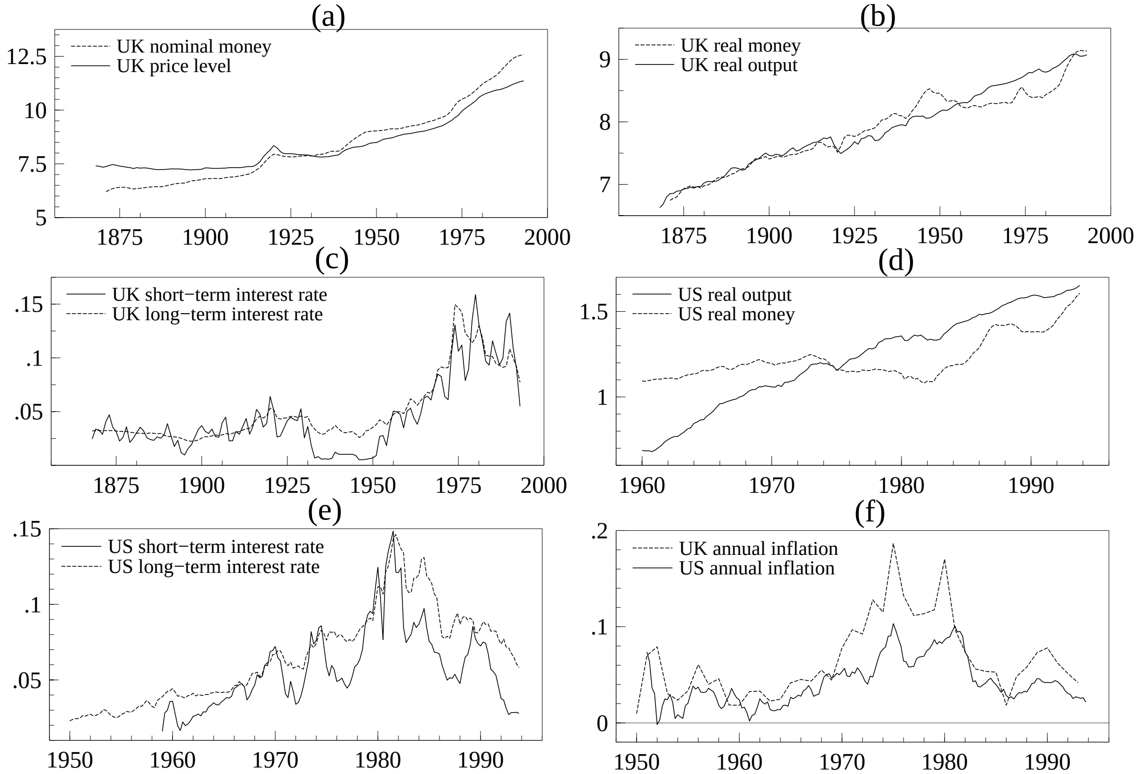

Figure 1: Some 'representative' time series

the relationships between them: all quiet till 1929, the two variables diverge markedly till the early 1950s, then rise together with considerable variation. Nor is the non-stationarity problem specific to the UK: panels d and e show the comparative graphs to b and c for the USA, using post-war quarterly data (from Baba, Hendry and Starr, 1992). Again, there is considerable evidence of change, although the last panel f comparing UK and US annual inflation rates suggests the UK may exhibit greater instability. It is hard to imagine any 'revamping' of the statistical assumptions such that these outcomes could be construed as drawings from stationary processes.[2]

Intermittent episodes of forecast failure (a significant deterioration in forecast performance relative to the anticipated outcome) confirm that economic data are not stationary: even poor models of stationary data would forecast on average as accurately as they fitted, yet that manifestly does not occur empirically. The practical problem facing econometricians is not a plethora of congruent models from which to choose, but to find any relationships that survive long enough to be useful. It seems clear that stationarity assumptions must be jettisoned for most observable economic time series.

Four issues immediately arise:

1. how important is the assumption of stationarity for modeling and inference?
2. what are the effects of incorrectly assuming it?
3. what are the sources of non-stationarity?

[2]It is sometimes argued that economic time series could be stationary around a deterministic trend, and we will comment on that hypothesis later.

4. can empirical analyses be transformed so stationarity becomes a valid assumption?

Essentially, the answers are 'very'; 'potentially hazardous'; 'many and varied'; and 'sometimes, depending on the source of non-stationarity'. Only intuitive explanations for these answers can be offered at this stage of our paper, but roughly:

1. when data means and variances are non-constant, observations come from different distributions over time, posing difficult problems for empirical modeling;

2. assuming constant means and variances when that is false can induce serious statistical mistakes, as we will show;

3. non-stationarity can be due to evolution of the economy, legislative changes, technological change, and political turmoil *inter alia*;

4. some forms of non-stationarity can be eliminated by transformations, and much of our paper concerns an important case where that is indeed feasible.

We expand on all four of these issues below, but to develop the analysis, we first discuss some necessary econometric concepts based on a simple model (in Section 2), and define the properties of a stationary and a non-stationary process (Section 3). As embryology often yields insight into evolution, we next review the history of regression analyses with trending data (Section 4) and consider the possibilities of obtaining stationarity by transformation (called cointegration). Section 5 uses simulation experiments to look at the consequences of data being non-stationary in regression models, building on the famous study in Yule (1926). We then briefly review tests to determine the presence of non-stationarity in the class noted in Section 5 (called univariate unit-root processes: Section 6), as well as the validity of the transformation in Section 4 (cointegration tests: Section 7). Finally, we empirically illustrate the concepts and ideas for a data set consisting of gasoline prices in two major locations (Section 8). Section 9 concludes.

2 Addressing non-stationarity

Non-stationarity seems a natural feature of economic life. Legislative change is one obvious source of non-stationarity, often inducing structural breaks in time series, but it is far from the only one. Economic growth, perhaps resulting from technological progress, ensures secular trends in many time series, as Figure 1 illustrated. Such trends need to be incorporated into statistical analyses, which could be done in many ways, including the venerable linear trend. Our focus here will be on a type of stochastic non-stationarity induced by persistent cumulation of past effects, called unit-root processes (an explanation for this terminology is provided below).[3] Such processes can be interpreted as allowing a different 'trend' at every point in time, so are said to have stochastic trends. Figure 2 provides an artificial example: 'joining' the successive observations as shown produces a 'jumpy' trend as compared to the (best-fitting) linear trend.

There are many plausible reasons why economic data may contain stochastic trends. For example, technology involves the persistence of acquired knowledge, so that the present level of technology is the cumulation of past discoveries and innovations. Economic variables depending closely on technological progress are therefore likely to have a stochastic trend. The impact of structural changes in the world oil market is another example of non-stationarity. Other variables related to

[3] Stochastic means the presence of a random variable.

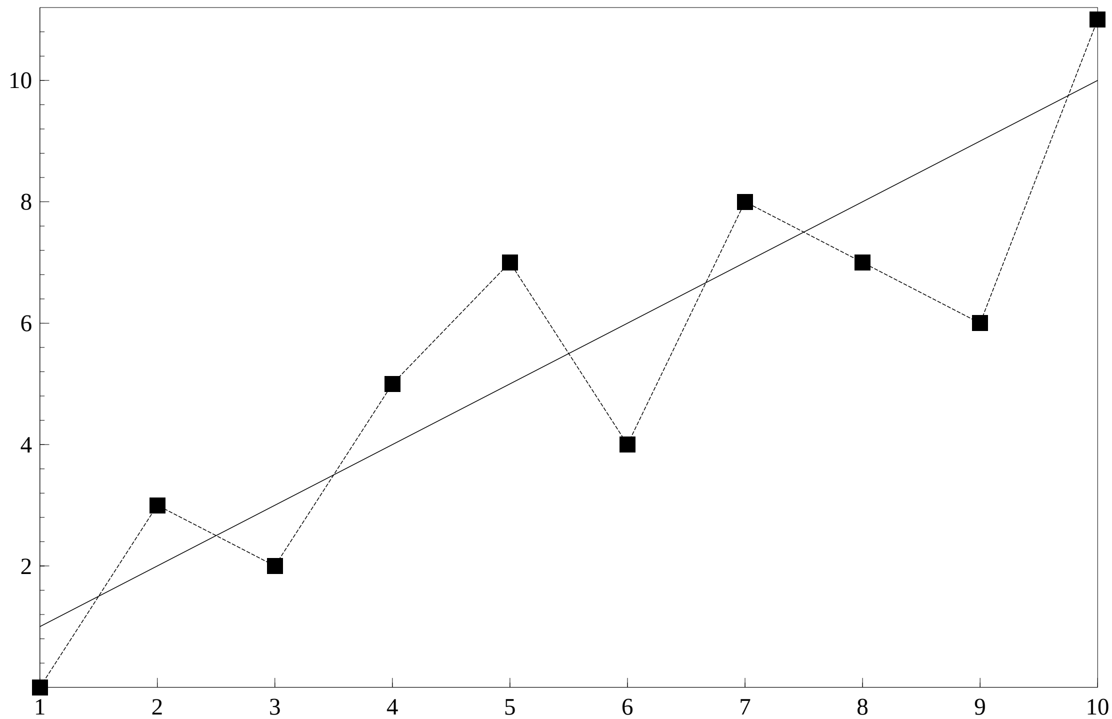

Figure 2: An artificial variable with a stochastic trend

the level of any variable with a stochastic trend will 'inherit' that non-stationarity, and transmit it to other variables in turn: nominal wealth and exports spring to mind, and therefore income and expenditure, and so employment, wages etc. Similar consequences follow for every source of stochastic trends, so the linkages in economies suggest that the levels of many variables will be non-stationary, sharing a set of common stochastic trends.

A non-stationary process is, by definition, one which violates the stationarity requirement, so its means and variances are non-constant over time. For example, a variable exhibiting a shift in its mean is a non-stationary process, as is a variable with a heteroscedastic variance over time. We will focus here on the non-stationarity caused by stochastic trends, and discuss its implications for empirical modeling.

To introduce the basic econometric concepts, we consider a simple regression model for a variable y_t containing a fixed (or deterministic) linear trend with slope β generated from an initial value y_0 by:

$$y_t = y_0 + \beta t + u_t \text{ for } \quad t = 1, ..., T. \tag{1}$$

To make the example more realistic, the error term u_t is allowed to be a first-order autoregressive process:

$$u_t = \rho u_{t-1} + \varepsilon_t. \tag{2}$$

That is, the current value of the variable u_t is affected by its value in the immediately preceding period (u_{t-1}) with a coefficient ρ, and by a stochastic 'shock' ε_t. We discuss the meaning of the autoregressive parameter ρ below. The stochastic 'shock' ε_t is distributed as $\mathsf{IN}[0, \sigma_\varepsilon^2]$, denoting an independent (I), normal (N) distribution with a mean of zero ($\mathsf{E}[\varepsilon_t] = 0$) and a variance $\mathsf{V}[\varepsilon_t] = \sigma_\varepsilon^2$: since these are constant parameters, an identical distribution holds at every point in time. A process such as $\{\varepsilon_t\}$ is often called a normal 'white-noise' process. Of the three desirable properties of independence, identical distributions, and normality, the first two are clearly the most important. The process u_t in (2) is not independent, so we first consider its properties. In the following, we will use the notation u_t for an autocorrelated process, and ε_t for a white-noise process. The assumption of only first-order autocorrelation in u_t, as shown in (2), is for notational simplicity, and all arguments generalize to higher-order autoregressions. Throughout the paper, we will use lower-case letters to indicate logarithmic scale, so $x_t = \log(X_t)$.

Dynamic processes are most easily studied using the lag operator L (see e.g., Hendry, 1995, chapter 4) such that $Lx_t = x_{t-1}$. Then, (2) can be written as $u_t = \rho L u_t + \varepsilon_t$ or:

$$u_t = \frac{\varepsilon_t}{1-\rho L}. \tag{3}$$

When $|\rho| < 1$, the term $1/(1-\rho L)$ in (3) can be expanded as $(1 + \rho L + \rho^2 L^2 + \cdots)$. Hence:

$$u_t = \varepsilon_t + \rho\varepsilon_{t-1} + \rho^2\varepsilon_{t-2} + \cdots. \tag{4}$$

Expression (4) can also be derived after repeated substitution in (2). It appears that u_t is the sum of all previous disturbances (shocks) ε_{t-i}, but that the effects of previous disturbances decline with time because $|\rho| < 1$. However, now think of (4) as a process directly determining u_t – ignoring our derivation from (2) – and consider what happens when $\rho = 1$. In that case, $u_t = \varepsilon_t + \varepsilon_{t-1} + \varepsilon_{t-2} + \cdots$, so each disturbance persists indefinitely and has a permanent effect on u_t. Consequently, we say that u_t has the 'stochastic trend' $\Sigma_{i=1}^t \varepsilon_i$. The difference between a linear stochastic trend and a deterministic trend is that the increments of a stochastic trend are random, whereas those of a deterministic trend are constant over time as Figure 2 illustrated. From (4), we notice that $\rho = 1$ is equivalent to the summation of the errors. In continuous time, summation corresponds to integration, so such processes are also called integrated, here of first order: we use the shorthand notation $u_t \sim \mathsf{I}(1)$ when $\rho = 1$, and $u_t \sim \mathsf{I}(0)$ when $|\rho| < 1$.

