

CENTRE FOR ECONOMETRIC ANALYSIS
CEA@Cass



<http://www.cass.city.ac.uk/cea/index.html>

Cass Business School
Faculty of Finance
106 Bunhill Row
London EC1Y 8TZ

Co-breaking: Recent Advances and a Synopsis of the Literature

David F. Hendry and Michael Massmann

CEA@Cass Working Paper Series

WP-CEA-06-2006

Co-breaking: Recent Advances and a Synopsis of the Literature

David F. Hendry, Department of Economics,
University of Oxford, Oxford OX1 3UQ, UK
(david.hendry@economics.ox.ac.uk)

Michael Massmann, Department of Economics,
University of Bonn, 53113 Bonn, Germany
(michael.massmann@uni-bonn.de)

February 15, 2006

Abstract

This paper has two aims: first, to provide a synopsis of the literature on co-breaking which has developed in several, seemingly disconnected, strands. We establish a consistent terminology, collect theoretical results, delimit co-breaking to cointegration and common features, and review recent contributions to co-breaking regressions and the budding analysis of co-breaking rank. Secondly, we present new results in the field, in particular, on the importance of co-breaking for policy analysis, with special emphasis on impulse-response functions. Moreover, a new procedure for co-breaking rank testing is presented, evaluated by means of Monte Carlo experiments, and illustrated using UK macroeconomic data.

KEY WORDS: structural break, common feature, rank test, policy analysis

1 Introduction and selective historical overview

This section overviews much of the literature over the past three decades on modelling non-stationary time series. It is indicative of various themes and practices that have prevailed, and continue to prevail, in macroeconometrics.

Many macroeconomic time series are generated by non-stationary random processes since various moments of observed data are not time invariant. Examples include trending GDP, the aggregate price level, especially in hyperinflation periods, or time series affected by political, legislative or technological structural change. Of particular interest in the present study are non-stationarities due to breaks in the coefficients of the deterministic variables characterizing the random process, for these irregularities play a decisive rôle in economic policy analysis and economic forecasting. For instance, Barrell (2001) illustrates the frequency of breaks by six major episodes of change during the 1990s alone while Hendry (2000c) establishes that a major source of forecast failure seems to be breaks in deterministic variables, with other breaks being much less pernicious. See also Burns (1986), Pain and Britton (1992), Wallis (1989) or Stock and Watson (1996) for a discussion of data irregularities in the context of forecasting.

There has been considerable progress in the econometric modelling of non-stationarities. Two prominent approaches to mapping data to stationarity involve prior regression on a polynomial time trend, and the taking of first, or higher order, differences. The sequence of papers by Courakis (1978), Hendry and Mizon (1978) and Williams (1978) on econometric modelling practices at the Bank of England provides a snapshot of the debate on the pros and cons of differencing in the 1970s. Subsequently, the seminal paper by Nelson and Plosser (1982) on some long-range US macroeconomic time series further kindled

the controversy on whether macroeconomic time series are trend-stationary or difference-stationary, i.e., whether disturbances have a transitory or permanent effect on the level of the process. As a result, two important developments since the mid-1980s may be noted.

First, much research effort has gone into the development of test procedures that aim at differentiating between a unit root process and deterministic non-stationarity. The early contributions by Dickey and Fuller (1979, 1981) led to the eponymous Dickey–Fuller test which examines the null hypothesis of a unit root against the alternative of a constant deterministic trend. However, developments in the statistics literature, see Box and Tiao (1965, 1975) on intervention analysis and Gallant and Fuller (1973) on segmented polynomial regression models, motivated a more flexible specification of a model’s deterministic component. Specifically, Perron (1989, 1990) as well as Rappoport and Reichlin (1989) were instrumental in devising modelling and testing procedures that allowed for indicator variables to represent shifts in the model’s intercept or linear trend. The former author proposed to augment the Dickey–Fuller test by intervention dummies and thus test the null of a unit root against the alternative of an intercept and a linear trend, either or both of which may be subject to a structural break. Applying his tests to the Nelson-Plosser data, he could reject the unit-root hypothesis for most of the time series involved. This conclusion was indeed supported by Andreou and Spanos (2003), who re-estimated the models specified by Nelson and Plosser (1982) and Perron (1989), ensuring they are data congruent. Rappoport and Reichlin (1989) followed Gallant and Fuller (1973) in modelling the Nelson-Plosser data by an $I(0)$ model around a segmented trend rather than by a unit root model. The authors argue that the true generating process of the data might go unnoticed since, as they showed in a set of Monte Carlo experiments, the standard Dickey-Fuller test favours the null of a unit root over the alternative of a stationary model around a constant trend when the true process is subject to segmented trends.

The second noteworthy development in the modelling of non-stationarity has been the contribution of cointegration by Granger (1981), according to which a linear combination of unit-root processes may be integrated of a lower order. Engle and Granger (1987) subsequently operationalized cointegration, building on the equilibrium-correction representation of an autoregressive process introduced by Sargan (1964) and popularized by Davidson, Hendry, Srba and Yeo (1978). Establishing a relationship that is stable over time between several stochastic variables, even though its constituents are individually non-stationary, added a new dimension to econometric modelling. Now the mapping of the time series to stationarity could be achieved within a system, i.e., without resorting to purely exogenous deterministic factors and without losing information by differencing. The concept was subsequently extended to linear combinations that cancelled other types of stationary or non-stationary ‘common feature’, to use the term coined by Engle and Kozicki (1993). For instance, Kang (1990) considers a deterministic time trend as a putative common feature, while Engle and Kozicki (1993) provide examples of the feature being an ARCH component or a cycle. Indeed, co-breaking as in Hendry (1996) is another member of this last class, which we now consider in more detail.

1.1 Common features

The literature on common features may be broadly classified into two strands. The first concerns the estimation of relationships that cancel the common feature of interest under the assumption that there is a known number of such relationships, usually one. In contrast, the second strand deals with procedures that simultaneously test for the number of common features and estimate the corresponding linear relationships. The following two sub-sections deal with these approaches in turn. In accordance with standard terminology, they will be called the single equation and the system approach to common feature analysis.

1.1.1 Single equation analysis

This approach amounts to what Engle and Kozicki (1993) call a common feature regression. Generally, it proceeds in two steps. In the first step, the relationship between the variables believed to have a feature in common is estimated. The second step then investigates whether or not the first step residuals continue to display the feature of interest. The most prominent example of a common feature regression is the original ‘cointegrating regression’ of Engle and Granger (1987). In a similar development to the extension of the Dickey–Fuller test by deterministic components, or nuisance parameters, the residual-based cointegration test has also been augmented by dummy variables or made robust against structural breaks; see, for instance, Gregory and Hansen (1996) and Arranz and Escribano (2000), respectively.

Another example of a common feature regression is the so-called ‘co-trending regression’ suggested by Ogaki and Park (1998) and Kang (1990). In these studies, each of the time series of interest is assumed to be generated by a linear deterministic time trend, which is no longer displayed in a linear combination thereof. This setup was subsequently extended by Chapman and Ogaki (1993) to include structural breaks, where the intercept and linear trend of individual time series may be subject to an arbitrary number of breaks, whose number is reduced in the residuals of the estimated co-trending relationship. Chapman and Ogaki (1993), and subsequent authors, call the cancellation of breaks ‘co-trending’, since trend breaks and the trend itself are varieties on a continuum of deterministic terms. Empirical examples of what might more aptly be called a ‘co-breaking regression’ are provided by Hendry and Mizon (1998) and Clements and Hendry (1999) on the one hand, and Morana (2002) on the other, who use dummy variables and a stochastic break process based on a Markov chain, respectively, to model irregularities in the data, and examine whether they vanish in linear combinations of the series.

1.1.2 System analysis

A more general approach to common feature analysis was taken by Johansen (1988), who suggested a procedure for estimating cointegrating vectors as well as for determining their number in a system of equations, the latter now not being implicitly assumed known as in cointegrating regressions. Johansen (1988) used a vector version of the equilibrium correction model by Davidson *et al.* (1978), in which the number of cointegrating vectors is determined by the rank of a particular coefficient matrix, for a purely stochastic model. Since processes whose levels are solely determined by their initial values cannot describe all macroeconomic time series, Johansen (1994) added a constant and linear trend to the analysis. The standard rank test, like the Dickey–Fuller unit-root and Engle–Granger cointegration tests, turned out not to be similar with respect to deterministic terms but Doornik, Hendry and Nielsen (1998) and Nielsen and Rahbek (2000), building on Kiviet and Phillips (1992), argue that similarity may be obtained by augmenting the rank hypothesis to include the intercept or the linear trend parameter, depending on the setting. Tests for cointegrating rank that allow for structural breaks include Inoue (1999), who proposes a rank test when there is a single trend break under the alternative hypothesis, and Johansen, Mosconi and Nielsen (2000) who generalize the approach in Perron (1989) by employing a segmented trend model that allows for an arbitrary number of simultaneous shifts in the model’s intercepts and linear trends.

Using the terminology of Engle and Kozicki (1993), the concept of cointegrating rank may be generalized to that of ‘co-feature rank’. Similarly, since many macroeconomic variables appear subject to structural breaks, Hendry (1996) suggests the idea of ‘co-breaking rank’ as the number of linear combinations of the variables that no longer display the breaks. However, while co-breaking regressions are by now firmly established in the literature, co-breaking rank analysis is still in its infancy. In particular, only three procedures that aim to estimate the co-breaking rank appear to have been suggested in the literature: First, Bierens (2000) develops a system test determining how many relationships no longer contain a non-linear trend function; secondly, Krolzig and Toro (2002) estimate the number of linear

combinations that cancel deterministic components in a vector autoregressive model; and Hatanaka and Yamada (2003) suggest a parametric procedure for testing for the co-breaking as well as cointegrating rank in a segmented trend model.

1.2 Co-breaking

The purpose of this paper is twofold. First, it aims to provide a synopsis of the literature, which has been developing in several, seemingly disconnected, strands. This includes establishing a consistent terminology, collecting theoretical results, delimiting co-breaking to cointegration and common features, and reviewing recent contributions to co-breaking regressions and the budding analysis of co-breaking rank. Its second purpose is to present new results in the field. In particular, the importance of co-breaking for policy analysis takes centre stage, with special emphasis on impulse-response functions. Moreover, a new procedure for co-breaking rank testing is presented, evaluated by means of Monte Carlo experiments and illustrated using UK macroeconomic data.

The remainder of the paper is structured as follows: Section 2 outlines the theory of co-breaking, section 3 relates co-breaking to empirical econometric research, section 4 reviews the literature on co-breaking regressions before section 5 discusses the analysis of co-breaking rank. Finally, section 6 concludes.

2 The concept of co-breaking

Section 2.1 clarifies our terminology. In section 2.2, the concept of co-breaking is defined as a linear combination of time-varying parameters which is constant over time. This might be thought impossible in the case of frequent breaks in the series concerned: however, in some important special situations, co-breaking arises naturally. Section 2.3 discusses such cases, dealing with co-breaking in a cointegrated framework.

2.1 Terminology

Central to our analysis are location shifts of random variables, i.e., changes in the unconditional expectations of the non-integrated transformations of the variables. Since unconditional expectations are functions of deterministic variables – variables whose future values are known with certainty but the parameters of which could change – location shifts are represented by changes in the parameters of deterministic terms. To illustrate, consider the simplest location model of a variable y_t :

$$y_t = \alpha + u_t \quad (1)$$

with $\alpha \neq 0$ and where $\{u_t\}$ is $I(0)$ such that $E(u_t) = 0$. A location shift to α^* from time period T onwards induces:

$$y_{T+h} = \alpha^* + u_{T+h} \equiv \alpha\mu + u_{T+h} \quad (2)$$

where $\mu = \alpha^*/\alpha \neq 1$ is the shifted intercept for $h > 0$. Location shifts may be mimicked by other factors, such as mis-estimating or mis-specifying deterministic components in models. The simulation evidence in Hendry (2000c) confirms their pernicious effects on forecasts, as well as highlights the difficulty of detecting other forms of break, especially those associated with mean-zero changes, which have no effects on unconditional means.

To relate the notion of location shifts to econometric models, consider the following illustrative static bivariate system:

$$\begin{aligned} y_t &= \alpha + \beta z_t + v_t \\ z_t &= \mu_z + \varepsilon_{z,t} \end{aligned} \quad (3)$$

$$\begin{pmatrix} v_t \\ \varepsilon_{z,t} \end{pmatrix} \sim \text{IN}_2 \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_v^2 & 0 \\ 0 & \sigma_z^2 \end{pmatrix} \right]$$

where $(\alpha, \beta, \mu_z, \sigma_v^2, \sigma_z^2) \in \mathbb{R}^3 \times \mathbb{R}_+^2$. The first equation relating y_t to z_t is deemed to be causal, i.e. $\partial y_t / \partial z_t = \beta$, with z_t strongly exogenous for $(\alpha, \beta, \sigma_v^2)$; whereas the second equation for z_t is a policy process. Thus, the system in (3) consists of one structural and one policy equation. However, its specification does not directly entail how changes in any subset of parameters might affect the remainder, so additional assumptions are required. Note that (3) can be solved for its reduced form to yield:

$$\begin{aligned} y_t &= \mu_y + \varepsilon_{y,t} \\ z_t &= \mu_z + \varepsilon_{z,t} \end{aligned} \quad (4)$$

$$\begin{pmatrix} \varepsilon_{y,t} \\ \varepsilon_{z,t} \end{pmatrix} \sim \text{IN}_2 \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_y^2 & \rho \\ \rho & \sigma_z^2 \end{pmatrix} \right]$$

such that

$$\alpha = \mu_y - \beta \mu_z, \quad (5)$$

$$\beta \sigma_z^2 = \rho \text{ and } \sigma_y^2 = \sigma_v^2 + \beta \rho.$$

Using ∇ to denote a parameter change, we define the following terminology: a *location shift* refers to the reduced form in (4) and is defined as a change in the unconditional mean, e.g., $\nabla \mu_y \neq 0$ or $\nabla \mu_z \neq 0$. Regarding the structural system in (3), a *structural break* occurs when the parameters of the causal model change, here either $\nabla \beta \neq 0$ or $\nabla \alpha \neq 0$ or both. A *policy regime shift*, on the other hand, occurs when the parameters of the policy process change, here either $\nabla \mu_z \neq 0$ or $\nabla \sigma_z^2 \neq 0$ or both. Finally, a change where $\nabla \mu_y \neq 0$ is called a *target shift* when it is the result of a policy regime shift. Thus, both regime and target shifts are special cases of location shifts.

Structural invariance occurs when a policy regime shift does not entail a structural break, i.e., $\nabla \mu_z \neq 0$ or $\nabla \sigma_z^2 \neq 0$ leaves $\nabla \alpha = 0$, $\nabla \beta = 0$ and $\nabla \sigma_v^2 = 0$. In particular, contemporaneous mean *co-breaking* occurs when α is invariant to changes in μ_z because there is a proportionate change in μ_y . To see this, consider (5) where β does not change, such that $\nabla \alpha = 0$ entails that:

$$\nabla \mu_y = \beta \nabla \mu_z. \quad (6)$$

Put differently, co-breaking occurs when a regime shift causes a target shift of proportion β . In general, changes in α may be due to changes in any of μ_z , μ_y and β , since $\nabla \alpha = \nabla \mu_y - \beta \nabla \mu_z - \mu_z \nabla \beta$. It is thus conceivable that a regime shift is exactly off-set by a structural break in β , such that $\beta \nabla \mu_z + \mu_z \nabla \beta = 0$, implying that for $\nabla \alpha = 0$, there need be no shift in the mean of y_t .

Finally, to distinguish parametric changes in the unconditional expectation of a random variable from those in its expectation conditional upon its own history, the term *deterministic shift* will refer exclusively to changes in the intercept of the autoregressive representation of the process. To illustrate, suppose that a scalar random variable w_t follows an autoregressive process of first order, so $w_t = \gamma_0 + \gamma_1 w_{t-1} + \varepsilon_t$ where $\varepsilon_t \sim \text{IN} [0, \sigma_\varepsilon^2]$ and $|\gamma_1| < 1$ such that $\{w_t\}$ is covariance-stationary. The distribution of w_t given w_{t-1} is $w_t | w_{t-1} \sim \text{IN} [\gamma_0 + \gamma_1 w_{t-1}, \sigma_\varepsilon^2]$ and the unconditional distribution is $w_t \sim \text{N} [\mu_w, \sigma_w^2]$, with $\mu_w = (1 - \gamma_1)^{-1} \gamma_0$ and $\sigma_w^2 = (1 - \gamma_1)^{-2} \sigma_\varepsilon^2$. The conditional distribution could change in response to shifts in γ_0 , γ_1 or σ_ε^2 , whereas a deterministic shift will denote $\nabla \gamma_0 \neq 0$. This may or may not lead to a location shift $\nabla \mu_w \neq 0$, depending on whether there is a compensating shift in the dynamic parameter γ_1 .

2.2 Contemporaneous mean co-breaking

We now review some established results to clarify what is known about co-breaking, and set the scene for the new developments.

2.2.1 The basic definition

Consider an n -dimensional stochastic process $\{\mathbf{x}_t\}$ over $t \in \{0, \dots, T\}$ and assume \mathbf{x}_t has a well-defined unconditional expectation around an initial parameterization $E[\mathbf{x}_0] = \boldsymbol{\rho}_0$. In particular, $\boldsymbol{\rho}_0$ depends only on deterministic variables whose parameters do not change: e.g., $\boldsymbol{\rho}_0$ may be a constant-parameter polynomial in t of order unity, i.e., $\boldsymbol{\rho}_0 = \boldsymbol{\rho}_{c,0} + \boldsymbol{\rho}_{\ell,0}t$, where $\boldsymbol{\rho}_{c,0}$ and $\boldsymbol{\rho}_{\ell,0}$ are the intercept and linear trend parameter, respectively. Note that the process $\{\mathbf{x}_t\}$ could be $I(0)$ or a derived function from an $I(1)$ process such as first differences or cointegrated combinations.

Following the terminology of section 2.1, a location shift in $\{\mathbf{x}_t\}$ is said to occur if, for any $t \in \mathcal{T} = \{1, \dots, T\}$,

$$E[\mathbf{x}_t - \boldsymbol{\rho}_0] = \boldsymbol{\rho}_t < \infty \quad (7)$$

and $\boldsymbol{\rho}_t \neq \boldsymbol{\rho}_{t-1}$. Put differently, the location shift occurs if the expected value of \mathbf{x}_t around its initial parameterisation in one time period deviates from that in the previous one.

Co-breaking is defined as the cancellation of location shifts across linear combinations of variables.

Definition 1 *The $(n \times r)$ matrix $\boldsymbol{\Phi}$ of rank $r < n$ is said to be contemporaneous mean co-breaking of order r for $\{\mathbf{x}_t\}$ in (7), denoted $CMC(r)$, if $\boldsymbol{\Phi}'\boldsymbol{\rho}_t = \mathbf{0}_{(r \times 1)} \forall t \in \mathcal{T}$.*

It follows from this definition that $\boldsymbol{\Phi}'E[\mathbf{x}_t - \boldsymbol{\rho}_0] = \boldsymbol{\Phi}'\boldsymbol{\rho}_t = \mathbf{0}_{(r \times 1)}$, so that the parameterization of the reduced set of the r co-breaking relationships $\boldsymbol{\Phi}'\mathbf{x}_t$ is independent of the location shifts.

