EUROPEAN JOURNAL OF PURE AND APPLIED MATHEMATICS
Vol. 10, No. 1, 2017, 58-81
ISSN 1307-5543 – www.ejpam.com
Published by New York Business Global

Sir Clive W.J. Granger Memorial Special Issue on Econometrics

Clive W.J. Granger and Cointegration

Jennifer L. Castle[1] and David F. Hendry[2,*]

[1] *Magdalen College and Institute for New Economic Thinking at the Oxford Martin School, University of Oxford, UK.*
[2] *Economics Department and Institute for New Economic Thinking at the Oxford Martin School, University of Oxford, UK.*

Abstract. Clive Granger developed the fundamental concept of cointegration for linking variables within non-stationary vector time series. Granger discovered cointegration while trying to refute a critique by Hendry of his research with Paul Newbold on 'nonsense regressions' between non-stationary data. Although the initial estimation and testing approach in his paper with Robert F. Engle has been superceded by a plethora of methods, the concept of cointegration has led to a merger of economic analyses of long-run equilibrium relations with empirical dynamic systems. The multivariate cointegration method of Søren Johansen extended Nobel Laureate Trygve Haavelmo's earlier formulation of an economy as a system of simultaneous stochastic relationships to non-stationary time series. Clive Granger was awarded The Sveriges Riksbank Prize in Economic Science in Memory of Alfred Nobel in 2003 for his contribution, sharing it with Rob Engle, whose citation was for developing methods for analyzing changing variances.

2010 Mathematics Subject Classifications: 62M10, 91B84

Key Words and Phrases: Clive Grange, cointegration, nonsense regression, equilibrium correction, Granger Representation Theorem.

1. Introduction

How cointegration was discovered is similar to many other scientific discoveries, namely the outcome of exploring numerous paths leading to useful but incomplete insights plagued by anomalies, yet eventually resolved, in this case by Clive Granger's major breakthrough.

*Corresponding author.

Email addresses: jennifer.castle@magd.ox.ac.uk (J. L. Castle),
david.hendry@nuffield.ox.ac.uk (D. F. Hendry)

Its history has been recounted a number of times, including by Hendry (2004) and Hendry and Timo Teräsvirta (2013) (see [56], [66]) on both of which we draw. For general histories of econometrics and time series, see Mary Morgan (1990) [81], Duo Qin (1993, 2013) [94][95], Hendry and Morgan (1995) [62], who also reprint many of the salient original historical publications, and Judy Klein (1997) [77].

'Odd' correlations between pairs of variables were found almost as soon as co-relation[†] analysis was invented by Francis Galton (1889) (see [31]) in the late 19th century, immediately followed by efforts to understand why these turned up surprisingly often. Two explanations soon appeared. The first was called 'spurious correlation', attributed by Udny Yule (1897) (see [110]) to the two variables being correlated with a third not included in the analysis. The second was called 'nonsense correlations', namely high correlations lacking sensible explanations, such as between the numbers of church marriages and mortality in the United Kingdom. Yule (1926) [111] argued such correlations were due to the variables involved being non-stationary, where some features of the distributions changed over time. This was the first important breakthrough in helping to understand why a specific type of non-stationarity could distort statistical inference so badly. A glance at many time-series suggests that stationarity, in the weak sense of constant unconditional means and variances, is not a reasonable starting point for empirical modeling.

Figure 1 (a) shows plots of the logs of annual nominal wages and prices in the UK over 1860–2011, which have trended dramatically (nominal wages have risen by 70,000% during that time): see Castle and Hendry (2014) [11]. Panel (b) records two key real variables over the same epoch, namely logs of productivity (measured by output per person per year) and of the associated capital stock (matched by mean values), which have moved in tandem except for the inter-war period, but also display changing trends. Panel (c) switches to the post-war period showing logs of quarterly UK real consumers' expenditure and disposable income, 1955–2004, which exhibit changing trends and evolving seasonality, as well as an increasing divergence. Finally, panel (d) illustrates that non-stationarity is not merely of interest in economics by showing a consequence of energy driven economic growth, namely greatly increasing atmospheric levels of carbon dioxide (CO_2) at Mauna Loa, with the marked seasonality driven by vegetation growth and decay (see Hendry and Felix Pretis, 2013, [63]).

Nevertheless, most theoretical and empirical econometric analyses assumed stationarity until the mid-to-late 1970s. For example, the 'error-correction' formulation by Denis Sargan (1964) (see [98]), which will play an important role later, did not address non-stationarity. Then there was a resurgence of interest in nonsense regressions beginning with the papers by Granger and Paul Newbold (1974,1977) [45][46], and the derivations of critical values for tests of unit roots in autoregressions by Wayne Fuller (1976) (see [30]). The attempt to rebut the critique of Granger and Newbold (1977) (see [46]) by Hendry (1977) (see [54]) prompted a counter critique by Granger, which led him to the discovery of cointegration. There are now more than **200,000** citations to publications that have included the word 'cointegration' in their title, based on Anne-Wil Harzing's invaluable

[†]Later renamed correlation, just as co-integration became cointegration: will co-breaking follow to be renamed cobreaking?

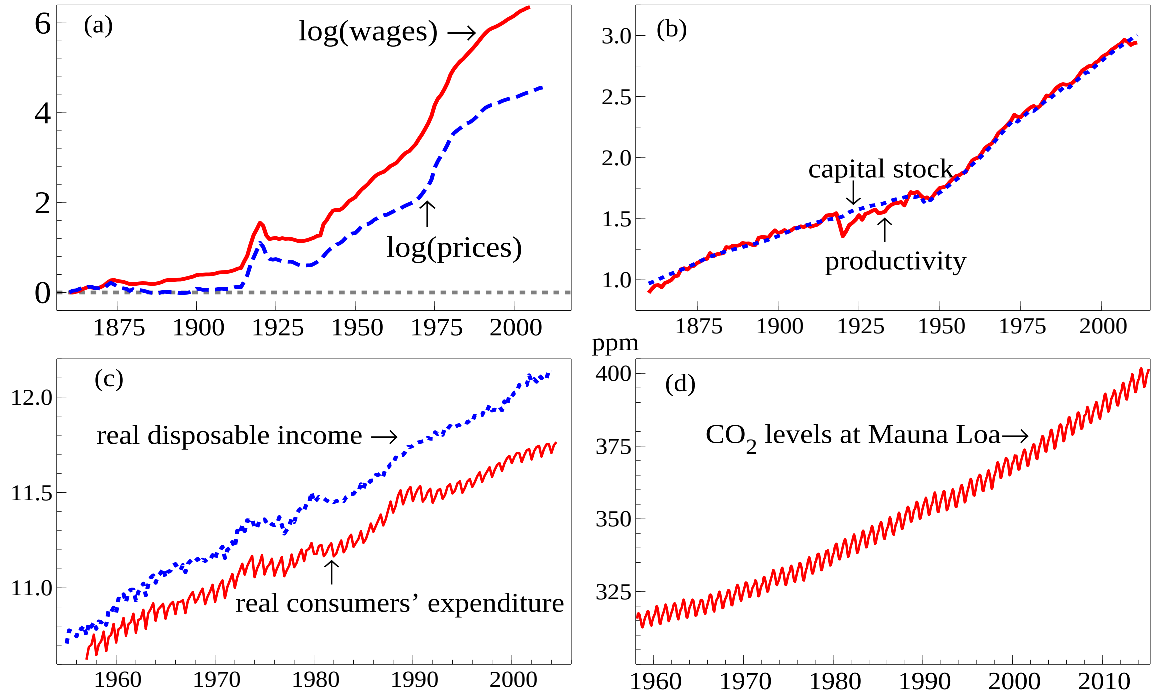

Figure 1: (a) Logs of annual wages and prices in the UK over 1860–2011; (b) Logs of annual productivity and the capital stock in the UK over 1860–2011; (c) Logs of quarterly UK real consumers' expenditure and disposable income, 1955–2004; (d) CO_2 levels at Mauna Loa in parts per million

Publish or Perish, Melbourne (2016),‡ a testimony to Clive's massive impact.

Section 2 first summarizes the intellectual history leading from nonsense regressions to equilibrium-correction models. The next phase was moving from long-run equilibrium-correction relations to cointegration between non-stationary time series, considered in section 3. Section 4 discusses the intrusion of unit-root distributions into statistical estimation and inference about the parameters of equilibrium-correction models. Then section 5 describes Granger's general formulation of cointegration and the key role of the Granger Representation Theorem. Section 6 discusses some of the implications of cointegration both for economic analyses and climate modeling. Section 7 concludes.