From (4), when $|\rho| < 1$, we can derive the properties of u_t as:

$$\mathsf{E}[u_t] = 0 \;\text{ and }\; \mathsf{V}[u_t] = \frac{\sigma_\varepsilon^2}{1-\rho^2}. \tag{5}$$

Hence, the larger the value of ρ, the larger the variance of u_t. When $\rho = 1$, the variance of u_t becomes indeterminate and u_t becomes a random walk. Interpreted as a polynomial in L, (3) has a factor of $1 - \rho L$, which has a root of $1/\rho$: when $\rho = 1$, (2) is called a unit-root process. While there may appear to be many names for the same notion, extensions yield important distinctions: for example, longer lags in (2) preclude u_t being a random walk, and processes can be integrated of order 2 (i.e., $\mathsf{I}(2)$), so have several unit roots.

Returning to the trend regression example, by substituting (2) into (1) we get:

$$y_t = \beta t + \frac{\varepsilon_t}{1-\rho L} + y_0$$

and by multiplying through the factor $(1 - \rho L)$:

$$(1-\rho L)y_t = (1-\rho L)\beta t + (1-\rho L)y_0 + \varepsilon_t. \tag{6}$$

From (6), it is easy to see why the non-stationary process which results when $\rho = 1$, is often called a unit-root process, and why an autoregressive error imposes a common-factor dynamics on a static regression model (see e.g., Hendry and Mizon, 1978). When $\rho = 1$ in (6), the root of the lag polynomial is unity, so it describes a linear difference equation with a unit coefficient.

Rewriting (6) using $Lx_t = x_{t-1}$, we get:

$$y_t = \rho y_{t-1} + \beta(1-\rho)t + \rho\beta + (1-\rho)y_0 + \varepsilon_t, \tag{7}$$

and it appears that the 'static' regression model (1) with autocorrelated residuals is equivalent to the following dynamic model with white-noise residuals:

$$y_t = b_1 y_{t-1} + b_2 t + b_0 + \varepsilon_t \tag{8}$$

where

$$\begin{aligned} b_1 &= \rho \\ b_2 &= \beta(1-\rho) \\ b_0 &= \rho\beta + (1-\rho)y_0. \end{aligned} \tag{9}$$

We will now consider four different cases, two of which correspond to non-stationary (unit-root) models, and the other two to stationary models:

Case 1. $\rho = 1$ and $\beta \neq 0$. It follows from (7) that $\Delta y_t = \beta + \varepsilon_t$, for $t = 1, \ldots, T$, where $\Delta y_t = y_t - y_{t-1}$. This model is popularly called a 'random walk with drift'. Note that $\mathsf{E}[\Delta y_t] = \beta \neq 0$ is equivalent to y_t having a linear trend, since although the coefficient of t in (7) is zero, the coefficient of y_{t-1} is unity so it 'integrates' the 'intercept' β, just as u_t cumulated past ε_t in (4).

Case 2. $\rho = 1$ and $\beta = 0$. From Case 1, it follows immediately that $\Delta y_t = \varepsilon_t$. This is called a pure random walk model: since $\mathsf{E}[\Delta y_t] = 0$, y_t contains no linear trend.

Case 3. $|\rho| < 1$ and $\beta \neq 0$ gives us (8), i.e., a 'trend-stationary' model. The interpretation of the coefficients b_1, b_2, and b_0 must be done with care: for example, b_2 is not an estimate of the trend in y_t, instead $\beta = b_2/(1-\rho)$ is the trend in the process.[4]

Case 4. $|\rho| < 1$ and $\beta = 0$ delivers $y_t = \rho y_{t-1} + (1-\rho)y_0 + \varepsilon_t$, which is the usual stationary autoregressive model with a constant term.

Hence:

- in the static regression model (1), the constant term is essentially accounting for the unit of measurement of y_t, i.e., the 'kick-off' value of the y series;
- in the dynamic regression model (8), the constant term is a weighted average of the growth rate β and the initial value y_0;
- in the differenced model ($\rho = 1$), the constant term is solely measuring the growth rate, β.

We will now provide some examples of economic variables that can be appropriately described by the above simple models.

3 Properties of a non-stationary and a stationary process

Unit-root processes can also arise as a consequence of plausible economic behavior. As an example, we will discuss the possibility of a unit root in the long-term interest rate. Similar arguments could apply to exchange rates, other asset prices, and prices of widely-traded commodities, such as gasoline.

[4] We doubt that GNP (say) has a deterministic trend, because of the thought experiment that output would then continue to grow if we all ceased working.....

If changes to long-term interest rates (R_l) were predictable, and $R_l > R_s$ (the short-term rate) – as usually holds, to compensate lenders for tying up their money – one could create a money machine. Just predict the forthcoming change in R_l, and borrow at R_s to buy bonds if you expect a fall in R_l (a rise in bond prices) or sell short if R_l is likely to rise. Such a scenario of boundless profit at low risk seems unlikely, so we anticipate that the expected value of the change in the long-term interest rate at time $t-1$, given the relevant information set $\mathcal{I}_{t-1}$, is zero, i.e., $\mathsf{E}_{t-1}[\Delta R_{l,t}|\mathcal{I}_{t-1}] = 0$ (more generally, the change should be small on a risk-adjusted basis after transactions costs). As a model, this translates into:

$$R_{l,t} = R_{l,t-1} + \epsilon_t \tag{10}$$

where $\mathsf{E}_{t-1}[\epsilon_t|\mathcal{I}_{t-1}] = 0$ and ϵ_t is an $\mathsf{ID}[0, \sigma_\varepsilon^2]$ process (where D denotes the relevant distribution, which need not be normal). The model in (10) has a unit coefficient on $R_{l,t-1}$, and as a dynamic relation, is a unit-root process. To discuss the implications for empirical modeling of having unit roots in the data, we first need to discuss the statistical properties of stationary and non-stationary processes.[5]

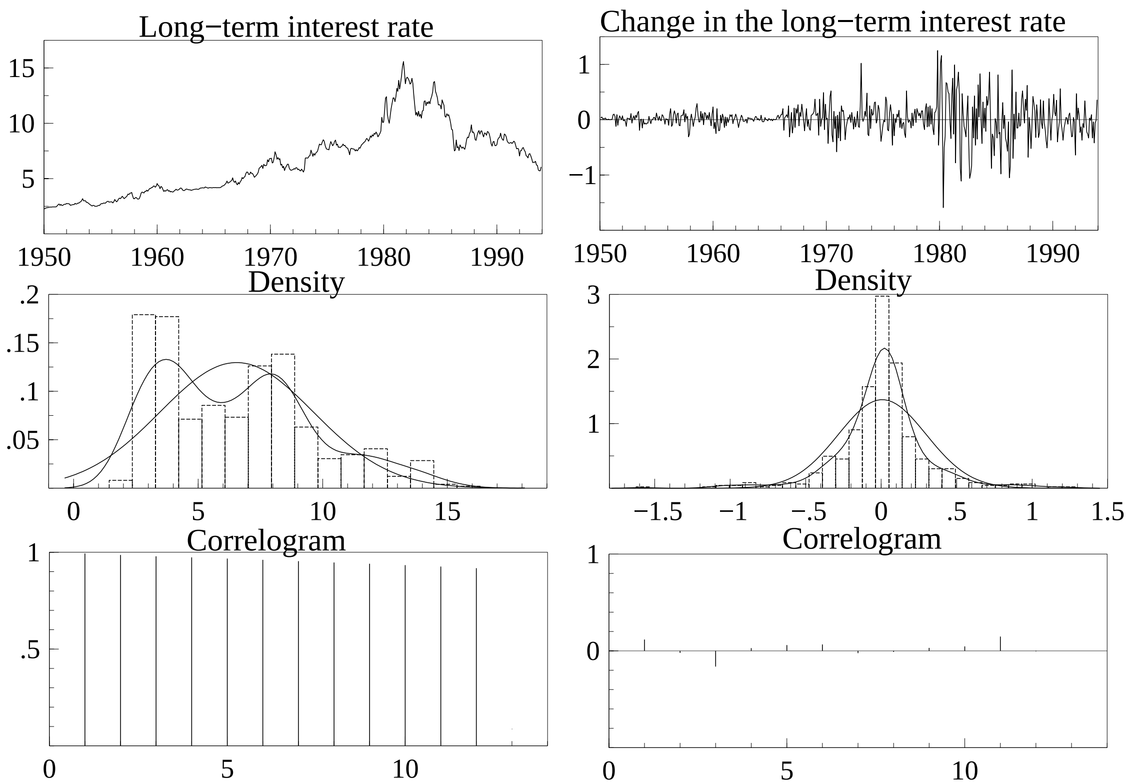

Figure 3: Monthly long-term interest rates in the USA, in levels and differences, 1950–1993

3.1 A non-stationary process

Equation (10) shows that the whole of $R_{l,t-1}$ and ϵ_t influence $R_{l,t}$, and hence, in the next period, the whole of $R_{l,t}$ influences $R_{l,t+1}$ and so on. Thus, the effect of ϵ_t persists indefinitely, and past

[5]Empirically, for the monthly data over 1950(1)–1993(12) on 20-year bond rates in the USA shown in Figure 3, the estimated coefficient of $R_{l,t-1}$ in (10) is 0.994 with an estimated standard error of 0.004.

errors accumulate with no 'depreciation', so an equivalent formulation of (10) is:

$$R_{l,t} = \epsilon_t + \epsilon_{t-1} + \cdots + \epsilon_1 + \epsilon_0 + \epsilon_{-1} \cdots \tag{11}$$

or alternatively:

$$R_{l,t} = \epsilon_t + \epsilon_{t-1} + \cdots + \epsilon_1 + R_{l,0} \tag{12}$$

where the initial value $R_{l,0} = \epsilon_0 + \epsilon_{-1}...$ contains all information of the past behavior of the long-term interest rate up to time 0. In practical applications, time 0 corresponds to the first observation in the sample. Equation (11) shows that theoretically, the unit-root assumption implies an ever-increasing variance to the time series (around a fixed mean), violating the constant-variance assumption of a stationary process. In empirical studies, the conditional model (12) is more relevant as a description of the sample variation, and shows that $\{R_{l,t}|R_{l,0}\}$ has a finite variance, $t\sigma_\epsilon^2$, but this variance is non-constant since it changes with $t = 1, \ldots, T$.

Cumulating random errors will make them smooth, and in fact, induces properties like those of economic variables, as first discussed by Working (1934) (so $R_{l,t}$ should be smooth, at least in comparison to its first difference, and is, as illustrated in Figure 3, panels a and b). From (12), taking $R_{l,0}$ as a fixed number, one can see that:

$$\mathsf{E}\left[R_{l,t}\right] = R_{l,0} \tag{13}$$

and that:

$$\mathsf{V}\left[R_{l,t}\right] = \sigma_\epsilon^2 t. \tag{14}$$

Further, perhaps not so easily seen, when $t > s$, the covariance between drawings $t - s$ periods apart is:

$$\mathsf{C}\left[R_{l,t}, R_{l,t-s}\right] = \mathsf{E}\left[\left(R_{l,t} - R_{l,0}\right)\left(R_{l,s} - R_{l,0}\right)\right] = \sigma_\epsilon^2 s \tag{15}$$

and so:

$$\mathsf{corr}^2\left[R_{l,t}, R_{l,t-s}\right] = 1 - \frac{s}{t}. \tag{16}$$

Consequently, when t is large, all the serial correlations for a random-walk process are close to unity, a feature of $R_{l,t}$ as illustrated in Figure 3, panel e. Finally, even if $\{R_{l,t}\}$ is the sum of a large number of errors it will not be approximately normally distributed. This is because each observation, $R_{l,t}|R_{l,0}$, $t = 1, ..., T$, has a different variance. Figure 3, panel c shows the histogram of approximately 80 quarterly observations for which the observed distribution is bimodal, and so does not look even approximately normal. We will comment on the properties of $\Delta R_{l,t}$ below.

To summarize, the variance of a unit-root process increases over time, and successive observations are highly interdependent. The theoretical mean of the conditional process $R_{l,t}|R_{l,0}$ is constant and equal to $R_{l,0}$. However, the theoretical mean and the empirical sample mean $\overline{R}_l$ are not even approximately equal when data are non-stationary (surprisingly, the sample mean divided by $\sqrt{3T}$ is distributed as $\mathsf{N}[0,1]$ in large samples: see Hendry, 1995, chapter 3).

3.2 A stationary process

We now turn to the properties of a stationary process. As argued above, most economic time series are non-stationary, and at best become stationary only after differencing. Therefore, we will from the outset discuss stationarity either for a differenced variable $\{\Delta y_t\}$ or for the IID errors $\{\varepsilon_t\}$.

