University of Leicester

DEPARTMENT OF ECONOMICS

Generalized Cointegration: A New Concept with an Application to Health Expenditure and Health Outcomes

Stephen Hall, University of Leicester, UK

P. A. V. B. Swamy, Retired from Federal Reserve Board, USA

George S. Tavlas, Bank of Greece, Economics Research Department, Greece

Working Paper No. 11/22
March 2011

Generalized Cointegration: A New Concept with an Application to Health Expenditure and Health Outcomes

Stephen G. Hall, P. A. V. B. Swamy and George S. Tavlas

Abstract We propose a new generalization of the concept of cointegration that allows for the possibility that a set of variables are involved in an unknown nonlinear relationship. Although these variables may be unit-root non-stationary, there exists a nonlinear combination of them that takes account of such non-stationarity. We then introduce an estimation technique that allows us to test for the presence of this generalized cointegration in the absence of knowledge as to the true nonlinear functional form and the full set of regressors. We outline the basic stages of the technique and discuss how the issue of unit-root non-stationarity and cointegration affects each stage of the estimation procedure. We then apply this technique to the relationship between health expenditure and health outcomes, which is an important but controversial issue. A number of studies have found very little or no relationship between the level of health expenditure and outcomes. In econometric terms, if there is such a relationship then there should exist a cointegrating relationship between these two variables and possibly many others. The problem that arises is that we may be either unable to measure these other variables or that we do not know about them, in which case we may incorrectly find no relationship between health expenditures and outcomes. We then apply the concept of generalized cointegration; we obtain a highly significant relationship between health expenditure and health outcomes.

Keywords Generalized cointegration, non-stationarity, time-varying coefficient model, coefficient driver
JEL Classification Numbers C130 · C190 · C220

The views expressed in this paper are the authors' own and do not necessarily represent those of their respective institutions. We would like to thank an anonymous referee for comments on this paper.

P. A. V. B. Swamy
Retired from Federal Reserve Board,
Washington, DC,
6333 Brocketts Crossing
Kingstowne, VA 22315
e-mail: swamyparavastu@hotmail.com

George S. Tavlas
Director General
Bank of Greece
21, El. Venizelos Ave.
102 50 Athens, Greece
e-mail: GTavlas@bankofgreece.gr

Stephen G. Hall
Leicester University Bank of Greece and NIESR,
Room: Astley Clarke 116,
University Road, Leicester, :
LEI 7RH, UK, e-mail: s.g.hall@le.ac.uk

1. Introduction

Developments in cointegration have dominated time series econometrics for the last 20 years. These developments have been almost exclusively within a linear framework and, while there have been various extensions into a nonlinear framework, these have generally been limited to very particular nonlinear functional forms. The reason for this circumstance is quite straightforward. In light of the standard definition of cointegration, given in Engle and Granger (1987), cointegration would become a trivial tautology unless we restrict the functional form of the relationship in a very strict way. Therefore while it is relatively straightforward to ask if a specific functional form links two or more variables together to produce a co-integrating combination, it is not generally possible to ask the more interesting question: 'Is there an unknown functional form, with possibly omitted variables, that would link two or more variables together in a structural relationship so as to yield a stationary error process?' Clearly, the spirit of that question is precisely what was being addressed in the above cited Engle and Granger paper, as well as in other earlier works on cointegration. However, there was no way to make this general question tractable. Therefore, a much more limited linear framework was adopted.

In this paper we propose a more general definition of cointegration. We also depart from the standard definition of integration of a variable, which is an inherently linear concept, to work more generally within a nonlinear framework. The implementation of our definition of generalized cointegration requires a new way to estimate the cointegrating parameters. We outline such a procedure and the way to conduct inference within this framework.

We then apply this technique to the important issue of the effect of additional health expenditure in promoting better health in the general population, which has been an important, but controversial issue. In this regard, Fisher *et al*. (2003), Skinner, Fisher and Wennberg (2005), Fisher, Bynum and Skinner (2009), and the U.S. Dartmouth project (Fisher *et al.,* 2009) all draw attention to the fact that there is no apparent correlation between health expenditure across states of the U.S. and the life expectancy of individuals in those states. Garber and Skinner (2008) highlight the fact that U.S. health care appears to be much more expensive than health care in other countries and yet the outcome does not seem to be obviously better. Baicker and Chandra (2004) demonstrate that high-spending states in the U.S. seem to experience worse outcomes than low spending states.

The application in this paper is designed to examine if convincing evidence may be found for a positive relationship between health expenditure and health outcomes, in particular, life expectancy. In standard econometric terms, both of these variables are trended; this implies that there should exist a cointegrating relationship between expenditure and outcomes. Researchers who have investigated this relationship have been confronted with two major problems: (1) this relationship is almost certainly nonlinear, as we cannot seriously believe that if health expenditure expanded to infinity we would produce infinite life expectancy, and, (2) there are certainly many missing variables that should prevent us from finding cointegration from a conventional perspective. These missing variables would include such factors as trends in smoking and exercise, developments in health technology, and even such influences as developments in working patterns and lifestyle, among many other factors. Given these two problems, conventional tests of cointegration between health expenditures and outcomes are unlikely to tell us anything useful in this context. In this paper, we introduce a new concept of cointegration, which allows for both unknown functional form and potentially important missing variables. We then illustrate the technique by testing the notion of generalized cointegration between the aggregate of total health expenditures undertaken by 19 OECD countries and the average life expectancy in the entire group of those countries over the period 1979 to 2008.