2. From nonsense regressions to equilibrium-correction models

In their historical review of econometrics, Hendry and Morgan (1995) (see [62]) record that one of the first analyses of 'problematic' correlations between time series when data are not stationary was by R.H. Hooker (1901) (see [67]), who suggested the difficulties arose from what he viewed as 'common trends' in the variables. Hooker in fact empirically modeled the relation between marriage rates and 'trade', and sorted out effects due

‡See http://www.harzing.com/publications/publish-or-perish-book?source=pop_4.28.1.6105

to trends and those due to 'oscillations', as well as considering the impacts of 'regime shifts'. A quarter century later, Udny Yule (1926) [111] showed that singly, and especially doubly, integrated data ($\mathsf{I}(1)$ and $\mathsf{I}(2)$ processes) generated 'nonsense correlations' between unconnected time series.[§] In the same year, but apparently unaware of each other's contributions, Bradford Smith (1926) (see [101]) discussed how to nest models in first differences with those in levels, a precursor to the 'error-correction' formulation in Sargan (1964) [98]: see Terence Mills (2011) [80].

The model Yule used in his 1926 Presidential Address to the Royal Statistical Society to understand correlation coefficients between $\mathsf{I}(1)$ or $\mathsf{I}(2)$ variables used two independent homoskedastic, random variables $\{e_t\}$ and $\{u_t\}$ independently drawn from normal distributions with zero means and variances σ_e^2 and σ_u^2, denoted by $e_t \sim \mathsf{IN}[0, \sigma_e^2]$ and $u_t \sim \mathsf{IN}[0, \sigma_u^2]$. Generate y_t and x_t from $y_0 = x_0 = 0$ to be $\mathsf{I}(1)$ by:

$$y_t = \sum_{i=1}^{t} e_i \quad \text{and} \quad x_t = \sum_{j=1}^{t} u_j. \tag{1}$$

In modern terms, the data generation process in Yule's (hand) simulations was the bivariate random walk:

$$\begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix} + \begin{pmatrix} e_t \\ u_t \end{pmatrix} \quad \text{where} \quad \begin{pmatrix} e_t \\ u_t \end{pmatrix} \sim \mathsf{IN}_2 \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_e^2 & 0 \\ 0 & \sigma_u^2 \end{pmatrix} \right]. \tag{2}$$

Yule found that the distribution of their correlation coefficient was approximately uniform and that the null of no relation between y_t and x_t could be rejected most of the time (around 70% at a nominal 5% significance level for a sample size of $T = 100$), even though the series were independently generated. Regressing either of these two $\mathsf{I}(1)$ variables on the other, as in (say):

$$y_t = \beta_0 + \beta_1 x_t + v_t. \tag{3}$$

reveals massively excess rejections of the correct null on a Student-t test of $\mathsf{H}_0{:}\beta_1 = 0$. That test is derived under the assumption that the observations are independent, identically distributed (IID), and over-rejection is due to serious underestimation of the estimated coefficient standard error, rather than really high correlations *per se*. The residuals, $\widehat{v}_t$, of an estimated equation like (3) are almost bound to be strongly positively autocorrelated because y_t is, and x_t cannot capture that phenomenon. Herman Wold (1952) (see [109]) had proved that least-squares conventionally-calculated estimated coefficient standard errors would be badly downward biased when there is strong positive residual serial correlation. In turn, that would inflate calculated t-statistics and lead to excess rejections of correct null hypotheses.

The problem of spurious rejection actually gets worse in larger samples even though unrelated random walks should wander apart: at $T = 1000$ in (3), using conventional

[§] Although Yule (1897) (see [110]) had discussed 'spurious' correlations, where two variables were apparently related because each was related to a third, that research seems to have been forgotten, and 'nonsense correlations' are sometimes called 'spurious' correlations.

t-tests assuming stationarity, the null hypothesis H_0: $\beta_1 = 0$ is incorrectly rejected 90% of the time at a 1% significance level; that is, the actual rejection percentage is 90% and not 1% as one might mistakenly believe from the use of the theoretical t-distribution. To explain high correlations, rather than over-rejection Yule found that when both variables were $\mathsf{I}(2)$, generated by separately cumulating y_t and x_t, the distribution of their correlation coefficient was approximately U-shaped, with the most likely values under the null being ± 1.

During the late 1960s and early 1970s, the large empirical macroeconomic systems had been found by Phillip Cooper (1972) and Charles Nelson (1972) (see [15][82]) to forecast less accurately than a-theoretic scalar time-series models, such as autoregressions or autoregressive moving-averages of the form advocated by George Box and Gwilym Jenkins (1970) (see [9]). Those large simultaneous-equations macroeconomic systems focused on being derived from economic theory analyses (see e.g. Duesenberry, Klein, Fromm, and Kuh, 1965, [21]), so were not very dynamic.[¶] Building on these findings, Granger and Newbold (1974,1977) (see [45][46]) proposed that nonsense regressions could be relatively common in macroeconomics as judged by their high correlations (measured by the value of R^2, the squared coefficient of determination), yet exhibiting substantial residual serial correlation (as measured by the statistic formulated by James Durbin and Geoffrey Watson (1950,1951) [22][23], denoted DW, but already shown by Kenneth Wallis (1967) [104], to *understate* the extent of residual autocorrelation in dynamic models). Granger and Newbold (1974) (see [45]) argued that nonsense regressions were probably present when $\mathsf{R}^2 > DW$, a not uncommon finding in large macroeconomic models of the time. By stressing both the need to account for the dynamic properties of economic relationships and the importance of testing estimated models, [45] became widely cited.

Alexander Aitken (1935) (see [1]) had earlier proposed generalized least squares (GLS) to take account of residual heteroskedasticity or residual autocorrelation for regression models with 'strongly exogenous' regressors, namely explanatory variables that were valid conditioning variables and were not dependent on lagged values of the dependent variable (i.e., not Granger caused by the dependent variable: see Granger, 1969, [34], and the paper by Hendry in this volume). However, GLS could hardly solve the problem in (3) where the residual autocorrelation was generated by omitting the non-stationary regressor y_{t-1}.

The alternative solution proposed by Granger and Newbold (1974) (see [45]) was to use first differences of variables in regressions like (3), rather than levels, an idea Hooker had explored earlier. From (2), $\Delta y_t = y_t - y_{t-1} = e_t$ and $\Delta x_t = u_t$, so differencing reduces these variables to a non-integrated form (denoted $\mathsf{I}(0)$) sustaining conventional statistical inference. Thus, they postulated that nonsense regressions could be eliminated by instead estimating relationships like (for some error term w_t):

$$\Delta y_t = \lambda_0 + \lambda_1 \Delta x_t + w_t. \tag{4}$$

When the original variables are $\mathsf{I}(1)$, their differences are $\mathsf{I}(0)$, so nonsense-regression problems from unit roots in data generation processes (DGPs) like (2) should be removed.

[¶]Much later, Hendry and Jean-François Richard (1982) (see [64]) showed that choosing the formulations of models by dynamic simulation falsely suggested such specifications performed better.

Although differencing also attenuates large positive residual serial correlation, as shown by Donald Cochrane and Guy Orcutt (1949) (see [84]), only data can be differenced, not equations. Differencing (3) would generate a negative moving average process for $\{w_t\}$, whereas when $\beta_1 = 0$ in (3), then $\lambda_0 = \lambda_1 = 0$ in (4) and $w_t = e_t$ so is actually white noise. Moreover, if (4) were a valid solution, and $\{x_t\}$ was weakly exogenous (see Engle, Hendry and Richard, 1983, [25]) then the Bradford Smith approach of formulating the nesting equation:

$$y_t = \gamma_0 + \gamma_1 x_t + \gamma_2 x_{t-1} + \gamma_3 y_{t-1} + w_t. \tag{5}$$

where $\gamma_0 = \gamma_1 = \gamma_2 = 0$ and $\gamma_3 = 1$ when (2) is the DGP should also work (although the legitimacy of conventional inference may be in doubt as we will see in §4).

Moreover, when the null is not true because $\beta_1 \neq 0$, then an estimate of λ_1 in (4) could be very far from β_1—which is precisely the problem Hooker found comparing estimates in levels and differences in 1901! Nevertheless, when it came to forecasting, Granger and Newbold (1974) [45] felt that it was important to use differenced-data dynamic models rather than systems of simultaneous-equations in levels, an issue discussed by Michael Clements in this volume and one we return to below.

Granger and Newbold (1977) (see [46]) then proposed a 'time-series approach to econometric model building', emphasizing their 1974 critique of nonsense regressions, and their solution of analyzing only data in differences. They also showed that applying to (3) GLS-based residual serial-correlation corrections (wrongly attributed to Cochrane and Orcutt) did not resolve the nonsense-regression problem. Clive felt he had amassed strong arguments for this approach, further supported by the findings that so-called 'naive' forecasting devices like a random walk often outperformed the large macroeconomic models. However, all was not what it seemed: indeed, much remained to be discovered at this stage.

Clive was taken aback by the critique in Hendry (1977) (see [54]) to the effect that 'nonsense regressions were a nonsense problem', easily resolved for relationships between levels by including appropriate lags of the regressors. He had expected economists to criticize his approach, but not time-series econometricians, and anyway he did not at first accept the validity of the critique. To explain the issues, and what they precipated, we turn to the origins and formulations of equilibrium-correction models.