A variable Δy_t is weakly stationary when its first two moments are constant over time, or more precisely, when $\mathsf{E}[\Delta y_t] = \mu$, $\mathsf{E}[(\Delta y_t - \mu)^2] = \sigma^2$, and $\mathsf{E}[(\Delta y_t - \mu)(\Delta y_{t-s} - \mu)] = \gamma(s)\ \forall s$, where μ, σ^2, and $\gamma(s)$ are finite and independent of t.[6]

As an example of a stationary process take the change in the long-term interest rate from (10):

$$\Delta R_{l,t} = \epsilon_t \quad \text{where} \quad \epsilon_t \sim \mathsf{ID}\left[0, \sigma_\epsilon^2\right]. \tag{17}$$

It is a stationary process with mean $\mathsf{E}[\Delta R_{l,t}] = 0$, $\mathsf{E}[(\Delta R_{l,t} - 0)^2] = \sigma_\varepsilon^2$, and $\mathsf{E}[(\Delta R_{l,t} - 0)(\Delta R_{l,t-s} - 0)] = 0\ \forall s$, because $\Delta R_{l,t} = \varepsilon_t$ is assumed to be an IID process. This is illustrated by the graph of the change in the long-term bond rate in Figure 3, panel b. We note that the autocorrelations have more or less disappeared (panel f), and that the density distribution is approximately normal except for an outlier corresponding to the deregulation of capital movements in 1983 (panel d).

An IID process is the simplest example of a stationary process. However, as demonstrated in (2), a stationary process can be autocorrelated, but such that the influence of past shocks dies out. Otherwise, as demonstrated by (4), the variance would not be constant.

4 Regression with trending variables: a historical review

One might easily get the impression that the unit-root literature is a recent phenomenon. This is clearly not the case: already in 1926, Udny Yule analyzed the hazards of regressing a trending variable on another unrelated trending variable – the so-called 'nonsense regression' problem. However, the development of 'unit-root econometrics' that aims to address this problem in a more constructive way, is quite recent.

A detailed history of econometrics has also developed over the past decade and full coverage is provided in Morgan (1990), Qin (1993), and Hendry and Morgan (1995). Here, we briefly review the evolution of the concepts and tools underpinning the analysis of non-stationarity in economics, commencing with the problem of 'nonsense correlations', which are extremely high correlations often found between variables for which there is no ready causal explanation (such as birth rates of humans and the number of storks' nests in Stockholm).

Yule (1926) was the first to formally analyze such 'nonsense correlations'. He thought that they were not the result of both variables being related to some third variable (e.g., population growth). Rather, he categorized time series according to their serial-correlation properties, namely, how highly correlated successive values were with each other, and investigated how their cross-correlation coefficient r_{xy} behaved when two unconnected series x and y had:

A] random levels;
B] random first differences;
C] random second differences.

For example, in case B, the data take the form $\Delta x_t = \varepsilon_t$ (Case 2. in Section 3) where ε_t is IID. Since the value of x_t depends on all past errors with equal weights, the effects of distant shocks persist, so the variance of x_t increases over time, making it non-stationary. Therefore, the level of x_t contains information about all permanent disturbances that have affected the variable, starting from the initial level x_0 at time 0, and contains the linear stochastic trend $\Sigma\varepsilon_i$. Similarly for the y_t series, although its stochastic trend depends on cumulating errors that are independent of those entering x_t.

[6] When the moments depend on the initial conditions of the process, stationarity holds only asymptotically (see e.g. Spanos, 1986), but we ignore that complication here.

Yule found that r_{xy} was almost normally distributed in case A, but became nearly uniformly distributed (except at the end points) in B. He was startled to discover that r_{xy} had a U-shaped distribution in C, so the correct null hypothesis (of no relation between x and y) was virtually certain to be rejected in favor of a near-perfect positive or negative link. Consequently, it seemed as if inference could go badly wrong once the data were non-stationary. Today, his three types of series are called integrated of orders zero, one, and two respectively (I(0), I(1), and I(2) as above). Differencing an I(1) series delivers an I(0), and so on. Section 5 replicates a simulation experiment that Yule undertook.

Yule's message acted as a significant discouragement to time-series work in economics, but gradually its impact faded. However, Granger and Newbold (1974) highlighted that a good fit with significant serial correlation in the residuals was a symptom associated with nonsense regressions. Hendry (1980) constructed a nonsense regression by using cumulative rainfall to provide a better explanation of price inflation than did the money stock in the UK. A technical analysis of the sources and symptoms of the nonsense-regressions problem was finally presented by Phillips (1986).

As economic variables have trended over time since the Industrial Revolution, ensuring non-stationarity resulted in empirical economists usually making careful adjustments for factors such as population growth and changes in the price level. Moreover, they often worked with the logarithms of data (to ensure positive outcomes and models with constant elasticities), and thereby implicitly assumed constant proportional relations between non-stationary variables. For example, if $\beta \neq 0$ and u_t is stationary in the regression equation:

$$y_t = \beta_0 + \beta_1 x_t + u_t \tag{18}$$

then y_t and x_t must contain the same stochastic trend, since otherwise u_t could not be stationary. Assume that y_t is aggregate consumption, x_t is aggregate income, and the latter is a random walk, i.e., $x_t = \Sigma\varepsilon_i + x_0$.[7] If aggregate income is linearly related to aggregate consumption in a causal way, then y_t would 'inherit' the non-stationarity from x_t, and u_t would be stationary unless there were other non-stationary variables than income causing consumption.

Assume now that, as before, y_t is non-stationary, but it is not caused by x_t and instead is determined by another non-stationary variable, say, $z_t = \Sigma\nu_i + z_0$, unrelated to x_t. In this case, $\beta_1 = 0$ in (18) is the correct hypothesis, and hence what one would like to accept in statistical tests. Yule's problem in case B can be seen clearly: if β_1 were zero in (18), then $y_t = \beta_0 + u_t$; i.e., u_t contains $\Sigma\nu_i$ so is non-stationary and, therefore, inconsistent with the stationarity assumption of the regression model. Thus, one cannot conduct standard tests of the hypothesis that $\beta_1 = 0$ in such a setting. Indeed, u_t being autocorrelated in (18), with u_t being non-stationary as the extreme case, is what induces the non-standard distributions of r_{xy}.

Nevertheless, Sargan (1964) linked static-equilibrium economic theory to dynamic empirical models by embedding (18) in an autoregressive-distributed lag model:

$$y_t = b_0 + b_1 y_{t-1} + b_2 x_t + b_3 x_{t-1} + \varepsilon_t. \tag{19}$$

The dynamic model (19) can also be formulated in the so-called equilibrium-correction form by subtracting y_{t-1} from both sides and subtracting and adding $b_2 x_{t-1}$ to the right-hand side of (19):

$$\Delta y_t = \alpha_0 + \alpha_1 \Delta x_t - \alpha_2 \left(y_{t-1} - \beta_1 x_{t-1} - \beta_0\right) + \varepsilon_t \tag{20}$$

where $\alpha_1 = b_2$, $\alpha_2 = (1-b_1)$, $\beta_1 = (b_2+b_3)/(1-b_1)$, and $\alpha_0+\alpha_2\beta_0 = b_0$. Thus, all coefficients in (20) can be derived from (19). Models such as (20) explain growth rates in y_t by the growth in x_t and the

[7] Recall lower case letters are in logaritmic form.

past disequilibrium between the levels. Think of a situation where consumption changes (Δy_t) as a result of a change in income (Δx_t), but also as a result of previous period's consumption not being in equilibrium (i.e., $y_{t-1} \neq \beta_0 + \beta_1 x_{t-1}$). For example, if previous consumption was too high, it has to be corrected downwards, or if it was too low, it has to be corrected upwards. The magnitude of the past disequilibrium is measured by $(y_{t-1} - \beta_1 x_{t-1} - \beta_0)$ and the speed of adjustment towards this steady-state by α_2.

Notice that when ε_t, Δy_t and Δx_t are $\mathsf{I}(0)$, there are two possibilities:

- $\alpha_2 \neq 0$ and $(y_{t-1} - \beta_1 x_{t-1} - \beta_0) \sim \mathsf{I}(0)$, or
- $\alpha_2 = 0$ and $(y_{t-1} - \beta_1 x_{t-1} - \beta_0) \sim \mathsf{I}(1)$.

The latter case can be seen from (19). When $\alpha_2 = 0$, then $\Delta y_t = \alpha_0 + \alpha_1 \Delta x_t + \varepsilon_t$, and by integrating we get $y_t = \alpha_1 x_t + \sum \varepsilon_t$. Notice also that by subtracting $\beta_1 \Delta x_t$ from both sides of (20) and collecting terms, we can derive the properties of the equilibrium error $u_t = (y - \beta_1 x - \beta_0)_t$:

$$(y - \beta_1 x - \beta_0)_t = \alpha_0 + (\alpha_1 - \beta_1)\Delta x_t + (1 - \alpha_2)(y - \beta_1 x - \beta_0)_{t-1} + \varepsilon_t \qquad (21)$$

or:

$$u_t = \rho u_{t-1} + \alpha_0 + (\alpha_1 - \beta_1)\Delta x_t + \varepsilon_t.$$

where $\rho = (1 - \alpha_2)$.[8] Thus, the equilibrium error is an autocorrelated process; the higher the value of ρ (equivalently the smaller the value of α_2), the slower is the adjustment back to equilibrium, and the longer it takes for an equilibrium error to disappear. If $\alpha_2 = 0$, there is no adjustment, and y_t does not return to any equilibrium value, but drifts as a non-stationary variable. To summarize: when $\alpha_2 \neq 0$ (so $\rho \neq 1$), the 'equilibrium error' $u_t = (y - \beta_1 x - \beta_0)_t$ is a stationary autoregressive process.

$\mathsf{I}(1)$ 'nonsense-regressions' problems will disappear in (20) because Δy_t and Δx_t are $\mathsf{I}(0)$ and, therefore, no longer trending. Standard t-statistics will be 'sensibly' distributed (assuming that ε_t is IID), irrespective of whether the past equilibrium error, u_{t-1}, is stationary or not.[9] This is because a stationary variable, Δy_t, cannot be explained by a non-stationary variable, and $\widehat{\alpha}_2 \simeq 0$ if $u_{t-1} \sim \mathsf{I}(1)$. Conversely, when $u_{t-1} \sim \mathsf{I}(0)$, then $\widehat{\alpha}_2$ measures the speed of adjustment with which Δy_t adjusts (corrects) towards each new equilibrium position.

Based on equations like (20), Hendry and Anderson (1977) noted that 'there are ways to achieve stationarity other than blanket differencing', and argued that terms like u_{t-1} would often be stationary even when the individual series were not. More formally, Davidson, Hendry, Srba and Yeo (1978) introduced a class of models based on (20) which they called 'error-correction' mechanisms (denoted ECMs). To understand the status of equations like (20), Granger (1981) introduced the concept of cointegration where a genuine relation exists, despite the non-stationary nature of the original data, thereby introducing the obverse of nonsense regressions. Further evidence that many economic time series were better construed as non-stationary than stationary was presented by Nelson and Plosser (1982), who tested for the presence of unit roots and could not reject that hypothesis. Closing this circle, Engle and Granger (1987) proved that ECMs and cointegration were actually two names for

[8] The change in the equilibrium $y_t = \beta_1 x_t - \beta_0$ is $\Delta y_t = \beta_1 \Delta x_t$, so these variables must have steady-state growth rates g_y and g_x related by $g_y = \beta_1 g_x$. But from (20), $g_y = \alpha_0 + \alpha_1 g_x$, hence we can derive that $\alpha_0 = -(\alpha_1 - \beta_1) g_x$, as occurs in (21).

[9] The resulting distribution is not actually a t-statistic as proposed by Student (1908): Section 6 shows that it depends in part on the Dickey–Fuller distribution. However, t is well behaved, unlike the 'nonsense regressions' case.

the same thing: cointegration entails a feedback involving the lagged levels of the variables, and a lagged feedback entails cointegration.[10]

5 Nonsense regression illustration

Using modern software, it is easy to demonstrate nonsense regressions between unrelated unit-root processes. Yule's three cases correspond to generating uncorrelated bivariate time series, where the data are cumulated zero, once and twice, before regressing either series on the other, and conducting a t-test for no relationship. The process used to generate the $\mathsf{I}(1)$ data mimics that used by Yule, namely:

$$\Delta y_t = \epsilon_t \quad \text{where} \quad \epsilon_t \sim \mathsf{IN}\left[0, \sigma_\epsilon^2\right] \tag{22}$$

$$\Delta x_t = \nu_t \quad \text{where} \quad \nu_t \sim \mathsf{IN}\left[0, \sigma_\nu^2\right] \tag{23}$$

and setting $y_0 = 0$, $x_0 = 0$. Also:

$$\mathsf{E}\left[\epsilon_t \nu_s\right] = 0 \ \forall t, s. \tag{24}$$

The economic equation of interest is postulated to be:

$$y_t = \beta_0 + \beta_1 x_t + u_t \tag{25}$$

where β_1 is believed to be the derivative of y_t with respect to x_t:

$$\frac{\partial y_t}{\partial x_t} = \beta_1.$$

Equations like (25) estimated by OLS wrongly assume $\{u_t\}$ to be an IID process independent of x_t. A t-test of H_0: $\beta_1 = 0$ (as calculated by a standard regression package, say) is obtained by dividing the estimated coefficient by its standard error:

$$\mathsf{t}_{\beta_1=0} = \frac{\widehat{\beta}_1}{\mathsf{SE}\left[\widehat{\beta}_1\right]} \tag{26}$$

where:

$$\widehat{\beta}_1 = \left(\sum (x_t - \overline{x})^2\right)^{-1} \sum (x_t - \overline{x})(y_t - \overline{y}), \tag{27}$$

and

$$\mathsf{SE}\left[\widehat{\beta}_1\right] = \frac{\widehat{\sigma}_u}{\sqrt{\sum (x_t - \overline{x})^2}}. \tag{28}$$

When u_t correctly describes an IID process, the t-statistic satisfies:

$$\mathsf{P}\left(\left|\mathsf{t}_{\beta_1=0}\right| \geq 2.0 \mid \mathsf{H}_0\right) \simeq 0.05. \tag{29}$$

[10] Some references to the vast literature on the theory and practice of testing for both unit roots and cointegration include: Banerjee and Hendry (1992), Banerjee, Dolado, Galbraith and Hendry (1993), Chan and Wei (1988), Dickey and Fuller (1979, 1981), Hall and Heyde (1980), Hendry (1995), Johansen (1988, 1991, 1992a, 1992b, 1995b), Johansen and Juselius (1990, 1992), Phillips (1986, 1987a, 1987b, 1988), Park and Phillips (1988, 1989), Phillips and Perron (1988), and Stock (1987).