Contemporaneous mean co-breaking seems unlikely when the $\{\boldsymbol{\rho}_t\}$ can change in any possible way from period to period, since a single matrix $\boldsymbol{\Phi}$ is required to annihilate all changes. Nevertheless, there are many cases where co-breaking can occur in principle even though no $\rho_{i,t}$ stays constant. To see this possibility, we next link CMC to a reduced-rank condition. Consider the $(n \times T)$ matrix $\mathbf{P} = (\boldsymbol{\rho}_1 : \boldsymbol{\rho}_2 : \dots : \boldsymbol{\rho}_T)$ where $T > n$. The condition that $\boldsymbol{\phi}'\boldsymbol{\rho}_t = 0 \forall t \in \mathcal{T}$ when $\boldsymbol{\phi} \neq \mathbf{0}_{(n \times 1)}$ can be written as $\boldsymbol{\phi}'\mathbf{P} = \mathbf{0}_{(1 \times T)}$, so that Theorem 2 follows from the equivalence of reduced rank and linear dependence.

Theorem 2 *A necessary and sufficient condition for $\boldsymbol{\phi}'\boldsymbol{\rho}_t = 0 \forall t \in \mathcal{T}$ when $\boldsymbol{\phi} \neq \mathbf{0}_{(n \times 1)}$, i.e. $CMC(1)$, is that:*

$$\text{rk}(\mathbf{P}) < n. \quad (8)$$

This theorem is tantamount to saying that the matrix \mathbf{P} must have reduced row rank for CMC to be feasible. The rank-plus-nullity theorem, see for instance Meyer (2000), then allows the deduction of the following corollary.

Corollary 3 *CMC is at least of order r if there exist r linearly independent vectors satisfying $\boldsymbol{\phi}'_i\boldsymbol{\rho}_t = 0$ such that the $(n \times r)$ matrix $\boldsymbol{\Phi} = (\boldsymbol{\phi}_1 : \dots : \boldsymbol{\phi}_r)$ has rank $r < n$. Then $\boldsymbol{\Phi}'\mathbf{P} = \mathbf{0}_{(r \times T)}$ so $\text{rk}(\mathbf{P}) \leq n - r$, and the nullity of \mathbf{P} , denoted $\text{nul}(\mathbf{P}) \geq r$, determines the order of CMC.*

Note that the matrix $\boldsymbol{\Phi}$ in Corollary 3 is not unique without suitable normalization, since if \mathbf{H} is any $(r \times r)$ non-singular matrix, then $\boldsymbol{\Phi}'\mathbf{P} = \mathbf{0}_{(r \times T)}$ implies that $\mathbf{H}\boldsymbol{\Phi}'\mathbf{P} = \mathbf{0}_{(r \times T)}$ as well.

Given that the order of co-breaking is equivalent to a rank condition, the following terminology is introduced.

Definition 4 The number r of linearly independent vectors ϕ_i , $i = 1, \dots, r$, in $\Phi = (\phi_1 : \dots : \phi_r)$ or, equivalently, the nullity of $\mathbf{P} = (\rho_1 : \rho_2 : \dots : \rho_T)$ is referred to as co-breaking rank.

This terminology is particularly useful when rank tests, employed to determine the order of CMC from the data, are discussed in section 5.

Since $\text{CMC}(r)$ implies $\text{CMC}(r - 1)$, and unmodelled breaks entail model mis-specification, it seems natural to seek the maximum degree of CMC. However, an upper bound on the order of CMC is established in the following theorem.

Corollary 5 CMC cannot exceed order $n - 1$ if breaks occur, i.e., if $\rho_s \neq \mathbf{0}_{(n \times 1)}$ for some s , since when Φ is $(n \times n)$ and non-singular, the equations $\Phi' \mathbf{P} = \mathbf{0}_{(r \times T)}$ would entail $\mathbf{P} = \mathbf{0}_{(n \times T)}$.

Conversely, for CMC not to be trivial, no $(m \times 1)$ -dimensional sub-vector $\rho_{1,t}$ of ρ_t , $m < n$, must remain constant $\forall t \in \mathcal{T}$ since otherwise an obvious co-breaking vector ϕ would consist of zeros apart from one unity value in the position that corresponds to a constant element in ρ_t .

Corollary 6 When any $(m \times 1)$, $m < n$, sub-vector $\rho_{1,t}$ of ρ_t is zero $\forall t \in \mathcal{T}$ then m linearly independent $(n \times 1)$ vectors ϕ_i may be found such that $\phi_i' \rho_t = 0$, $i = 1 \dots, m$, and CMC of at least order m occurs.

Finally, the fact that \mathbf{P} must have reduced row rank may be alternatively illustrated by, first, reformulating the breaks in terms of row vectors $\rho^{(i)'} = (\rho_1^{(i)} : \dots : \rho_T^{(i)})$, $i = 1, \dots, n$, of dimension $(1 \times T)$, so that $\mathbf{P} = (\rho^{(1)} : \dots : \rho^{(n)})'$ and, secondly, noting Corollary 7:

Corollary 7 CMC cannot occur when the $\rho^{(i)}$ are linearly independent, since then, for all $\phi \neq \mathbf{0}_{(n \times 1)}$, $\mathbf{P}' \phi = \sum_{i=1}^n \phi_i \rho^{(i)} \neq \mathbf{0}_{(T \times 1)}$.

2.2.2 The reduced-rank restriction

In addition to the instance mentioned in Corollary 6, at least two further cases may be identified that induce co-breaking. These correspond to scenarios in which the reduced row rank of \mathbf{P} , see Theorem 2, is not a restriction. First, an undersized sample induces a reduced rank by default:

Corollary 8 When $T < n$, then $\text{rk}(\mathbf{P}) \leq T < n$ and there must be CMC of at least order $n - T$.

Secondly, and relatedly, if ρ_t does not vary in every period $t \in \mathcal{T}$ but in fact only changes $k < n$ times, so that \mathbf{x}_t is subject to only that number of location shifts over the course of the sample as in (7), then at least $n - k$ co-breaking vectors can always be found. To see this, assume that ρ_t has $k < T$ distinct values, leading to the decomposition of \mathbf{P} into

$$\mathbf{P} = \boldsymbol{\theta} \mathbf{D} \quad (9)$$

where $\boldsymbol{\theta}$ is $(n \times k)$ and comprises the $k < T$ distinct columns of \mathbf{P} , while \mathbf{D} is $(k \times T)$ and specifies how the remaining $(T - k)$ columns of \mathbf{P} are generated as linear combinations of $\boldsymbol{\theta}$. In this setting, the reduced row rank of \mathbf{P} translates into an equivalent condition on $\boldsymbol{\theta}$ such that co-breaking occurs if $\boldsymbol{\theta}$ has reduced row rank. This, however, must happen if $\boldsymbol{\theta}$ has fewer columns than rows, i.e., if there are fewer distinct shifts in the sample than variables in the system:

Corollary 9 If (9) holds and $\text{rk}(\boldsymbol{\theta}) \leq k < n$ then CMC occurs for at least order $n - k$.

A case in point would be a single simultaneous blip in all elements of \mathbf{x}_t for which $(n - 1)$ co-breaking vector could always be found.

More interesting than the scenario described in Corollary 9 is the decomposition in (9) coupled with the assumption that there are more distinct values in $\boldsymbol{\rho}_t$ than variables in the system, so $n \leq k$. As Engle and Kozicki (1993) point out, it is only in this setting that a reduced rank condition on $\boldsymbol{\theta}$ becomes a restriction, namely $\text{rk}(\boldsymbol{\theta}) < n$. Assuming that $\boldsymbol{\theta}$ is of rank $\text{rk}(\boldsymbol{\theta}) = n - r < k$ with $r > 0$, it may be decomposed into the product of two full-rank matrices $\boldsymbol{\zeta}$ and $\boldsymbol{\nu}$ such that:

$$\boldsymbol{\theta} = \boldsymbol{\zeta}\boldsymbol{\nu}' \quad (10)$$

where $\boldsymbol{\zeta}$ is of dimension $(n \times (n - r))$ and of rank $\text{rk}(\boldsymbol{\zeta}) = n - r$, while $\boldsymbol{\nu}$ is of dimension $(k \times (n - r))$ and of rank $\text{rk}(\boldsymbol{\nu}) = n - r$; see, for instance, Rao (1973, Section 1b). This scenario corresponds closely to the idea that co-breaking involves shifts being related across variables *and* over time, so the shifts are not merely distinct but also linearly dependent. The setting in Corollary 9, however, obfuscates whether CMC occurs because the shifts are ‘truly’ related over time, i.e., $\text{rk}(\boldsymbol{\theta}) < k$, or whether it occurs despite $\text{rk}(\boldsymbol{\theta}) = k$. The assumption that $n \leq k$ avoids the possibility of such ‘spurious’ co-breaking.

Substituting (10) into (9) yields $\mathbf{P} = \boldsymbol{\zeta}\boldsymbol{\nu}'\mathbf{D}$ so the r co-breaking vectors $\boldsymbol{\Phi} = (\phi_1 : \dots : \phi_r)$ are given by the orthogonal complement of $\boldsymbol{\zeta}$, namely the $(n \times r)$ matrix $\boldsymbol{\zeta}_\perp$ such that $(\boldsymbol{\zeta} : \boldsymbol{\zeta}_\perp)$ is of full rank n and $\boldsymbol{\zeta}'_\perp \boldsymbol{\zeta} = \mathbf{0}_{(r \times (n-r))}$.

Corollary 10 *If $\boldsymbol{\rho}_t$ has $n \leq k < T$ distinct values over the sample such that $\mathbf{P} = \boldsymbol{\theta}\mathbf{D}$ as in (9), and $\text{rk}(\boldsymbol{\theta}) = n - r$, with $r > 0$, so that $\boldsymbol{\theta} = \boldsymbol{\zeta}\boldsymbol{\nu}'$ as in (10), then there exists an $(n \times r)$ matrix $\boldsymbol{\Phi} = \boldsymbol{\zeta}_\perp$ such that the r linear combinations $\boldsymbol{\zeta}'_\perp \mathbf{E}[\mathbf{x}_t - \boldsymbol{\rho}_0] = \mathbf{0}_{(r \times 1)}$ are the co-breaking relationships.*

2.2.3 Extensions

Various extensions of the concept of contemporaneous mean co-breaking are conceivable. For instance, co-breaking could be defined for higher moments by direct extension, leading to, e.g., variance co-breaking. Analogously, inter-temporal mean co-breaking (IMC) could be defined as the cancellation of deterministic shifts across variables and time periods, see Hendry (1996). Moreover, the class of processes $\{\mathbf{x}_t\}$ in Definition 1 may be extended to conditional functions of past variables. Defining $\mathbf{X}_{t-1}^1 = \sigma(\mathbf{x}_1, \dots, \mathbf{x}_{t-1})$ as the σ -field generated by $\{\mathbf{x}_i\}_{i=1}^{t-1}$, a deterministic shift in $\{\mathbf{x}_t | \mathbf{X}_{t-1}^1\}$ occurs if, for any $t \in \mathcal{T}$,

$$\mathbf{E}[(\mathbf{x}_t - \boldsymbol{\pi}_0) | \mathbf{X}_{t-1}^1] = \sum_{i=1}^g \boldsymbol{\Pi}_i \mathbf{x}_{t-i} + \boldsymbol{\pi}_t \quad (11)$$

and $\boldsymbol{\pi}_t \neq \boldsymbol{\pi}_{t-1}$. While changes in the dynamic parameters $\boldsymbol{\Pi}_i$ are possible, previous research has established that the main impacts on policy variables and forecast biases seem to derive from entailed changes in the series’s location; see, for instance, Clements and Hendry (1998). The equivalent to the condition $\boldsymbol{\Phi}'\boldsymbol{\rho}_t = \mathbf{0}_{(r \times 1)}$ in Definition 1 is seen to be $\boldsymbol{\Phi}'\boldsymbol{\pi}_t = \mathbf{0}_{(r \times 1)}$, leading to the following distinction:

Definition 11 *If there exists an $(n \times r)$ matrix $\boldsymbol{\Phi}$ of rank $r < n$ such that $\boldsymbol{\Phi}'\boldsymbol{\rho}_t = \mathbf{0}_{(r \times 1)} \forall t \in \mathcal{T}$ in (7), then the transformations $\boldsymbol{\Phi}'\mathbf{E}[\mathbf{x}_t - \boldsymbol{\rho}_0]$ are said to be unconditional co-breaking relationships. Conversely, if $\boldsymbol{\Phi}$ is such that $\boldsymbol{\Phi}'\boldsymbol{\pi}_t = \mathbf{0}_{(r \times 1)} \forall t \in \mathcal{T}$ in (11), then $\boldsymbol{\Phi}'\mathbf{E}[(\mathbf{x}_t - \boldsymbol{\pi}_0) | \mathbf{X}_{t-1}^1]$ are said to be conditional co-breaking relationships.*

It is instructive to decompose the shift vector $\boldsymbol{\rho}_t$ into intercept and trend shifts such that:

$$\mathbf{E}[\mathbf{x}_t - \boldsymbol{\rho}_0] = \boldsymbol{\rho}_t = \boldsymbol{\rho}_{c,t} + \boldsymbol{\rho}_{\ell,t}. \quad (12)$$

Concatenating the values of $\boldsymbol{\rho}_{c,t}$ and those of $\boldsymbol{\rho}_{\ell,t}$ into $(n \times T)$ matrices yields $\mathbf{P}_c = (\boldsymbol{\rho}_{c,1} : \cdots : \boldsymbol{\rho}_{c,T})$ and $\mathbf{P}_\ell = (\boldsymbol{\rho}_{\ell,1} : \cdots : \boldsymbol{\rho}_{\ell,T})$, respectively. The analysis in section 2.2.1 now carries over directly to both \mathbf{P}_c and \mathbf{P}_ℓ such that conditions for intercept co-breaking or trend co-breaking are, according to Theorem 2, $\text{rk}(\mathbf{P}_c) < n$, and $\text{rk}(\mathbf{P}_\ell) < n$. The following definition summarizes.

Definition 12 *If there exists an $(n \times r)$ matrix $\boldsymbol{\Phi}$ of rank $r < n$ such that $\boldsymbol{\Phi}'\boldsymbol{\rho}_{c,t} = \mathbf{0}_{(r \times 1)}$ or $\boldsymbol{\Phi}'\boldsymbol{\rho}_{\ell,t} = \mathbf{0}_{(r \times 1)}$, $\forall t \in \mathcal{T}$ in (12), then the transformations $\boldsymbol{\Phi}'\mathbf{E}[\mathbf{x}_t - \boldsymbol{\rho}_0]$ are said to be intercept, or trend, co-breaking relationships, respectively.*

This definition is easily generalized to other forms of location shifts. To see this, decompose $\boldsymbol{\rho}_t$ into the sum of p deterministic terms, $\boldsymbol{\rho}_t = \sum_{i=1}^p \boldsymbol{\rho}_{i,t}$ or, in full-sample notation, $\mathbf{P} = \sum_{i=1}^p \mathbf{P}_i$. Suppose that each $\boldsymbol{\rho}_{i,t}$ in fact has only $k_i < T$ distinct values $\boldsymbol{\theta}_i$, $i \in \{1, \dots, p\}$, such that, in analogy to (9), $\mathbf{P} = \sum_{i=1}^p \boldsymbol{\theta}_i \mathbf{D}_i$. Further, let \mathcal{I} be any subset of $\{1, \dots, p\}$ and \mathcal{I}_c its complement. Then, concatenate the $\boldsymbol{\theta}_i$ into $\boldsymbol{\theta}_{\mathcal{I}}$ and $\boldsymbol{\theta}_{\mathcal{I}_c}$ such that $\boldsymbol{\theta}_{\mathcal{I}}$ is $(n \times k_{\mathcal{I}})$ and $\boldsymbol{\theta}_{\mathcal{I}_c}$ is $(n \times k_{\mathcal{I}_c})$, where $k_{\mathcal{I}} = \sum_{i \in \mathcal{I}} k_i$ and $k_{\mathcal{I}_c} = \sum_{i \in \mathcal{I}_c} k_i$. As a consequence, \mathbf{P} may be written as

$$\mathbf{P} = \boldsymbol{\theta}_{\mathcal{I}} \mathbf{D}_{\mathcal{I}} + \boldsymbol{\theta}_{\mathcal{I}_c} \mathbf{D}_{\mathcal{I}_c} \quad (13)$$

where $\mathbf{D}_{\mathcal{I}}$ and $\mathbf{D}_{\mathcal{I}_c}$ are of dimension $(k_{\mathcal{I}} \times T)$ and $(k_{\mathcal{I}_c} \times T)$, respectively. Imposing the reduced-rank restriction of (10) on $\boldsymbol{\theta}_{\mathcal{I}}$, it is clear that $\boldsymbol{\Phi} = \boldsymbol{\zeta}_{\perp}$ are co-breaking vectors for the deterministic terms in $\mathbf{D}_{\mathcal{I}}$.

Finally, examine more closely the rôle played by the initial parameterization $\mathbf{E}[\mathbf{x}_0] = \boldsymbol{\rho}_0$. It was defined above to be a constant-parameter expression of deterministic variables such as the polynomial $\boldsymbol{\rho}_0 = \boldsymbol{\rho}_{c,0} + \boldsymbol{\rho}_{\ell,0}t$, where $\boldsymbol{\rho}_{c,0}$ is a constant-parameter baseline intercept and $\boldsymbol{\rho}_{\ell,0}t$ is a constant-parameter baseline trend. If the parameter vector $\boldsymbol{\rho}_t$ comprises deterministic shifts in these two components, as in (12), then (7) may be re-written as:

$$\mathbf{E}[\mathbf{x}_t] = \boldsymbol{\rho}_0 + \boldsymbol{\rho}_t = \boldsymbol{\varrho}_t \quad (14)$$

such that the expected value $\mathbf{E}[\mathbf{x}_t]$ becomes a *piecewise* polynomial in t . The concepts outlined in section 2.2.1 now equally apply to (14). In particular, the following is a direct extension of Definition 1:

Definition 13 *An $(n \times r)$ matrix $\boldsymbol{\Phi}$ such that $\boldsymbol{\Phi}'\boldsymbol{\varrho}_t = \mathbf{0}_{(r \times 1)} \forall t \in \mathcal{T}$ is said to be mean-zero co-breaking for $\{\mathbf{x}_t\}$, with the result that $\mathbf{E}[\boldsymbol{\Phi}'\mathbf{x}_t] = \mathbf{0}_{(r \times 1)}, \forall t \in \mathcal{T}$.*

Moreover, following Theorem 2, a necessary and sufficient condition for mean-zero co-breaking to occur is that the $(n \times T)$ matrix $(\boldsymbol{\varrho}_1 : \cdots : \boldsymbol{\varrho}_T)$ has reduced row rank. The terminology in Definition 13 is intended to take account of the fact that it is not only the shifts $\boldsymbol{\rho}_t$ that are cancelled by $\boldsymbol{\Phi}$ but the entire piecewise polynomial function $\boldsymbol{\varrho}_t$. Chapman and Ogaki (1993) and Hatanaka and Yamada (2003), *inter alia*, consider similar settings, and call $\boldsymbol{\Phi}'\mathbf{x}_t$ *co-trending* relationships. However, since it seems more natural in empirical modelling, and especially in policy analysis and forecasting exercises, to examine location shifts where there is no need for $\boldsymbol{\Phi}'\mathbf{x}_t$ to have mean zero we disregard this possibility in the analysis below, except where otherwise indicated.