3. From equilibrium correction to cointegration

A large control theory literature existed by the early 1950s, applicable to many situations from running chemical plants to stabilizing moving mechanisms. Bill Phillips (1954,1957) (see [88][89]) discussed its application to economies, proposing roles for derivative, proportional and integral control mechanisms to stabilize economic fluctuations based on feedbacks of departures from the planned path. Denis Sargan (1964) (see [98]) then formulated an important variant of (5), which transpired to be related to derivative and proportional control mechanisms. Transform (5) to Sargan's formulation with $\epsilon_t \sim \mathsf{IN}[0, \sigma_\epsilon^2]$ as:

$$\Delta y_t = \gamma_0 + \gamma_1 \Delta x_t + (\gamma_2 + \gamma_1) x_{t-1} + (\gamma_3 - 1) y_{t-1} + \epsilon_t. \tag{6}$$

Provided $\gamma_3 \neq 1$, the lagged terms in (6) can be combined using $\kappa_1 = (\gamma_1 + \gamma_2)/(1 - \gamma_3)$ as:

$$\Delta y_t = \gamma_0 + \gamma_1 \Delta x_t + (\gamma_3 - 1)(y_{t-1} - \kappa_1 x_{t-1}) + \epsilon_t \tag{7}$$

The derivative control is provided by $\gamma_1 \Delta x_t$, and the proportional control by $(\gamma_3 - 1)(y_{t-1} - \kappa_1 x_{t-1})$. In his model of the determination of wages in the UK, where y_t denotes the log of nominal wages and x_t is the log of prices, so $(y_t - x_t)$ measures real wages, Sargan (1964) [98] considered the 'homogeneous' case that $\gamma_1 + \gamma_2 = 1 - \gamma_3 \neq 0$, so $\kappa_1 = 1$. He also explicitly included a 'planned path', which we denote by $(y - x)^*_{t-1}$, towards which the model would equilibrate. The resulting formulation can be expressed as:

$$\Delta y_t = \gamma_0 + \gamma_1 \Delta x_t + (\gamma_3 - 1)\left(y_{t-1} - x_{t-1} - (y - x)^*_{t-1}\right) + \epsilon_t. \tag{8}$$

In (8), $(y - x)^*_{t-1}$ could represent productivity, so real wages would equilibrate to that: an updated example is provided in Castle and Hendry (2014) (see [11]). The third term in (8) was originally called an error-correction mechanism (ECM), in line with its control-theory origins, as it corrected past discrepancies between y_t and x_t, assumed to arise from past mistakes, but is now usually called an equilibrium-correction mechanism (EqCM) for reasons explained below.

Assuming (8) is a viable representation (we will consider issues of statistical inference in §4), several important features are revealed. First, since Δx_t and $(y_{t-1} - x_{t-1} - (y - x)^*_{t-1})$ are unlikely to be highly correlated, the estimate of γ_1 will be essentially unaffected by whether or not the equilibrium-correction mechanism (EqCM) is included in the model. Comparing (8) with (4) reveals the estimate of λ_1 is the short-run response of Δy_t to Δx_t. However, the long-run relation between y and x is κ_1 given in equation (7), which could differ radically from γ_1, especially when $\kappa_1 = 1$. The role of the EqCM is to ensure y_t and x_t do not drift apart, which would occur if $\gamma_1 \neq \kappa_1$ without the EqCM term in (8). This requires $-1 < (\gamma_3 - 1) < 0$. Second, Lawrence Klein (1953) (see [78]) had argued for the existence of constant 'great ratios' like consumers' expenditure to income, capital to output, etc., which in log form would be differentials like $(y_t - x_t)$. To economists, such stabilizing relationships seemed natural: for example, unless expenditure and income were closely related, savings would diverge. Formulations like (7) or (8) seemed to ensure divergence would not occur. Although it may seem anachronistic because they were published later, the papers by Hendry and Gordon Anderson (1977) [58] and James Davidson, Hendry, Frank Srba and Stephen Yeo (1978) [16] (usually known as DHSY, and pronounced 'daisy') were prepared well before Hendry (1977) [54] and both included what they called ECMs. Indeed, the former adopted a control theory approach to deriving their dynamic model with 'disequilibrium' adjustment between past levels, and as an indirect criticism of Granger and Newbold (1974) [45], (to quote) 'there are ways to achieve stationarity other than blanket differencing'.

These papers provided the basis for Hendry (1977) [54], who unknowingly used the same formulation as Bradford Smith and cited the ECM in Sargan (1964) [98] to argue that the 'nonsense regressions problem' was easily resolved. Instead of fitting static models like (3), use specifications like (5), or better still (7). However, the stationarity of EqCM log-ratios required that any non-stationarity in the two series would have to cancel in the differential—which at that time, Clive Granger believed could not occur. When each time series was driven by an $\mathsf{I}(1)$ process, they would wander, and hence could not stay connected. Hendry's critique provoked Clive to try and formally disprove the possibility of substantive relations between the levels of $\mathsf{I}(1)$ variables: reconciling the divergent views of economists and statisticians would lead to cointegration. But before that, the story takes another twist.

4. The intrusion of non-standard unit-root distributions

Analyses of potentially non-stationary autoregressive processes have a distinguished pedigree including those by John White (1958,1959) (see [106][107]) for both random walks and explosive roots. [106] also derived the Laplace transform of the denominator and numerator in the t-statistic discussed above. Wayne Fuller (1976) (see [30]) demonstrated the need for different critical values when testing null hypotheses involving unit roots, followed up by David Dickey and Fuller (1979,1981) (see [18][19]) and Gwyn Aneuryn-Evans and Gene Savin (1981,1984) (see [3][4]) although those papers post date the debate in the previous section. The required critical values depended on whether the DGP and/or the model included deterministic terms like constants, dummy variables and/or trends, and how they are modeled, making for a rather complicated testing problem, later clarified (see e.g., Anders Rahbek and Bent Nielsen, 2000, [83]). Nevertheless, even before 1975, it was known that non-standard distributions occurred when estimating models of processes with unit roots.

Initially, this complication was ignored by those building EqCM models for $\mathsf{I}(1)$ data. While theoretical derivations needed to use tools like integrals of Brownian motion that in 1975 were new to econometricians, Monte Carlo simulation is the weak-theorist's friend, or perhaps not... Hendry and Grayham Mizon (1978) (see [60]) used a mainframe version of what is now called *PcNaive*[‖] to simulate the distributions of estimators and tests in models like (7) when the DGP had a unit root, and found a relatively negligible difference from conventional (limiting Normal) distributions. We were completely misled. In stationary processes, the value of the constant term is relatively innocuous, but not when there are unit roots, as then it can act as a drift term as in, say, a random walk. Unknowingly, Hendry and Mizon (1978) [60] had set a value for the constant term that entailed rapid growth in the levels (this was before it was easy to obtain graphs of the data, which would have revealed the problem). Kenneth West (1988) (see [105]) proved that in such a case, limiting Normal distributions did indeed result. In most settings, however, growth is not sufficiently fast to avoid using non-standard distributions. Inference procedures need to deliver correct decisions under both the null hypothesis of no relation and the alternative, and in unit-root processes, conventional critical values can over-reject the null when it is true. For models like (7), James Stock (1987) (see [102]) derived the limiting distributions

[‖]For Numerical Analysis of Instrumental Variables Estimators: see Jurgen Doornik and Hendry (2013) [20].

of parameter estimators under the alternative that the $\mathsf{I}(1)$ variables under analysis are genuinely linked.

Thus, the story twists again: were the findings of apparently 'well behaved' EqCMs an artefact of using incorrect critical values, and Granger was right that no such relations could exist? On the one hand, nonsense regressions seemed to depend on dynamic mis-specification, which could be avoided by including lagged values of all the variables in the relation. Transforming such regressions to equilibrium-correction models both provided a link to long-run economic equilibria that were essential to explain 'great ratios', and seemed a good characterization of $\mathsf{I}(1)$ data, whereas differencing all variables was not a viable solution for economics. On the other hand, estimation and testing in models of DGPs with unit roots required non-standard critical values, casting doubt on earlier inferences based on conventional significance levels. Moreover, single equation EqCMs simply assumed the unmodeled variables like x_t had the appropriate properties to generate the observed non-stationarity. If the model postulated for x_t also had an EqCM feedback with a coefficient μ different from κ_1 above, the solved long-run equilibrium of the bivariate process for y_t and x_t must be a point, contradicting the evidence that the data were $\mathsf{I}(1)$.

To look ahead, the deliberate creation and detection of nonsense regressions in Hendry (1980) (see [55]) summarizes the understanding at the time Granger began to formulate the concept of cointegration. However, the 'nonsense regressions problem' was only eventually clarified in the formal analyses by Peter Phillips (1986,1998) (see [90][92]). Also Yoon Park and Phillips (1988,1989) (see [85][86]) analyzed statistical inference when regressions were fitted to integrated processes, Christopher Sims, James Stock and Mark Watson (1990) [100] showed that conventional inference about the parameters in models like (5) was valid when the variables could be transformed such that the parameter was that of an $\mathsf{I}(0)$ combination, and Neil Ericsson and James MacKinnon (1999) [28] derived the distributions of 'error-correction' tests for cointegration based on EqCMs like (7), thereby providing appropriate critical values for tests of $(\gamma_3 - 1) = 0$ in $\mathsf{I}(1)$ processes. However, in the early 1980s, the analysis was about to move to a system formulation for vector $\mathsf{I}(1)$ processes, heralding the arrival of cointegration methods.