This, however, is not the case if u_t is autocorrelated: in particular, if u_t is $\mathsf{I}(1)$. In fact, at $T = 100$, from the Monte Carlo experiment, we would need a critical value of **14.8** to define a 5% rejection frequency under the null because:

$$\mathsf{P}\left(\left|\mathsf{t}_{\beta_1=0}\right| \geq 14.8 \mid \mathsf{H}_0\right) \simeq 0.05,$$

so serious over-rejection occurs using (29). Instead of the conventional critical value of 2, we should use 15.

Why does such a large distortion occur? We will take a closer look at each of the components in (26) and expand their formulae to see what happens when u_t is non-stationary. The intuition is that although $\widehat{\beta}_1$ is an unbiased estimator of β_1, so $\mathsf{E}[\widehat{\beta}_1] = 0$, it has a very large variance, but the calculated $\mathsf{SE}[\widehat{\beta}_1]$ dramatically under-estimates the true value.

From (28), $\mathsf{SE}[\widehat{\beta}_1]$ consists of two components, the residual standard error, $\widehat{\sigma}_u$, and the sum of squares, $\sum(x_t - \overline{x})^2$. When $\beta_1 = 0$, the estimated residual variance $\widehat{\sigma}_u^2$ will in general be lower than $\sigma_u^2 = \sum(y_t - \overline{y})^2/T$. This is because the estimated value $\widehat{\beta}_1$ is usually different from zero (sometimes widely so) and, hence, will produce smaller residuals. Thus:

$$\sum(y_t - \widehat{y}_t)^2 \leq \sum(y_t - \overline{y})^2, \tag{30}$$

where $\widehat{y}_t = \widehat{\beta}_0 + \widehat{\beta}_1 x_t$. More importantly, the sum of squares $\sum(x_t - \overline{x})^2$ is not an appropriate measure of the variance in x_t when x_t is non-stationary. This is so because $\overline{x}$ (instead of x_{t-1}) is a very poor 'reference line' when x_t is trending, as is evident from the graphs in Figures 1 and 3, and our artificial example 2. When the data are stationary, the deviation from the mean is a good measure of how much x_t has changed, whereas when x_t is non-stationary, it is the deviation from the previous value that measures the stochastic change in x_t. Therefore:

$$\sum(x_t - \overline{x})^2 \gg \sum(x_t - x_{t-1})^2. \tag{31}$$

so both (30) and (31) work in the same direction of producing a serious downward bias in the estimated value of $\mathsf{SE}[\widehat{\beta}_1]$.

It is now easy to understand the outcome of the simulation study: because the correct standard error is extremely large, i.e., $\sigma_{\beta_1} \gg \mathsf{SE}[\widehat{\beta}_1]$, the dispersion of $\widehat{\beta}_1$ around zero is also large, big positive and negative values both occur, inducing many big 't-values'.

Figure 4 reports the frequency distributions of the t-tests from a simulation study by *PcNaive* (see Doornik and Hendry, 1998), using $M = 10,000$ drawings for $T = 50$. The shaded boxes are for ± 2, which is the approximate 95% conventional confidence interval. The first panel (a) shows the distribution of the t-test on the coefficient of x_t in a regression of y_t on x_t when both variables are white noise and unrelated. This is numerically very close to the correct distribution of a t-variable. The second panel (denoted b, in the top row) shows the equivalent distribution for the nonsense regression based on (22)–(26). The third panel (c, left in the lower row) is for the distribution of the t-test on the coefficient of x_t in a regression of y_t on x_t, y_{t-1} and x_{t-1} when the data are generated as unrelated stationary first-order autoregressive processes. The final panel (d) shows the t-test on the equilibrium-correction coefficient α_2 in (20) for data generated by a cointegrated process (so $\alpha_2 \neq 0$ and β is known).

The first and third panels are close to the actual distribution of Student's t; the former is as expected from statistical theory, whereas the latter shows that outcome is approximately correct in dynamic models once the dynamics have been included in the equation specification. The second panel shows an outcome that is wildly different from t, with a distributional spread so wide that

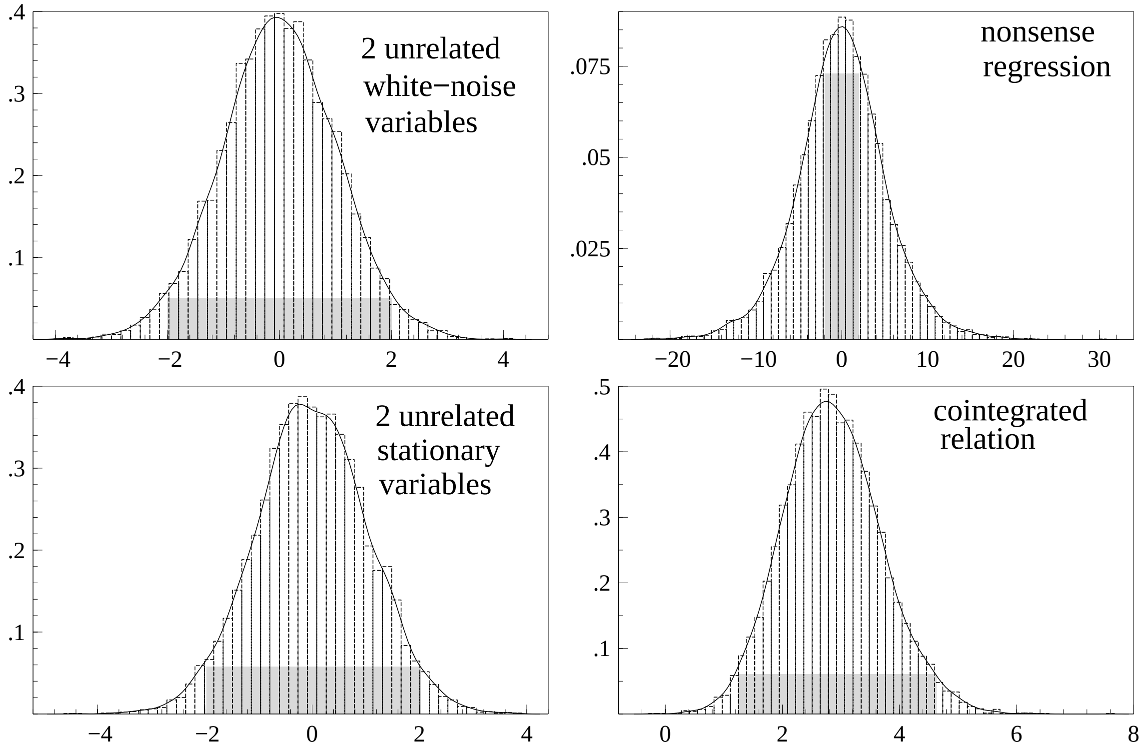

Figure 4: Frequency distributions of nonsense-regression t-tests

most of the probability lies outside the usual region of ± 2. While the last distribution is not centered on zero – because the true relation is indeed non-null – it is included to show that the range of the distribution is roughly correct.

Thus, panels (c) and (d) show that, by themselves, neither dynamics nor unit-root non-stationarity induce serious distortions: the nonsense-regressions problem is due to incorrect model specification. Indeed, when y_{t-1} and x_{t-1} are added as regressors in case (b), the correct distribution results for $\mathsf{t}_{\beta_1=0}$, delivering a panel very similar to (c), so the excess rejection is due to the wrong standard error being used in the denominator (which as shown above, is badly downwards biased by the untreated residual autocorrelation).

Almost all available software packages contain a regression routine that calculates coefficient estimates, t-values, and R^2 based on OLS. Since the computer will calculate the coefficients independently of whether the variables are stationary or not (and without issuing a warning when they are not), it is important to be aware of the following implications for regressions with trending variables:

(i) Although $\mathsf{E}[\widehat{\beta}_1] = 0$, nevertheless $\mathsf{t}_{\beta_1=0}$ diverges to infinity as T increases, so that conventionally-calculated critical values are incorrect (see Hendry, 1995, chapter 3).

(ii) R^2 cannot be interpreted as a measure of goodness-of-fit.

The first point means that one will too frequently reject the null hypothesis ($\beta_1 = 0$) when it is true. Even in the best case, when $\beta_1 \neq 0$, i.e., when y_t and x_t are causally related, standard t-tests will be biased with too frequent rejections of a null hypothesis such as $\beta_1 = 1$, when it is true.

Hence statistical inference from regression models with trending variables is unreliable based on standard OLS output. The second point will be further discussed and illustrated in connection with the empirical analysis.

All this points to the crucial importance of *always* checking the residuals of the empirical model for (unmodeled) residual autocorrelation. If autocorrelation is found, then the model should be re-specified to account for this feature, because many of the conventional statistical distributions, such as Student's t, the F, and the χ^2 distributions become approximately valid once the model is re-specified to have a white-noise error. So even though unit roots impart an important non-stationarity to the data, reformulating the model to have white-noise errors is a good step towards solving the problem; and transforming the variables to be $\mathsf{I}(0)$ will complete the solution.

6 Testing for unit roots

We have demonstrated that stochastic trends in the data are important for statistical inference. We will now discuss how to test for the presence of unit roots in the data. However, the distinction between a unit-root process and a near unit-root process need not be crucial for practical modeling. Even though a variable is stationary, but with a root close to unity (say, $\rho > 0.95$), it is often a good idea to act as if there are unit roots to obtain robust statistical inference. An example of this is given by the empirical illustration in Section 8.

We will now consider unit-root testing in a univariate setting. Consider estimating β in the autoregressive model:

$$y_t = \beta y_{t-1} + \epsilon_t \quad \text{where} \quad \epsilon_t \sim \mathsf{IN}\left[0, \sigma_\epsilon^2\right] \tag{32}$$

under the null of $\beta = 1$ and $y_0 = 0$ (i.e., no determinist trend in the levels), using a sample of size T. Because $\mathsf{V}[y_t] = \sigma_\epsilon^2 t$, the data second moments (like $\sum_{t=1}^{T} y_{t-1}^2$) grow at order T^2, so the distribution of $\widehat{\beta} - \beta$ 'collapses' very quickly. Again, we can illustrate this by simulation, estimating (32) at $T = 25$, 100, 400, and 1000. The four panels for the estimated distribution in Figure 5 have been standardized to the same x-axis for visual comparison – and the convergence is dramatic. For comparison, the corresponding graphs for $\beta = 0.5$ are shown in Figure 6, where second moments grow at order T. Thus, to obtain a limiting distribution for $\widehat{\beta} - \beta$ which neither diverges nor degenerates to a constant, scaling by T is required (rather than $\sqrt{T}$ for $\mathsf{I}(0)$ data). Moreover, even after such a scaling, the form of the limiting distribution is different from that holding under stationarity.

The 't-statistic' for testing H_0: $\beta = 1$, often called the Dickey–Fuller test after Dickey and Fuller (1979), is easily computed, but does not have a standard t-distribution. Consequently, conventional critical values are incorrect, and using them can lead to over-rejection of the null of a unit root when it is true. Rather, the Dickey–Fuller (*DF*) test has a skewed distribution with a long left tail, making it hard to discriminate the null of a unit root from alternatives close to unity. The more general test, the Augmented Dickey-Fuller test is defined in the next section.

Unfortunately, the form of the limiting distribution of the *DF* test is also altered by the presence of a constant or a trend in either the DGP or the model. This means that different critical values are required in each case, although all the required tables of the correct critical values are available. Worse still, wrong choices of what deterministic terms to include – or which table is applicable – can seriously distort inference. As demonstrated in Section 2, the role and the interpretation of the constant term and the trend in the model changes as we move from the stationary case to the non-stationary unit-root case. It is also the case for stationary data that incorrectly omitting (say) an intercept can be disastrous, but mistakes are more easily made when the data are non-stationary.