The remainder of this paper is divided into five sections. Section 2 generalizes the conventional notion of cointegration to include a general class of nonlinear economic relationships and a general type of non-stationarity without resorting to variable differencing. For this purpose, a general time-varying coefficient (TVC) model for some available data with precise interpretations of its coefficients and with appropriate assumptions is utilized in Section 3. Section 4 presents the conditions under which a parameterized TVC model is consistently estimable. An empirical relationship between the aggregate of health expenditures by 19 OECD countries and the average life expectancy in the entire group of those countries is given in Section 5. Finally, we draw some conclusions in Section 6.

2. Generalized Cointegration and the Definition of Integratedness

Cointegration is defined in terms of integrated variables. A variable is said to be integrated of order, say *d*, denoted by *I(d)*, if it becomes stationary after being first differenced *d* times. The idea underlying the simplest notion of cointegration is that if there is a structural relationship linking a group of *I(d)* variables together, then, there should be a combination of them with the disturbance which is integrated to an order less than *d*. . This concept is usually expressed within a linear framework in terms of variables integrated to the same order *d* which combine to produce a disturbance integrated to an order d - b, $d \geq b > 0$, smaller than *d*. Note that the order $d – b$ may not be equal to zero. Thus, cointegration may not always lead to models with stationary disturbances. There are extensions of the simplest notion of cointegration where the possibility of having variables with different orders of integration can be explored. Because we are dealing with a potentially nonlinear true model, which is assumed to be unknown, we need a slightly more general definition of non-stationarity and cointegration than is usually used in the literature. Typically, we focus on the order of integration of a variable; however, in the presence of general nonlinearity, variables may not be integrated at all. When a variable is integrated of order d = 0, such a variable may not be stationary but may be (weakly or strongly) non-unit-root non-stationary, and when d > 0, it can be unit-root non-stationary by virtue of our assumption. However, it is easy to demonstrate that there are also non-stationary variables that are not unit-root non-stationary. For example, let

$$x_t = f_t(x_{t-1}) + \varepsilon_t \qquad (1.$$

where $\frac{\partial f_t(x_{t-1})}{\partial x_{t-1}} \geq 1$ or time dependent so that model (1) implies that x_t is non-stationary but is not necessarily unit root non-stationary. Equivalently, this relationship can be expressed as[1]

$$x_t = \gamma_{0t} + \gamma_{1t} x_{t-1} \qquad (2.$$

[1] This is true simply because we may think of γ_{1t} as a function of x_{t-1}. If (1) is linear the coefficient γ_{1t} would be constant. If (1) is nonlinear, then the coefficient will be time-varying.

where γ_{0t} and γ_{0t} are time-dependent with profiles determined by the functional form of γ_{0t}. Then, in general, equation (1) is nonlinear and the first difference of x_t may be expressed as,

$$\begin{aligned}\Delta x_t &= x_t - x_{t-1} = \gamma_{0t} + \gamma_{1t}x_{t-1} - \gamma_{0t-1} - \gamma_{1t-1}x_{t-2} \\ &= \Delta\gamma_{0t} + \gamma_{1t}x_{1t-1} - \gamma_{1t-1}x_{t-2} + \gamma_{1t}x_{t-2} - \gamma_{1t}x_{t-2} \\ &= \Delta\gamma_{0t} + \gamma_{1t}\Delta x_{1t-1} + \Delta\gamma_{1t}x_{t-2}\end{aligned} \qquad (3.$$

which is, in general, neither stationary nor unit-root non-stationary since the last term in (3) contains the level of x_t; hence, x_t is non-unit-root non-stationary and is not integrated. Also, Δx_t does not possess a finite unconditional mean if x_t and/or γ_t follow random walk processes. Thus, each time equation (2) is differenced additional terms enter into it giving a non-parsimonious form unless equation (1) is linear or its intercept and slopes (excluding its error term) are constant, which will not generally be the case. There are a number of possible definitions of cointegration. According to the conventional definition, a set of *I(1)* variables is said to be cointegrated if they follow a linear model in which (i) the error term is *I(0)* with mean zero such that it is mean independent of the included explanatory variables and (ii) the coefficients are free of specification biases (see Greene 2008, p. 756). This definition of cointegration implies a linear framework; to make it operational we must assume that we know all the unit-root non-stationary elements of the set of variables under consideration. In practice, this is a situation that rarely, if ever, applies.

We, therefore, propose the following, more-general definition that allows for nonlinearity and omitted variables: the variables y_t and x_t are cointegrated in a general sense if the bias-free component of the coefficient of x_t in the relation of y_t to x_t is nonzero. (The meaning of bias-free component will be made clear below.) To explain, consider the following general relationship between *y*, *x* and a set of other variables *w*, all of which are assumed to be non-stationary, not necessarily of unit-root type.

$$y_t = f_t(x_t, w_t) \qquad (4.$$

which according to economic theories, is a valid relationship. Then, under our definition of generalized cointegration *y* and *x* are cointegrated if in equation (4), x_t enters with a nonzero coefficient, which may not be constant. That is:

$$\frac{\partial y_t}{\partial x_t} \neq 0 \qquad (5.$$

Cointegration should only arise if there is a (possibly nonlinear) structural relationship holding a set of variables together. If there is such a relationship, this implies that the bias-free effect of *x* on *y* will be nonzero. It follows that if

$$\frac{\partial y_t}{\partial x_t} = 0 \qquad (6.$$

we have a spurious relationship between the two variables. However, if we run a standard regression between *x* and *y*, we may falsely find a significant coefficient, which is spurious.

To make this definition of cointegration operational, we need an estimation technique that (i) will yield bias-free estimates of coefficients, (ii) accounts for the fact that the true functional form is unknown, and (iii) accounts for the fact that there may be omitted variables and measurement errors. We turn to such a technique in the next section.