5. Cointegrated variables

Fortune favours the prepared mind when it comes to discoveries (or inventions), and in Clive's case the background preparation included Granger (1966) (see [33]), which showed that many economic time series had most of their spectral mass at the lowest frequencies of the spectrum. One possible explanation for that phenomenon was that the variables trended; another of course was that they were non-stationary from a unit root. Thus, Granger was already thinking about trends, then viewed as comprising frequencies with a period at least as long as the length of the series. Such a notion must have helped him towards (common) stochastic trends. His final published paper, Granger and Halbert White (2011) [51], again concerned trends.

Clive Granger did not accept the claims that the class of EqCM models like (8) were the end game. In his discussion of Hendry and Richard (1983) (see [65]), he questioned whether some of the purported equilibrium-correction terms were genuinely $\mathsf{I}(0)$: is the UK wage share over 1860–2011 graphed in Figure 2 (taken from Hendry (2015) [57]) actually $\mathsf{I}(0)$, and how could that be reliably established? The simplest test of the claim that the wage share in Figure 2 is not $\mathsf{I}(1)$ would be to use an augmented Dickey–Fuller (ADF) test, with sufficient lagged differences, to test the null of a unit root. Here, with two lags (which are in fact insignificant) the test rejects the null of a unit root at 5%. Although that rejected, the linear combinations of the log-level variables involved had assumed known coefficients, so the homogeneity hypotheses also needed to be tested.

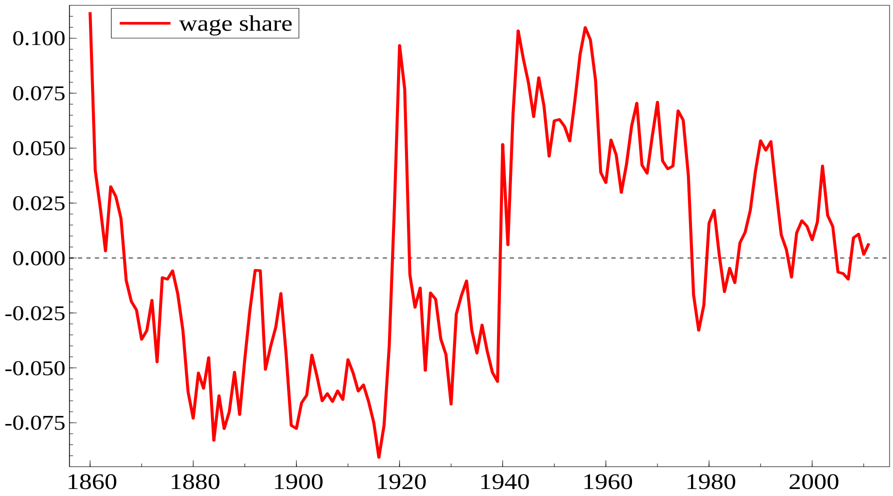

Figure 2: Wage share in the UK over 1860–2011.

5.1. Granger Representation Theorem

An important insight by Granger was that earlier analyses had *assumed* the data were $\mathsf{I}(1)$, somehow 'inherited' from unmodeled variables, rather than that property being endogenously generated by the process under analysis, as in (2). Instead, Granger commenced with processes akin to (1) with common errors to give a system representation in vector notation (denoted by bold) where $\mathbf{z}_t' = (y_t : x_t)$:

$$\Delta \mathbf{z}_t = \mathbf{C}\epsilon_t + \sum_{i=0}^{\infty} \mathbf{C}_i \Delta \epsilon_{t-i}. \tag{9}$$

By assuming $\mathbf{C}$ has reduced rank, (9) can be inverted to obtain the infinite order AR process:

$$\Delta \mathbf{z}_t = \mathbf{\Pi} \mathbf{z}_{t-1} + \sum_{i=0}^{\infty} \mathbf{\Gamma}_i \Delta \mathbf{z}_{t-i}. \tag{10}$$

with $\mathbf{\Pi}$ of reduced rank and $\mathbf{\Pi C} = \mathbf{C\Pi} = \mathbf{0}$. This is the Granger representation theorem, which enables transition from the moving average representation to the autoregressive representation (and back again) when there are $\mathsf{I}(1)$ variables in the system, although in practice the lag length is truncated.

While Granger's starting point was the moving average representation, the more frequent use of the Granger Representation Theorem is to derive the moving average representation from the autoregressive respresentation assuming $\mathbf{\Pi}$ has reduced rank. Given the symmetry, both directions of proof are feasible, but it is notable that the more common direction is for the converse to Granger's original proof (see Engle and Granger, 1987, p.255, [24]). Hence, we explore cointegration commencing from the autoregressive representation, before noting Granger's original theorem in (19) and (20).

Let us assume that y_t is generated by (7) and the equation for x_t is also of the form:

$$\Delta x_t = \theta_0 + \theta_1 \Delta x_{t-1} + \theta_2 \left(y_{t-1} - \kappa_1 x_{t-1}\right) + \eta_t, \quad \text{where} \quad \eta_t \sim \mathsf{IN}\left[0, \sigma^2_\eta\right] \tag{11}$$

with $|\theta_1| < 1$ and $\theta_2 \neq 0$, so the *same* feedback as in (7) occurs in **both** equations, then the data must be $\mathsf{I}(1)$. That this occurs is most easily seen by looking at the simplest special cases of (7) and (11) written as:

$$\Delta y_t = \phi_0 + \phi_1 (y_{t-1} - \kappa_1 x_{t-1}) + \epsilon_t \tag{12}$$

$$\Delta x_t = \theta_0 + \theta_2 \left(y_{t-1} - \kappa_1 x_{t-1}\right) + \eta_t \tag{13}$$

Then the combination of θ_2 times Δy_t minus ϕ_1 times Δx_t eliminates the EqCM:

$$\theta_2 \Delta y_t - \phi_1 \Delta x_t = (\theta_2 \phi_0 - \phi_1 \theta_0) + \theta_2 \epsilon_t - \phi_1 \eta_t \tag{14}$$

so is a random walk with drift, whereas y_t minus $\kappa_1 x_t$ delivers:

$$y_t - \kappa_1 x_t = (\phi_0 - \kappa_1 \theta_0) + (1 + \phi_1 - \kappa_1 \theta_2)\left(y_{t-1} - \kappa_1 x_{t-1}\right) + \epsilon_t - \eta_t \tag{15}$$

which is a stationary autoregression when $|1 + \phi_1 - \kappa_1 \theta_2| < 1$, so that condition is also required. Consequently, from (15) there is one $\mathsf{I}(0)$ relation, and from (14), one $\mathsf{I}(1)$.

The long-run relation must be a trajectory (rather than a point) along which $(y - \kappa_1 x - \kappa_0) = 0$ on average, where $\kappa_0 = (\phi_0 - \kappa_1 \theta_0) / (\kappa_1 \theta_2 - \phi_1)$ is the long-run equilibrium mean (dependent on the units of measurement of y_t and x_t). Suddenly, $\mathsf{I}(1)$ is not needed as a separate postulate: if there are fewer EqCM feedbacks in the levels (here one) than variables (here two), the process cannot be stationary. Nevertheless, the feedback term $(y_{t-1} - \kappa_1 x_{t-1})$ must be stationary (subject to some parametric restrictions) and consequently eliminates one unit root. As Granger records in both Phillips (1997) and Granger (2004) (see [91] and [40]), he tried to prove that linear combinations of $\mathsf{I}(1)$ variables like $(y_t - \kappa_1 x_t)$ must still be $\mathsf{I}(1)$, so EqCMs were not a viable model class. Instead, he established conditions under which co-integration would occur, namely some linear combinations of $\mathsf{I}(1)$ variables were $\mathsf{I}(0)$, or more generally, of a lower order of integration than the original variables.