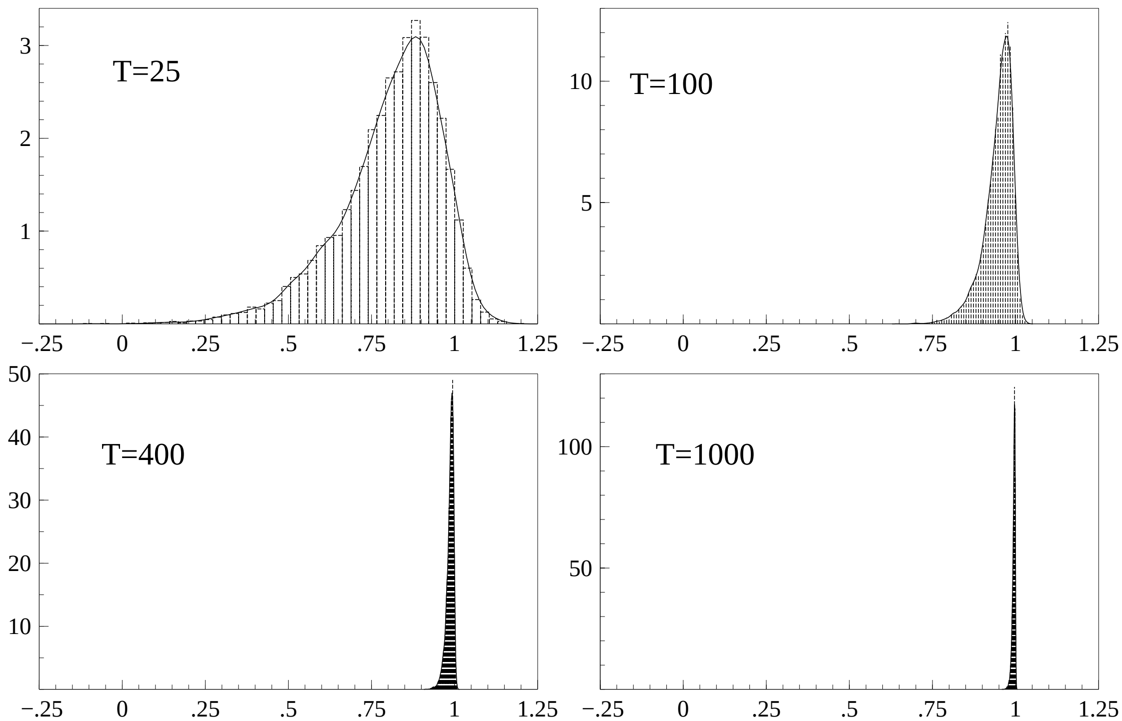

Figure 5: Frequency distributions of unit-root estimators

However, always including a constant and a trend in the estimated model ensures that the test will have the correct rejection frequency under the null for most economic time series. The required critical values have been tabulated using Monte Carlo simulations by Dickey and Fuller (1979, 1981), and most time-series econometric software (e.g., *PcGive*) automatically provides appropriate critical values for unit-root tests in almost all relevant cases, provided the correct model is used (see discussion in the next section).

7 Testing for cointegration

When data are non-stationary purely due to unit roots, they can be brought back to stationarity by linear transformations, for example, by differencing, as in $x_t - x_{t-1}$. If $x_t \sim \mathsf{I}(1)$, then by definition $\Delta x_t \sim \mathsf{I}(0)$. An alternative is to try a linear transformation like $y_t - \beta_1 x_t - \beta_0$, which induces cointegration when $y_t - \beta_1 x_t - \beta_0 \sim \mathsf{I}(0)$. But unlike differencing, there is no guarantee that $y_t - \beta_1 x_t - \beta_0$ is $\mathsf{I}(0)$ for any value of $\boldsymbol{\beta}$, as the discussion in Section 4 demonstrated.

There are many possible tests for cointegration: the most general of them is the multivariate test based on the vector autoregressive representation (VAR) discussed in Johansen (1988). These procedures will be described in Part II. Here we only consider tests based on the static and the dynamic regression model, assuming that x_t can be treated as weakly exogenous for the parameters of the conditional model (see e.g., Engle, Hendry and Richard, 1983).[11] As discussed in Section 5, the condition that there exists a genuine causal link between $\mathsf{I}(1)$ series y_t and x_t is that the residual

[11] Regression methods can be applied to model $\mathsf{I}(1)$ variables which are in fact linked (i.e., cointegrated). Most tests still have conventional distributions, apart from that corresponding to a test for a unit root.

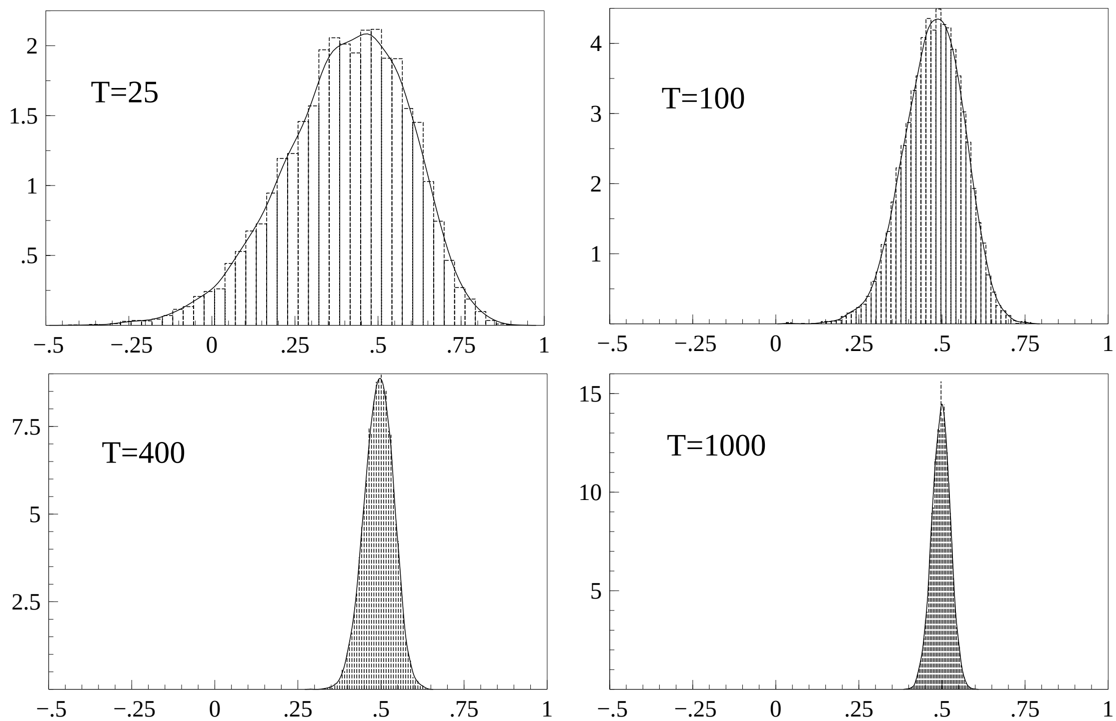

Figure 6: Frequency distributions of stationary autoregression estimators

$u_t \sim \mathsf{I}(0)$, otherwise a 'nonsense regression' has been estimated. Therefore, the Engle–Granger test procedure is based on testing that the residuals u_t from the static regression model (18) are stationary, i.e., that $\rho < 1$ in (2). As discussed in the previous section, the test of the null of a unit coefficient, using the *DF* test, implies using a non-standard distribution.

Let $\widehat{u}_t = y_t - \widehat{\beta}_1 x_t - \widehat{\beta}_0$ where $\widehat{\boldsymbol{\beta}}$ is the OLS estimate of the long-run parameter vector $\boldsymbol{\beta}$, then the null hypothesis of the *DF* test is H_0: $\rho = 1$, or equivalently, H_0: $1 - \rho = 0$ in:

$$\widehat{u}_t = \rho \widehat{u}_{t-1} + \varepsilon_t \tag{33}$$

or:

$$\Delta \widehat{u}_t = (1 - \rho) \widehat{u}_{t-1} + \varepsilon_t.$$

The test is based on the assumption that ε_t in (33) is white noise, and if the AR(1) model in (33) does not deliver white-noise errors, then it has to be augmented by lagged differences of residuals:

$$\Delta \widehat{u}_t = (1 - \rho) \widehat{u}_{t-1} + \psi_1 \Delta \widehat{u}_{t-1} + \cdots + \psi_m \Delta \widehat{u}_{t-m} + \varepsilon_t. \tag{34}$$

We call the test of H_0: $1-\rho = 0$ in (34) the augmented Dickey–Fuller test (*ADF*). A drawback of the *DF*-type test procedure (see Campos, Ericsson and Hendry, 1996, for a discussion of this drawback) is that the autoregressive model (33) for $\widehat{u}_t$ is the equivalent of imposing a common dynamic factor on the static regression model:

$$(1 - \rho L) y_t = \beta_0 (1 - \rho) + \beta_1 (1 - \rho L) x_t + \varepsilon_t. \tag{35}$$

For the DF test to have high power to reject H_0: $1-\rho=0$ when it is false, the common-factor restriction in (35) should correspond to the properties of the data. Empirical evidence has not produced much support for such common factors, rendering such tests non-optimal. Instead, Kremers, Ericsson and Dolado (1992) contrast them with a direct test for H_0: $\alpha_2=0$ in:

$$\Delta y_t = \alpha_0 + \alpha_1 \Delta x_t + \alpha_2 \left(y_{t-1} - \beta_1 x_{t-1} - \beta_0 \right) + \varepsilon_t, \quad (36)$$

where the parameters $\{\alpha_0, \alpha_1, \alpha_2, \beta_0, \beta_1\}$ are not constrained by the common-factor restriction in (35). Unfortunately, the null rejection frequency of their test depends on the values of the 'nuisance' parameters α_1 and σ_ε^2, so Kiviet and Phillips (1992) developed a test which is invariant to these values. The test reported in the empirical application in the next section is based on this test, and its distribution is illustrated in Figure 4, panel d. Although non-standard, so its critical values have been separately tabulated, its distribution is much closer to the Student t-distribution than the Dickey–Fuller, and correspondingly Banerjee *et al.* (1993) find the power of $\mathsf{t}_{\alpha_2=0}$ can be high relative to the DF test. However, when x_t is not weakly exogenous (i.e., when not only y_t adjusts to the previous equilibrium error as in (36), but also x_t does), the test is potentially a poor way of detecting cointegration. In this case, a multivariate test procedure is needed.

8 An empirical illustration

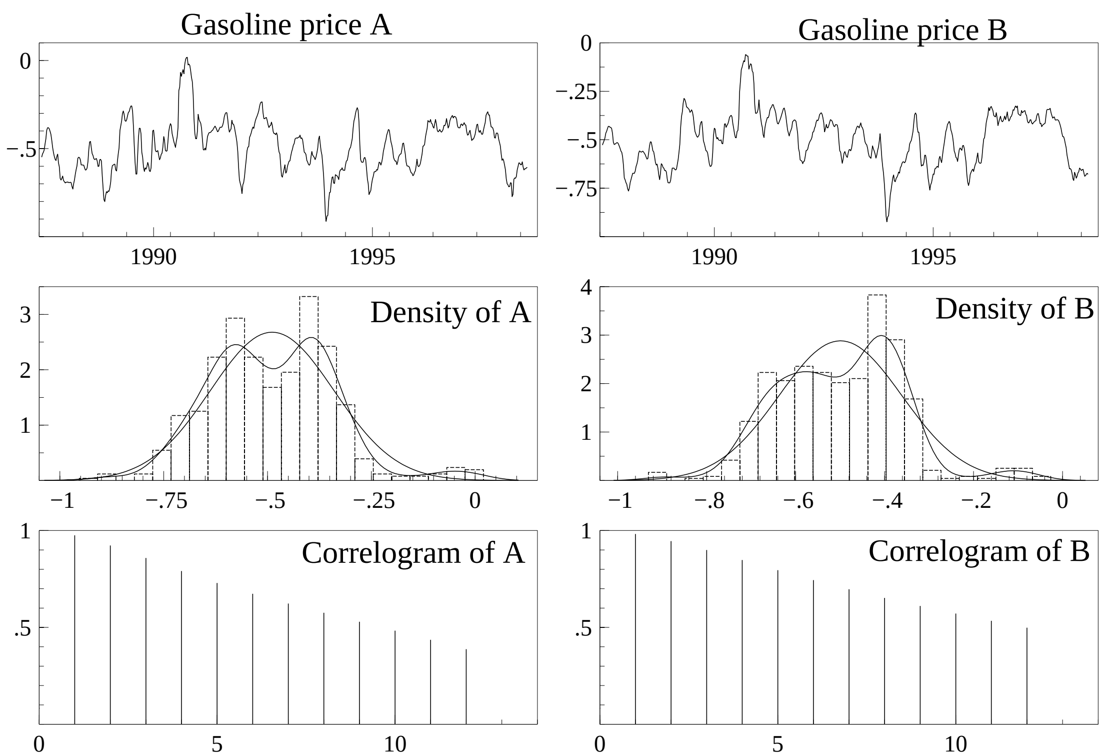

Figure 7: Gasoline prices at two locations in (log) levels, their empirical densities and autocorrelograms

In this section, we will apply the concepts and ideas discussed above to a data set consisting of two weekly gasoline prices ($P_{a,t}$ and $P_{b,t}$) at different locations over the period 1987 to 1998. The data in levels are graphed in Figure 7 on a log scale, and in (log) differences in Figure 8. The price levels exhibit 'wandering' behavior, though not very strongly, whereas the differenced series seem to fluctuate randomly around a fixed mean of zero, in a typically stationary manner. The bimodal frequency distribution of the price levels is also typical of non-stationary data, whereas the frequency distribution of the differences is much closer to normality, perhaps with a couple of outliers. We also notice the large autocorrelations of the price levels at long lags, suggesting non-stationarity, and the lack of such autocorrelations for the differenced prices, suggesting stationarity (the latter are shown with lines at $\pm 2\mathsf{SE}$ to clarify the insignificance of the autocorrelations at longer lags).