3. The Interpretations of Model Coefficients and Appropriate Assumptions

Conventional econometrics is to a large extent the study of individual causes of biased coefficient estimates: 'non-sufficient sets' of omitted variables, measurement errors, incorrect functional forms, etc. These problems are usually dealt with one at a time in a textbook context, but, of course, practical work is plagued by all these problems at once. In what follows, we outline (i) the basic problem of interpreting coefficients when these problems are present and (ii) our proposed procedure for dealing with these problems simultaneously. In particular, we are concerned with the case in which the dependent variable of an economic relationship is non-stationary (not necessarily of unit-root type) and at least two sets of its determinants are also non-stationary of any type and where there is a (possibly) nonlinear relationship between these variables which produces a parameterized version of a combination of them with non-constant coefficients. That is, we outline a general nonlinear form of cointegration. We restrict ourselves here to the case of two sets of non-stationary variables simply because this situation allows for all the cases we believe are of interest. These two sets of variables could be equally thought of as two individual (unit-root or other) non-stationary variables. We also allow for measurement error and omitted variables that may be either stationary or one or both of the two sets of (unit-root or other) non-stationary independent variables.

Denote the dependent variable by y_t^*; it is related to a hypothesized set of $K - 1$ of its determinants, denoted by x_{1t}^*, …, $x_{K-1,t}^*$, where K-1 may be only a subset of the complete set of determinates of y_t^*, in which case the relation of y_t^* to x_{1t}^*, …, $x_{K-1,t}^*$ may be subject to omitted-variable biases. Any specific functional form may be incorrect and may lead to specification errors. In addition to these problems, the available data on y_t^*, x_{1t}^*, …, $x_{K-1,t}^*$ may not be perfect measures of the underlying true variables, causing an errors-in-variables problem.

Suppose that T measurements on y_t^*, x_{1t}^*, …, $x_{K-1,t}^*$ are made and these measurements are actually the sums of "true" values and measurement errors: $y_t = y_t^* + \mathrm{v}_{0t}$, $x_{jt} = x_{jt}^* + \mathrm{v}_{jt}$, j = 1, …, K-1, t = 1, …, T, where the variables y_t, x_{1t}, …, x_{Kt} without an asterisk are the observable variables, the variables with an asterisk are the unobservable "true" values, and the v's are measurement errors. Given the possibility that the true functional form we are estimating may be unknown and that there may be some important variables missing from x_{1t}, …, $x_{K-1,t}$, we need a model that will capture all these potential problems.

It is useful at this point to clarify what we believe to be the main objective of econometric estimation. In our view, the objective is to obtain consistent estimates of the bias-free effect on a correctly measured dependent variable of changing one of its correctly measured determinants, holding all of its other correctly measured determinants constant. That is, we aim to find an estimate of the bias-free component of the coefficient of any x^*_{jt} in a general nonlinear regression of y^*_t on x^*_{1t}, …, $x^*_{K-1,t}$. This view, of course, is the interpretation that is usually placed on the coefficients of a standard econometric model. The interpretation depends crucially on the assumption that the conventional model has bias-free coefficients, which is not the case in the presence of model misspecification. Note that the term "bias-free" here means without both omitted-variable and measurement-error bias components.

We begin by specifying a model, which provides a complete explanation of the dependent variable y.

$$y_t = \gamma_{0t} + \gamma_{1t}x_{1t} + \cdots + \gamma_{K-1,t}x_{K-1,t} \ (t = 1, \ldots, T) \qquad (7.$$

which we call "the time-varying coefficient (TVC) model".[2] The explanatory variables of this model are called the included regressors. As this model provides a complete explanation of y, all the misspecifications in the model, as well as the true coefficients, must be captured by the time-varying coefficients. Note that if the true functional form is nonlinear the time-varying coefficients may be thought of as being the true nonlinear structure and so they are able to capture any possible function. These coefficients will also capture the effects of measurement error and omitted variables.

Equation (7) is called the observation equation and its coefficients are called the state variables if it is embedded in a state-space model. We now apply a formal decomposition of these time-varying coefficients which illustrates the various components they contain.

Notation and Assumptions Let m_t denote the total number of the determinants of y^*_t. The exact value of m_t is usually unknown at any time. We assume that m_t is larger than K-1 (that is, the number of determinants is greater than the determinants for which we have observations) and possibly varies over time. This assumption means that there are

[2] It is worth noting that in a recent paper Granger (2008) suggested that he believed that the next major development in econometrics would be time-varying parameter models, and he quoted a theorem which he attributed to White from unpublished work in 2006 which demonstrated that a time-varying parameter model might represent any unknown functional form. This theorem was first established by Swamy and Mehta (1975). Even the models Granger considered were not new. They were considered previously in Swamy, Chang, Mehta and Tavlas (2003). Thus, we refer to this theorem as the Swamy Theorem.

determinants of y_t^* that are excluded from equation (7). Let x_{gt}^*, g = K, …, m_t, denote these excluded determinants. Let α_{0t}^* denote the intercept and let both α_{jt}^*, j = 1, …, K-1, and α_{gt}^*, g = K, …, m_t, denote the other coefficients of the regression of y_t^* on all of its determinants. The true functional form of this regression determines the time profiles of α^*'s, each of which is a function of both the sets of variables x_{jt}^*'s and x_{gt}^*'s. These time profiles are unknown, since the true functional form is unknown. For g = K, …, m_t, let $x_{gt}^* = \lambda_{0gt}^* + \lambda_{1gt}^* x_{1t}^* + \cdots + \lambda_{K-1,gt}^* x_{K-1,t}^*$. The true functional forms of these regressions determine the time profiles of λ^*'s. Each of these λ^*'s is a function of the x_{jt}^*. Let a set, denoted by S_1, consist of those x_{jt}, j = 1, ..., K-1, that take the value zero with zero probability and let another set, denoted by S_2, consist of those x_{jt}'s that take the value zero with positive probability.