Granger (1981) (see [36]) provided the first analysis. When the dependent variables of the bivariate representation are the $\mathsf{I}(0)$ first differences, Δy_t and Δx_t, of the $\mathsf{I}(1)$ variables y_t and x_t, where the regressor variables include (say) Δx_t, Δx_{t-1}, Δy_{t-1}, y_{t-1} and x_{t-1}, of which (6) and (11) are special cases, then the equations are balanced if and only if a linear combination of y_{t-1} and x_{t-1} is $\mathsf{I}(0)$. If there is a combination like $(y_t - \kappa_1 x_t)$ that is $\mathsf{I}(0)$, then both sides of the equations are stationary, and the formulations are internally consistent. If that combination is $\mathsf{I}(1)$, however, then either the coefficients of the levels variables must be zero, or at least one of Δy_t and Δx_t become $\mathsf{I}(1)$, contradicting the starting assumptions. Writing equations (7) and (11) as a system with $\gamma_4 = \gamma_3 - 1$:

$$\begin{pmatrix} 1 & -\gamma_1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \Delta y_t \\ \Delta x_t \end{pmatrix} = \begin{pmatrix} \gamma_0 \\ \theta_0 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & \theta_1 \end{pmatrix} \begin{pmatrix} \Delta y_{t-1} \\ \Delta x_{t-1} \end{pmatrix} + \begin{pmatrix} \gamma_4 \\ \theta_2 \end{pmatrix} (y_{t-1} - \kappa_1 x_{t-1}) + \begin{pmatrix} \epsilon_t \\ \eta_t \end{pmatrix} \tag{16}$$

then (16) generalizes (2) in four ways by including: (*i*) a contemporaneous relation in the first equation; (*ii*) potentially non-zero intercepts; (*iii*) lagged reactions to the first differences; and most importantly, (*iv*) the same levels feedback in both equations. When y_t and x_t are $\mathsf{I}(1)$ and $(y_{t-1} - \kappa_1 x_{t-1})$ is $\mathsf{I}(0)$, then the representation in (16) is $\mathsf{I}(0)$ and conventional statistical inference applies. However, to establish that a linear combination of $\mathsf{I}(1)$ variables is $\mathsf{I}(0)$ involves non-standard distributions.

In vector notation (16) can be written as:

$$\mathbf{\Gamma}_0 \Delta \mathbf{z}_t = \delta_0 + \mathbf{\Gamma}_1 \Delta \mathbf{z}_{t-1} + \alpha \left(\beta' \mathbf{z}_{t-1}\right) + \xi_t \quad \text{where} \quad \xi_t \sim \mathsf{IN}_2\left[\mathbf{0}, \mathbf{\Sigma}_\xi\right] \tag{17}$$

with $\xi_t' = (\epsilon_t : \eta_t)$. In (17), the coefficient matrix for the lagged levels $\mathbf{z}_{t-1}$ has been written as a singular product of $\alpha = (\gamma_4 : \theta_2)'$ and $\beta' = (1 : -\kappa_1)$ which are 2×1 and 1×2 so both are (at most) of rank 1. A stationary linear combination of two $\mathsf{I}(1)$ variables is a special case, but when it holds, such a pair of level variables is now called cointegrated. Thus, the $\mathsf{I}(1)$ component must be common to y_t and x_t (Hooker's 'common trends') and it must cancel to achieve a balanced formulation. In general, if there are n $\mathsf{I}(1)$ levels variables, $\mathbf{z}_t$, connected by $r < n$ feedbacks $\beta' \mathbf{z}_{t-1}$, then the system generates both $n - r$ integrated variables and r cointegrated combinations. For simplicity of notation, we set $\mathbf{\Gamma}_0 = \mathbf{I}$, noting that it is non-singular, so rewrite (17) as the n-dimensional vector process:

$$\Delta \mathbf{z}_t = \pi_0 + \mathbf{\Pi}_0 \Delta \mathbf{z}_{t-1} + \mathbf{\Pi}_1 \mathbf{z}_{t-1} + \zeta_t \quad \text{where} \quad \zeta_t \sim \mathsf{IN}_n\left[\mathbf{0}, \mathbf{\Omega}_\zeta\right] \tag{18}$$

When $\mathbf{\Pi}_0$ has a value consistent with $\mathbf{z}_t$ being either $\mathsf{I}(0)$ or $\mathsf{I}(1)$ as required for the following special cases, then the rank of the $n \times n$ matrix $\mathbf{\Pi}_1$ determines the properties of the data. When rank$(\mathbf{\Pi}_1) = n$, the process in (18) must be stationary, because the change in every variable is related to its level. This was implicitly the assumption of earlier EqCM equations like DHSY, although as discussed below, it was later established that many inferences remain valid using conventional critical values even when the system is $\mathsf{I}(1)$. Conversely, when rank$(\mathbf{\Pi}_1) = 0$, (18) describes a vector autoregression (VAR) in $\Delta \mathbf{z}_t$ so all variables are $\mathsf{I}(1)$. In between, when rank$(\mathbf{\Pi}_1) = r < n$, then $\mathbf{z}_t$ is $\mathsf{I}(1)$ and there are r $\mathsf{I}(0)$ linear combinations $\beta' \mathbf{z}_t$, where β is $n \times r$ of rank r. In that case, as explained in a series of papers starting with Søren Johansen (1988), and his book Johansen (1995) (see [69][71]), $\mathbf{\Pi}_1 = \alpha\beta'$ where α is $n \times r$ of rank r. This reduced-rank of the long-run matrix of the dynamics is what determines cointegration. Moreover, α represents the 'strength' of the feedbacks of the various cointegrating combinations onto each change. However, the role of $\mathbf{\Pi}_0\Delta\mathbf{z}_{t-1}$ in (18) is not innocuous, as we discuss in §5.2.

Sticking to the case when $\Delta\mathbf{z}_t$ is a vector of n stationary $\mathsf{I}(0)$ variables and setting $\mathsf{E}[\Delta\mathbf{z}_t] = \mathbf{0}\ \forall t$ for simplicity, from the famous decomposition theorem in Herman Wold (1938) (see [108]), we have the moving-average representation:

$$\Delta\mathbf{z}_t = \sum_{i=0}^{\infty} \mathbf{R}_i \epsilon_{t-i} \quad \text{where} \quad \epsilon_t \sim \mathsf{IID}_n\left[\mathbf{0}, \mathbf{\Omega}_\epsilon\right] \tag{19}$$

with $\mathbf{R}_0 = \mathbf{I}_n$ and $\sum_{i=0}^{s} \mathbf{R}_i = \mathbf{R}(1)$ has rank $n - r$, truncation the lag length to s. The Granger representation theorem, in the direction that Granger originally solved, states that (19) can be inverted to the vector autoregressive formulation:

$$\Delta\mathbf{z}_t = \sum_{i=1}^{k} \mathbf{A}_i \Delta\mathbf{z}_{t-i} + \alpha\beta'\mathbf{z}_{t-1} + \epsilon_t \tag{20}$$

($k = \infty$ but in practice the lag length is truncated) where α and β' are of rank r as in (18), delivering a vector EqCM. Johsen (1988) [69] provides a clear general formulation, so the representation is often called 'Granger–Johansen'.

Clive linked cointegration with Granger causality in Granger (1986) (see [37]) by showing that if two series are cointegrated then at least one of them must cause the other, tying together two of his main ideas. Indeed, a non-zero reduced rank $n > r > 0$ of $\mathbf{\Pi}_1$ in (18) entails that some EqCM(s) must enter at least one equation, so ensuring both cointegration and Granger causality for those variables.

Granger (1981) (see [36]) precipitated a vast literature, although he did not suggest any statistical tests for cointegration to make the notion operational. His first proposal was in Granger and Andrew Weiss (1983) (see [50]), generalizing previous unit-root tests and extending cointegration to more than a bivariate framework. Engle and Granger (1987) (see [24]) developed a two-step estimator for the parameters of cointegrated relationships (also see Anindya Banerjee, Juan Dolado, Hendry and Gregor Smith, 1986, [5]), closely followed by many approaches including full maximum likelihood methods (see e.g., Johansen, 1988, [69]).

An earlier literature had considered reduced-rank conditions both for multiple regression models (see Maurice Bartlett, 1938, [7]), and when estimating simultaneous equations systems (see e.g., Ted Anderson and Herman Rubin, 1949, [2]), then later for multivariate statistical models (see e.g., Box and George Tiao, 1977, [10]). In such settings, eigenvalues play a key role, and indeed the eigenvalues of the companion matrix also determine the dynamic properties of (18) as discussed in §5.2. However, the eigenvalues can be real or complex, making statistical derivations complicated. The beauty of the formulation in Johansen (1988) [69] is to construct the analysis such that the eigenvalues must be real and lie between zero and unity, facilitating the derivations of the limiting distributions of their estimators. Moreover, using the [69] formulation will enable a more formal statement of the conditions that will induce an I(2) process, as discussed in §5.2.

An alternative interpretation of why time series are cointegrated is by the cancellation of their 'common stochastic trends': see e.g., Stock and Watson, 1988, [103]. That approach focuses on α in (20) rather than β. Let $\alpha_\perp$ be the $n \times (n-r)$ matrix such that $\alpha'_\perp \alpha = \mathbf{0}$ and $(\alpha_\perp : \alpha)$ is full rank, where α is the adjustment matrix in (20). Jesus Gonzalo and Granger (1995) (see [32]) propose estimating the common stochastic trends by the linear combination $\alpha'_\perp \mathbf{z}_t$, since $\alpha'_\perp \alpha\beta' = \mathbf{0}$, thereby eliminating the cointegration vector. However, Johansen (1995, p.62, exercise 4.3) [72] gives an example of a model with two lags in which $\alpha'_\perp \mathbf{z}_t$ is stationary with $\alpha_\perp = \beta$. So while the approach is popular, one must be cautious of estimating common trends by $\alpha'_\perp \mathbf{z}_t$ because $\alpha'_\perp \sum_{i=0}^{t} \epsilon_i$ is eliminated by $\alpha'_\perp \beta_\perp = \mathbf{0}$. Instead, calculating $\alpha'_\perp \mathbf{\Gamma} \mathbf{z}_t$, where $\mathbf{\Gamma}$ is the coefficient on the lagged difference vector, delivers $\alpha'_\perp \sum_{i=0}^{t} \epsilon_i$ as $\alpha'_\perp \mathbf{\Gamma} \beta_\perp (\alpha'_\perp \mathbf{\Gamma} \beta_\perp)^{-1} \alpha'_\perp = \alpha'_\perp$.