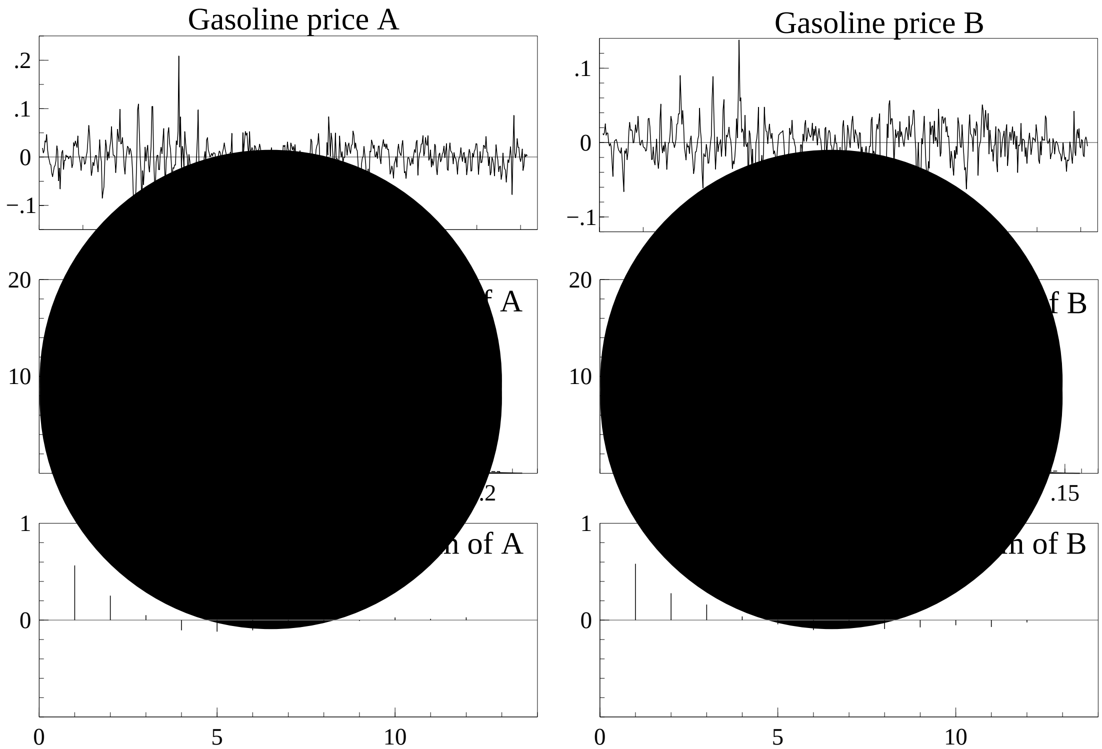

Figure 8: Gasoline prices at two locations in differences, their empirical density distribution and autocorrelogram

We first report the estimates from the static regression model:

$$y_t = \beta_0 + \beta_1 x_t + u_t,$$

and then consider the linear dynamic model:

$$y_t = a_0 + a_1 x_t + a_2 y_{t-1} + a_3 x_{t-1} + \epsilon_t. \tag{38}$$

Consistent with the empirical data, we will assume that $\mathsf{E}[\Delta y_t] = \mathsf{E}[\Delta x_t] = 0$, i.e., there are no linear trends in the data. Without changing the basic properties of the model, we can then rewrite (38) in the equilibrium-correction form:

$$\Delta y_t = a_1 \Delta x_t + (a_2 - 1)\left(y - \beta_0 - \beta_1 x\right)_{t-1} + \epsilon_t \tag{39}$$

where $a_2 \neq 1$, and:

$$\beta_0 = \frac{a_0}{1 - a_2} \quad \text{and} \quad \beta_1 = \frac{a_1 + a_3}{1 - a_2}. \tag{40}$$

In formulation (39), the model embodies the lagged equilibrium error $(y - \beta_0 - \beta_1 x)_{t-1}$, which captures departures from the long-run equilibrium as given by the static model. As demonstrated in (21), the equilibrium error will be a stationary process if $(a_2 - 1) \neq 0$ with a zero mean:

$$\mathsf{E}\left[y_t - \beta_0 - \beta_1 x_t\right] = 0, \tag{41}$$

whereas if $(a_2 - 1) = 0$, there is no adjustment back to equilibrium and the equilibrium error is a non-stationary process. The link of cointegration to the existence of a long-run solution is manifest here, since $y_t - \beta_0 - \beta_1 x_t = u_t \sim \mathsf{I}(0)$ implies a well-behaved equilibrium, whereas when $u_t \sim \mathsf{I}(1)$, no equilibrium exists. In (39), $(\Delta y_t, \Delta x_t)$ are $\mathsf{I}(0)$ when their corresponding levels are $\mathsf{I}(1)$, so with $\epsilon_t \sim \mathsf{I}(0)$, the equation is 'balanced' if and only if $(y - \beta_0 - \beta_1 x)_t$ is $\mathsf{I}(0)$ as well.

This type of 'balancing' occurs naturally in regression analysis when the model formulation permits it: we will demonstrate empirically in (43) that one does not need to actually write the model as in (39) to obtain the benefits. What matters is whether the residuals are uncorrelated or not.

The estimates of the static regression model over 1987(24)–1998(29) are:

$$\begin{aligned} p_{a,t} &= \underset{(2.2)}{0.018} + \underset{(67.4)}{1.01}\, p_{b,t} + u_t \\ \mathsf{R}^2 &= 0.89,\ \widehat{\sigma}_u = 0.050,\ DW = 0.18 \end{aligned} \tag{42}$$

where DW is the Durbin–Watson test statistic for first-order autocorrelation, and the 't-statistics' based on (26) are given in parentheses. Although the DW test statistic is small and suggests non-stationarity, the DF test of u_t in (42) supports stationarity ($DF = -8.21^{**}$).

Furthermore, the following mis-specification tests were calculated:

$$\begin{aligned} &AR(1\text{–}7),\ \mathsf{F}(7, 569) &&= 524.6\ [0.00]^{**} \\ &ARCH(7),\ \mathsf{F}(7, 562) &&= 213.2\ [0.00]^{**} \\ &\textit{Normality},\ \chi^2(2) &&= 22.9\ [0.00]^{**} \end{aligned}$$

The $AR(1\text{–}m)$ is a test of residual autocorrelation of order m distributed as $\mathsf{F}(m, T)$, i.e. a test of $\mathsf{H}_0 : u_t = \varepsilon_t$ against $\mathsf{H}_1 : u_t = \rho_1 u_{t-1} + \cdots + \rho_m u_{t-m} + \varepsilon_t$. The test of autocorrelated errors of order 1-7 is very large and the null of no autocorrelation is clearly rejected. The $ARCH(m)$ (see Engle, 1982) is a test of autoregressive residual heteroscedasticity of order m distributed as $\mathsf{F}(m, T - m)$. *Normality* denotes the Doornik and Hansen (1994) test of residual normality, distributed as $\chi^2(2)$. It is based on the third and the fourth moments around the mean, i.e., it tests for skewness and excess kurtosis of the residuals. Thus, the normality and homoscedasticity of the residuals are also rejected. So the standard assumptions underlying the static regression model are clearly violated. In Figure 9, panel (a), we have graphed the actual and fitted values from the static model (42), in (b) the residuals $\widehat{u}_t$, in (c) their correlogram and in (d) the residual histogram compared with the normal distribution. The residuals show substantial temporal correlation, consistent with a highly autocorrelated process. This is further confirmed by the correlogram, exhibiting large, though declining, autocorrelations. This can explain why the DF test rejected the unit-root hypothesis above: though ρ is close to unity, the large sample size improves the precision of the test and, hence, allows us to reject the hypothesis. Furthermore, the residuals seem to be symmetrically distributed around the mean, but are leptokurtic to some extent.

To evaluate the forecasting performance of the model, we have calculated the one-step ahead forecasts and their (calculated) 95% confidence intervals over the last two years. The outcome is illustrated in Figure 10a, which shows periods of consistent over- and under- predictions. In

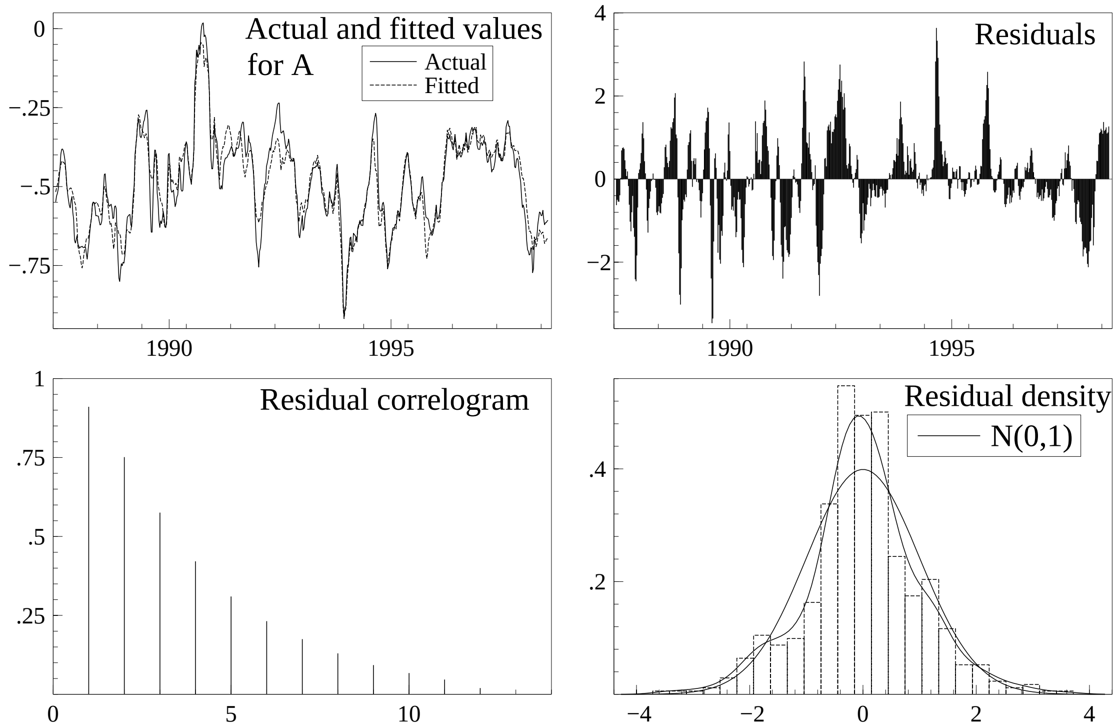

Figure 9: Actual and fitted values, residuals, their correlogram and histogram, from the static model

particular, at the end of 1997 until the beginning of 1998, predictions were consistently below actual prices, and thereafter consistently above.[12]

Finally, we have recursively calculated the estimated β_0 and $\beta_1$coefficients in (42) from 1993 onwards. Figure 11 shows these recursive graphs. It appears that even if the recursive estimates of β_0 and $\beta_1$are quite stable, at the end of the period they are not within the confidence band at the beginning of the recursive sample (and remember the previous footnote). We also report the 1-step residuals with $\pm 2\mathsf{SE}$, and the sequence of constancy tests based on Chow (1964) statistics (scaled by their 1% critical values, so values > 1 reject).

Altogether, most empirical economists would consider this econometric outcome as quite unsatisfactory. We will now demonstrate how much the model can be improved by accounting for the left-out dynamics.

The estimate of the dynamic regression model (38) with two lags is:

$$p_{a,t} = \underset{(0.3)}{0.001} + \underset{(36.1)}{1.33}\, p_{a,t-1} - \underset{(12.5)}{0.46}\, p_{a,t-2} + \underset{(23.9)}{0.91}\, p_{b,t} - \underset{(14.8)}{1.12}\, p_{b,t-1} + \underset{(6.4)}{0.34} p_{b,t-2} + \varepsilon_t \tag{43}$$
$$\mathsf{R}^2 = 0.99,\ \widehat{\sigma}_\varepsilon = 0.018,\ DW = 2.03$$

Compared with the static regression model, we notice that the DW statistic is now close to 2, and that the residual standard error has decreased from 0.050 to 0.018, so the precision has increased

[12]The software assumes the residuals are homoscedastic white-noise when computing confidence intervals, standard errors, etc., which is flagrantly wrong here.