2.3 Cointegration co-breaking

This section considers two special instances of co-breaking arising naturally in a cointegrated vector equilibrium-correction model (VEqCM). Recall the n -dimensional conditional process $\{\mathbf{x}_t | \mathbf{X}_{t-1}^1\}$ in equation (11) of the previous section, and define $\boldsymbol{\varpi}_t = \boldsymbol{\pi}_0 + \boldsymbol{\pi}_t$ by analogy with $\boldsymbol{\varrho}_t$ in (14). The entire distribution may then be specified as:

$$\mathbf{x}_t | \mathbf{X}_{t-1}^1 \sim \mathbf{N}_n(\boldsymbol{\varpi}_t, \boldsymbol{\Sigma}), \quad (15)$$

where $\varpi_t = E[\mathbf{x}_t | \mathbf{X}_{t-1}^1]$ is the conditional mean and $\Sigma = V[\mathbf{x}_t | \mathbf{X}_{t-1}^1]$ the conditional variance. Specifically, assume for simplicity that $\mathbf{X}_{t-1}^1 = \sigma(\mathbf{x}_{t-1}, \mathbf{x}_{t-2})$ and, initially, that $\varpi_t = \varpi$, such that the data generation process (DGP) over the period $T = \{1, \dots, T\}$ is given by a second-order vector autoregression (VAR):

$$\mathbf{x}_t = \varpi + \Pi_1 \mathbf{x}_{t-1} + \Pi_2 \mathbf{x}_{t-2} + \epsilon_t \quad (16)$$

where $\epsilon_t | \mathbf{X}_{t-1}^1 \sim \text{IN}_n[\mathbf{0}, \Sigma]$ denotes an independent normal error. In (16), none of the roots of the autoregressive polynomial $\Pi(z)$ lies inside the unit circle. The DGP is assumed integrated of order unity, and satisfies $r < n$ cointegration relations such that $\Pi_1 + \Pi_2 - \mathbf{I}_n = \alpha\beta'$, where the following assumptions are made: first, α and β are $(n \times r)$ matrices of rank r such that α_\perp and β_\perp are $(n \times (n-r))$ matrices of rank $(n-r)$ with $\alpha'\alpha_\perp = \beta'\beta_\perp = \mathbf{0}_{(r \times (n-r))}$. Secondly, $\text{rk}(\alpha'_\perp \Psi \beta_\perp) = (n-r)$, where $\Psi = \mathbf{I}_n - \Psi_1$, with $\Psi_1 = -\Pi_2$, is the mean-lag matrix, to rule out the possibility that $\mathbf{x}_t \sim I(2)$. Then (16) can be reparametrized as the VEqCM:

$$\Delta \mathbf{x}_t = \varpi + \alpha\beta' \mathbf{x}_{t-1} + \Psi_1 \Delta \mathbf{x}_{t-1} + \epsilon_t \quad (17)$$

where $\Delta \mathbf{x}_t$ and $\beta' \mathbf{x}_t$ are $I(0)$. Pre-multiplying the intercept ϖ by the identity $\mathbf{I}_n = \alpha(\beta'\alpha)^{-1}\beta' + \beta_\perp(\alpha'_\perp\beta_\perp)^{-1}\alpha_\perp$, where $\alpha_\perp \neq \beta$, yields $\varpi = \gamma - \alpha\mu$, where $\gamma = \beta_\perp(\alpha'_\perp\beta_\perp)^{-1}\alpha_\perp\varpi$ is $(n \times 1)$ and $\mu = -(\beta'\alpha)^{-1}\beta'\varpi$ is $(r \times 1)$, and $\beta'\gamma = \mathbf{0}$ by construction. The VEqCM in deviations about means is therefore:

$$\Delta \mathbf{x}_t - \gamma = \alpha(\beta' \mathbf{x}_{t-1} - \mu) + \Psi_1 \Delta \mathbf{x}_{t-1} + \epsilon_t. \quad (18)$$

The system grows at the unconditional rate $E[\Delta \mathbf{x}_t] = \delta$ where $\gamma = (\mathbf{I}_n - \Psi_1)\delta$ and $\beta'\delta = \mathbf{0}$, with long-run solution $E[\beta' \mathbf{x}_t] = \mu$. Although the decomposition of ϖ is not orthogonal, since $\gamma'\alpha\mu \neq \mathbf{0}$, (18) is isomorphic to (17).

Any of the parameters α , β , γ , or μ could be subject to change, but we focus on the last two such that $\Pi_1 + \Pi_2 = \mathbf{I}_n + \alpha\beta'$ remains constant. Thus, suppose that up until time period $t = T$, the intercepts ϖ , and hence the equilibrium means and the unconditional growth rates, are determined by the relation $\pi_0 = \gamma_0 - \alpha\mu_0$ while, for $t > T$, they are subject to location shifts such that $\pi_t = \gamma_t - \alpha\mu_t$. Thus, the post-break intercept is given by $\varpi_t = \pi_0 + \pi_t = (\gamma_0 + \gamma_t) - \alpha(\mu_0 + \mu_t)$ such that, for $t > T$, the VEqCM in (18) is:

$$\begin{aligned} \Delta \mathbf{x}_t &= (\gamma_0 + \gamma_t) + \alpha(\beta' \mathbf{x}_t - (\mu_0 + \mu_t)) \\ &\quad + \Psi_1 \Delta \mathbf{x}_{t-1} + \epsilon_t \\ &= \widetilde{\Delta \mathbf{x}}_t + (\gamma_t - \alpha\mu_t), \end{aligned} \quad (19)$$

where $\widetilde{\Delta \mathbf{x}}_t$ is the constant-parameter value of $\Delta \mathbf{x}_t$ in (18), and $\gamma_t - \alpha\mu_t$ is the composite intercept shift. In terms of the system in (19), consider the $(n \times r)$ matrix Φ and form the r linear combinations $\Phi' \Delta \mathbf{x}_t$:

$$\Phi' \Delta \mathbf{x}_t = \Phi' \widetilde{\Delta \mathbf{x}}_t + \Phi' (\gamma_t - \alpha\mu_t). \quad (20)$$

Then, r -dimensional equilibrium-mean co-breaking requires that $\Phi' \alpha\mu_t = \mathbf{0}_{(r \times 1)}$, $\forall t > T$, while r -dimensional drift co-breaking requires $\Phi' \gamma_t = \mathbf{0}_{(r \times 1)}$, $\forall t > T$. Two instances of Φ meet these conditions naturally.

First, the orthogonal complement of α , namely α_\perp , satisfies $\alpha'_\perp \alpha = \mathbf{0}_{((n-r) \times r)}$ by definition. In (20), choose $\Phi = \alpha_\perp$ so that:

$$\alpha'_\perp \Delta \mathbf{x}_t = \alpha'_\perp \gamma_0 + \alpha'_\perp \Psi_1 \Delta \mathbf{x}_{t-1} + \alpha'_\perp \epsilon_t + \alpha'_\perp \gamma_t. \quad (21)$$

This result is stated in the following theorem.

Theorem 14 ‘Common trends’ are equilibrium-mean co-breaking.

Since α_{\perp} is the ‘selector’ of the equations which are not dependent on EqCMs, this result is close to explaining the effectiveness of differencing as a ‘solution’ to equilibrium-mean shifts as in Clements and Hendry (1998). Second differencing, or further co-breaking, however, would be required to remove the shift in the drift coefficient as well. To illustrate that this is feasible, suppose that $\alpha' = (\mathbf{a}'_{(r \times r)}; \mathbf{0}'_{((n-r) \times r)})$, such that an orthogonal complement α_{\perp} satisfying $\alpha' \alpha_{\perp} = \mathbf{0}$ is $\alpha'_{\perp} = (\mathbf{0}'_{(r \times (n-r))}; \mathbf{I}_{n-r})$. Then the co-breaking relationship in (21) would effectively be the lower block of $(n - p)$ equations, denoted by $\mathbf{x}_{b,t}$:

$$\Delta \mathbf{x}_{b,t} = \gamma_{0,b} + \Psi_{1,b} \Delta \mathbf{x}_{t-1} + \epsilon_{b,t} + \gamma_{t,b}. \quad (22)$$

It is now conceivable that there exists some linear combinations $\Phi' \Delta \mathbf{x}_{b,T+1}$ that are drift as well as equilibrium-mean break free. In particular, when $n - r > 1$ such a combination will always exist for a one-off change $\gamma_{t,b} \neq \mathbf{0}_{((n-r) \times 1)}$. Consequently, the drift and equilibrium-mean co-breaking vectors would be given by $\alpha_{\perp} \Phi$.

Secondly, recall that γ lies in the column space of β_{\perp} , denoted by $\gamma \in c(\beta_{\perp})$, such that $\beta' \gamma = \mathbf{0}$, implying that the cointegrating vectors are trend free. As long as $\beta' \alpha \neq \mathbf{0}_{(r \times r)}$, therefore, choosing in (20) $\Phi = \beta$ yields:

$$\begin{aligned} \beta' \Delta \mathbf{x}_t &= \beta' \alpha (\beta' \mathbf{x}_t - \mu_0) + \beta' \Psi_1 \Delta \mathbf{x}_{t-1} \\ &+ \beta' \epsilon_t - \beta' \alpha \mu_t, \end{aligned} \quad (23)$$

eliminating the shifts in the drift parameter. The following theorem summarizes.

Theorem 15 The cointegration vector is drift co-breaking.

3 Co-breaking in empirical research

The concept of co-breaking is relevant to many aspects of empirical research. Three such instances will be considered in this section. In section 3.1, the relationship between co-breaking and super exogeneity in policy analysis is examined, with the discussion drawing on Engle and Hendry (1993), Banerjee, Hendry and Mizon (1996) and Hendry and Mizon (1998). Secondly, building on recent advances in Hendry (2003), the rôle of co-breaking in impulse-response analysis is investigated in section 3.2. Finally, in section 3.3, we consider the importance of co-breaking in forecasting macroeconomic time series.

3.1 Policy analysis and super exogeneity

Recall the unconditional mean of the n -dimensional independent Normal random variable \mathbf{x}_t as given in equation (14) of section 2.2.3, and assume its entire distribution is:

$$\mathbf{x}_t \sim \text{IN}_n [\boldsymbol{\varrho}_t, \boldsymbol{\Sigma}_t], \quad (24)$$

where the mean vector $\boldsymbol{\varrho}_t$ and the covariance matrix $\boldsymbol{\Sigma}_t$ are not constant over time, compare (24) to the conditional distribution in (15) of the previous section to clarify the notation. Suppose that an economic theory involves the two random variables \mathbf{y}_t and \mathbf{z}_t , of dimension $(n_1 \times 1)$ and $(n_2 \times 1)$, $n_1 + n_2 = n$, respectively, obtained by appropriate partitioning of \mathbf{x}_t :

$$\begin{pmatrix} \mathbf{y}_t \\ \mathbf{z}_t \end{pmatrix} \sim \text{IN}_n \left[\begin{pmatrix} \boldsymbol{\varrho}_{1,t} \\ \boldsymbol{\varrho}_{2,t} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{11,t} & \boldsymbol{\Sigma}_{12,t} \\ \boldsymbol{\Sigma}'_{12,t} & \boldsymbol{\Sigma}_{22,t} \end{pmatrix} \right]. \quad (25)$$

In particular, the theory stipulates that the mean of \mathbf{y}_t is proportionally related to that of \mathbf{z}_t :

$$\boldsymbol{\varrho}_{1,t} = \Upsilon_t \boldsymbol{\varrho}_{2,t} \quad (26)$$

where Υ_t is the parameter of interest. Although (26) does not involve a constant term, the subsequent analysis would be easily modified to that effect. An example of (26) would be the Permanent Income Hypothesis (PIH) as in Friedman (1957), according to which $\boldsymbol{\varrho}_{1,t}$ and $\boldsymbol{\varrho}_{2,t}$ denote permanent consumption and permanent income, respectively. The basic issue of the present analysis is to ask under what conditions the theory in (26) may be investigated by the empirical model:

$$\mathbf{y}_t = \Upsilon_t \mathbf{z}_t + \boldsymbol{\epsilon}_t, \quad (27)$$

where $\boldsymbol{\epsilon}_t \sim \text{IN}_n[\mathbf{0}, \boldsymbol{\Omega}]$, for $t = 1, \dots, T$.

To answer that question it is necessary to derive the expectation of \mathbf{y}_t conditional on \mathbf{z}_t , given their joint distribution in (25) and the relationship in (26). Following Engle and Hendry (1993), the conditional mean is:

$$\text{E}[\mathbf{y}_t | \mathbf{z}_t] = \mathbf{b}_t + \boldsymbol{\Gamma}_t \mathbf{z}_t \quad (28)$$

where $\boldsymbol{\Gamma}_t = \boldsymbol{\Sigma}_{12,t} \boldsymbol{\Sigma}_{22,t}^{-1}$ and $\mathbf{b}_t = \boldsymbol{\varrho}_{1,t} - \boldsymbol{\Gamma}_t \boldsymbol{\varrho}_{2,t}$, while the conditional variance is denoted by $\boldsymbol{\Omega}_t = \boldsymbol{\Sigma}_{11,t} - \boldsymbol{\Sigma}_{12,t} \boldsymbol{\Sigma}_{22,t}^{-1} \boldsymbol{\Sigma}_{21,t}$. Incorporating (26) in (28) yields $\mathbf{b}_t = (\Upsilon_t - \boldsymbol{\Gamma}_t) \boldsymbol{\varrho}_{2,t}$ or rearranging:

$$\text{E}[\mathbf{y}_t | \mathbf{z}_t] = \Upsilon_t \mathbf{z}_t + (\boldsymbol{\Gamma}_t - \Upsilon_t) (\mathbf{z}_t - \boldsymbol{\varrho}_{2,t}). \quad (29)$$

Thus, the parameters of the conditional and marginal densities respectively are $\phi_{1,t} = (\mathbf{b}_t : \boldsymbol{\Gamma}_t : \boldsymbol{\Omega}_t)'$ and $\phi_{2,t} = (\boldsymbol{\varrho}_{2,t} : \boldsymbol{\Sigma}_{22,t}^\epsilon)'$, where $(\cdot)^\epsilon$ denotes column vectoring, retaining only non-redundant elements.

Three conditions must be satisfied for (27) to be a constant-parameter conditional empirical model:

- (a) $\Upsilon_t = \boldsymbol{\Gamma}_t$,
- (b) $\phi_{1,t}$ is invariant to $\phi_{2,t}$,
- (c) $\boldsymbol{\Gamma}_t = \boldsymbol{\Gamma}$ and $\boldsymbol{\Omega}_t = \boldsymbol{\Omega}$, $\forall t$.

To see this, consider equation (29): Setting up an empirical model of \mathbf{y}_t given \mathbf{z}_t and $(\mathbf{z}_t - \boldsymbol{\varrho}_{2,t})$ would be fraught with perils since the two parameter sets $\phi_{1,t}$ and $\phi_{2,t}$ are not variation-free. This is essentially due to the cross-equation restriction given by the theory model in (26). Variation-freeness, however, may be ensured by imposing condition (a), implying $\mathbf{b}_t = \mathbf{0}_{(n_1 \times 1)}$ and thus $\text{E}[\mathbf{y}_t | \mathbf{z}_t] = \Upsilon_t \mathbf{x}_t$. The marginal variable \mathbf{z}_t is thus weakly exogenous for $\phi_{1,t}$ and may safely be disregarded in the empirical analysis of \mathbf{y}_t given \mathbf{z}_t without loss of information. Condition (b) stipulates that $\phi_{1,t}$, i.e., Υ_t and $\boldsymbol{\Omega}_t$, must not change in response to changes in $\phi_{2,t}$. This is crucial since, for instance, $\boldsymbol{\Gamma}_t$ is a function of $\boldsymbol{\Sigma}_{22,t}$ by definition, and if changes in the latter entailed changes in the former, analyzing the conditional model would be impossible despite weak exogeneity. In particular, if changes in $\phi_{2,t}$ are interpreted as policy interventions and the elements in $\phi_{1,t}$ as structural parameters, it is condition (b) that makes policy analysis feasible; otherwise the conditional model may be prone to the Lucas critique, see Lucas (1976). The conjunction of (a) and (b) hence ensures that \mathbf{z}_t is super-exogenous for Υ_t . Finally, condition (c) demands that $\boldsymbol{\Sigma}_{12,t} \boldsymbol{\Sigma}_{22,t}^{-1}$ and $\boldsymbol{\Sigma}_{11,t} - \boldsymbol{\Sigma}_{12,t} \boldsymbol{\Sigma}_{22,t}^{-1} \boldsymbol{\Sigma}_{21,t}$ be constant over time. Note that this constancy is not synonymous with the invariance in (b), since non-constancy of $\boldsymbol{\Gamma}_t$ and $\boldsymbol{\Omega}_t$ could be inherited from, e.g., $\boldsymbol{\Sigma}_{12,t}$. Condition (c) is not strictly necessary for super exogeneity to hold, but it is not clear how an empirical analysis could proceed if the values of $\boldsymbol{\Gamma}$ and $\boldsymbol{\Omega}$ changed over time in an unknown fashion. Conversely, $\boldsymbol{\Gamma}$ being constant may in fact be an indicator for it being invariant to changes in $\phi_{2,t}$. Note also that the existence of feedback from $\mathbf{Y}_{t-1}^1 = \sigma(\mathbf{y}_1, \dots, \mathbf{y}_{t-1})$ onto \mathbf{z}_t is irrelevant, so

the issue of strong exogeneity of \mathbf{z}_t for Υ does not arise. See Engle, Hendry and Richard (1983) for a thorough discussion of exogeneity and Hendry and Santos (2006) for a new testing procedure linked to co-breaking.