Theorem 1 *The intercept of (7) satisfies the equation,*

$$\gamma_{0t} = \alpha_{0t}^* + \sum_{g=K}^{m_t} \alpha_{gt}^* \lambda_{0gt}^* + \mathrm{v}_{0t} - \sum_{j \in S_2} \left(\alpha_{jt}^* + \sum_{g=K}^{m_t} \alpha_{gt}^* \lambda_{jgt}^* \right) \mathrm{v}_{jt} \qquad (8.$$

and the coefficients of (7) other than the intercept satisfy the equations,

$$\gamma_{jt} = \alpha_{jt}^* + \sum_{g=K}^{m_t} \alpha_{gt}^* \lambda_{jgt}^* - \left(\alpha_{jt}^* + \sum_{g=K}^{m_t} \alpha_{gt}^* \lambda_{jgt}^* \right) \left(\frac{\mathrm{v}_{jt}}{x_{jt}} \right) \quad \text{if } x_{jt} \in S_1$$

and

$$= \alpha_{jt}^* + \sum_{g=K}^{m_t} \alpha_{gt}^* \lambda_{jgt}^* \;\; \text{if } x_{jt}^* \in S_2 \qquad (9.$$

where the λ_{0gt}^* *are a 'sufficient set' of excluded variables in the sense that they in conjunction with the* x_{jt}^*, *are at least sufficient to determine* y_t^*.

Proof See Swamy and Tavlas (2001, 2007). □

Thus, we interpret the TVC's of (7) in terms of the underlying correct coefficients, a 'sufficient set' of excluded variables, the observed explanatory variables and measurement errors in both dependent and explanatory variables. By assuming that the α^*'s and λ^*'s are possibly time varying, we do not *a priori* rule out the possibility that the relationship of y_t^* with all of its determinants and the regressions of the determinants of y_t^* excluded from (7) on the determinants of y_t^* included in (7) are nonlinear.

In terms of non-stationarity and nonconstancy we can consider three cases, assuming that x_{jt}^*, j = 1, …, K – 1, and λ_{0gt}^*, g = K, …, m_t, are the two sets of the determinants of y_t^*.

1. Both the function $\sum_{g=K}^{m_t} \alpha_{gt}^* \lambda_{0gt}^*$ of a 'sufficient set' of excluded variables and the function $\mathrm{v}_{0t} - \sum_{j\in S_2}\left(\alpha_{jt}^* + \sum_{g=K}^{m_t} \alpha_{gt}^* \lambda_{jgt}^*\right)\mathrm{v}_{jt}$ of measurement errors are white noise processes with means zero and the true intercept α_{0t}^* is constant for all *t*, both the components α_{jt}^* and $\sum_{g=K}^{m_t} \alpha_{gt}^* \lambda_{jgt}^*$ of the coefficient of x_{jt}^* are constant and the measurement error v_{jt} = 0 for all *j* and *t*, in which case the non-stationarity will be confined to the mean of y_t, as in the standard regression models.

2. If x_{jt}^*, j = 1, …, K-1, are non-stationary but the other non-stationary variables are λ_{0gt}^*, g = K, …, m_t, a sufficient set of excluded variables, then even if the α_{jt}^* are constant for all *j* and *t*, the γ_{jt} will be nonconstant if their other components are nonconstant.

3. Both the sets of variables x_{jt}^*'s and λ_{0gt}^*'s may be non-stationary, in which case again the γ_{jt} will be nonconstant if any or all of their components in (8) and (9) are nonconstant.

Theorem 2 *For j* = 1, …, *K*-1, *the component* α_{jt}^* *of* γ_{jt} *in (9) is bias-free and unique.*

Proof It can be seen from equation (9) that the component α^*_{jt} of γ_{jt} is free of omitted-variables bias ($=\sum_{g=K}^{m_t}\alpha^*_{gt}\lambda^*_{jgt}$), measurement-error bias ($=-\left(\alpha^*_{jt}+\sum_{g=K}^{m_t}\alpha^*_{gt}\lambda^*_{jgt}\right)\times\left(\mathrm{v}_{\mathrm{jt}}/x_{jt}\right)$), and of functional-form bias, since we allow the α^*s and λ^*s to have the correct time profiles. These biases are not unique being dependent on what determinants of y^*_t are excluded from (7) and on the v_{jt}. Only α^*_{jt} is unique being the coefficient of x^*_{jt} in the correctly specified relation of y^*_t to all of its determinants. Alternatively stated, the component α^*_{jt} is bias-free and unique because it represents a property of the real world that remains invariant against mere changes in the language we use to describe it (see Basmann 1988, p. 73; Pratt and Schlaifer 1984, p. 13; Zellner 1979, 1988). □

This is true irrespective of the type of non-stationarity of the variables under consideration as the 'sufficient set' of omitted variables can fully reflect the omitted non-stationary variables. An issue that we have not addressed, however, is how to correctly identify the bias-free component and omitted-variable and measurement-error biases in the case of non-stationarity. We turn to this issue in the next section.

4. Identification and Consistent Estimation of Time-Varying Coefficient Model

4.1 Identification

As noted above, generalized cointegration takes place if the bias-free component of the coefficient linking two variables is non-zero. In order to test whether this situation applies, we are interested in the bias-free components α^*'s, not in the omitted-variable and measurement-error biases. To obtain accurate estimates of the α^*_{jt} using the observations in (7), we need to first decompose each γ_{jt} with $j > 0$ into its components in (9). Our method of identifying these components and performing the decomposition is based on the following assumptions.