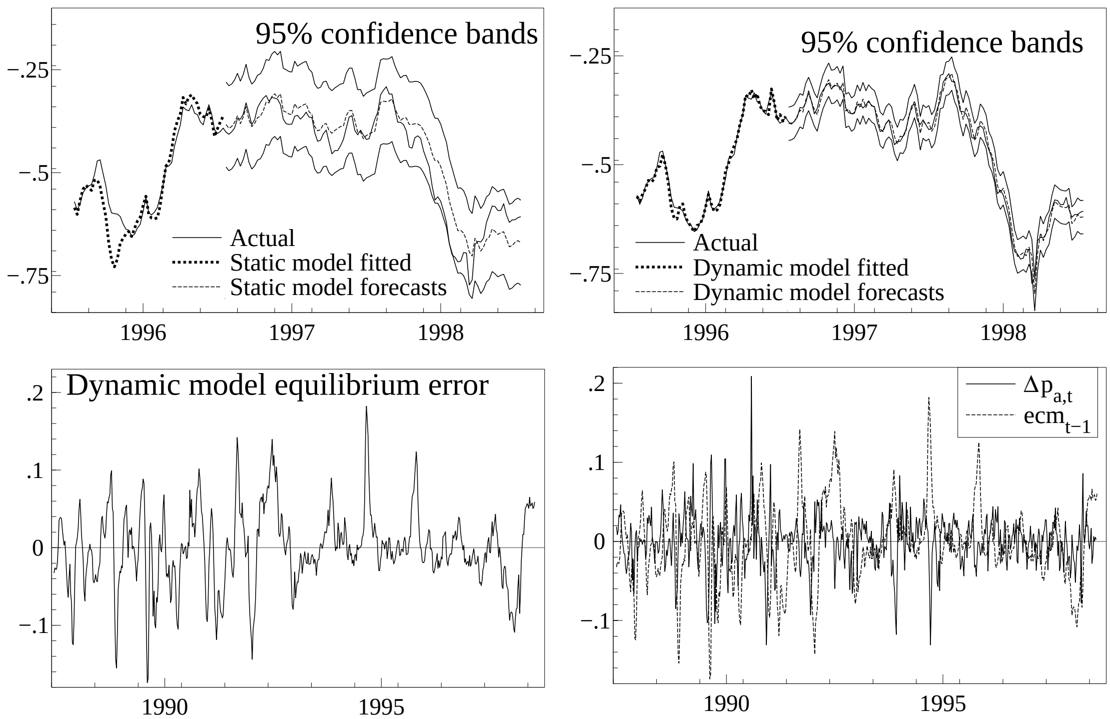

Figure 10: Static and dynamic model forecasts, the equilibrium error, and its relation to $\Delta p_{a,t}$

approximately 2.5 times. The mis-specification tests are:

$$\begin{aligned} AR(1\text{–}7),\ \mathsf{F}(7,563) &= 1.44\ [0.19] \\ ARCH(7),\ \mathsf{F}(7,556) &= 10.6\ [0.00]^{**} \\ Normality,\ \chi^2(2) &= 79.8\ [0.00]^{**} \end{aligned}$$

We note that residual autocorrelation is no longer a problem, but there is still evidence of the residuals being heteroscedastic and non-normal. Figure 12 show the graphs of actual and fitted, residuals, residual correlogram, and the residual histogram compared to the normal distribution. Fitted and actual values are now so close that it is no longer possible to visually distinguish between them. Residuals exhibit no temporal correlation. However, with the increased precision (the smaller residual standard error), we can now recognize several 'outliers' in the data. This is also evident from the histogram exhibiting quite long tails, resulting in the rejection of normality. The rejection of homoscedastic residual variance seems to be explained by a larger variance in the first part of the sample, prevailing approximately till the end of 1992, probably due to the Gulf War.

As for the static regression model, the one-step ahead prediction errors with their 95% confidence intervals have been graphed in Figure 10b. It appears that there is no systematic under- or over-prediction in the dynamic model. Also the prediction intervals are much narrower, reflecting the large increase in precision as a result of the smaller residual standard error in this model.

It is now possible to derive the static long-run solution as given by (39) and (40):

$$p_{a,t} = \underset{(0.3)}{0.008} + \underset{(22.8)}{0.99}\, p_{b,t} \quad \mathsf{t}_{ur} = 8.25^{**} \tag{44}$$

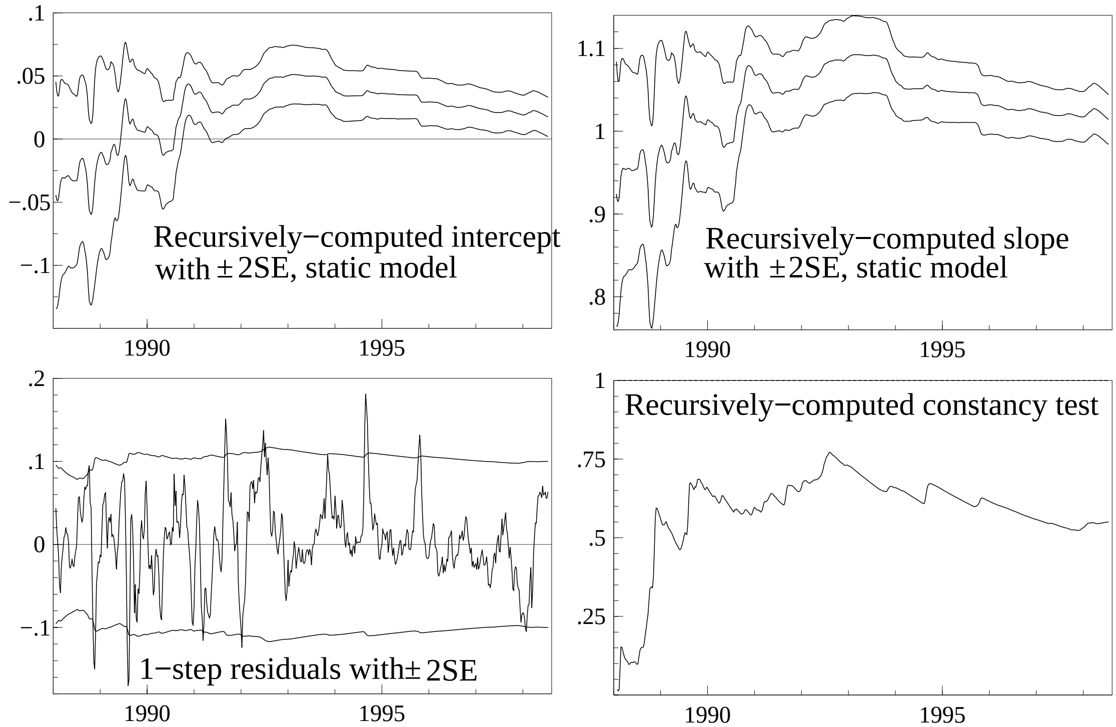

Figure 11: Recursively-calculated coefficients of β_0 and β_1 in (41) with 1-step residuals and Chow tests

Note that the coefficient estimates of β_0 and β_1 are almost identical to the static regression model, illustrating the fact that the OLS estimator was unbiased. However, the correctly-calculated standard errors of estimates produce much smaller t-values, consistent with the downward bias of $\mathsf{SE}[\widehat{\beta}]$ in the static regression model, revealing an insignificant intercept. The graph of the equilibrium error u_t (shown in Figure 10c) is essentially the log of the relative price, $p_{a,t} - p_{b,t}$. Note that u_t has a zero mean, consistent with (41), and that it is strongly autocorrelated, consistent with (21). From the coefficient estimate of the ecm_{t-1} below in (45), we find that the autocorrelation coefficient $(1 - \alpha_2)$ in (21) corresponds to 0.86, which is a fairly high autocorrelation. In the static model, this autocorrelation was left in the residuals, whereas in the dynamic model, we have explained it by short-run adjustment to current and lagged changes in the two gasoline prices. Nevertheless, the values of the DF test on the static residuals, and the unit-root test in the dynamic model (t_{ur} in (44)) are closely similar here, even though the test for two common factors in (43) rejects (one common factor is accepted).

Finally, we report the estimates of the model reformulated in equilibrium-correction form (39), suppressing the constant of zero due to the equilibrium-correction term being mean adjusted:

$$\Delta p_{a,t} = \underset{(12.5)}{0.46}\,\Delta p_{a,t-1} + \underset{(24.3)}{0.91}\,\Delta p_{b,t} - \underset{(6.5)}{0.34}\,\Delta p_{b,t-1} - \underset{(8.3)}{0.13}\,ecm_{t-1} + \varepsilon_t \qquad (45)$$
$$\mathsf{R}^2 = 0.69,\ \widehat{\sigma}_\varepsilon = 0.018,\ DW = 2.05$$

Notice that the corresponding estimated coefficients, the residual standard error, and DW are identical with the estimates in (43), demonstrating that the two models are statistically equivalent, though one is formulated in non-stationary and the other in stationary variables. The coefficient

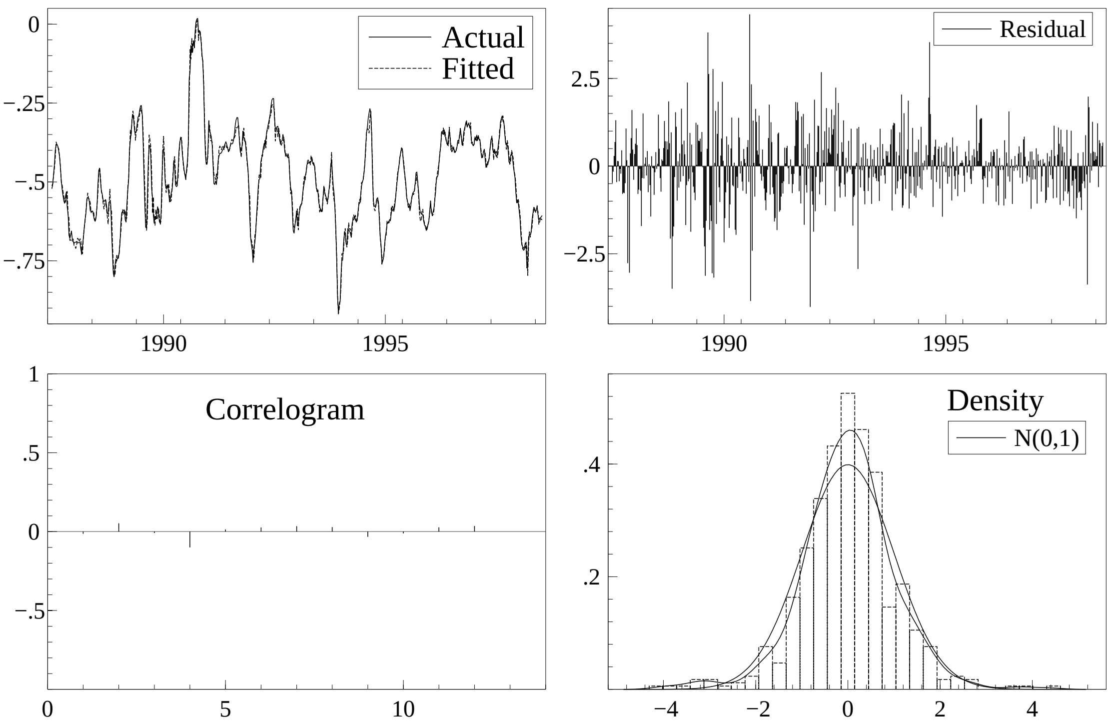

Figure 12: Actual and fitted values, residuals, their correlogram and histogram for the dynamic model

of -0.13 on ecm_{t-1} suggests moderate adjustment, with 13% of any disequilibrium in the relative prices being removed each week.[13] Figure 10d shows the relation of $\Delta p_{a,t}$ to ecm_{t-1}.

However, R^2 is lower in (45) than in (42) and (43), demonstrating the hazards of interpreting R^2 as a measure of goodness of fit. When we calculate $\mathsf{R}^2 = \Sigma(y_t - \widehat{y}_t)^2/\Sigma(y_t - \overline{y})^2$, we essentially compare how well our model can predict y_t compared to a straight line. When y_t is trending, any other trending variable can do better than a straight line, which explains the high R^2 often obtained in regressions with non-stationary variables. Hence, as already discussed in Section 5, the sum of squares $\Sigma(y_t - \overline{y})^2$ is not an appropriate measure of the variation of a trending variable. In contrast, $\mathsf{R}^2 = \Sigma(\Delta y_t - \widehat{\Delta y_t})^2/\Sigma(\Delta y_t - \overline{\Delta y})^2$ from (45) measures the improvement in model fit compared to a random walk as a reasonable measure of how good our model is (although that R^2 would also change if the dependent variable became ecm_t in another statistically-equivalent version of the same model).

Finally, we report the recursively-calculated coefficients of the model parameters in (45) in Figure 13. The estimated coefficients are reasonably stable over time, although a few of the recursive estimates fall outside the confidence bands (especially around the Gulf War). Such non-constancies reveal remaining non-stationarities in the two gasoline prices, but resolving this issue would necessitate another paper.....

[13] The equilibrium-correction error ecm_{t-1} does not influence $\Delta p_{b,t}$ in a bivariate system, so that aspect of weak exogeneity is not rejected.