Given that the three conditions (a)–(c) are satisfied, the conditional expectation in (29) becomes:

$$E[\mathbf{y}_t | \mathbf{z}_t] = \Upsilon \mathbf{z}_t, \quad (30)$$

the conditional variance is Ω , $\forall t$, and the joint density in (25) takes the form:

$$\begin{pmatrix} \mathbf{y}_t \\ \mathbf{z}_t \end{pmatrix} \sim \text{IN}_n \left[\begin{pmatrix} \Upsilon \boldsymbol{\varrho}_{2,t} \\ \boldsymbol{\varrho}_{2,t} \end{pmatrix}, \begin{pmatrix} \Omega + \Upsilon \Sigma_{22,t} \Upsilon' & \Upsilon \Sigma_{22,t} \\ \Upsilon \Sigma_{22,t} & \Sigma_{22,t} \end{pmatrix} \right]. \quad (31)$$

Under these conditions, therefore, the parameters $(\boldsymbol{\varrho}_{2,t}, \Sigma_{22,t})$ can change in the marginal model $\mathbf{z}_t \sim \text{IN}_{n_2}[\boldsymbol{\varrho}_{2,t}, \Sigma_{22,t}]$, with the agents who determine \mathbf{y}_t continuing to act as they planned, and as it is thought they would. Thus, the empirical analogue of (30) is indeed given by the model in (27), which will deliver correct policy conclusions, and not suffer from the Lucas critique. Conversely, if \mathbf{z}_t is not super exogenous for Υ , then policy scenario analyses may yield misleading conclusions; Hendry (1995) elaborates on this issue.

To emphasize the link between super exogeneity and co-breaking, recall that an equivalent requirement to condition (a) above was that $\mathbf{b}_t = \mathbf{0}_{(n_1 \times 1)}$, or $\boldsymbol{\varrho}_{1,t} - \Upsilon \boldsymbol{\varrho}_{2,t} = \mathbf{0}_{(n_1 \times 1)}$. In addition, conditions (b) and (c) establish that Υ_t is invariant to changes in $\boldsymbol{\phi}_{2,t} = (\boldsymbol{\varrho}_{2,t} : \Sigma_{22,t}^c)'$ and is also otherwise constant, so that an implication of conditions (a)–(c) is:

$$\boldsymbol{\varrho}_{1,t} - \Upsilon \boldsymbol{\varrho}_{2,t} = \mathbf{0}_{(n_1 \times 1)}. \quad (32)$$

This is precisely the condition for contemporaneous mean co-breaking established in section 2.2, with $\Phi' = (\mathbf{I}_{n_1} : -\Upsilon)$ being the co-breaking vectors, see Definition 13 in particular. Put differently, although \mathbf{x}_t as a whole follows a multivariate Normal distribution with time-varying parameters $\boldsymbol{\varrho}_t$ and Σ_t , the linear combination $\Phi' \mathbf{x}_t$ has constant parameters Υ and Ω under the same conditions as are necessary to establish the super exogeneity of \mathbf{z}_t . Thus, although the marginal distributions of \mathbf{y}_t and \mathbf{z}_t are not constant, see (25) and (31), the conditional model of \mathbf{y}_t given \mathbf{z}_t is, as in (30). It is interesting that, in the present setting, conditions (a)–(c) also imply that, $\forall t$:

$$\Sigma_{12,t} - \Upsilon \Sigma_{22,t} = \mathbf{0}_{(n_1 \times n_2)}, \quad (33)$$

so that the means in (32) are interrelated by the same parameter Υ as the covariances $\Sigma_{12,t}$ are with the variances $\Sigma_{22,t}$. As a consequence, mean co-breaking and what may be called variance co-breaking occur simultaneously.

3.2 Impulse-response analyses

Impulse-response analysis is a widely-used method for evaluating the response of one set of variables to ‘shocks’ in another set of variables; see, e.g., Lütkepohl (1993), Runkle (1987), and Sims (1980). The finding that shifts in the parameters of dynamic reactions are not readily detectable is potentially disastrous for impulse-response analyses of economic policy based on closed systems, such as VARs. Since changes in VAR intercepts and dynamic coefficient matrices may not be detected – even when tested for – yet the full-sample estimates are a weighted average across different regimes, the resulting impulse responses need not represent the policy outcomes that will in fact occur. Indeed, this problem may be exacerbated by specifying VARs in first differences (as often occurs), since deterministic factors play a small role in such models: benefits in winning forecasting competitions are paid for by losses

in coherent policy evaluation. Moreover, this is only one of a sequence of drawbacks to using impulse responses from models to evaluate policy, as emphasized over recent years: see, *inter alia*, Banerjee *et al.* (1996), Ericsson, Hendry and Mizon (1998a), and Hendry and Mizon (1998).

The results of impulse-response analyses are well known not to be invariant to the ordering of variables: yet ‘avoiding’ that problem by reporting orthogonalized impulse responses violates weak exogeneity for most orderings. Specifying a variable to be weakly or strongly exogenous alters the impulse responses, irrespective of whether or not that variable actually has the appropriate exogeneity status.

As an illustration, partition the n -dimensional \mathbf{x}_t as in section 3.1 above, assume that Σ is not time-variant, and suppose for simplicity that $n_1 = n_2 = 1$ such that y_t and z_t are scalar random variables; cf. equations (24) and (25). The objective of the analysis is to examine the effect of an ‘impulse’ to z_t on y_t , with $\Sigma_{12} \neq 0$. There are two ‘Choleski’ factorizations, and we take the first, based on the order in which the variables appear in \mathbf{x}_t . The conditional-marginal factorization is:

$$\begin{pmatrix} y_t|z_t \\ z_t \end{pmatrix} \sim \text{IN}_2 \left[\begin{pmatrix} \varrho_{1,t} + \Gamma(z_t - \varrho_{2,t}) \\ \varrho_{2,t} \end{pmatrix}, \begin{pmatrix} \Sigma_{yy} - \Gamma\Sigma_{yz} & 0 \\ 0 & \Sigma_{zz} \end{pmatrix} \right], \quad (34)$$

where $\Gamma = \Sigma_{yz}/\Sigma_{zz}$: compare the discussion surrounding equation (28). A perturbation to the $\{z_t\}$ process is not uniquely identifiable as deriving from a change to $\varrho_{2,t}$ or to the ‘error’ $z_t - \varrho_{2,t} = \epsilon_{2,t}$, although most impulse-response analyses ignore the former possibility. However, the effect on y_t need not be the same for these two possible sources.

To see the problem, let the underlying economics derive from agents’ optimization behaviour as the constant relationship $\varrho_{1,t} = \Upsilon\varrho_{2,t}$, an example of which is the PIH in (26) above. In other words, y_t and z_t co-break. Then from (34):

$$\begin{pmatrix} y_t|z_t \\ z_t \end{pmatrix} \sim \text{IN}_2 \left[\begin{pmatrix} \Upsilon\varrho_{2,t} + \Gamma\epsilon_{2,t} \\ \varrho_{2,t} \end{pmatrix}, \begin{pmatrix} \Sigma_{yy} - \Gamma\Sigma_{yz} & 0 \\ 0 & \Sigma_{zz} \end{pmatrix} \right]. \quad (35)$$

Thus, unless $\Upsilon = \Gamma$, it matters which ‘part’ of the $\{z_t\}$ process is perturbed. But $\Upsilon = \Gamma$ is the condition for the weak exogeneity of z_t for Γ in the conditional model, see section 3.1, so unless weak exogeneity holds, impulse-response outcomes are not unique. Moreover, at most one direction is meaningful. The converse Choleski factorization is:

$$\begin{pmatrix} y_t \\ z_t|y_t \end{pmatrix} \sim \text{IN}_2 \left[\begin{pmatrix} \varrho_{1,t} \\ \varrho_{2,t} + \Xi(y_t - \varrho_{1,t}) \end{pmatrix}, \begin{pmatrix} \Sigma_{yy} & 0 \\ 0 & \Sigma_{zz} - \Upsilon\Sigma_{yz} \end{pmatrix} \right], \quad (36)$$

where $\Xi = \Sigma_{yz}/\Sigma_{yy}$. When, however, $\Upsilon = \Gamma$ then $\Xi \neq 1/\Gamma$ by the non-degeneracy of the joint distribution of $\mathbf{x}_t = (y_t : z_t)'$, so an incorrect reaction will be inferred.

The literature on ‘structural VARs’, see, e.g., Bernanke (1986), and Blanchard and Quah (1989), which also analyzes impulse responses for a transformed system, faces a similar difficulty. The lack of understanding of the crucial role of weak exogeneity in impulse-response analyses is puzzling in view of the obvious feature that any given ‘shock’ to the error and to the intercept are indistinguishable, yet the actual reaction in the economy will be the same only if the means and variances are linked in the same way – which is the weak exogeneity condition in the present setting. But since policy changes are involved, super exogeneity, and hence co-breaking as just discussed, is really needed.

3.3 Forecasting

Perhaps the role of greatest potential importance for co-breaking is in the arena of forecasting. Hendry and Massmann (1999) consider its application, and note the added value of having a block of constant relations even if other aspects of the system being modelled are non-constant, so all variables cannot be easily forecast. Factorize \mathbf{x}_t into $\Phi' \mathbf{x}_t$ and its orthogonal component $\Phi'_\perp \mathbf{x}_t$, where the former are the constant causal links and the latter isolate the non-constant relations. Then $\Phi' \mathbf{x}_{T+h}$ should be forecastable based on its econometric model and data up until $t = T$, without any of the adaptive modifications, such as intercept corrections or additional differencing, needed following location shifts, as discussed in Hendry (2006). Even though these adaptations will be needed after breaks for $\Phi'_\perp \widehat{\mathbf{x}}_{T+h|T}$, there will either be fewer intercept corrections, or less of a loss from unnecessary over-differencing.

Most econometric models fall in the class of equilibrium corrections (including regressions, VARs, VEqCMs, and simultaneous systems, as well as GARCH processes for variances), so the most pernicious break is a shift in the equilibrium mean: the data will move location, but the model will revert to the in-built equilibrium. Above, we considered the relation between co-breaking and cointegration, and established that ‘common trends’ based on α_\perp are equilibrium-mean co-breaking, whereas cointegration vectors β are drift co-breaking. Thus, following any shifts in equilibrium means, the unfortunate implication is either they have to be eliminated (which α_\perp induces), or the new equilibrium mean has to be determined rapidly (which intercept corrections can facilitate). Conversely, as noted in the previous section, other parameter shifts are less likely to induce forecast failure.

4 Co-breaking regressions

This section describes an intuitive way of operationalizing co-breaking. The basic idea is that regression models may be used, first, to ascertain whether a given vector of random variables is subject to location shifts and, secondly, to investigate whether the shifts vanish in a linear combination of the variables. These models are henceforth referred to as co-breaking regressions, in analogy to cointegrating regressions in Engle and Granger (1987) and co-feature regressions in Engle and Kozicki (1993). Importantly, co-breaking regressions allow one to investigate whether location shifts in individual variables remain present in a *given number* of linear combinations, be they predicted by economic theory or estimated. They are not concerned with the question of *how many* co-breaking relationships exist among the variables of interest. This issue will be dealt with in section 5.

To formalize the co-breaking regression approach, consider a variant of the distribution of $\mathbf{x}_t | \mathbf{X}_{t-1}^1$ given in (15) of section 2.3, namely $\mathbf{x}_t | \mathcal{F}_t \sim \text{ID}[\varpi_t, \Sigma]$, where \mathcal{F}_t is a general conditioning information set that may contain, for instance, \mathbf{X}_{t-1}^1 or the σ -field generated by a set of exogenous variables \mathbf{w}_t . Note that $\mathbf{x}_t | \mathcal{F}_t$ is not necessarily Normally distributed. The conditional mean is again decomposed into the sum of an initial parameterization and deterministic shifts, i.e. $\varpi_t = \pi_0 + \pi_t$. In line with the discussion of section 2.2.2, the shift vector π_t is assumed not to change in every time period but to comprise only $k > n$ distinct values such that $(\pi_1 : \dots : \pi_T) = \kappa \mathbf{D}$, where κ is $(n \times k)$ and $\mathbf{D} = (\mathbf{d}_1 : \dots : \mathbf{d}_T)$ is $(k \times T)$; cf. equation (9). A linear regression model for \mathbf{x}_t may now be formulated as:

$$\mathbf{x}_t = \pi_0 + \kappa \mathbf{d}_t + \delta \mathbf{w}_t + \varepsilon_t \quad (37)$$

where $\mathbf{w}_t \in \mathcal{F}_t$ and $\varepsilon_t | \mathcal{F}_t \sim \text{IID}(\mathbf{0}_{(n \times 1)}, \Sigma)$. Assuming that $\text{rk}(\kappa) = n - 1$, κ may be decomposed into the product $\kappa = \xi \eta'$, where ξ is $(n \times (n - 1))$, η is $(k \times (n - 1))$ and both ξ and η are of full rank $\text{rk}(\xi) = \text{rk}(\eta) = n - 1$, see (10). As a consequence, there exists an $(n \times 1)$ vector $\xi'_\perp \xi = \mathbf{0}_{(1 \times (n - 1))}$. Thus, the shifts will no longer be present in the linear combination $\xi'_\perp \mathbf{x}_t = \xi'_\perp \pi_0 + \xi'_\perp \delta \mathbf{w}_t + \xi'_\perp \varepsilon_t$. Put differently, the co-breaking regression:

$$y_t = \xi'_{\perp, 1} \mathbf{z}_t + \tilde{\pi}_0 + \tilde{\delta} \mathbf{w}_t + \tilde{\varepsilon}_t \quad (38)$$

will be free of structural breaks. In (38), \mathbf{x}_t has been partitioned into a scalar component y_t and an $(n - 1)$ -dimensional vector \mathbf{z}_t , while the corresponding components of $\boldsymbol{\xi}_\perp$ are $(\xi_{\perp,0} : -\boldsymbol{\xi}'_{\perp,1})'$ and $\xi_{\perp,0}$ has been normalized to unity; the remaining quantities are defined as, for instance, $\tilde{\boldsymbol{\pi}}_0 = \boldsymbol{\xi}'_{\perp} \boldsymbol{\pi}_0$.

In practice, the co-breaking regression procedure may be implemented in two steps: First, test whether the k shifts \mathbf{d}_t are in fact present in every one of the n components of \mathbf{x}_t , that is, estimate a regression model such as (37) and test for the significance of $\boldsymbol{\kappa}$. Secondly, augment the conditional model of y_t on \mathbf{z}_t in (38) by \mathbf{d}_t and test whether the shifts are now insignificant, with the co-breaking vector $\boldsymbol{\xi}_\perp$ either estimated or imposed.

A number of specific implementations of this general procedure have been suggested in the literature: see Chapman and Ogaki (1993), Hendry and Mizon (1998) and Morana (2002). These approaches differ in three main aspects. First, the conditional distribution $\mathbf{x}_t | \mathcal{F}_t$ is specified with respect to different information sets, resulting, correspondingly, in conditional or unconditional co-breaking relationships. Secondly, the definition of the shifts \mathbf{d}_t varies and, thirdly, the models place different restrictions upon the dependence between y_t and \mathbf{z}_t , and thus different estimators and tests are employed to make inference in the conditional model in (38). Details of the three models will be provided in sections 4.1–4.3, emphasizing the generality of the regression approach to co-breaking, before they are assessed in section 4.4.

4.1 A piecewise-polynomial model

Setting $\mathcal{F}_t = \emptyset$ and thus effectively considering the unconditional distribution $\mathbf{x}_t \sim \text{ID}[\boldsymbol{\varrho}_t, \boldsymbol{\Sigma}]$, Chapman and Ogaki (1993) specify the regression model in (37) as the components model $\mathbf{x}_t = \boldsymbol{\varrho}_t + \boldsymbol{\varepsilon}_t$ or:

$$\mathbf{x}_t = \boldsymbol{\rho}_0 + \boldsymbol{\theta} \mathbf{d}_t + \boldsymbol{\varepsilon}_t, \quad (39)$$

such that $\boldsymbol{\varrho}_t = \boldsymbol{\rho}_0 + \boldsymbol{\rho}_t = \boldsymbol{\rho}_0 + \boldsymbol{\theta} \mathbf{d}_t$ is a piecewise polynomial function in t , see equation (14). If, for instance, the order of the polynomials is $s = 2$, then $\boldsymbol{\rho}_0$ is the sum of a baseline intercept $\boldsymbol{\rho}_{0,0}$, a baseline linear trend $\boldsymbol{\rho}_{1,0}$, and a baseline quadratic trend $\boldsymbol{\rho}_{2,0}$. Moreover, with q location shifts in each of the s components of $\boldsymbol{\rho}_t$ at times $T_1 < \dots < T_q$, the sample is effectively divided into $q + 1$ sub-samples, each potentially having a segmented trend of up to order s . As a consequence:

$$\boldsymbol{\theta} \mathbf{d}_t = \boldsymbol{\theta}_0 \mathbf{d}_{0,t} + \boldsymbol{\theta}_1 \mathbf{d}_{1,t} + \dots + \boldsymbol{\theta}_s \mathbf{d}_{s,t} \quad (40)$$

where $\mathbf{d}_{i,t} = (t_{T_1}^i : \dots : t_{T_q}^i)'$ is a $(q \times 1)$ vector with typical element $t_{T_j}^i = (t - T_j)^i 1_{\{t > T_j\}}$ while $\boldsymbol{\theta}_i = (\boldsymbol{\theta}_{i,1} : \dots : \boldsymbol{\theta}_{i,q})$ is $(n \times q)$, for $i \in \{0, 1, \dots, s\}$ and $j \in \{1, \dots, q\}$. It follows that $\boldsymbol{\theta}$ in (39) is of dimension $(n \times (s + 1)q)$ and \mathbf{d}_t is $((s + 1)q \times 1)$. The function in (40) is, without any zero restrictions on elements in $\boldsymbol{\theta}$ in (39), discontinuous at every single shift point; see the discussion in Massmann (2003a). Also, Chapman and Ogaki (1993) generalize this setup to piecewise polynomial functions in t that have different orders s_j in the $q + 1$ sub-samples.

To estimate s , Chapman and Ogaki (1993) suggest fitting polynomials of order $s + \varsigma$, $\varsigma > 0$, by means of least squares (LS) and using χ^2 tests to decide on the significance of the corresponding coefficient vectors. In order to implement the second stage of the procedure, concatenate first the $\boldsymbol{\theta}_i$ into $\boldsymbol{\theta}_{\mathcal{I}}$ and $\boldsymbol{\theta}_{\mathcal{I}^c}$ where \mathcal{I} is a subset of $\{1, \dots, (s + 1)q\}$ and \mathcal{I}^c is its complement, as in equation (13) of section 2.2.3, with \mathbf{d}_t partitioned accordingly into $\mathbf{d}_{\mathcal{I},t}$ and $\mathbf{d}_{\mathcal{I}^c,t}$. Assume then that $\text{rk}(\boldsymbol{\theta}_{\mathcal{I}}) = n - 1$ such that it may be decomposed into $\boldsymbol{\theta}_{\mathcal{I}} = \boldsymbol{\zeta} \boldsymbol{\nu}'$, and estimate the following variant of the co-breaking regression in (38) by LS:

$$y_t = \boldsymbol{\zeta}'_{\perp,1} \mathbf{z}_t + \boldsymbol{\zeta}'_{\perp} \boldsymbol{\rho}_0 + \tilde{\boldsymbol{\theta}} \mathbf{d}_t + \tilde{\boldsymbol{\varepsilon}}_t \quad (41)$$

where $\tilde{\boldsymbol{\theta}} = \boldsymbol{\zeta}'_{\perp} \boldsymbol{\theta}$ and $\tilde{\boldsymbol{\varepsilon}}_t = \boldsymbol{\zeta}'_{\perp} \boldsymbol{\varepsilon}_t$. If, using the χ^2 test, the coefficients of $\mathbf{d}_{\mathcal{I}^c,t}$ are shown to be insignificant then $\boldsymbol{\zeta}_{\perp}$ is orthogonal to $\boldsymbol{\theta}_{\mathcal{I}}$ and co-breaking for the elements of $\mathbf{d}_{\mathcal{I}^c,t}$ has occurred.