On the one hand, testing for cointegration remains important to clarify the properties of models and the validity of inference in empirical estimates. On the other hand, the implications of the Granger Representation Theorem for economics suggest that economies are high-dimensional integrated-cointegrated systems, albeit subject to evolution and shifts, as most economic agents use fewer decision variables (such as bank balances, incomes, and wealth) than the huge number of decisions they make, inducing reduced rank in the equilibrium correction feedbacks. Thus, taking economic data as being generated by at least an integrated-cointegrated system seems a better starting point for empirical analyses than assuming stationarity.

5.2. Doubly integrated processes

Methods for handling cointegration in doubly integrated data have been extensively investigated: among many others, see Johansen (1992,1995) [70][73], Katarina Juselius (2006) [76], Rahbek, Hans-Christian Kongsted and Clara Jørgensen, (1999) [96], and Paulo Paruolo and Rahbek (1999) [87]. Granger's counter critique of Hendry and Richard (1983) [65] questioned whether their 'error-correction' term, based on theory and not tested for stationarity, was indeed I(0). For the much studied UK demand for money relation, his complaint has transpired to be well founded, as the EqCM was still I(1), having cointegrated from I(2) to I(1), but still needing combinations with differences of the I(2) variables (making them I(1)) to finally become I(0). To demonstrate, assume that $(1 + \phi_1 - \kappa_1\theta_2) = 1$ in (15), such that $y_t - \kappa_1 x_t$ is I(1). Then (12) and (13) must imply that y_t and x_t are I(2). Extensions to multicointegration, where the cumulated sum of the stationary linear combinations of a vector integrated series is cointegrated with itself, links stocks and flows, as in Granger and Tae-Hwy Lee (1989,1991) (see [43][44]) and Tom Engsted and Niels Haldrup (1999) [27].

The accuracy in finite samples of critical values for testing cointegration depends on how close the I(1) process is to I(2), so a number of 'correction' approaches have been proposed. These include Bartlett corrections (see Johansen, 2002, [75]), and bootstrap-based critical values (see e.g., Giuseppe Cavaliere, Rahbek and Robert Taylor, 2012, [12]).

5.3. Other developments

Generalizations in many directions soon followed, including among others, to non-stationary seasonal processes (see Svend Hylleberg, Engle, Granger and Byung Sam Yoo, 1990, [68]), non-linear cointegration (see e.g., Granger, 1993, [38], and Alvaro Escribano and Gerard Pfann, 1998, [29]) and stochastic unit roots (see Granger and Norman Swanson, 1997, [49]).

Cointegration is basically a linear concept, and the classical assumption in empirical work has been that drift towards the equilibrium, postulated by models with cointegrated variables, is symmetric: the strength of attraction is a linear function of the distance of the system from the equilibrium. Granger has also investigated non-linear cointegration, loosening the symmetry assumption. This is sometimes necessary in macroeconomics: planned inventories and orders, for example, help to smooth production so have asymmetric effects as in Granger and Lee (1989), Granger and Swanson (1996), and Granger (1996) (see [43] [48] and [39]), as well as Frédérique Bec and Rahbek (2004) and Dennis Kristensen and Rahbek (2013) (see [8] and [79]). The paper by Timo Teräsvirta in this volume discusses Granger's contributions to understanding non-linear relationships.

Forecasting in cointegrated systems has also attracted a lot of interest from Engle and Yoo (1987) (see [26]) onwards, including Michael Clements and Hendry (1995,1999) (see [13][14]) as well as in threshold cointegrated systems by (e.g.) Jan De Gooijer and Antoni Vidiella-i-Anguera (2004) [17]. Optimal forecasts of cointegrated variables should also be cointegrated, helping to improve long-term forecasting. However, forecast failure revealed two problems with standard linear cointegration analyses.

First, other forms of non-stationarity needed to be tackled jointly with cancelling unit roots, and co-breaking offered one possibility for doing so (see Hendry and Michael Massmann, 2007, [59]). Co-breaking is analogous to cointegration by cancelling location shifts across variables, rather than cancelling common trends. Like cointegration, it leads to only a subset of stable relations (the ones which co-break) with the remainder still exhibiting shifts. In both cases, differencing provides a 'solution' for the remaining variables in that it removes unit roots for integrated processes, or converts location shifts to impulses.

Secondly, by construction, equilibrium-correction representations converge back to the imposed underlying equilibrium, but do so even if the actual equilibrium has shifted. This problem clarified the real distinction between equilibrium correction, and error correction, mechanisms: the latter would track the new location of the data even after shifts.** Generalizations to modeling breaks in cointegrated models are provided by Johansen, Rocco Mosconi and Nielsen (2000) and Peter Hansen (2003) among others (see [74] and [53]).

As discussed by Ryoko Ito in this volume, Granger pioneered research into fractionally-integrated processes with Roselyne Joyeux in 1980. In Granger (1980) (see [35]), he also showed that the aggregation of autoregressive processes generates such fractionally-

**Clive had in fact thought of using 'equilibrium correction' rather than 'error correction' in the title of Engle and Granger (1987) [24].

integrated processes. Given long time series data for many financial variables, conditional variances have been shown to be persistent, consistent with long memory in volatility. Granger demonstrated empirically that on decomposing a high-frequency return time series (such as a stock return) into the product of its absolute value and sign, only the former had long-memory: its autocorrelation function decays slowly, whereas the sign is almost unpredictable – see Granger and Zhuanxin Ding (1995), Granger and Chor-Yiu Sin (2000), and Tina Rydberg and Neil Shephard (2003) ([41], [47] and [97]). This started another fertile field, since the forecastability of the absolute-value series is interesting for evaluating financial risk.

6. Implications of cointegration for empirical analyses

The debate about econometric methods for modeling economic time series in the 1970s began with the need to reconcile the economists desire for relationships connecting the (log) levels of variables with the statisticians desire to difference data given their worries about unit roots in time series leading to nonsense regressions. Cointegration provided a valid way of incorporating economic theory about relationships in levels into dynamic econometric systems of endogenous variables in first differences, nesting the two approaches, a major generalization of the (forgotten) proposal by Bradford Smith (1926) [101]. The stationary linear combinations in levels are interpretable as long-run equilibrium relationships, defining steady states, with short-run adjustments determined by lagged differences. Granger threw immense light on this problem by his formulation of cointegration and its links to equilibrium correction, making his contribution one of the most important developments in time-series econometrics since the probability foundations by Trygve Haavelmo (1944) (see [52]).

General equilibrium analysis remains central to economic reasoning, and cointegration provides a statistical formulation of long-run economic relationships, and conversely when long-run relations genuinely exist, then their variables should be cointegrated. For example, cointegration implications can sometimes be derived from the optimizing behavior of economic agents subject to their budget constraints, or from intertemporal optimization plans such as lifetime savings for pension provision. However, some macroeconomic theories still need to be revamped to be relevant to integrated rather than stationary processes, and to take account of location shifts must move to modeling general disequilibria rather than general equilibria (see e.g., Hendry and Mizon, 2014, [61]).

Cointegration can also arise as the empirical implementation of physical laws: Pretis (2015) (see [93]) establishes the equivalence between a two-component energy-balance model and a cointegrated vector autoregression (CVAR) like (18). Indeed, in any observational discipline facing non-stationary time series, cointegration should be investigated as better representing any long-run relations that may be present, facilitating a more orthogonal parametrization, enabling clearer interpretations of the evidence through jointly analyzing short-run and long-run properties, and sustaining viable statistical inference.

7. Conclusion

Clive Granger's unraveling of cointegration and common trends and their key properties was a major advance, buttressed by many later important insights. As with other key intellectual developments, false paths and joint complications had to be overcome by new thinking to understand the interacting problems. Doing so, in turn opened the door to previously unsuspected issues and yet further advances. The wealth of citations, further insights, and empirical applications such as the successful cointegration models in Gunnar Bårdsen, Øyvind Eitrheim, Eilev Jansen, and Ragnar Nymoen (2005) and Juselius (2006) (see [6] and [76]), among many other empirical studies, bear witness to the fecundity of Granger's ideas. Hendry and Teräsvirta (2013) (see [66]) discuss Clive Granger as a person, his career and the honours he received, as well as explaining his many research contributions.