Assumption 1 (Auxiliary information) *Each coefficient of (7) is linearly related to certain drivers plus a random error,*

$$\gamma_{jt} = \pi_{j0}z_{0t} + \sum_{d=1}^{p-1}\pi_{jd}z_{dt} + \varepsilon_{jt} \quad (j = 0, 1, \ldots, K\text{-}1), \qquad (10.$$

where $z_{0t} \equiv 1$, *the* π s *are fixed parameters, the* z_{dt} *are what are called the coefficient drivers, and different coefficients of (7) can be functions of different sets of coefficient drivers.*

Here, the issue of identification is important. If both γ_{jt} and α^*_{jt} are constant, then clearly they are not identifiable. Regardless of whether α^*_{jt} is constant, if γ_{jt} is nonconstant (due to its nonconstant omitted-variables and measurement-error bias components) we need to include a set of coefficient drivers to identify its components.

Assumption 2 *For j* = 0, 1, …, *K*-1, *the set of p coefficient drivers* z_{0t}, z_{1t}, …, $z_{p-1,t}$ *in (10) divides into three disjoint subsets* A_{1j}, A_{2j}, *and* A_{3j} *so that for j = 0,* $\sum_{d\in A_{1j}} \pi_{jd} z_{dt}$, $\sum_{d\in A_{2j}} \pi_{jd} z_{dt}$, *and* $(\sum_{d\in A_{3j}} \pi_{jd} z_{dt} + \varepsilon_{jt} + \sum_{j'\in S_2}(\sum_{d\in A_{3j'}} \pi_{j'd} z_{dt} + \varepsilon_{j't}))$ *have the same sign, magnitude, and same pattern of time variation as* α^*_{0t}, $\sum_{g=K}^{m_t} \alpha^*_{gt}\lambda^*_{0gt}$, *and* v_{0t} - $\sum_{j'\in S_2}\left(\alpha^*_{j't} + \sum_{g=K}^{m_t} \alpha^*_{gt}\lambda^*_{j'gt}\right)\mathrm{v}_{j't}$, *respectively; for* $j \in S_1$, $\sum_{d\in A_{1j}} \pi_{jd} z_{dt}$, $\sum_{d\in A_{2j}} \pi_{jd} z_{dt}$, *and* $\sum_{d\in A_{3j}} \pi_{jd} z_{dt} + \varepsilon_{jt}$ *have the same sign, magnitude, and same pattern of time variation as* α^*_{jt}, $\sum_{g=K}^{m_t} \alpha^*_{gt}\lambda^*_{jgt}$, *and* $-\left(\alpha^*_{jt} + \sum_{g=K}^{m_t} \alpha^*_{gt}\lambda^*_{jgt}\right)\left(\frac{\mathrm{v}_{jt}}{x_{jt}}\right)$, *respectively, and for* $j \in S_2$, $\sum_{d\in A_{1j}} \pi_{jd} z_{dt}$ *and* $\sum_{d\in A_{2j}} \pi_{jd} z_{dt}$ *have the same sign, magnitude, and same pattern of time variation as* α^*_{jt} *and* $\sum_{g=K}^{m_t} \alpha^*_{gt}\lambda^*_{jgt}$, *respectively, over the relevant estimation and forecasting periods.*

Here we are assuming that for each *j*, the drivers in the sets A_{1j}, A_{2j}, and A_{3j} separate the direct effect α^*_{jt} from the specification biases in the model. Here, again, we can draw out some important implications for the assignment of coefficient drivers to the sets A_{1j}, A_{2j}, and A_{3j}. It is appropriate to assign the constant z_{0t} to A_{1j} for all *j > 0*. If the component α^*_{jt} is constant while γ_{jt} is non-stationary (or nonconstant), then z_{0t} should be the only variable in A_{1j}. In this case, we know that all non-stationary (or nonconstant) drivers should be in the set A_{2j}. Of course, it is possible that α^*_{jt} itself is non-stationary (or nonconstant), due to the unknown nonlinear functional form, in which case we have a difficult problem of splitting out just the correct amount of non-stationarity (or nonconstancy) between the sets A_{1j} and A_{2j}. However, the assumption that α^*_{jt} is constant is not a very strong one if y^*_t is nonlinearly related to all of its determinants.

Assumption 3 *The K-vector* $\varepsilon_t = (\varepsilon_{0t}, \varepsilon_{1t}, ..., \varepsilon_{K-1,t})'$ *of errors in (10) follows the stochastic equation,*

$$\varepsilon_t = \Phi \varepsilon_{t-1} + u_t , \qquad (11.$$

where Φ *is a* $K \times K$ *(not necessarily diagonal) matrix whose eigenvalues are less than 1 in absolute value, the K-vector* $u_t = (u_{0t}, u_{1t}, ..., u_{K-1,t})'$ *is distributed with* $E(u_t \mid z_{1t}, \ldots, z_{p-1,t}) = 0$ *and*

$$E(u_t u'_{t'} \mid z_{1t}, \ldots, z_{p-1,t}) = \begin{cases} \sigma_u^2 \Delta_u & \text{if } t = t' \\ 0 & \text{if } t \neq t' \end{cases} , \qquad (12.$$

where Δ_u *may not be diagonal.*

This assumption considerably generalizes (10). If we assumed that the errors in (10) were serially independent, this would imply a very simple dynamic structure. By making the assumption that the errors in fact have a serial correlation structure we are allowing a much richer dynamic structure, although we are imposing some common factors in this structure to keep the model tractable.

In terms of non-stationarity, by assuming that all the eigenvalues are less than 1 in absolute terms we are ruling out the possibility that non-stationarity in γ_{jt} is generated by the error process ε_t. This, then, isolates the non-stationarity as coming from the coefficient drivers.