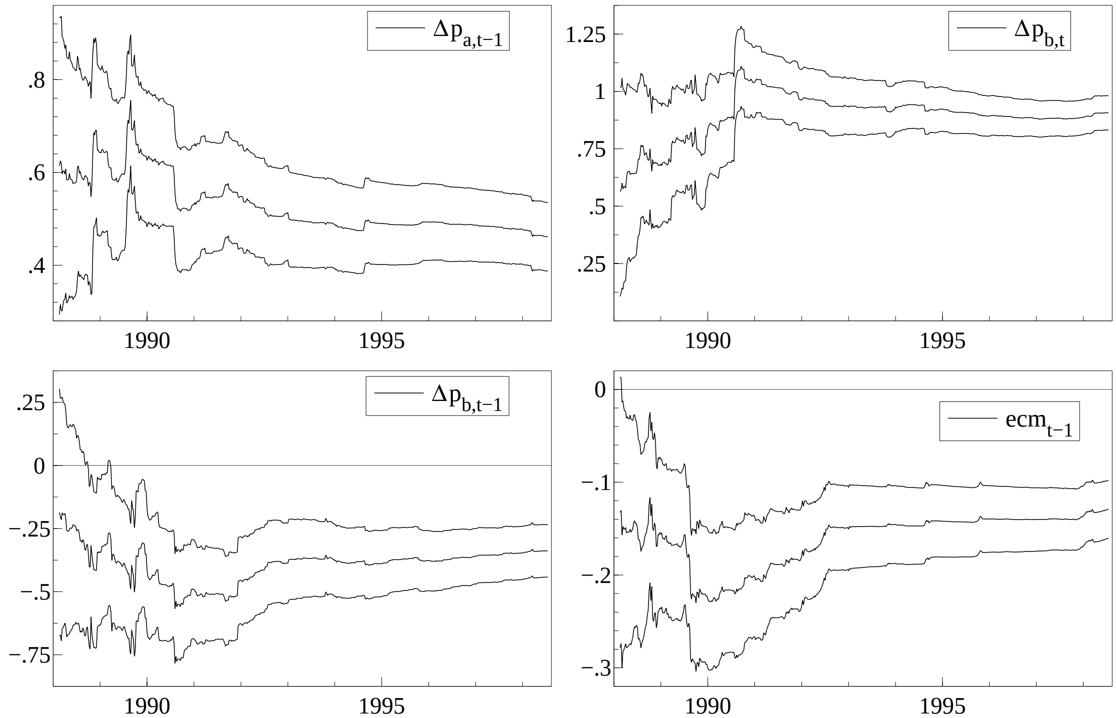

Figure 13: Recursively-calculated parameter estimates of the ECM model

9 Conclusion

Although 'classical' econometric theory generally assumed stationary data, particularly constant means and variances across time periods, empirical evidence is strongly against the validity of that assumption. Nevertheless, stationarity is an important basis for empirical modeling, and inference when the stationarity assumption is incorrect can induce serious mistakes. To develop a more relevant basis, we considered recent developments in modeling non-stationary data, focusing on autoregressive processes with unit roots. We showed that these processes were non-stationary, but could be transformed back to stationarity by differencing and cointegration transformations, where the latter comprised linear combinations of the variables that did not have unit roots.

We investigated the comparative properties of stationary and non-stationary processes, reviewed the historical development of modeling non-stationarity and presented a re-run of a famous Monte Carlo simulation study of the dangers of ignoring non-stationarity in static regression analysis. Next, we described how to test for unit roots in scalar autoregressions, then extended the approach to tests for cointegration. Finally, an extensive empirical illustration using two gasoline prices implemented the tools described in the preceding analysis.

Unit-root non-stationarity seems widespread in economic time series, and some theoretical models entail unit roots. Links between variables will then 'spread' such non-stationarities throughout the economy. Thus, we believe it is sensible empirical practice to assume unit roots in (log) levels until that is rejected by well-based evidence. Cointegrated relations and differenced data both help model unit roots, and can be related in equilibrium-correction equations, as we illustrated. For modeling purposes, a unit-root process may also be considered as a statistical approximation when serial correlation is high. Monte Carlo studies have demonstrated that treating near-unit roots as unit roots in situations where the unit-root hypothesis is only approximately correct makes statistical inference more reliable than otherwise.

Unfortunately, other sources of non-stationarity may remain, such as changes in parameters (particularly shifts in the means of equilibrium errors and growth rates) or data distributions, so careful empirical evaluation of fitted equations remains essential. We reiterate the importance of having white-noise residuals, preferably homoscedastic, to avoid mis-leading inferences. This emphasizes the advantages of accounting for the dynamic properties of the data in equilibrium-correction equations, which not only results in improved precision from lower residual variances, but delivers empirical estimates of adjustment parameters.

Part II of our attempt to explain cointegration analysis will address system methods. Since cointegration inherently links several variables, multivariate analysis is natural, and recent developments have focused on this approach. Important new insights result, but new modeling decisions also have to made in practice. Fortunately, there is excellent software available for implementing the methods discussed in Johansen (1995a), including CATS in RATS (see Hansen and Juselius, 1995) and *PcFiml* (see Doornik and Hendry, 1997), and we will address the application of such tools to the gasoline price data considered above.

References

Attfield, C. L. F., Demery, D., and Duck, N. W. (1995). Estimating the UK demand for money function: A test of two approaches. Mimeo, Economics department, University of Bristol.

Baba, Y., Hendry, D. F., and Starr, R. M. (1992). The demand for M1 in the U.S.A., 1960–1988. *Review of Economic Studies*, **59**, 25–61.

Banerjee, A., Dolado, J. J., Galbraith, J. W., and Hendry, D. F. (1993). *Co-integration, Error Correction and the Econometric Analysis of Non-Stationary Data*. Oxford: Oxford University Press.

Banerjee, A., and Hendry, D. F. (1992). Testing integration and cointegration: An overview. *Oxford Bulletin of Economics and Statistics*, **54**, 225–255.

Campos, J., Ericsson, N. R., and Hendry, D. F. (1996). Cointegration tests in the presence of structural breaks. *Journal of Econometrics*, **70**, 187–220.

Chan, N. H., and Wei, C. Z. (1988). Limiting distributions of least squares estimates of unstable autoregressive processes. *Annals of Statistics*, **16**, 367–401.

Chow, G. C. (1964). A comparison of alternative estimators for simultaneous equations. *Econometrica*, **32**, 532–553.

Davidson, J. E. H., Hendry, D. F., Srba, F., and Yeo, J. S. (1978). Econometric modelling of the aggregate time-series relationship between consumers' expenditure and income in the United Kingdom. *Economic Journal*, **88**, 661–692. Reprinted in Hendry, D. F. (1993), *Econometrics: Alchemy or Science?* Oxford: Blackwell Publishers.

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.

Dickey, D. A., and Fuller, W. A. (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica*, **49**, 1057–1072.

Doornik, J. A., and Hansen, H. (1994). A practical test for univariate and multivariate normality. Discussion paper, Nuffield College.

Doornik, J. A., and Hendry, D. F. (1997). *Modelling Dynamic Systems using PcFiml 9 for Windows.* London: Timberlake Consultants Press.

Doornik, J. A., and Hendry, D. F. (1998). Monte Carlo simulation using PcNaive for Windows. Unpublished typescript, Nuffield College, University of Oxford.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity, with estimates of the variance of United Kingdom inflations. *Econometrica*, **50**, 987–1007.

Engle, R. F., and Granger, C. W. J. (1987). Cointegration and error correction: Representation, estimation and testing. *Econometrica*, **55**, 251–276.

Engle, R. F., Hendry, D. F., and Richard, J.-F. (1983). Exogeneity. *Econometrica*, **51**, 277–304. Reprinted in Hendry, D. F., *Econometrics: Alchemy or Science?* Oxford: Blackwell Publishers, 1993; and in Ericsson, N. R. and Irons, J. S. (eds.) *Testing Exogeneity*, Oxford: Oxford University Press, 1994.

Friedman, M., and Schwartz, A. J. (1982). *Monetary Trends in the United States and the United Kingdom: Their Relation to Income, Prices, and Interest Rates, 1867–1975.* Chicago: University of Chicago Press.

Granger, C. W. J. (1981). Some properties of time series data and their use in econometric model specification. *Journal of Econometrics*, **16**, 121–130.

Granger, C. W. J., and Newbold, P. (1974). Spurious regressions in econometrics. *Journal of Econometrics*, **2**, 111–120.

Hall, P., and Heyde, C. C. (1980). *Martingale Limit Theory and its Applications.* London: Academic Press.

Hansen, H., and Juselius, K. (1995). CATS in RATS: Cointegration analysis of time series. Discussion paper, Estima, Evanston, IL.

Hendry, D. F. (1980). Econometrics: Alchemy or science?. *Economica*, **47**, 387–406. Reprinted in Hendry, D. F. (1993), *Econometrics: Alchemy or Science?* Oxford: Blackwell Publishers.

Hendry, D. F. (1995). *Dynamic Econometrics.* Oxford: Oxford University Press.

Hendry, D. F., and Anderson, G. J. (1977). Testing dynamic specification in small simultaneous systems: An application to a model of building society behaviour in the United Kingdom. In Intriligator, M. D. (ed.), *Frontiers in Quantitative Economics*, Vol. 3, pp. 361–383. Amsterdam: North Holland Publishing Company. Reprinted in Hendry, D. F. (1993), *Econometrics: Alchemy or Science?* Oxford: Blackwell Publishers.

Hendry, D. F., and Mizon, G. E. (1978). Serial correlation as a convenient simplification, not a nuisance: A comment on a study of the demand for money by the Bank of England. *Economic Journal*, **88**, 549–563. Reprinted in Hendry, D. F. (1993), *Econometrics: Alchemy or Science?* Oxford: Blackwell Publishers.

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

Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, **12**, 231–254.

Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica*, **59**, 1551–1580.

Johansen, S. (1992a). Determination of cointegration rank in the presence of a linear trend. *Oxford Bulletin of Economics and Statistics*, **54**, 383–398.

Johansen, S. (1992b). A representation of vector autoregressive processes integrated of order 2. *Econometric Theory*, **8**, 188–202.

Johansen, S. (1995a). *Likelihood-based Inference in Cointegrated Vector Autoregressive Models*. Oxford: Oxford University Press.

Johansen, S. (1995b). *Likelihood based Inference on Cointegration in the Vector Autoregressive Model*. Oxford: Oxford University Press.

Johansen, S., and Juselius, K. (1990). Maximum likelihood estimation and inference on cointegration – With application to the demand for money. *Oxford Bulletin of Economics and Statistics*, **52**, 169–210.

Johansen, S., and Juselius, K. (1992). Testing structural hypotheses in a multivariate cointegration analysis of the PPP and the UIP for UK. *Journal of Econometrics*, **53**, 211–244.

Kiviet, J. F., and Phillips, G. D. A. (1992). Exact similar tests for unit roots and cointegration. *Oxford Bulletin of Economics and Statistics*, **54**, 349–367.

Kremers, J. J. M., Ericsson, N. R., and Dolado, J. J. (1992). The power of cointegration tests. *Oxford Bulletin of Economics and Statistics*, **54**, 325–348.

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

Nelson, C. R., and Plosser, C. I. (1982). Trends and random walks in macroeconomic time series: some evidence and implications. *Journal of Monetary Economics*, **10**, 139–162.

Park, J. Y., and Phillips, P. C. B. (1988). Statistical inference in regressions with integrated processes. part 1. *Econometric Theory*, **4**, 468–497.

Park, J. Y., and Phillips, P. C. B. (1989). Statistical inference in regressions with integrated processes. part 2. *Econometric Theory*, **5**, 95–131.

Phillips, P. C. B. (1986). Understanding spurious regressions in econometrics. *Journal of Econometrics*, **33**, 311–340.

Phillips, P. C. B. (1987a). Time series regression with a unit root. *Econometrica*, **55**, 277–301.

Phillips, P. C. B. (1987b). Towards a unified asymptotic theory for autoregression. *Biometrika*, **74**, 535–547.

Phillips, P. C. B. (1988). Regression theory for near-integrated time series. *Econometrica*, **56**, 1021–1043.

Phillips, P. C. B., and Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika*, **75**, 335–346.

Qin, D. (1993). *The Formation of Econometrics: A Historical Perspective*. Oxford: Clarendon Press.

Sargan, J. D. (1964). Wages and prices in the United Kingdom: A study in econometric methodology (with discussion). In Hart, P. E., Mills, G., and Whitaker, J. K. (eds.), *Econometric Analysis for National Economic Planning*, Vol. 16 of *Colston Papers*, pp. 25–63. London: Butterworth Co. Reprinted as pp. 275–314 in Hendry D. F. and Wallis K. F. (eds.) (1984). *Econometrics and Quantitative Economics*. Oxford: Basil Blackwell, and as pp. 124–169 in Sargan J. D. (1988), *Contributions to Econometrics*, Vol. 1, Cambridge: Cambridge University Press.

Spanos, A. (1986). *Statistical Foundations of Econometric Modelling*. Cambridge: Cambridge University Press.

Stock, J. H. (1987). Asymptotic properties of least squares estimators of cointegrating vectors. *Econometrica*, **55**, 1035–1056.

Student (1908). On the probable error of the mean. *Biometrika*, **6**, 1–25.

Working, H. (1934). A random difference series for use in the analysis of time series. *Journal of the American Statistical Association*, **29**, 11–24.

Yule, G. U. (1926). Why do we sometimes get nonsense-correlations between time-series? A study in sampling and the nature of time series (with discussion). *Journal of the Royal Statistical Society*, **89**, 1–64.