Acknowledgements

Financial support from the Robertson Foundation (award 9907422), Institute for New Economic Thinking (grant 20029822) and Statistics Norway through Research Council of Norway Grant 236935, are all gratefully acknowledged, as are helpful comments on an earlier draft from Vanessa Berenguer-Rico, Michael P. Clements, Ryoko Ito, Søren Johansen, Andrew B. Martinez, Bent Nielsen, Felix Pretis and Timo Teräsvirta.

References

[1] A. C. Aitken. On least square and linear combination of observations. *Proceedings of the Royal Society of Edinburgh*, 55:42–48, 1935.

[2] T. W. Anderson and H. Rubin. Estimation of the parameters of a single equation in a complete system of stochastic equations. *Annals of Mathematical Statistics*, 20:46–63, 1949.

[3] G. B. Aneuryn-Evans and N. E. Savin. Testing for unit roots: 1. *Econometrica*, 49:753–779, 1981.

[4] G. B Aneuryn-Evans and N. E. Savin. Testing for unit roots: 2. *Econometrica*, 52:1241–1269, 1984.

[5] A. Banerjee, J. J. Dolado, D. F. Hendry, and G. W. Smith. Exploring equilibrium relationships in econometrics through static models: Some Monte Carlo evidence. *Oxford Bulletin of Economics and Statistics*, 48:253–277, 1986.

[6] G. Bårdsen, Ø. Eitrheim, E. S. Jansen, and R. Nymoen. *The Econometrics of Macroeconomic Modelling.* Oxford University Press, Oxford, 2005.

[7] M. S. Bartlett. Further aspects of the theory of multiple regression. *Proceedings of the Cambridge Philosophical Society*, 34:33–40, 1938.

[8] F. Bec and A. C. Rahbek. Vector equilibrium correction models with non-linear discontinuous adjustments. *Econometrics Journal*, 7:628–651, 2004.

[9] G. E. P. Box and G. M. Jenkins. *Time Series Analysis, Forecasting and Control.* Holden-Day, San Francisco, 1976. First published, 1970.

[10] G. E. P. Box and G. C. Tiao. A canonical analysis of multiple time series. *Biometrika*, 64:355–365, 1977.

[11] J. L. Castle and D. F. Hendry. Semi-automatic non-linear model selection. In N. Haldrup, M. Meitz, and P. Saikkonen, editors, *Essays in Nonlinear Time Series Econometrics*, pages 163–197. Oxford University Press, Oxford, 2014.

[12] G. Cavaliere, A. C. Rahbek, and A. M. R. Taylor. Bootstrap determination of the co-integration rank in vector autoregressive models. *Econometrica*, 80:1721–1740, 2012.

[13] M. P. Clements and D. F. Hendry. Forecasting in cointegrated systems. *Journal of Applied Econometrics*, 10:127–146, 1995.

[14] M. P. Clements and D. F. Hendry. *Forecasting Non-stationary Economic Time Series.* MIT Press, Cambridge, Mass., 1999.

[15] J. P. Cooper. The predictive performance of quarterly econometric models of the United States. In B. G. Hickman, editor, *Econometric Models of Cyclical Behaviour*, number 36 in National Bureau of Economic Research Studies in Income and Wealth, pages 813–947. Columbia University Press, New York, 1972.

[16] J. E. H. Davidson, D. F. Hendry, F. Srba, and J. S. Yeo. Econometric modelling of the aggregate time-series relationship between consumers' expenditure and income in the United Kingdom. *Economic Journal*, 88:661–692, 1978.

[17] J. G. De Gooijer and A. Vidiella-i-Anguera. Forecasting threshold cointegrated systems. *International Journal of Forecasting*, 20(2):237 – 253, 2004.

[18] D. A. Dickey and W. A. Fuller. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74:427–431, 1979.

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

[20] J. A. Doornik and D. F. Hendry. *Interactive Monte Carlo Experimentation in Econometrics using PcNaive: OxMetrics 7.10: Volume IV.* Timberlake Consultants Press, London, 2013.

[21] J. S. Duesenberry, L. R. Klein, G. Fromm, and E. Kuh, editors. *Brookings Quarterly Econometric Model of the United States.* North-Holland Publishing Company, Amsterdam, 1965.

[22] J. Durbin and G. S. Watson. Testing for serial correlation in least squares regression I. *Biometrika*, 37:409–428, 1950.

[23] J. Durbin and G. S. Watson. Testing for serial correlation in least squares regression II. *Biometrika*, 38:159–178, 1951.

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

[25] R. F. Engle, D. F. Hendry, and J-F. Richard. Exogeneity. *Econometrica*, 51:277–304, 1983.

[26] R. F. Engle and B. S. Yoo. Forecasting and testing in co-integrated systems. *Journal of Econometrics*, 35:143–159, 1987.

[27] T. Engsted and N. Haldrup. Multicointegration in stock-flow models. *Oxford Bulletin of Economics and Statistics*, 61:237–254, 1999.

[28] N. R. Ericsson and J. G. MacKinnon. Distributions of error correction tests for cointegration. *Econometrics Journal*, 5:285–318, 2002.

[29] A. Escribano and G. A. Pfann. Nonlinear error correction, asymmetric adjustment and cointgration. *Economic Modelling*, 15:197–216, 1998.

[30] W. A. Fuller. *Introduction to Statistical Time Series.* John Wiley & Sons, New York, 1976.

[31] F. Galton. Co-relations and their measurement, chiefly from anthropometric data. *Proceedings of the Royal Society of London*, 45:135–145, 1889.

[32] J. Gonzalo and C. W. J. Granger. Estimation of common long memory components in cointegrated systems. *Journal of Business and Economic Statistics*, 13:27–35, 1995.

[33] C. W. J. Granger. The typical spectral shape of an economic variable. *Econometrica*, 34:150–161, 1966.

[34] C. W. J. Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37:424–438, 1969.

[35] C. W. J. Granger. Long memory relationships and the aggregation of dynamic models. *Journal of Econometrics*, 14:227–238, 1980.

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

[37] C. W. J. Granger. Developments in the study of cointegrated economic variables. *Oxford Bulletin of Economics and Statistics*, 48:213–228, 1986.

[38] C. W. J. Granger. Strategies for modelling nonlinear time series relationships. *Economic Record*, 60:233–238, 1993.

[39] C. W. J. Granger. Introducing nonlinearity into cointegration. *Revista de Econometria*, 16:25–36, 1996.

[40] C. W. J. Granger. Time series analysis, cointegration, and applications. In T. Frängsmyr, editor, *The Nobel Prizes 2003*, pages 360–366. Almqvist and Wiksell International, Stockholm, 2004.

[41] C. W. J. Granger and Z. Ding. Some properties of absolute return: An alternative measure of risk. *Annals of Economics and Statistics*, 40:67–91, 1995.

[42] C. W. J. Granger and R. Joyeux. An introduction to long memory time series models and fractional differencing. *Journal of Time Series Analysis*, 1:15–30, 1980.

[43] C. W. J. Granger and T-H. Lee. Investigation of production, sales and inventory relationships using multicointegration and non-symmetric error correction models. *Journal of Applied Econometrics*, 4:145–159, 1989.

[44] C. W. J. Granger and T-H. Lee. Multicointegration. In R. F. Engle and C. W. J. Granger, editors, *Long-Run Economic Relationships*, pages 179–190. Oxford University Press, Oxford, 1991.

[45] C. W. J. Granger and P. Newbold. Spurious regressions in econometrics. *Journal of Econometrics*, 2:111–120, 1974.

[46] C. W. J. Granger and P. Newbold. The time series approach to econometric model building. In Sims [99], pages 7–21.

[47] C. W. J. Granger and C.-Y Sin. Modelling the absolute returns of different stock indices: Exploring the forecastability of an alternative measure of risk. *Journal of Forecasting*, 19:277–298, 2000.

[48] C. W. J. Granger and N. R. Swanson. Further developments in the study of cointegrated variables. *Oxford Bulletin of Economics and Statistics*, 58:537–553, 1995.

[49] C. W. J. Granger and N. R. Swanson. An introduction to stochastic unit-root processes. 80:35–62, 1997.

[50] C. W. J. Granger and A. A. Weiss. Time series analysis of error-correction models. In S. Karlin, T. Amemiya, and L. A. Goodman, editors, *Studies in Econometrics, Time Series, and Multivariate Statistics*, pages 255–278. Academic Press, New York, 1983.

[51] C. W. J. Granger and H. White. Consideration of trends in time series. *Journal of Time Series Econometrics*, 3 (1):DOI: 10.2202/1941–1928.1092, 2011.

[52] T. Haavelmo. The probability approach in econometrics. *Econometrica*, 12:1–118, 1944. Supplement.

[53] P. R. Hansen. Structural changes in the cointegrated vector autoregressive model. 114:261–295, 2003.

[54] D. F. Hendry. On the time series approach to econometric model building. In Sims [99], pages 183–202.