The trending behaviour of y_t , \mathbf{z}_t and \mathbf{d}_t will guarantee consistent estimation of the parameters in (41). Yet a problem arises because the LS approach suggested by Chapman and Ogaki (1993) neglects the endogeneity of \mathbf{z}_t . In the absence of a condition ensuring weak exogeneity of \mathbf{z}_t for $\zeta_{\perp,1}$ and $\tilde{\theta}$, the correlation between \mathbf{z}_t and $\tilde{\varepsilon}_t$ will cause the asymptotic distribution of the least-squares estimator of $\zeta_{\perp,1}$ to depend on nuisance parameters. This is analogous to the asymptotic median bias arising in cointegrating regressions; see Phillips and Loretan (1991) and Phillips (1989). As simulation studies by, for instance, Banerjee, Dolado, Hendry and Smith (1986) have shown, this bias may be severe in small samples and estimation or testing procedures corresponding to those suggested by Phillips and Hansen (1990) and Phillips and Park (1988), respectively, may be needed to ensure correct inference in (41).

Chapman and Ogaki (1993) use their algorithm to re-evaluate the stationarity of the real interest rate, as predicted by the Fisher relation. Using monthly US data for the nominal yield and the inflation rate between 1959-02 and 1990-12, the model in (39) was fitted to the data and the specification of \mathbf{d} was found to be a linear trend up until a location shift in $T_1 = 1979-10$ and a quadratic trend in the second sub-sample. In the second step, χ^2 tests examined the significance of estimates of $\tilde{\theta}$ with a view to reducing the order of the polynomials in t in the two sub-samples. Chapman and Ogaki (1993) circumvent the endogeneity problem by assuming the co-breaking vector to be known: $\zeta_{\perp} = (1 : -1)'$. However, the null hypothesis that the linear combination $\zeta_{\perp}' \mathbf{x}_t$ is stationary around (a) a breaking intercept, thus implying linear and quadratic trend co-breaking, and (b) a constant intercept, this being tantamount to intercept, linear trend and quadratic trend co-breaking, was rejected in both instances.

4.2 A multivariate Normal model

Hendry and Mizon (1998) suggest an alternative approach to the modelling of co-breaking regressions motivated by the general-to-specific methodology as outlined in, for instance, Hendry (1995) and Mizon (1995). The basic idea of this methodology may be described by two modelling stages. At a first stage, a multivariate Normal model is estimated unrestrictedly and considered a valid baseline for subsequent analysis if it explains the salient data features sufficiently well. One such feature would be location shifts that are apparent in constituent elements of the vector of random variables. Once this data congruency is achieved, the second modelling stage proceeds by reducing the general unrestricted model by means of restrictions to yield a statistically acceptable, yet more parsimonious, model which is easier to interpret. An example restriction is that implied by the weak exogeneity of a subset of variables for the parameters of interest, implying that the equations describing these variables within that sub-system may be disregarded without loss of information. The result is a model of the remaining variables conditional on the weakly exogenous ones.

In the present context, to implement the first stage of this strategy, consider the model in (37) and specify the information set as $\mathcal{F}_t = \mathbf{X}_{t-1}^{t-g}$ such that $\mathbf{w}_t = (\mathbf{x}'_{t-1} : \dots : \mathbf{x}'_{t-g})'$ and $\delta = (\mathbf{\Pi}_1 : \dots : \mathbf{\Pi}_g)$. The complete system is then estimated by means of maximum likelihood and the shifts \mathbf{d}_t determined in the process. The second stage of the modelling procedure then consists in testing the weak exogeneity of \mathbf{z}_t for the parameters of interest and, if the reduction is valid, in transforming the system into:

$$y_t = \xi_{\perp,1} \mathbf{z}_t + \tilde{\pi}_0 + \tilde{\kappa} \mathbf{d}_t + \sum_{i=1}^g \tilde{\Pi}_i \mathbf{x}_{t-i} + \tilde{\varepsilon}_t, \quad (42)$$

where the transformed parameters and variables are defined analogously to (38) above. If, in addition, \mathbf{z}_t is super exogenous in (42), then fewer dummy variables in \mathbf{d}_t are necessary to describe the data in the conditional model than in the original joint model and evidence for co-breaking has been found: section 3.1 provides a formal account of co-breaking in conditional models and exogeneity conditions.

The feasibility of economic policy analysis using co-breaking regressions is illustrated by Hendry and Mizon (1998), who estimate a bivariate model in money supply and interest rates. In particular, annual

UK data ranging from 1871–1993 were available for the following variables: broad money stock M , real net national income Y , the price level P , and the short-term nominal interest rate Rna . With lower case letters denoting the logarithm of the data, the variables of interest were $mpy = m - p - y$ and Rna , where mpy may be interpreted either as the log of inverse velocity or as long-run equilibrium, money demand with unit income elasticity imposed; see Hendry and Ericsson (1991) and Ericsson, Hendry and Prestwich (1998b). For the period between 1871 and 1975, the data set was compiled by Friedman and Schwartz (1982), and for the remaining years by Attfield, Demery and Duck (1995): these data series are also used in the analysis discussed in section 5.4.

The modelling procedure followed the steps outlined at the beginning of this section. First, a bivariate VAR(5) in mpy and Rna , supplemented by dummy variables, was estimated and shown to be data-congruent. Specifically, impulse dummies to account for potential outliers as well as step dummies for ‘policy regimes’ were included in the initial specification; see Hendry and Mizon (1998, Section 9). The only step dummies remaining significant in the data-congruent specification were those for 1876–1913, 1934–1943, and 1984–1993. Next, following results obtained in Ericsson *et al.* (1998b), one cointegration relationship was imposed such that the VAR representation could be mapped into a cointegrated VEqCM form. Since weak exogeneity of ΔRna for the parameters in the Δmpy equation was found, the analysis continued on the basis of the conditional model of Δmpy given ΔRna . Importantly, only one of the three step dummies, viz. that for 1984–1993, remained significant in that model, so that the final parsimonious and data-congruent model is interpreted to be a co-breaking relationship for the other two regimes.

4.3 A Markov-switching model

A third instance of the general procedure in Engle and Kozicki (1993) may be implemented in a Markov-switching framework. To that end, reconsider the i th component of the n -dimensional open system in (37):

$$x_{it} = \pi_{i0} + \boldsymbol{\kappa}^{(i)} \mathbf{d}\mathbf{i}_t + \boldsymbol{\delta}^{(i)} \mathbf{w}_t + \varepsilon_{it}, \quad (43)$$

where π_{i0} is the i th element of $\boldsymbol{\pi}_0$ while $\boldsymbol{\kappa}^{(i)}$ and $\boldsymbol{\delta}^{(i)}$ are the i th row of $\boldsymbol{\kappa}$ and $\boldsymbol{\delta}$, respectively. The $(k \times 1)$ variable $\mathbf{d}\mathbf{i}_t$ represents the mean shifts of x_{it} , but rather than being composed of deterministic indicator functions, it is now a stochastic break process based on a Markov chain. In particular, $\boldsymbol{\kappa}^{(i)} \mathbf{d}\mathbf{i}_t = \sum_{j=1}^k \boldsymbol{\kappa}_j^{(i)} di_{jt}$ where $di_{jt} = 1_{(s_t=j)}$ and the regime variable $s_t = 1, \dots, k$ follows an unobserved discrete Markov process with transition probabilities $\Pr(s_{t+1} = j_2 | s_t = j_1) = p_{j_1 j_2}$, and $\sum_{j_2=1}^k p_{j_1 j_2} = 1, \forall j_1$. Hamilton (1989) and Kim (1994) derive filtering and smoothing algorithms for the unobserved break process, while the parameters in (43) as well as the transition probabilities may be estimated by means of maximum likelihood; see also Krolzig (1997) for an exposition.

Let a significance test indicate that the estimated break process $\{\widehat{\mathbf{d}\mathbf{i}_t}\}$ is indeed a feature of x_{it} . Since $\{\widehat{\mathbf{d}\mathbf{i}_t}\}$ is, a priori, different for each component x_{it} , $i = 1, \dots, n$, a break process shared by all components can be extracted by estimating the full system in (37), namely $\mathbf{x}_t = \boldsymbol{\pi}_0 + \boldsymbol{\kappa} \mathbf{d}_t + \boldsymbol{\delta} \mathbf{w}_t + \boldsymbol{\varepsilon}_t$, with the states being restricted to be perfectly correlated across the variables. In order to see whether the so-derived $\{\widehat{\mathbf{d}}_t\}$ is a common feature, the null hypothesis $H_0: \text{rk}(\boldsymbol{\kappa}) = n - 1$ is again examined by estimating the model in (38), reproduced here for convenience:

$$y_t = \boldsymbol{\xi}'_{\perp,1} \mathbf{z}_t + \tilde{\boldsymbol{\pi}}_0 + \tilde{\boldsymbol{\delta}} \mathbf{w}_t + \tilde{\boldsymbol{\varepsilon}}_t. \quad (44)$$

The estimation algorithm is 2SLS, with $\{\widehat{\mathbf{d}}_t, \mathbf{w}_t\}$ being the set of instruments. Subsequently, the validity of the $\dim(\widehat{\mathbf{d}}_t) - \dim(\mathbf{z}_t) = 1$ over-identifying restrictions in the system may be tested using the procedure due to Sargan (1958) and Hansen (1982).

In an analysis of Eurozone inflation, Morana (2002) implements a variant of this procedure. While the endogeneity of \mathbf{z}_t in (44) is fully taken into account by his use of 2SLS, Morana (2002) does not discuss the problem of unidentified nuisance parameters under the null hypothesis of the feature test, thereby casting doubt on his empirical findings. To be more specific, a test of significance of $\hat{\kappa}^{(i)}$ in (43) is effectively a test of the number of regimes in the Markov chain. It is well-known, however, that under the restriction of fewer regimes, the transition probabilities pertaining to the regimes under the alternative are not identified, with the consequence that the distribution of the usual Wald, likelihood-ratio or Lagrange multiplier statistics is non-standard. Procedures accounting for that complication have been proposed by Hansen (1992) and Garcia (1992) and have, in different contexts, been successfully applied by, inter alia, Garcia and Perron (1996) and Evans (2003).

4.4 Discussion

This section discussed how co-breaking regressions may be used to estimate co-breaking relationships, taking their number as given. Co-breaking regressions are analogous to cointegrating regressions, and are effectively a special case of common feature regressions conceived by Engle and Kozicki (1993). Three approaches to co-breaking regressions were discussed in this section, which differed in three aspects. They considered a stochastic process $\{\mathbf{x}_t\}$ with respect to different information sets, they made recourse to different specifications of the break function, and they handled the endogeneity problem in the regression model in alternative fashions. First, Chapman and Ogaki (1993) model the unconditional process $\{\mathbf{x}_t\}$, assume the deterministic terms to be represented by a piecewise trend polynomial, and avoid the endogeneity issue by assuming the unconditional co-breaking vector to be known. Secondly, Hendry and Mizon (1998) consider a VAR model of \mathbf{x}_t , use impulse and step dummies to model location shifts, and test the exogeneity of the marginal process before estimating the co-breaking relationship. Finally, the Markov-switching approach advocated by Morana (2002) examines an open system of \mathbf{x}_t given $\mathbf{w}_t \in \mathcal{F}_t$, employs a Markov-switching model to estimate a stochastic break process, and uses instrumental variables estimation to allow for endogeneity in the conditional co-breaking regression.

These three approaches reflect different modelling traditions in econometrics. At the one end of the spectrum, the analysis of Chapman and Ogaki (1993) in terms of polynomial trend functions and ARMA processes is mainly descriptive, and thus in the Box–Jenkins time-series tradition. At the other end, Hendry and Mizon (1998) and Morana (2002) attempt both descriptive and prescriptive data modelling, emphasizing data-driven inference that does not neglect economic interpretability. The latter two approaches differ, however, in the way they conduct inference. Hendry and Mizon (1998) establish approximate Normality of the residuals in the system, and thus engage in exact inference, while the Markov-switching model relies on asymptotic distributions, both of the feature test by Hansen (1992) and Garcia (1992), and of the test for over-identifying restrictions.

5 Co-breaking rank analysis

In the present section, we relax the assumption made in the previous section that the number of co-breaking relationships is known and consider procedures that allow the estimation of the co-breaking rank. Once the rank has been inferred from the information contained in the data, a co-breaking regression analysis could in principle be used to estimate parameters of interest. As opposed to co-breaking regressions, however, co-breaking rank analysis is still in its infancy and, to the best of our knowledge, only three procedures that aim at estimating the order of co-breaking have so far been suggested in the literature. First, Bierens (2000) examines a model that under the null hypothesis of co-breaking does not contain a certain, unspecified, non-linear trend function. The non-parametric nature of this procedure,

however, is outside the scope of the present paper and thus not considered further. Instead, interest focuses on the analyses conducted by Krolzig and Toro (2002) and by Hatanaka and Yamada (2003). In particular, the procedure by Krolzig and Toro (2002) is based on a stationary vector autoregressive (VAR) model and involves a likelihood ratio (LR) test of the co-breaking rank hypothesis based on canonical correlations. This method will be reviewed in section 5.1. The procedure by Hatanaka and Yamada (2003), to be discussed in section 5.2, tests for the co-breaking rank using principal component analysis in an unobserved components model with an $I(1)$ stochastic term. In the course of the discussion in section 5.3, it will become clear, however, that both procedures are not satisfactory in all respects. In an attempt to improve on them, therefore, we suggest a new parametric procedure for co-breaking rank testing in section 5.4 that combines aspects of the two methodologies. In particular, we also employ an unobserved components model, albeit with an $I(0)$ stochastic component, and use canonical correlations to perform a LR test of the co-breaking rank hypothesis. We provide theoretical details of the procedure, Monte Carlo simulation evidence on the rank test performance, and an empirical application that re-visits the Stolper–Samuelson theorem.

5.1 A vector autoregressive approach

In their approach to co-breaking rank analysis, Krolzig and Toro (2002) analyze the n -dimensional conditional random variables $\mathbf{x}_t | \mathbf{X}_{t-1}^{t-g} \sim \text{NID}_n[\boldsymbol{\varpi}_t, \Sigma]$ with time varying conditional mean $\boldsymbol{\varpi}_t = \boldsymbol{\pi}_0 + \boldsymbol{\pi}_t = \boldsymbol{\pi}_0 + \boldsymbol{\kappa} \mathbf{d}_t$. The corresponding closed-form VAR(g) model in \mathbf{x}_t is:

$$\mathbf{x}_t = \boldsymbol{\pi}_0 + \boldsymbol{\kappa} \mathbf{d}_t + \sum_{i=1}^g \boldsymbol{\Pi}_i \mathbf{x}_{t-i} + \boldsymbol{\varepsilon}_t, \quad (45)$$

as, for instance, in Hendry and Mizon (1998); see section 4.2. The term $\boldsymbol{\pi}_0$ again comprises the deterministic components that have constant parameter representations while $\boldsymbol{\pi}_t = \boldsymbol{\kappa} \mathbf{d}_t$ models q level shifts in $\boldsymbol{\varpi}_t$, occurring in time periods $T_1 < \dots < T_q$. For convenience, these shifts are assumed to affect the intercept in (45) such that $\boldsymbol{\kappa}$ is of dimension $(n \times q)$ and \mathbf{d}_t is $(q \times 1)$ with typical element $1_{\{t > T_1\}}$. Importantly, the $(n \times n)$ autoregressive coefficient matrices $\boldsymbol{\Pi}_i$, $i = 1, \dots, q$, are such that the roots of the VAR polynomial are outside the unit circle so the vector process $\mathbf{x}_t \sim I(0)$.

As before, co-breaking in this setting is synonymous with $\boldsymbol{\kappa}$ being of reduced row rank $\text{rk}(\boldsymbol{\kappa}) = n - r$, since this allows the decomposition of $\boldsymbol{\kappa}$ into $\boldsymbol{\kappa} = \boldsymbol{\xi} \boldsymbol{\eta}'$, such that the orthogonal complement $\boldsymbol{\xi}_\perp$ of $\boldsymbol{\xi}$ yields the r linear combinations $\boldsymbol{\xi}'_\perp \mathbf{x}_t$ that, suitably normalized, no longer contain the deterministic shifts \mathbf{d}_t . To estimate the rank of $\boldsymbol{\kappa}$ in (45), Krolzig and Toro (2002) suggest using canonical correlation analysis, similar to the strategy pursued by Johansen (1988) for estimating the cointegrating rank in a VEqCM. Rewrite (45) as the reduced-rank regression model:

$$\mathbf{x}_t = \boldsymbol{\pi}_0 + \boldsymbol{\kappa} \mathbf{d}_t + \boldsymbol{\Gamma} \mathbf{w}_t + \boldsymbol{\varepsilon}_t, \quad (46)$$

where $\boldsymbol{\kappa} = \boldsymbol{\xi} \boldsymbol{\eta}'$ and the autoregressive component is captured by $\boldsymbol{\Gamma} \mathbf{w}_t$. Given the normality of $\boldsymbol{\varepsilon}_t$, the null hypothesis of $H_0: \text{rk}(\boldsymbol{\kappa}) \leq s$ may be tested against the unrestricted alternative of $H_1: \text{rk}(\boldsymbol{\kappa}) = \min(n, q)$ by means of a likelihood ratio (LR) test. Anderson (1951) shows that the test statistic reduces to the following expression:

$$\text{LR} = -T \sum_{i=s+1}^{\min(n,q)} \log(1 - \hat{\lambda}_i) \xrightarrow{d} \chi^2[(n-s)(q-s)] \quad (47)$$

where $\hat{\lambda}_1 \geq \dots \geq \hat{\lambda}_{\min(n,q)}$ are the estimated squared partial canonical correlations between \mathbf{x}_t and \mathbf{d}_t , corrected for the deterministic terms in $\boldsymbol{\pi}_0$ and for \mathbf{w}_t : also see Hotelling (1936) and Bartlett (1938). The

asymptotic distribution of LR in (47) is derived by Box (1949). Since the rank hypotheses are nested:

$$H_0(s = 0) \subset H_0(s \leq 1) \subset \cdots \subset H_0(s \leq s^* - 1) \quad (48)$$

where $s^* = \min(n, q)$, the test procedure commences with the most restrictive hypothesis and continues until the first null cannot be rejected.

Krolzig and Toro (2002) examine the small-sample properties of the LR test by conducting a number of Monte Carlo experiments. Specifically, they consider two generating processes for their data, both being given by a bivariate version of (45) with $q = 2$ breaks in the intercept: DGP1 specifies $\text{rk}(\kappa) = 2$, DGP2 imposes the restriction $\text{rk}(\kappa) = 1$. For a given degree of autocorrelation and given break magnitudes, Krolzig and Toro (2002) test the null hypothesis of $H_0: \text{rk}(\kappa) \leq 1$ for both DGPs. The results may be summarized by saying that although the test is generally oversized and has power problems at $T = 50$, these deficiencies are much alleviated as the sample size increases to $T = 150$. Similar results are obtained by varying the degrees of autocorrelation and the break dates in the DGP.