Assumption 4 *The regressor* x_{jt} *of (7) is conditionally independent of its coefficient* γ_{jt} *given the coefficient drivers in (10) for all j and t.*

A vector formulation of model (7) is

$$y_t = x_t'\gamma_t , \tag{13.}$$

where $x_t = (x_{0t} \equiv 1,\ x_{1t},\ \ldots,\ x_{K-1,t})'$ and $\gamma_t = (\gamma_{0t}, \gamma_{1t},\ \ldots,\ \gamma_{K-1,t})'$. A matrix formulation of (10) is

$$\gamma_t = \Pi z_t + \varepsilon_t , \tag{14.}$$

where $\Pi = \left[\pi_{jd}\right]_{0\le j\le K-1, 0\le d\le p-1}$ is a $K \times p$ matrix having π_{jd} as its $(j+1, d+1)$-th element and $z_t = (1, z_{1t}, ..., z_{p-1,t})'$. Substituting (14) into (13) gives

$$y_t = (z_t' \otimes x_t')\pi^{Long} + x_t'\varepsilon_t , \tag{15.}$$

where $\otimes$ denotes a Kronecker product, and π^{Long} is a *Kp*-vector, denoting a column stack of Π. The observations in (7) can be displayed in a matrix form as

$$y = X_z\pi^{Long} + D_x\varepsilon , \tag{16.}$$

where $y = (y_1, ..., y_T)'$ is a *T*-vector, $X_z = (z_1 \otimes x_1, ..., z_T \otimes x_T)'$ is $T \times Kp$, $D_x = diag_{1\le t\le T}(x_t')$ is $T \times KT$, and $\varepsilon = (\varepsilon_1', ..., \varepsilon_T')'$ is a *TK*-vector.

Theorem 3 *Under Assumptions 1-4,* $E(y \mid X_z) = X_z\pi^{Long}$ *and* $\mathrm{Var}(y \mid X_z) = D_x\sigma_u^2\Sigma_\varepsilon D_x'$ *where* $\sigma_u^2\Sigma_\varepsilon$ *is the covariance matrix of* ε.

Proof See Swamy, Yaghi, Mehta and Chang (2007, p. 3386). □

Under Assumptions 1 and 3, the variance of γ_{jt} is finite for all *j* and *t*. The Chebychev inequality shows that if γ_{jt} has a small variance, then its distribution is tightly concentrated about its mean implied by Assumptions 1 and 3 (see Lehmann 1999, p. 52). Assumptions 2 and 4 provide a prime consideration guiding the selection of coefficient drivers, especially in the presence of non-stationarity. The magnitude of ε_{jt} gets reduced as the number of correct coefficient drivers in (10) increases. The larger this number , the smaller the magnitude of ε_{jt} and the smaller the variance of γ_{jt}. Including many correct coefficient drivers in (10) may imply that the errors of equation (10) are white-noise variables or the matrix Φ in equation (11) is null. If Assumption 3 is replaced by the assumption that ε_t follows a random walk for all *t*, then the unconditional variance of γ_{jt} is not finite.

The fixed coefficient vector π^{Long} in (15) is identified if X_z has full column rank. A necessary condition for X_z to have full column rank is that $T > Kp$. The error vector ε is not identified because the necessary condition $T > TK$ for D_x to have full column rank is false. This result implies that ε is not consistently estimable (see Lehmann and Casella 1998, p. 57). Swamy and Tinsley (1980, p. 117) call this phenomenon "a form of Uncertainty Principle". Correct coefficient drivers should be used in (10) to reduce the unidentifiable portions (the ε_{jt}) of the coefficients of (7). However, $D_x\varepsilon$ being equal to *y* - $X_z\pi^{Long}$ with identifiable π^{Long} is identifiable, provided D_x has full row rank. The best linear unbiased predictor (BLUP) of $D_x\varepsilon$ can be used to obtain consistent estimators of Φ, Δ_u, and σ_u^2 in (11) and (12), as shown in Chang, Hallahan and Swamy (1992) and Chang, Swamy, Hallahan and Tavlas (2000). Under Assumptions 1-4, the BLUP of $D_x\varepsilon$ exists (see Swamy, Yaghi, Mehta and Chang 2007, p. 3387). So we make

Assumption 5 (i) X_z *has full column rank, (ii)* D_x *has full row rank, and (iii) (assumption 5(iii) is here)* $T \geq Kp$ + *(the number of unknown distinct elements of* Φ, Δ_u, *and* σ_u^2 + 4) *so that the degrees of freedom left unutilized after estimating all the unknown parameters of model (10) is at least 4.*

Assumptions 5(i) and 5(ii) make all the coefficients and $D_x\varepsilon$ of (16) statistically meaningful. Equation (10), which establishes a link between the coefficients of (7) and the coefficients and errors of (15), shows that if the coefficients and $D_x\varepsilon$ of (16) are statistically meaningful, then so are the coefficients of (7). In certain situations specified in Judge, Griffiths, Hill, Lütkepohl and Lee (1985, p. 612), the finite moments of the estimators of the coefficients of (15) exist up to the degrees of freedom that remain unutilized after the estimation of these coefficients. Assumption 5(iii) is made to guarantee the existence of at least finite fourth moments for the estimators of the coefficients of (15) in these situations.[3]

4.2 Consistent estimation

Under certain conditions, an iteratively rescaled generalized least squares (IRSGLS) estimator $\hat{\pi}^{Long}$ of π^{Long} and Φ and $\sigma_u^2\Delta_u$, estimated from the IRSGLS residuals $y - X_z\hat{\pi}^{Long}$, are consistent and asymptotically normal[4] (see Swamy, Tavlas, Hall and Hondroyiannis 2010). Essentially, this procedure simply minimizes the generalized least squares criterion formed from (16) in an iterative framework.