[55] D. F. Hendry. Econometrics: Alchemy or science? *Economica*, 47:387–406, 1980.

[56] D. F. Hendry. The Nobel Memorial Prize for Clive W.J. Granger. *Scandinavian Journal of Economics*, 106:187–213, 2004.

[57] D. F. Hendry. *Introductory Macro-econometrics: A New Approach.* Timberlake Consultants, London, 2015. http://www.timberlake.co.uk/macroeconometrics.html.

[58] D. F. Hendry and G. J. Anderson. Testing dynamic specification in small simultaneous systems: An application to a model of building society behaviour in the United Kingdom. In M. D. Intriligator, editor, *Frontiers in Quantitative Economics*, volume 3, pages 361–383. North Holland Publishing Company, Amsterdam, 1977.

[59] D. F. Hendry and M. Massmann. Co-breaking: Recent advances and a synopsis of the literature. *Journal of Business and Economic Statistics*, 25:33–51, 2007.

[60] D. F. Hendry and G. E. Mizon. 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, 1978.

[61] D. F. Hendry and G. E. Mizon. Unpredictability in economic analysis, econometric modeling and forecasting. *Journal of Econometrics*, 182:186–195, 2014.

[62] D. F. Hendry and M. S. Morgan, editors. *The Foundations of Econometric Analysis.* Cambridge University Press, Cambridge, 1995.

[63] D. F. Hendry and F. Pretis. Anthropogenic influences on atmospheric CO2. In R. Fouquet, editor, *Handbook on Energy and Climate Change*, pages 287–326. Edward Elgar, Cheltenham, 2013.

[64] D. F. Hendry and J-F. Richard. On the formulation of empirical models in dynamic econometrics. *Journal of Econometrics*, 20:3–33, 1982.

[65] D. F. Hendry and J-F. Richard. The econometric analysis of economic time series (with discussion). *International Statistical Review*, 51:111–163, 1983.

[66] D. F. Hendry and T. Teräsvirta. Sir Clive William John Granger 1934–2009. *Biographical Memoirs of Fellows of the British Academy*, XII:451–469, 2013.

[67] R. H. Hooker. Correlation of the marriage rate with trade. *Journal of the Royal Statistical Society*, 64:485–492, 1901.

[68] S. Hylleberg, R. F. Engle, C. W. J. Granger, and B. S. Yoo. Seasonal integration and cointegration. *Journal of Econometrics*, 44:215–238, 1990.

[69] S. Johansen. Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, 12:231–254, 1988. Reprinted in R.F. Engle and C.W.J. Granger (eds), *Long-Run Economic Relationships*, Oxford: Oxford University Press, 1991, 131–52.

[70] S. Johansen. A representation of vector autoregressive processes integrated of order 2. *Econometric Theory*, 8:188–202, 1992.

[71] S. Johansen. *Likelihood-based Inference in Cointegrated Vector Autoregressive Models*. Oxford University Press, Oxford, 1995.

[72] S. Johansen. The role of ancillarity in inference for nonstationary variables. *Economic Journal*, 105:302–320, 1995.

[73] S. Johansen. A statistical analysis of cointegration for I(2) variables. *Econometric Theory*, 11:25–59, 1995.

[74] S. Johansen, R. Mosconi, and B. Nielsen. Cointegration analysis in the presence of structural breaks in the deterministic trend. *Econometrics Journal*, 3:216–249, 2000.

[75] S.J.Johansen. A Small Sample Correction for the Test of Cointegrating Rank in the Vector Autoregressive Model. *Econometrica*, 70:1929–1961, 2002.

[76] K. Juselius. *The Cointegrated VAR Model: Methodology and Applications*. Oxford University Press, Oxford, 2006.

[77] J. L. Klein. *Statistical Visions in Time: A History of Time Series Analysis 1662-1938*. Cambridge University Press, Cambridge, 1997.

[78] L. R. Klein. *A Textbook of Econometrics*. Row, Peterson and Company, Evanston, Ill., 1953.

[79] D. Kristensen and A. C. Rahbek. Testing and inference in nonlinear cointegrating vector error correction models. *Econometric Theory*, 29:1238–1288, 2013.

[80] T. C. Mills. Bradford Smith: An econometrician decades ahead of his time. *Oxford Bulletin of Economics and Statistics*, 73:276–285, 2010.

[81] M. S. Morgan. *The History of Econometric Ideas*. Cambridge University Press, Cambridge, 1990.

[82] C. R. Nelson. The prediction performance of the FRB-MIT-PENN model of the US economy. *American Economic Review*, 62:902–917, 1972.

[83] B. Nielsen and A. Rahbek. Similarity issues in cointegration analysis. *Oxford Bulletin of Economics and Statistics*, 62:5–22, 2000.

[84] G. H. Orcutt and D. Cochrane. A sampling study of the merits of autoregressive and reduced form transformations in regression analysis. *Journal of the American Statistical Association*, 44:356–372, 1949.

[85] J. Y. Park and P. C. B. Phillips. Statistical inference in regressions with integrated processes. Part 1. *Econometric Theory*, 4:468–497, 1988.

[86] J. Y. Park and P. C. B. Phillips. Statistical inference in regressions with integrated processes. Part 2. *Econometric Theory*, 5:95–131, 1989.

[87] P. Paruolo and A. Rahbek. Weak exogeneity in I(2) systems. *Journal of Econometrics*, 93:281–308, 1999.

[88] A. W. H. Phillips. Stabilization policy in a closed economy. *Economic Journal*, 64:290–333, 1954.

[89] A. W. H. Phillips. Stabilization policy and the time form of lagged response. *Economic Journal*, 67:265–277, 1957.

[90] P. C. B. Phillips. Understanding spurious regressions in econometrics. *Journal of Econometrics*, 33:311–340, 1986.

[91] P. C. B. Phillips. The ET interview: Professor Clive Granger. *Econometric Theory*, 13:253–303, 1997.

[92] P. C. B. Phillips. New tools for understanding spurious regressions. *Econometrica*, 66:1299–1325, 1998.

[93] F. Pretis. Models of Climate Systems: The Equivalence of Two-Component Energy Balance Models and Cointegrated VARs. Working paper, 750, Economics Department, Oxford University.

[94] D. Qin. *The Formation of Econometrics: A Historical Perspective*. Clarendon Press, Oxford, 1993.

[95] D. Qin. *A History of Econometrics: The Reformation from the 1970s*. Clarendon Press, Oxford, 2013.

[96] A. Rahbek, H. C. Kongsted, and C. Jørgensen. Trend-stationarity in the I(2) cointegration model. *Journal of Econometrics*, 90:265–289, 1999.

[97] T. H. Rydberg and N. Shephard. Dynamics of trade-by-trade price movements: Decomposition and models. *Journal of Financial Econometrics*, 1:2–25, 2003.

[98] J. D. Sargan. Wages and prices in the United Kingdom: A study in econometric methodology (with discussion). In P. E. Hart, G. Mills, and J. K. Whitaker, editors, *Econometric Analysis for National Economic Planning*, volume 16 of *Colston Papers*, pages 25–63. Butterworth Co., London, 1964.

[99] C. A. Sims, editor. *New Methods in Business Cycle Research.* Federal Reserve Bank of Minneapolis, Minneapolis, 1977.

[100] C. A. Sims, J. H. Stock, and M. W. Watson. Inference in linear time series models with some unit roots. *Econometrica*, 58:113–144, 1990.

[101] B. B. Smith. Combining the advantages of first-difference and deviation-from-trend methods of correlating time series. *Journal of the American Statistical Association*, 21:55–59, 1926.

[102] J. H. Stock. Asymptotic properties of least squares estimators of cointegrating vectors. *Econometrica*, 55:1035–1056, 1987.

[103] J. H. Stock and M. W. Watson. Testing for common trends. *Journal of the American Statistical Association*, 83:1097–1107, 1988.

[104] K. F. Wallis. Lagged dependent variables and serially correlated errors: A reappraisal of three pass least squares. *Review of Economics and Statistics*, 69:555–567, 1967.

[105] K. D. West. Asymptotic normality when regressors have a unit root. *Econometrica*, 56:1397–1417, 1988.

[106] J. S. White. The limiting distribution of the serial correlation coefficient in the explosive case. *Annals of Mathematical Statistics*, 29:1188–1197, 1958.

[107] J. S. White. The limiting distribution of the serial correlation coefficient in the explosive case II. *Annals of Mathematical Statistics*, 30:831–834, 1959.

[108] H. O. A. Wold. *A Study in The Analysis of Stationary Time Series.* Almqvist and Wicksell, Stockholm, 1938.

[109] H. O. A. Wold. *Demand Analysis: A Study in Econometrics.* Almqvist and Wicksell, Stockholm, 1952.

[110] G. U. Yule. On the theory of correlation. *Journal of the Royal Statistical Society*, 60:812–838, 1897.

[111] G. U. Yule. 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, 1926. Reprinted in Hendry, D. F. and Morgan, M. S. (1995), *The Foundations of Econometric Analysis.* Cambridge: Cambridge University Press.