Given the rank test in (47), Krolzig and Toro (2002) devise two statistical tests for examining the hypothesis of whether the random variables conditioned upon in the co-breaking combinations $\xi'_\perp \mathbf{x}_t$ are super exogenous for the parameters of interest. Recall from section 3.1 that in addition to the reduced-rank condition on κ such that $\xi'_\perp \pi_t = \mathbf{0}, \forall t$, the crucial requirement for this to occur is that $\xi_\perp = (\mathbf{I}_r : -\mathbf{Y})$, where $\mathbf{Y} = C(\mathbf{y}_t, \mathbf{z}_t)V(\mathbf{z}_t)^{-1}$. The tests by Krolzig and Toro (2002) have slightly better small sample properties than the original test for super exogeneity suggested by Engle and Hendry (1993).

An empirical application of the Krolzig and Toro (2002) co-breaking rank procedure is provided by Schreiber (2004) who uses it to estimate US equilibrium unemployment. Using quarterly data on productivity growth and the unemployment rate with an effective sample ranging from 1950-3 to 2003-3 he finds that in a vector autoregressive model, both series are characterized by intercept shifts in 1973-1 and 1994-4. Due to the reduced rank of his estimate of κ in (46), however, a linear combination of the VAR equations is shown to be constant over time, thereby providing an empirical counterpart to the theoretical long-run concept of core unemployment.

5.2 An I(1) unobserved components model

Hatanaka and Yamada (2003) consider what we call an unobserved components (UC) representation of the n -dimensional random process $\{\mathbf{x}_t\}$:

$$\mathbf{x}_t = \boldsymbol{\varrho}_t + \mathbf{u}_t, \quad (49)$$

where $\boldsymbol{\varrho}_t = E[\mathbf{x}_t]$ is the time-varying unconditional mean of \mathbf{x}_t and \mathbf{u}_t is a mean-zero stochastic component. Referring to (49) as an UC model in fact slightly abuses terminology, since the deterministic $\boldsymbol{\varrho}_t$ is not unobserved. The model may, however, be generalized to include random break processes $\boldsymbol{\varrho}_t$ as in section 4.3, thus warranting this descriptor. The stochastic component \mathbf{u}_t is assumed to follow a general ARIMA process with order of integration of unity such that there exist r^* cointegrating relationships $\beta' \mathbf{u}_t \sim I(0)$, where β is $(n \times r^*)$. The deterministic component $\boldsymbol{\varrho}_t$ is generated by the equilibrium-correction process:

$$\Delta \boldsymbol{\varrho}_t = \chi \zeta'_\perp \boldsymbol{\varrho}_{t-1} + \boldsymbol{\omega} \mathbf{d}_{0,t} \quad (50)$$

where the parameters χ and ζ_\perp are $(n \times r)$ and of full rank $\text{rk}(\chi) = \text{rk}(\zeta_\perp) = r$ such that the eigenvalues of $(\mathbf{I}_r + \zeta'_\perp \chi)$ are inside the unit circle. The coefficient matrix $\boldsymbol{\omega}$ is $(n \times q)$ such that $\text{rk}(\boldsymbol{\omega}' \chi_\perp) = n - r$. The $(q \times 1)$ vector $\mathbf{d}_{0,t}$ is an ‘elementary trend function’ that takes the role of a disturbance term and will be specified further below. Solving (50) for $\boldsymbol{\varrho}_t$ yields:

$$\boldsymbol{\varrho}_t = \zeta \boldsymbol{\nu}' \mathbf{d}_{1,t} + \boldsymbol{\theta}_0 \mathbf{d}_{0,t} + \mathbf{e}_t \quad (51)$$

where $\boldsymbol{\nu}' = (\boldsymbol{\chi}'_{\perp} \boldsymbol{\zeta})^{-1} \boldsymbol{\chi}'_{\perp} \boldsymbol{\omega}$ is $((n-r) \times q)$,

$$\boldsymbol{\theta}_0 = \boldsymbol{\chi} (\boldsymbol{\zeta}'_{\perp} \boldsymbol{\chi})^{-1} \sum_{j=0}^{\infty} (\mathbf{I}_r + \boldsymbol{\zeta}'_{\perp} \boldsymbol{\chi})^j \boldsymbol{\zeta}'_{\perp} \boldsymbol{\omega},$$

is $(n \times q)$ while $\boldsymbol{\zeta}$ and $\boldsymbol{\chi}_{\perp}$ are the orthogonal complements of $\boldsymbol{\zeta}_{\perp}$ and $\boldsymbol{\chi}$, respectively. Finally, $\mathbf{d}_{1,t} = \sum_{s=1}^t \mathbf{d}_{0,s}$, and \mathbf{e}_t is an asymptotically negligible remainder term. If now $\mathbf{d}_{0,t}$ comprises a constant term and indicator variables representing the timing of $(q-1)$ location shifts in time periods $T_1 < \dots < T_{q-1}$ such that $\mathbf{d}_{0,t} = (1 : 1_{\{t \geq T_1\}} : \dots : 1_{\{t \geq T_{q-1}\}})'$, then $\mathbf{d}_{1,t} = (t : (t-T_1) 1_{\{t \geq T_1\}} : \dots : (t-T_{q-1}) 1_{\{t \geq T_{q-1}\}})'$ constitutes a linear trend and trend shifts. From the expression in (51) it is clear that there are $(n-r)$ common deterministic trends $\boldsymbol{\nu}' \mathbf{d}_{1,t}$ in the system while (50) shows that there are r co-trending combinations $\boldsymbol{\zeta}'_{\perp} \boldsymbol{\varrho}_t$. In the ensuing analysis, the possibility that $r = n$ is disallowed since this would imply that there is no deterministic trend and no trend shifts in $\boldsymbol{\varrho}_t$, thereby defeating the purpose of the exercise.

The task set by Hatanaka and Yamada (2003) is to determine the co-breaking rank r as well as to estimate the number of co-trending vectors that are also cointegrating, say r_1 . To that end, they note that the estimated principal components $\hat{c}_{i,t}$, $i = 1, \dots, n$, of \mathbf{x}_t , computed on the basis of the scaled data covariance matrix:

$$\boldsymbol{\Omega} = T^{-3} \sum_{t=1}^T (\mathbf{x}_t - \bar{\mathbf{x}}) (\mathbf{x}_t - \bar{\mathbf{x}})',$$

where $\bar{\mathbf{x}} = T^{-1} \sum_{t=1}^T \mathbf{x}_t$, have different orders of magnitude, depending on whether their leading term contains a deterministic trend, a unit root or neither of the two. Specifically, defining $q \sim O_p(T^c)$ as meaning that $\sum_{t=1}^T a_t^2 \sim O_p(T^c)$ as $T \rightarrow \infty$, the authors show that

$$\hat{c}_{i,t} \sim \begin{cases} O_p(T^{3/2}), & \text{for } i = 1, \dots, n-r \\ O_p(T), & \text{for } i = n-r+1, \dots, n-r_2 \\ O_p(T^{1/2}), & \text{for } i = n-r_2, \dots, n. \end{cases} \quad (52)$$

The number of principal components that belong in each of these three groups is determined by the dimension of $c(\boldsymbol{\zeta})$, $c(\boldsymbol{\zeta}_{\perp,2}) = c(\boldsymbol{\zeta}_{\perp}) \cap c(\boldsymbol{\beta}_{\perp})$ and $c(\boldsymbol{\zeta}_{\perp,1}) = c(\boldsymbol{\zeta}_{\perp}) \cap c(\boldsymbol{\beta})$, respectively.

Given the difference in the principal components' orders of magnitude, Hatanaka and Yamada (2003) suggest a two-stage procedure of tests for unit roots and deterministic trends to estimate $n-r$ and r_1 , and thereby r and r_2 . At the first stage, each principal component is examined for a unit root. It is shown that both univariate augmented Dickey–Fuller (ADF) tests, see Dickey and Fuller (1979), Said and Dickey (1984) and Ng and Perron (1995), as well as system tests developed in Johansen (1988, 1991, 1994) remain asymptotically well-behaved despite the non-standard deterministic terms the principal components contain; see also Perron (1989), Inoue (1999) and Johansen *et al.* (2000). That is to say, the null hypothesis of a unit root is rejected with asymptotic probability one when the principal component lies either in $c(\boldsymbol{\zeta})$ and $\boldsymbol{\zeta}$ are cointegrating vectors, or in $c(\boldsymbol{\zeta}_{\perp,1})$. Similarly, the asymptotic acceptance probability of the null is equal to the significance level when the principal component lies either in $c(\boldsymbol{\zeta})$ and $\boldsymbol{\zeta}$ are not cointegrating vectors, or in $c(\boldsymbol{\zeta}_{\perp,2})$. However, the tests are unable to differentiate between these two sub-cases under either null or alternative.

In order to sub-divide further each of these two categories, Hatanaka and Yamada (2003) devise two tests for a deterministic trend, to be used at the second stage of the analysis. The first of these tests uses the fact that the $I(0)$ principal components that lie in $c(\boldsymbol{\zeta})$, when $\boldsymbol{\zeta}$ are cointegrating vectors, and those that lie in $c(\boldsymbol{\zeta}_{\perp,1})$, have different orders of magnitude, namely $O(T^{3/2})$ and $O(T^{1/2})$, respectively. Similarly, the second test relies on the $I(1)$ principal components in $c(\boldsymbol{\zeta})$, when the columns of $\boldsymbol{\zeta}$ are not cointegrating vectors, to have a deterministic trend while those in $c(\boldsymbol{\zeta}_{\perp,2})$ have not. A complication

only arises because the asymptotic distribution of the second test statistic depends on $c(\nu')$ under the null. Estimating this quantity by minimizing the residual sum of squares of the reduced-rank regression in (51), with \mathbf{e}_t assumed IID, the asymptotic distribution of the second test statistic may be tabulated, although it is now data-dependent. Hatanaka and Yamada (2003) show, however, that both tests for a deterministic trend are consistent and have asymptotic power of unity.

As estimates of $n-r$ and r_1 , and thus r_2 , they suggest taking the values implied by the first hypothesis in the sequence $H_0(n-r=i \wedge r_1=j)$, where j takes values $j=0, \dots, n-i$, for every value $i=1, \dots, q$. This sequence lists all theoretically consistent hypotheses about the three-way partitioning of \mathbf{x}_t 's range space. In particular, the hypothesis $H_0(n-r=i \wedge r_1=j)$, for $i=1, \dots, q$ and $j=0, \dots, n-i$, means that $c(\zeta)$ is spanned by the principal components corresponding to the largest i eigenvalues, that $c(\zeta_{\perp,2})$ is spanned by the principal components corresponding to the $n-i-j$ next largest eigenvalues, and $c(\zeta_{\perp,1})$ is spanned by the remaining j principal components. Put differently, H_0 implies that each principal component in $c(\zeta)$ is $O(T^{3/2})$, each in $c(\zeta_{\perp,2})$ is $O_p(T)$ and each in $c(\zeta_{\perp,1})$ is $O_p(T^{1/2})$. Thus, to decide whether or not H_0 is true, each of the principal components needs to be subjected to the unit root and trend tests described above. Hatanaka and Yamada (2003) show that this decision rule rejects all false null hypotheses with probability one provided that T is sufficiently large.

In order to ascertain the merits of the above decision rule, Hatanaka and Yamada (2003) conduct some small-scale Monte Carlo simulations. Their DGPs are of dimension $n=2$, but differ in the values of r_1 and r^* . Since $n-r=0$ is ruled out by assumption, the first principal component is a basis vector of $c(\zeta)$ by default, and only the second principal component needs classifying into one of $c(\zeta)$, $c(\zeta_{\perp,2})$ and $c(\zeta_{\perp,1})$. Thus, their sequence of hypotheses only consists in subjecting the second principal component to the unit-root null and the null of no deterministic trend. For a sample size of $T=800$, the empirical rejection frequencies are close to their analytic asymptotic counterparts, under the null as well as under the alternative. For $T=100$, however, serious small-sample problems become apparent. As an example, in the case of the DGP whose second principal component lies in $c(\zeta)$ and for which ζ is not cointegrating, the unit-root hypothesis is accepted far too often. It may be suspected that this is a reflection of the well-known problem of weak power of unit-root tests in small samples due to the presence of a deterministic trend; see, for instance, DeJong, Nankervis, Savin and Whiteman (1992). Moreover, in the case of the DGP whose second principal component contains a stochastic as well as a deterministic trend, the null hypothesis of no deterministic trend is not rejected often enough. The reason for this could be that the data-dependent critical values for the test are derived using the wrong rank of ν' in the estimation of the reduced-rank regression in (51).

5.3 Discussion

In summary, the co-breaking rank procedure suggested by Krolzig and Toro (2002) has the attraction that it is based on a likelihood ratio test whose asymptotic properties are known and whose small sample performance is shown to be satisfactory. Note, however, that what the procedure derives are *conditional* co-breaking relationships, see Definition 11. If one were interested in *unconditional* co-breaking combinations, however, it is not immediately apparent how these may be obtained on the basis of the approach considered by Krolzig and Toro (2002). To see the difficulty, consider the VAR model of \mathbf{x}_t in (45), and suppose κ is of reduced rank r such that $\kappa = \xi\eta'$. Then the $(n-r)$ co-breaking relationship may be found by premultiplying the VAR equations $\Pi(L)\mathbf{x}_t = \pi_0 + \xi\eta'\mathbf{d}_t + \varepsilon_t$ by ξ'_{\perp} , where $\Pi(L)$ is the autoregressive lag polynomial. Yet solving this expression for a meaningful co-breaking combination in the process $\{\mathbf{x}_t\}$ is by no means trivial since the inverse of $\Pi(L)$ is, generally, a polynomial of infinite order, making $\mathbf{x}_t = [\Pi(L)]^{-1}(\pi_0 + \xi\eta'\mathbf{d}_t + \varepsilon_t)$ an intricate function in lagged values of \mathbf{d}_t .

Hatanaka and Yamada (2003), in turn, suggest a procedure for simultaneously determining the co-breaking rank and the cointegrating rank in an I(1) UC model. An attractive feature of their analysis

is that the deterministic component $\boldsymbol{\varrho}_t$ follows the equilibrium-correction mechanism in (50), rendering unnecessary the assumption that more breaks are needed than the system has variables; see the discussion in section 2.2.2. Three aspects of the suggested procedure, however, are not appealing. First, the setup of $\boldsymbol{\varrho}_t$ can only cope with location shifts in the trend component, and it is not clear how intercept breaks may be cancelled by co-breaking combinations. Secondly, the analysis relies on the assumption that the stochastic component of the model is $I(1)$ and that the deterministic component contains a linear trend, which may be seen as somewhat restrictive. Lastly, the tests for deterministic trend appear largely ad hoc and are not shown to be based on an optimality principle. More practically, one of them has a non-standard distribution whose critical values need to be tabulated for each given data set, and their Monte Carlo hints at some unsatisfactory small-sample properties of the proposed decision rule.

In an attempt to improve on Krolzig and Toro (2002) and Hatanaka and Yamada (2003), the following section suggests a new procedure for estimating the co-breaking rank. Our approach will follow the latter authors in using a UC model for the process $\{\mathbf{x}_t\}$ so as to estimate *unconditional* co-breaking relationships. Yet we employ an LR test based on canonical correlations to test for the co-breaking rank, as do the former authors.

5.4 An $I(0)$ unobserved components model

5.4.1 Interpretation and estimation

Consider the UC model $\mathbf{x}_t = \boldsymbol{\varrho}_t + \mathbf{u}_t$, where $\boldsymbol{\varrho}_t$ is a general deterministic component and \mathbf{u}_t is a purely stochastic component. The attraction of this representation is that it is the process $\{\mathbf{x}_t\}$ itself that is modelled, not a linear transformation of it, as in the case of a vector autoregressive representation of \mathbf{x}_t . Thus, the co-breaking vectors $\boldsymbol{\Phi}$ would cancel out all or some constituent parts of the mean of \mathbf{x}_t and are hence, in the terminology of Section 2.2.3, *unconditional*.

The specification of $\boldsymbol{\varrho}_t$ could be as general as a piecewise polynomial function in t , as in the discussion of Chapman and Ogaki (1993), but, with a view to the empirical analysis below, we will restrict ourselves to considering q shifts in the linear trend of \mathbf{x}_t , occurring at times $T_1 < \dots < T_q$. Hence, we assume that $\boldsymbol{\varrho}_t = \boldsymbol{\rho}_0 + \boldsymbol{\theta}_1 \mathbf{d}_{1,t}$, where $\boldsymbol{\rho}_0$ consists of a constant intercept and a constant linear time trend, and $\boldsymbol{\theta}_1 \mathbf{d}_{1,t}$ models the trend breaks such that $\boldsymbol{\theta}_1$ is $(n \times q)$ and $\mathbf{d}_{1,t} = ((t - T_1)1_{\{T_1+1 \leq t \leq T_2\}} : \dots : (t - T_q)1_{\{T_q+1 \leq t\}})'$ is $(q \times 1)$. This definition of trend shifts is a linear transformation of that used in section 5.2, and allows the coefficient pertaining to the j th element of $\mathbf{d}_{1,t}$, $j \in \{1, \dots, q\}$, to be interpreted as the trend difference relative to the baseline. To avoid the difficulties in distinguishing between deterministic and stochastic trends we assume that the random component of \mathbf{x}_t follows a stationary AR(g) process: $\mathbf{u}_t = \sum_{i=1}^g \boldsymbol{\Pi}_i \mathbf{u}_{t-i} + \boldsymbol{\varepsilon}_t$ where $\boldsymbol{\varepsilon}_t \sim \mathbf{N}[\mathbf{0}, \boldsymbol{\Sigma}]$, and $\det(\boldsymbol{\Pi}(z)) = 0$ implies $|z| > 1$. The process $\{\mathbf{x}_t\}$ is thus modelled as stationary fluctuations around a segmented trend function, see also Rappoport and Reichlin (1989) and Johansen *et al.* (2000).

Following the discussion in previous sections, a $(n \times r)$ -dimensional matrix $\boldsymbol{\Phi}$ of trend co-breaking vectors such that $\boldsymbol{\Phi}'\boldsymbol{\theta}_1 = \mathbf{0}$ exists if $\text{rk}(\boldsymbol{\theta}_1) = r < \min(n, q)$. As a result, $\boldsymbol{\theta}_1$ may be decomposed into $\boldsymbol{\theta}_1 = \boldsymbol{\zeta}\boldsymbol{\nu}'$, and the trend co-breaking relationships are given by $\boldsymbol{\zeta}'_{\perp}\mathbf{x}_t$. Imposing the reduced-rank restriction on the deterministic component, and re-writing the dynamics of the UC model in terms of observables, yields the isomorphism:

$$\begin{aligned} \mathbf{x}_t = & \sum_{i=1}^g \boldsymbol{\Pi}_i \mathbf{x}_{t-i} + \boldsymbol{\Pi}(1) \boldsymbol{\pi}_0 + \boldsymbol{\zeta}\boldsymbol{\nu}'\mathbf{d}_{1,t} \\ & + \sum_{i=1}^g \boldsymbol{\Pi}_i \boldsymbol{\zeta}\boldsymbol{\nu}'\mathbf{d}_{1,t-i} + \boldsymbol{\varepsilon}_t, \end{aligned} \tag{53}$$

where all parameters are freely varying. This representation will be referred to as a restricted VAR (RVAR) model due to the common factor restriction on the compound coefficients of the intercept and of $\mathbf{d}_{1,t-i}$, $i = 1, \dots, g$.