The model may also be formulated as a standard state space form and the Kalman filter may be used to provide a predictor $\hat{\varepsilon}$ of ε in (16) and $D_x\hat{\varepsilon}$, in turn, may be used with a distributional assumption about ε to provide maximum likelihood estimates of π^{Long}, Φ, and $\sigma_u^2\Delta_u$ using standard software such as E-VIEWS.[5] The state space form of (16) consists of the measurement equation which is given by equation (7) and the state equations which are given by (10). Once this model has been estimated either by the Kalman filter in conjunction with the maximum likelihood method or by an iteratively rescaled generalized least squares (IRSGLS) procedure, the results include estimates of γ_{jt}, ε_{jt}, and π_{jd} $j = 0,1..., K-1, d = 0,1,..., p-1$.[6]

---

[3] Swamy, Mehta and Singamsetti (1996) explain how model (16) might be estimated when X_z has less than full column rank and $D_x\Sigma_\varepsilon D_x'$ is singular.

[4] A computer program is freely available which implements this technique at http://www.le.ac.uk/ec/sh222/soft.htm

[5] However, the Kalman filter has the disadvantage that it does not provide estimators of Φ and $\sigma_u^2\Delta_u$.

[6] The IRSGLS procedure has the advantage that it is a distribution-free method, in contrast to the method of maximum likelihood.

Any simple spread sheet program can then be used to calculate ε_{jt} simply as the residual of (10), and the bias-free estimate can be calculated simply as

$$\gamma_{jt}^{u} = \gamma_{jt} - \sum_{d \in A_{2j}} \pi_{jd} z_{dt} - \sum_{j \in A_{3j}} \pi_{jd} z_{dt} - \varepsilon_{jt} \text{ if } j > 0 \text{ and } j \in S_1$$

and

$$= \gamma_{jt} - \sum_{d \in A_{2j}} \pi_{jd} z_{dt} \text{ if } j > 0 \text{ and } j \in S_2$$

which again may be easily calculated in any spreadsheet program.

5. An Example: Health Expenditure and Life Expectancy

The relationship, if there is one, between health expenditure and outcomes, is an extremely important one from a policy perspective. For example, if there is a positive relationship, then cuts in health expenditure as part of general fiscal consolidation will lead to falls in health outcomes and, therefore, inevitably to more deaths. Policy makers need to face this fact squarely at times of cuts if such a relationship exists. This relationship is, however, far from uncontentious as the papers cited in the Introduction make clear; there has been a fairly substantial literature debating the issue including seminal contributions from Jones (2005), Hall and Jones (2007) and Skinner and Staiger (2009), as well as the contributions cited in the Introduction.

From the perspective of a policy maker, we would argue that the crucial information which is needed is whether or not there is a link from health expenditure to life expectancy. It would, of course, be desirable to have a complete model of life expectancy, which included all real world determinants, but realistically such a model will always be impossible to achieve. Nevertheless, as long as the policy maker has clear evidence that such a link exists, and a good estimate of its empirical magnitude, relevant decisions can be made. The concept of generalized cointegration, discussed above, is particularly relevant in this case because there are many variables that affect life expectancy which could never be introduced into a study of this kind. Such variables include factors like social habits, including smoking, exercise, and diet, and many external factors such as wars, unusual diseases, and health technology and migration habits, among many others. Hence, the conventional definition of cointegration should always find that cointegration does not exist. Thus, it is not surprising that researchers have found it difficult to uncover a significant relationship between health expenditure and life expectancy. Generalised cointegration allows for the existence of these omitted variables and gives robust and consistent results; it is, therefore, an appropriate technique in this case.

In this section we use data for average life expectancy and average per capita real health expenditure for 19 OECD countries from the OECD Health Data 2010 data base.[7] We have aggregated these series into total health expenditure for the 19 countries and average life expectancy across the countries. We have gathered data on average total expenditure on health per capita in constant price US dollars (EXP) for 19 OECD countries from the OECD health data bank 2010 and data on average life expectancy (LIFE) over the period 1970 to 2008. Using the idea of generalized cointegration, outlined above, we will then test for a non-zero bias-free relationship between these two variables. If a significantly non-zero relationship is detected, we will have established the presence of generalized cointegration and, hence, we will have established the existence of a cointegrating relationship linking expenditure and outcomes. The specific model we estimated is:

$$Log(LIFE) = \gamma_{0t} + \gamma_{1t} Log(EXP) \qquad (17.$$

We have chosen to specify the relationship in logs so that the coefficients may be interpreted as elasticities; however, very comparable results to those presented below may be obtained by using the raw data. As coefficient drivers we used the lagged values of both the log(EXP) and Log(LIFE) and a deterministic trend. In order to derive the bias-free effect, we removed both the lagged effects from the total coefficient to obtain the bias-free result.

[7] The 19 countries are Australia, Austria, Belgium, Canada, Denmark, Finland, Germany, Iceland, Ireland, Japan, Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Switzerland, UK and the USA. These countries were selected on the basis that they had almost complete data for the required period.

Table 1 reports the average value for the intercept and the coefficient of Log(EXP), both in terms of the total effect and the bias-free effect. Figure 1 plots these two coefficients against time. The key result is the t-statistic (equal to 312.1) on the bias-free coefficient, which is highly significant; it is this value which represents our test for generalized cointegration. The result confirms the existence of a cointegrating relationship between health expenditure and health outcomes. As shown in Figure 1, the bias-free coefficient falls very gradually over time, suggesting that there is an underlying nonlinear relationship, but the elasticity of outcomes with respect to expenditure is falling only very slowly.

The overall conclusion to be drawn here is straightforward. Increased health expenditure produces increase in life expectancy with an elasticity of around 0.29 which although it is falling slowly, is quite stable.