There are, however, two problems with estimating this UC model or, equivalently, the RVAR model in (53). First, given $\text{rk}(\boldsymbol{\theta}_1) = r$, the parameters are poorly identified because the regressors $\mathbf{d}_{1,t-i}$, $i = 0, 1, \dots, g$, are highly collinear. Secondly, r is unknown and it is not clear which analytical method may be able to determine it. Numerical optimization procedures, in turn, are not appealing, given the multitude of reduced-rank conditions in (53).

To circumvent these problems, we suggest the following procedure: the RVAR in (53) is generalized to yield an unrestricted VAR (UVAR) in \mathbf{x}_t which is more easily estimated. Subsequently, by imposing restrictions on the UVAR parameters, the coefficients of interest may be retrieved. To be more specific, consider the UVAR model:

$$\mathbf{x}_t = \sum_{i=1}^g \boldsymbol{\Pi}_i \mathbf{x}_{t-i} + \boldsymbol{\psi}_0 + \boldsymbol{\psi}_{01} \mathbf{e}_{0,t} + \boldsymbol{\psi}_1 \mathbf{e}_{1,t} + \boldsymbol{\psi}_f \mathbf{f}_t + \boldsymbol{\varepsilon}_t \quad (54)$$

where all parameters are unrestricted for the time being. The regressors $\mathbf{e}_{0,t}$ and $\mathbf{e}_{1,t}$ are of dimension $(q \times 1)$ while \mathbf{f}_t is $(qg \times 1)$, and their respective representative elements are $e_{0,i,t} = 1_{\{T_i+g+1 \leq t \leq T_{i+1}\}}$, $e_{1,i,t} = t 1_{\{T_i+g+1 \leq t \leq T_{i+1}\}}$ and $f_{i,j,t} = 1_{\{t=T_i+j\}}$, where $i = 1, \dots, q$ and $j = 1, \dots, g$. The idea of this representation is to divide the sample into $q + 1$ sub-samples, and model each as potentially having its own intercept and linear trend; see $\mathbf{e}_{0,t}$ and $\mathbf{e}_{1,t}$. Importantly, the first g observations of each sub-sample are regarded as transition periods following a location shift and are dummied out by \mathbf{f}_t . As a result, $\mathbf{e}_{1,t}$ effectively only models the slopes of the broken trend, while the intercept shifts associated with the trend shifts are captured by $\mathbf{e}_{0,t}$ and \mathbf{f}_t . The UVAR in (54) may thus be considered a generalized version of the segmented trend model discussed above.

The UVAR being more general than the RVAR, restrictions are imposed on the parameters of the former to obtain the latter. Using simple algebraic manipulations, the RVAR is obtained by setting $\boldsymbol{\psi}_0 = \boldsymbol{\Pi}(1) \boldsymbol{\pi}_0$, $\boldsymbol{\psi}_{01} = \boldsymbol{\psi}_{01}(\boldsymbol{\pi}_0, \boldsymbol{\Pi}_1, \dots, \boldsymbol{\Pi}_g)$,

$$\boldsymbol{\psi}_1 = \boldsymbol{\Pi}(1) \boldsymbol{\theta}_1, \quad (55)$$

and $\boldsymbol{\psi}_f = \boldsymbol{\psi}_f(\boldsymbol{\pi}_0, \boldsymbol{\Pi}_1, \dots, \boldsymbol{\Pi}_g)$: Massmann (2003a) provides details of the transformations. The coefficient matrices $\boldsymbol{\psi}_{01}$ and $\boldsymbol{\psi}_f$ are complicated functions of RVAR parameters. If, however, $\boldsymbol{\psi}_1$ is of reduced rank r such that $\boldsymbol{\psi}_1 = \boldsymbol{\zeta} \boldsymbol{\nu}'$, where $\boldsymbol{\zeta}$ and $\boldsymbol{\nu}$ are of full rank r and of dimension $(n \times r)$ and $(q \times r)$, respectively, then from (55), $\boldsymbol{\theta}_1 = \boldsymbol{\zeta} \boldsymbol{\nu}'$ is obtained by imposing $\boldsymbol{\zeta} = \boldsymbol{\Pi}(1) \boldsymbol{\zeta}$, provided $\boldsymbol{\Pi}(1)$ is non-singular.

The estimation of the UVAR model in (54) may be conducted along the lines of Krolzig and Toro (2002), see section 5.1. In particular, the sequence of rank hypotheses given by $H_0: \text{rk}(\boldsymbol{\psi}_1) = r$, for $r = 0, \dots, \min(n, q) - 1$, may be implemented using the LR test based on the canonical correlations between \mathbf{x}_t and $\mathbf{e}_{1,t}$, corrected for the remaining explanatory variables in (54). The test statistic is analogous to LR given in equation (47). Once the co-breaking rank r has been determined, the parameters in the model may be estimated by maximum likelihood.

5.4.2 Monte Carlo experiments

To examine the small-sample distribution of the LR statistic, we carried out a set of Monte Carlo experiments. The artificial data were generated by the UC model in (54), with $n = 3$, $q = 4$, $g = 1$ and $r = 1$. Of particular interest is the impact on LR of varying degrees of autocorrelation in \mathbf{x}_t so that in addition to the baseline case of \mathbf{u}_t being a serially uncorrelated series with $\boldsymbol{\Pi}_1 = \mathbf{0}$, three magnitudes of positive

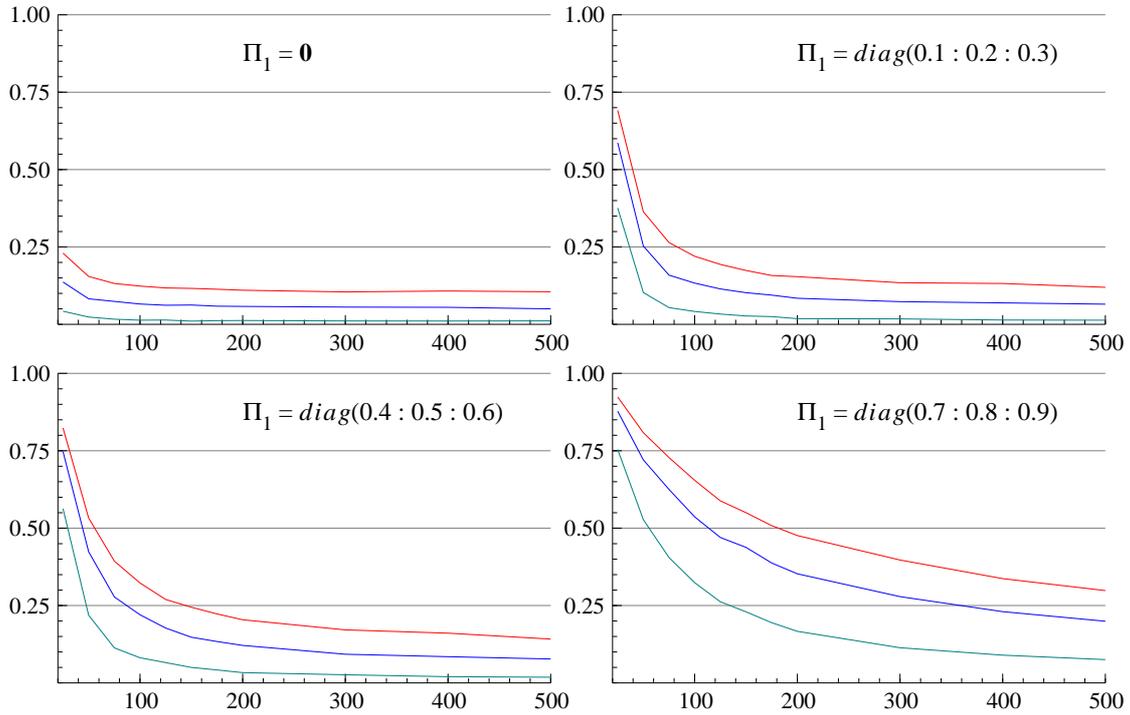


Figure 1: Results of Monte Carlo experiment. The four panels show the actual null rejection frequency of the LR test against the sample sizes in T^* for the specified autoregressive coefficient Π_1 . The three lines in each panel are for the 10%, 5% and 1% significance levels, respectively.

autocorrelation were examined, namely $\Pi_1 = \text{diag}(0.3 : 0.2 : 0.1)$, $\Pi_1 = \text{diag}(0.6 : 0.5 : 0.4)$, and $\Pi_1 = \text{diag}(0.9 : 0.8 : 0.7)$. The Monte Carols were recursive with the following sample sizes: $T^* = \{25, 50, 75, 100, 125, 150, 175, 200, 300, 400, 500\}$. The magnitudes of the trend breaks were chosen relative to the noise in the system, see Hendry and Ericsson (1991) and Ericsson *et al.* (1998b), but not scaled by the sample size. Their timing was proportional to the sample size: $\{0.2T, 0.4T, 0.6T, 0.8T\}$. The number of replications was $M = 5,000$ and one stream of random numbers was used for all replications to facilitate inter-estimator comparability. The results were generated using Ox 3.30, see Doornik (2001).

The statistical models that were estimated were special cases of (54) above. They were correctly specified in that trend break dummies were included and the correct lag length was selected for the endogenous variable. The LR statistic was computed, evaluated at the 10%, 5% and 1% nominal critical values of its asymptotic distribution and the rejection frequencies recorded. The null $H_0: \text{rk}(\psi_1) = 0$ was rejected in all replications at all significance levels. This is reassuring as it means that the trend breaks dummies are detected as significant regressors in explaining the data. As regards the ability of the test to reject the correct null $H_0: \text{rk}(\psi_\ell) \leq 1$, however, the test is oversized, and increasingly so as the degree of autocorrelation in the data increases: see Figure 1. Thus, the greater the autocorrelation in the data, the greater is the probability of concluding that the rank of ψ_ℓ is not reduced, i.e. that no co-breaking is present.

	$\hat{\lambda}_i$	LR	df	p-val
$H_0: r_0 = 0$	0.2147	57.9775	18	0.0000
$H_0: r_0 \leq 1$	0.1775	30.1863	10	0.0008
$H_0: r_0 \leq 2$	0.0648	7.7135	5	0.1027

Table 1: *Results of rank test. The table shows row-wise the results of the reduced-rank tests on θ_1 . The second column presents the estimated squared partial canonical correlations, the third column shows the value of the associated LR test statistic, the fourth column displays the degrees of freedom of the corresponding asymptotic distribution, with the resulting p-value listed in column five.*

5.4.3 Revisiting the Stolper–Samuelson Theorem

Our illustrative empirical application uses variables from the same UK data set as those used by Hendry and Mizon (1998) in section 4.2 above. The data series were also previously analyzed by Hendry (2000a, 2001). The variables of interest are the UK GDP deflator, an index of nominal UK wages, and world prices in sterling, such that the dimension of the system is $n = 3$. The three variables are measured in logarithms and denoted by pg , w , and pwe , respectively. Annual observations are available between 1872 and 1990 so that the sample size is $T = 119$. Modelling this data set has two attractions: first, it spans a time period that may arguably be described as ‘the long-term’, at least in an economic sense. It thus warrants the estimation of ‘long-run’ relations such as co-breaking combinations, and, at the same time, allows the identification of large ‘intermittent’ breaks. Secondly, the transformations $(w - pg)$ and $(pwe - pg)$ represent real UK wages and real world prices, respectively, two key ingredients in the Stolper–Samuelson Theorem. An interpretation of the results in this context, and a re-appraisal of the Theorem, may thus be attempted.

After graphing the data, $q = 6$ trend break points are identified in the following time periods, consistent with major ‘interventions’ in the past; see Box and Tiao (1975): 1914 and 1920 to denote World War I and its aftermath; 1932 to model the devaluation of the US dollar against gold; 1939 and 1949 for World War II; and, finally, 1974 to denote the oil crisis. In consequence, the break points are assumed known, i.e., deterministic, unlike in the analysis by, for instance, Hamilton (1989). For present purposes, testing for unknown break points as proposed in the literature by Bai and Perron (1998) is arguably not necessary since the model is meant to be illustrative: estimated break points with confidence intervals make the suggested procedure based on dummy variables difficult to adapt. The data are displayed in the first three panels of Figure 2, with the break points superimposed.

The empirical analysis proceeded in two stages. At the first stage, the UVAR model in (54) was fitted such that the residuals, individually and as vectors, were approximately Normally distributed and not autocorrelated; see Rao (1951) for an F-approximation to the asymptotic χ^2 distribution of the autocorrelation test, and Doornik and Hansen (1994) for the vector Normality test. This meant including four lags of the endogenous variable such that $g = 4$. At the second stage, the nested sequence of hypotheses $H_0: \text{rk}(\theta_1) \leq r$, for $r = 1, \dots, \min(n, q)$ was tested by applying partial canonical correlation analysis. The computations were carried out using COBRA, an OxPack module for GiveWin: see Massmann (2003b) and Doornik and Hendry (2001).

The estimated squared partial canonical correlations between \mathbf{x}_t and $\mathbf{d}_{1,t}$ as well as the resulting LR test statistics and associated p-values are shown in Table 1. Importantly, the residuals of the restricted model have remained approximately Normal and serially uncorrelated. The estimated rank of ψ_1 is $\hat{r} = 2$ such that the resulting dimensions of $\hat{\zeta}$ and $\hat{\nu}$ are (3×2) and (6×2) , respectively, so there are two common deterministic shifts $\nu' \mathbf{d}_{1,t}$ in the system. The eigenvalues of the estimated long-run matrix $\hat{\Pi}(1)$ are inside the unit circle, such that $\hat{\Pi}(1)$ is non-singular and the unconditional trend co-breaking vector $\hat{\zeta}_\perp$ may be derived as the orthogonal complement to $\hat{\zeta} = \hat{\Pi}(1)^{-1} \hat{\zeta}$, viz. $\hat{\zeta}_\perp = (-7.91 \ 4.72 \ 1.00)'$.

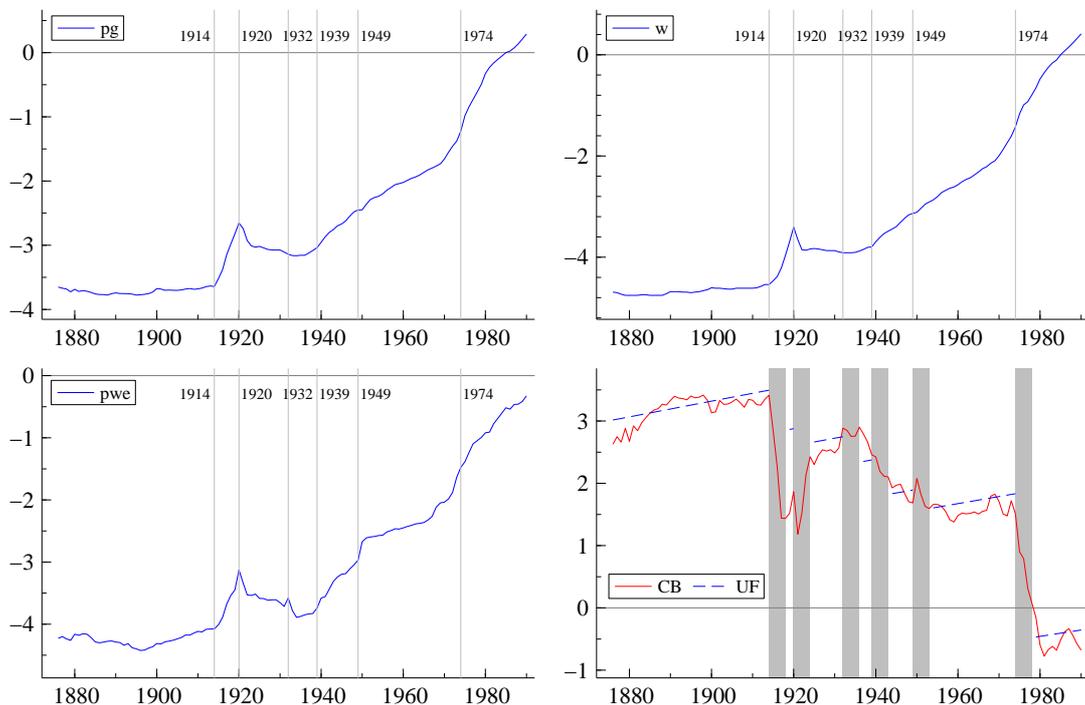


Figure 2: *Empirical results. The top-left, top-right and bottom-left panels show the log of the UK GDP deflator (pg), nominal UK wages (w) and world prices (pwe), respectively. The six common break points in 1914, 1920, 1932, 1939, 1949 and 1974 are superimposed. In the bottom-right panel the solid line (CB) displays the estimated trend co-breaking relationship $\tilde{\zeta}'_{\perp} \mathbf{x}_t$ while the dashed line (UF) shows its estimated mean unconditional fit $\tilde{\zeta}'_{\perp} \hat{\psi}_0 + \tilde{\zeta}'_{\perp} \hat{\psi}_{01} \mathbf{e}_{0,t}$ with the transition periods \mathbf{f}_t , shaded in grey, discarded so as not to clutter the graph, see equation (54).*

Consequently, the trend co-breaking relationship is given by:

$$\tilde{\zeta}'_{\perp} \mathbf{x}_t = -7.91pg_t + 4.72w_t + pwe_t \quad (56)$$

and is displayed in the fourth panel of Figure 2. The linear combination has a non-zero intercept and significant baseline trend. It is, however, no longer subject to slope breaks, but only affected by the intercept breaks implicit in the trend breaks. The hypothesis that the independent columns of ψ_{01} lie in the span of ζ was rejected.

For an economic interpretation of the estimated co-breaking relationship in (56), suppose there existed a procedure which allowed us to test the hypothesis that the unconditional co-breaking space is spanned by the vector $\tilde{\zeta}'_{\perp} = (-6 : 5 : 1)$. If that hypothesis were not rejected, then the restricted co-breaking relationship would be given by:

$$\tilde{\zeta}'_{\perp} \mathbf{x}_t = 5 \cdot (w_t - pg_t) + 1 \cdot (pwe_t - pg_t). \quad (57)$$

Three conclusions could be drawn on the basis of (57). First and foremost, it is a long-run relationship between real wages in the UK and real world prices measured in sterling which has remained stable, i.e., been unaffected by trend breaks, over the past 119 years. Secondly, the UK real wage and real

world prices are inversely related: when there is a 1% increase in real world prices in sterling, then real wages decrease by 5% on average. Finally, if world prices are taken to be a proxy for import prices, then an interpretation of (57) in terms of the well-known theorem by Stolper and Samuelson (1941) may be attempted. This theorem says that an increase in import prices, perhaps due to the imposition of a tariff, leads to an increase in that factor's price which is used intensively in the import-competing industry. This is a statement about the long term, since the factors, usually taken to be labour and capital, are assumed to be completely mobile; see Dixit and Norman (1980) for a detailed exposition. If the present dataset supports the hypothesis that an increase in the price of imports leads to a fall in real wages, doubt is cast on the applicability of the Stolper–Samuelson theorem. To see this, note that, on the one hand, imports into the UK are generally considered labour-intensive while, on the other hand, the UK labour share is reported by Bentolila and Saint-Paul (2003) to have fluctuated around 70% over the last 40 years and is also relatively constant in our data set. Thus the Stolper–Samuelson theorem would not predict an inverse relationship between world prices and the real wage, as found in the data.

6 Conclusion

Co-breaking relationships are linear combinations of the form $\Phi'x_t$ that depend on fewer breaks in their deterministic components than x_t on its own. Although a number of results pertaining to this idea appear in the literature, no comprehensive account of them has yet been given. A first objective of the paper was to provide a synopsis of the literature on co-breaking. In particular, we offered an overview of the theoretical concept and its properties, delimiting co-breaking to cointegration and common features, and reviewing recent contributions to co-breaking regressions and co-breaking rank analysis.

The second objective of the paper was to present new results in the field. Therefore, we discussed the rôle of co-breaking in empirical research in general, and in impulse-response analysis in particular. Moreover, in our discussion of co-breaking rank tests, we concluded that none of the existing procedures as yet is entirely satisfactory. This is partly because attempting to account for both constant and breaking components of deterministic as well as stochastic trends is ambitious, although developing congruent models which are to remain constant over time necessitates doing so. We presented an unobserved components model with an $I(0)$ stochastic component, and found that it provided a useful framework for modelling co-breaking, especially as its unrestricted VAR generalization is convenient for estimation and testing purposes. The Monte Carlo results showed that the test for co-breaking rank was oversized in the presence of autocorrelation, but nevertheless, this approach was applied to modelling UK prices over the past century in light of the Stolper–Samuelson Theorem.

7 Acknowledgements

The authors would like to thank Giovanni Urga for his encouragement, support and constructive comment. Further thanks are due to two anonymous referees whose criticism led to considerable improvement of the paper. The authors are also grateful to Jörg Breitung, Norbert Christopeit, Bent Nielsen as well as participants of the 2004 ‘Common Features’ conference in London for helpful discussions on the topic. The usual disclaimer applies. David Hendry gratefully acknowledges financial support from the ESRC under grant RES 051 270035.

References

Anderson, T. W. (1951). Estimating linear restrictions on regression coefficients for multivariate normal distributions. *Annals of Mathematical Statistics*, **22**, 327–351.

- Andreou, E., and Spanos, A. (2003). Statistical adequacy and the testing of trend versus difference stationarity. *Econometric Reviews*, **22**, 217–237.
- Arranz, M. A., and Escribano, A. (2000). Cointegration testing under structural breaks: a robust extended error correction model. *Oxford Bulletin of Economics and Statistics*, **62**, 23–52.
- 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.
- Bai, J., and Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica*, **66**, 47–78.
- Banerjee, A., Dolado, J. J., Hendry, D. F., and Smith, G. W. (1986). Exploring equilibrium relationships in econometrics through static models: Some Monte Carlo evidence. *Oxford Bulletin of Economics and Statistics*, **48**, 253–277.
- Banerjee, A., and Hendry, D. F. (eds.) (1997). *The Econometrics of Economic Policy*. Oxford: Blackwell.
- Banerjee, A., Hendry, D. F., and Mizon, G. E. (1996). The econometric analysis of economic policy. *Oxford Bulletin of Economics and Statistics*, **58**, 573–600. Reprinted in Banerjee and Hendry (1997).
- Barrell, R. (2001). Forecasting the world economy. In Hendry, D. F., and Ericsson, N. R. (eds.), *Understanding Economic Forecasts*, pp. 149–169. Cambridge, Mass.: MIT Press.
- Bartlett, M. S. (1938). Further aspects of the theory of multiple regression. *Proceedings of the Cambridge Philosophical Society*, **34**, 33–40.
- Bentolila, S., and Saint-Paul, G. (2003). Explaining movements in the labour share. *Contributions to Macroeconomics*, **3**, article 9.
- Bernanke, B. S. (1986). Alternative explanations of the money-income correlation. *Carnegie-Rochester Conference Series on Public Policy*, **25**, 49–99.
- Bierens, H. J. (2000). Nonparametric nonlinear cotrending analysis, with an application to interest and inflation in the United States. *Journal of Business and Economic Statistics*, **18**, 323–337.
- Blanchard, O., and Quah, D. (1989). The dynamic effects of aggregate demand and supply disturbances. *American Economic Review*, **79**, 655–673.
- Box, G. E. P. (1949). A general distribution theory for a class of likelihood criteria. *Biometrika*, **36**, 317–346.
- Box, G. E. P., and Tiao, G. C. (1965). A change in level of a non-stationary time series. *Biometrika*, **52**, 181–192.
- Box, G. E. P., and Tiao, G. C. (1975). Intervention analysis with applications to economic and environmental problems. *Journal of the American Statistical Association*, **70**, 70–79.
- Burns, T. (1986). The interpretation and use of economic predictions. *Proceedings of the Royal Society of London*, **A407**, 103–125. Reprinted in Mills (1999).
- Chapman, D. A., and Ogaki, M. (1993). Cotrending and the stationarity of the real interest rate. *Economics Letters*, **42**, 133–138.
- Clements, M. P., and Hendry, D. F. (1998). *Forecasting Economic Time Series*. The Marshall Lectures on Economic Forecasting. Cambridge: Cambridge University Press.
- Clements, M. P., and Hendry, D. F. (1999). *Forecasting Non-Stationary Economic Time Series*. Boston: MIT Press.
- Courakis, A. S. (1978). Serial correlation and a Bank of England study of the demand for money: an exercise in measurement without theory. *Economic Journal*, **88**, 537–548.
- Davidson, J. E. H., Hendry, D. F., Srba, F., and Yeo, S. (1978). Econometric modelling of the aggre-

- gate time-series relationship between consumer expenditure and income in the United Kingdom. *Economic Journal*, **88**, 661–692.
- DeJong, D. N., Nankervis, J. C., Savin, N. E., and Whiteman, C. H. (1992). The power problems of unit root tests in time series with autoregressive errors. *Journal of Econometrics*, **53**, 323–343.
- 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.
- Dixit, A. K., and Norman, V. (1980). *Theory of International Trade*. Cambridge: Cambridge University Press.
- Doornik, J. A. (2001). *Ox 3.0 – An Object-Oriented Matrix Programming Language* 4th edn. London: Timberlake Consultants Press.
- Doornik, J. A., and Hansen, H. (1994). An omnibus test for univariate and multivariate Normality. Mimeo, Nuffield College, University of Oxford.
- Doornik, J. A., and Hendry, D. F. (2001). *GiveWin Version 2: An Interface to Empirical Modelling*. London: Timberlake.
- Doornik, J. A., Hendry, D. F., and Nielsen, B. (1998). Inference in cointegrated models: UK M1 revisited. *Journal of Economic Surveys*, **12**, 533–572.
- Engle, R. F., and Granger, C. W. J. (1987). Cointegration and error correction: Representation, estimation and testing. *Econometrica*, **55**, 251–276. Reprinted in Engle and Granger (1991).
- Engle, R. F., and Granger, C. W. J. (eds.) (1991). *Long-run Economic Relationships*. Oxford: Oxford University Press.
- Engle, R. F., and Hendry, D. F. (1993). Testing super exogeneity and invariance in regression models. *Journal of Econometrics*, **56**, 119–139.
- Engle, R. F., Hendry, D. F., and Richard, J.-F. (1983). Exogeneity. *Econometrica*, **51**, 277–304. Reprinted in Ericsson and Irons (1994) and Hendry (2000b).
- Engle, R. F., and Kozicki, S. (1993). Testing for common features. *Journal of Business and Economic Statistics*, **11**, 369–395.
- Ericsson, N. R., Hendry, D. F., and Mizon, G. E. (1998a). Exogeneity, cointegration and economic policy analysis. *Journal of Business and Economic Statistics*, **16**, 370–387.
- Ericsson, N. R., Hendry, D. F., and Prestwich, K. M. (1998b). The demand for broad money in the United Kingdom, 1878-1993. *Scandinavian Journal of Economics*, **100**, 289–324.
- Ericsson, N. R., and Irons, J. S. (eds.) (1994). *Testing Exogeneity*. Oxford: Oxford University Press.
- Evans, M. D. D. (2003). Real risk, inflation risk and the term structure. *Economic Journal*, **113**, 345–389.
- Friedman, M. (1957). *A Theory of the Consumption Function*. Princeton: Princeton University Press.
- 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.
- Gallant, A. R., and Fuller, W. A. (1973). Fitting segmented polynomial regression models whose join points have been estimated. *Journal of the American Statistical Association*, **68**, 144–147.
- Garcia, R. (1992). Asymptotic null distribution of the likelihood-ratio test in markov-switching models. Mimeo, Université de Montréal.
- Garcia, R., and Perron, P. (1996). An analysis of the real interest rate under regime shifts. *Review of Economics and Statistics*, **78**, 111–125.

- Granger, C. W. J. (1981). Some properties of time series data and their use in econometric model specification. *Journal of Econometrics*, **16**, 121–130.
- Gregory, A. W., and Hansen, B. E. (1996). Tests for cointegration in models with regime and trend shifts. *Oxford Bulletin of Economics and Statistics*, **58**, 555–560.
- Hagemann, H., Landesman, M., and Scazzieri, R. (eds.)(2002). *The Economics of Structural Change*. Cheltenham: Edward Elgar.
- Hamilton, J. D. (1989). A new approach to the estimation of non-stationary economic time series and the business cycle. *Econometrica*, **57**, 357–384.
- Hansen, B. E. (1992). The likelihood-ratio test under non-standard conditions: Testing the Markov switching model of GNP. *Journal of Applied Econometrics*, **7**, S61–S82.
- Hansen, L. P. (1982). Large sample properties of generalised method of moments estimators. *Econometrica*, **50**, 1029–1054.
- Hatanaka, M., and Yamada, H. (2003). *Co-trending: A Statistical System Analysis of Economic Trends*. Tokyo: Springer.
- Hendry, D. F. (1995). *Dynamic Econometrics*. Oxford: Oxford University Press.
- Hendry, D. F. (1996). A theory of co-breaking. Mimeo, Nuffield College, University of Oxford.
- Hendry, D. F. (2000a). Does money determine UK inflation over the long run?. In Backhouse, R. E., and Salanti, A. (eds.), *Macroeconomics and the Real World*, Vol. 1: Econometric Techniques and Macroeconomics, pp. 85–114. Oxford: Oxford University Press.
- Hendry, D. F. (2000b). *Econometrics: Alchemy or Science* New edn. Oxford: Oxford University Press. 1st edn 1993.
- Hendry, D. F. (2000c). On detectable and non-detectable structural change. *Structural Change and Economic Dynamics*, **11**, 45–65. Reprinted in Hagemann, Landesman and Scazzieri (2002).
- Hendry, D. F. (2001). Modelling UK inflation, 1875–1991. *Journal of Applied Econometrics*, **16**, 255–275.
- Hendry, D. F. (2003). Econometric modelling in a policy context. Mimeo, Economics Department, University of Oxford.
- Hendry, D. F. (2006). Robustifying forecasts from equilibrium-correction models. *Journal of Econometrics*, forthcoming.
- Hendry, D. F., and Ericsson, N. R. (1991). Modeling the demand for narrow money in the United Kingdom and the United States. *European Economic Review*, **35**, 833–886.
- Hendry, D. F., and Massmann, M. (1999). Macroeconometric forecasting and co-breaking. Mimeo, paper presented at 19th International Symposium on Forecasting, Washington DC.
- 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.
- Hendry, D. F., and Mizon, G. E. (1998). Exogeneity, causality, and co-breaking in economic policy analysis of a small econometric model of money in the UK. *Empirical Economics*, **23**, 267–294.
- Hendry, D. F., and Santos, C. (2006). Automatic tests of super exogeneity. Mimeo, Economics Department, University of Oxford.
- Hendry, D. F., and Wallis, K. F. (eds.)(1984). *Econometrics and Quantitative Economics*. Oxford: Basil Blackwell.
- Hotelling, H. (1936). Relations between two sets of variates. *Biometrika*, **28**, 321–377.
- Inoue, A. (1999). Tests of cointegration rank with a trend break. *Journal of Econometrics*, **90**, 215–237.

- Johansen, S. (1988). Statistical analysis of cointegrating vectors. *Journal of Economic Dynamics and Control*, **12**, 231–254.
- Johansen, S. (1991). Estimation and hypothesis testing of cointegrating vectors in Gaussian vector autoregressive models. *Econometrica*, **59**, 1551–1580.
- Johansen, S. (1994). The role of the constant and linear terms in cointegration analysis of non-stationary variables. *Econometric Reviews*, **12**, 205–229.
- Johansen, S., Mosconi, R., and Nielsen, B. (2000). Cointegration analysis in the presence of structural breaks in the deterministic trend. *Econometrics Journal*, **3**, 216–249.
- Kang, H. (1990). Common deterministic trends, common factors and cointegration. In Fomby, T. B., and Rhodes, G. F. (eds.), *Advances in Econometrics*, Vol. 8: Co-integration, Spurious Regressions, and Unit Roots, pp. 249–269. Greenwich, CT: JAI Press.
- Kim, C. J. (1994). Dynamic linear models with markov-switching. *Journal of Econometrics*, **60**, 1–22.
- 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.
- Krolzig, H.-M. (1997). *Markov-Switching Vector Autoregressions*. Berlin: Springer-Verlag.
- Krolzig, H.-M., and Toro, J. (2002). Testing for super-exogeneity in the presence of common deterministic shifts. *Annales d'Economie et de Statistique*, **67/68**, 41–71.
- Lucas, R. E. (1976). Econometric policy evaluation: a critique. *Carnegie-Rochester Conference Series on Public Policy*, **1**, 19–46.
- Lütkepohl, H. (1993). *Introduction to Multiple Time Series Analysis* 2nd edn. Berlin: Springer-Verlag.
- Massmann, M. (2003a). Co-breaking: representation, estimation and testing. DPhil thesis, University of Oxford.
- Massmann, M. (2003b). Cobra: A package for co-breaking analysis. Mimeo, paper presented at the 1st OxMetrics user conference, London.
- Meyer, C. (2000). *Matrix Analysis and Applied Linear Algebra*. Philadelphia: Society for Industrial and Applied Mathematics.
- Mills, T. C. (ed.) (1999). *Economic Forecasting*. Cheltenham: Edward Elgar.
- Mizon, G. E. (1995). Progressive modelling of macroeconomic time series: The LSE methodology. In Hoover, K. D. (ed.), *Macroeconometrics: Developments, Tensions, and Progress*, Ch. 4. Boston: Kluwer.
- Morana, C. (2002). Common persistent factors in inflation and excess nominal money growth and a new measure of core inflation. *Studies in Nonlinear Dynamics and Econometrics*, **6**, article 3.
- Nelson, C. R., and Plosser, C. I. (1982). Trends and random walks in macroeconomic time series. *Journal of Monetary Economics*, **10**, 129–162.
- Ng, S., and Perron, P. (1995). Unit root tests in ARMA models with data-dependent methods for the selection of the truncation lag. *Journal of the American Statistical Association*, **90**, 268–281.
- Nielsen, B., and Rahbek, A. (2000). Similarity issues in cointegration analysis. *Oxford Bulletin of Economics and Statistics*, **62**, 5–22.
- Ogaki, M., and Park, J. Y. (1998). A cointegration approach to estimating preference parameters. *Journal of Econometrics*, **82**, 107–137. Available in 1989 as Rochester Center for Economic Research Working Paper No. 209, University of Rochester.
- Pain, N., and Britton, A. (1992). The recent experience of economic forecasting in Britain: some lessons from National Institute forecasts. Discussion paper (new series) 20, National Institute.
- Perron, P. (1989). The great crash, the oil price shock, and the unit root hypothesis. *Econometrica*, **57**,

1361–1401.

- Perron, P. (1990). Testing for a unit root in a time series with a changing mean. *Journal of Business and Economic Statistics*, **8**, 153–162.
- Phillips, P. C. B. (1989). Optimal inference in cointegrated systems. *Econometrica*, **59**, 283–306.
- Phillips, P. C. B., and Hansen, B. E. (1990). Statistical inference in instrumental variable regression with I(1) processes. *Review of Economic Studies*, **57**, 99–125.
- Phillips, P. C. B., and Loretan, M. (1991). Estimating long-run economic equilibria. *Review of Economic Studies*, **58**, 407–437.
- Phillips, P. C. B., and Park, J. Y. (1988). On the formulation of Wald tests of non-linear restrictions. *Econometrica*, **56**, 1065–1083.
- Rao, C. R. (1951). An asymptotic expansion of the distribution of Wilk's criterion. *Bulletin of the International Statistical Institute*, **33**, 177–180.
- Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*. New York: John Wiley.
- Rappoport, P., and Reichlin, L. (1989). Segmented trends and non-stationary time series. *Economic Journal*, **99**, 168–177.
- Runkle, D. E. (1987). Vector autoregressions and reality. *Journal of Business and Economic Statistics*, **5**, 437–442.
- Said, S. E., and Dickey, D. A. (1984). Testing for unit roots in autoregressive - moving average models of unknown order. *Biometrika*, **71**, 599–608.
- Sargan, J. D. (1958). The estimation of economic relationships using instrumental variables. *Econometrica*, **26**, 393–415.
- Sargan, J. D. (1964). Wages and prices in the United Kingdom: A study in econometric methodology. In Hart, P. E., Mills, G., and Whitaker, J. K. (eds.), *Econometric Analysis for National Economic Planning*. London: Butterworths. Reprinted in Hendry and Wallis (1984).
- Schreiber, S. (2004). Shifts in equilibrium unemployment: macroeconomic theory and evidence. Mimeo, University of Frankfurt.
- Sims, C. A. (1980). Macroeconomics and reality. *Econometrica*, **48**, 1–48. Reprinted in Granger, C. W. J. (ed.) (1990), *Modelling Economic Series*. Oxford: Clarendon Press.
- Stock, J. H., and Watson, M. W. (1996). Evidence on structural instability in macroeconomic time series relations. *Journal of Business and Economic Statistics*, **14**, 11–30.
- Stolper, W., and Samuelson, P. A. (1941). Protection and real wages. *Review of Economic Studies*, **9**, 58–73.
- Wallis, K. F. (1989). Macroeconomic forecasting: A survey. *Economic Journal*, **99**, 28–61. Reprinted in Mills (1999).
- Williams, D. (1978). Estimating in levels or first differences: a defence of the method used for certain demand-for-money equations. *Economic Journal*, **88**, 564–568.