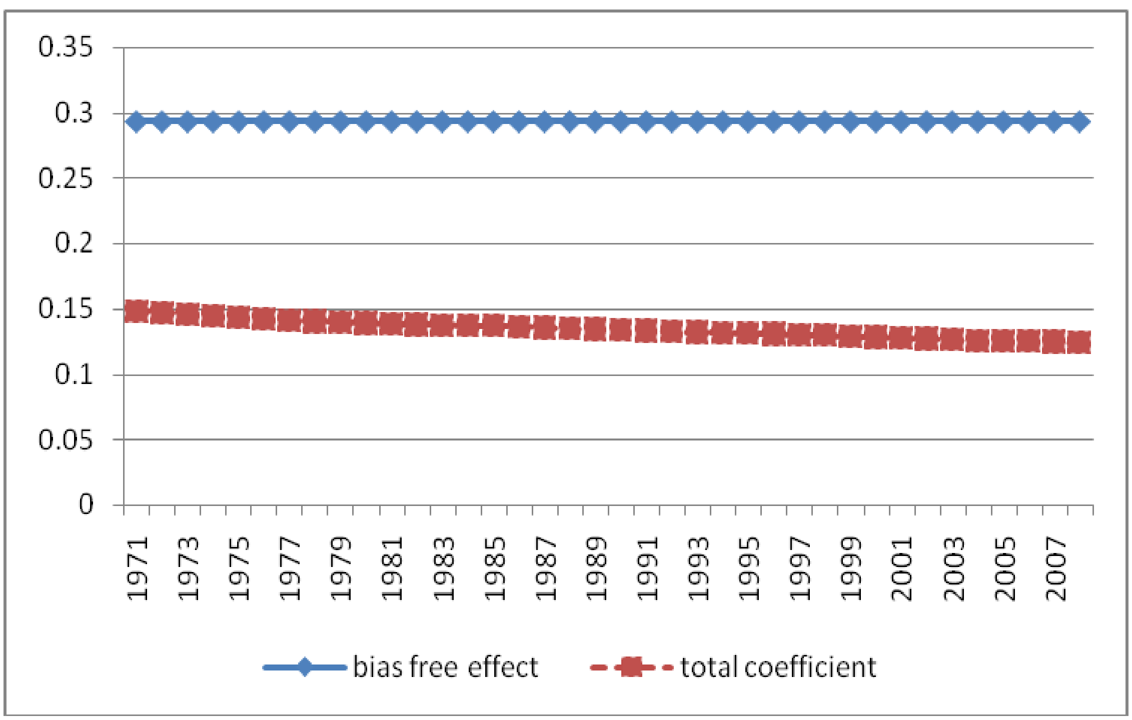
6 Conclusions

We have proposed a generalization of the standard definition of cointegration that allows for the existence of an unknown structural nonlinear relationship between a set of non-stationary variables. The idea underlying this definition is as follows. If a structural relationship exists between two or more variables, the implication of this is that there will be a non-zero coefficient attached to any of the independent variables. Therefore, the significance of an unbiased estimate of this coefficient becomes a simple direct test of generalized cointegration. Furthermore, we can estimate this coefficient and test its significance without knowing the true functional form of the relationship and/or the full set of variables that enter into it. This definition can be made operational by applying the TVC estimation technique, which provides just such an unbiased estimate of the coefficient. Non-stationarity does not pose any particular problem for TVC estimation. However, as in other modeling situations the explicit recognition of non-stationarity does offer advantages, in particular, in the identification of the correct set of coefficient drivers to identify bias-free component of the time-varying coefficient correctly. Finally, we have applied this technique to the issue of the linkage between health expenditure and life expectancy. We provided strong evidence for the existence of a cointegrating vector between these variables and possibly a set of additional unidentified variables.

Table 1		
TVC Estimation: The Effect of Health Care Expenditure on Life Expectancy		
Variables	**Equation**	
	Total coefficients (1)	Bias-free effects (2)
Constant	0.007	0.12
Log(EXP)	0.134	0.29*** [312.1]

Notes: Figures in brackets are t-ratios. ***, ** and * indicate significance at 1%, 5% and 10% level, respectively. The estimates are the time averages of the estimates of the time-varying coefficients over the entire sample period; the bias-free effects are the time averages of the estimates of the coefficients with the effects of the lagged driver variables removed.

Figure 1: The Total Coefficient on Log(EXP) and the Bias-Free effect



References

Baicker, K., Chandra, A.: The productivity of physicians specialization: Evidence from the medicare program. American Economic Review **94**, 357-361 (2004)

Basmann, R.L.: Causality tests and observationally equivalent representations of econometric models. J Econ Ann **39**, 69-104 (1988)

Chang, I-L., Hallahan, C., Swamy, P.A.V.B.: Efficient computation of stochastic coefficients models, pp. 43-54. In: Amman, H.M., Belsley, D.A., Pau, L.F. (eds.) Computational Economics and Econometrics. Boston, MA: Kluwer Academic Publishers (1992)

Chang, I-L., Swamy, P.A.V.B., Hallahan, C., Tavlas, G.S. (2000). A computational approach to finding causal economic laws. Computational Economics **16**, 105-136

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

Fisher E.S., Bynum, J.P., Skinner, J.: Slowing the growth of health care costs- Lessons from regional variation. New England Journal of Medicine **360**, 849-852 (2009)

Fisher, E.S., Goodman, D., Skinner, J., Bronner, K.: Health care spending, quality, and outcomes. The Dartmouth Institute Policy Brief (February).

Fisher E.S., Wennberg, D., Stukel, T., Gottlieb, D., Lucas, F.L., Pinder, E.L.: The implications of regional variations in medicare spending. Part 1: The content, quality and accessibility of care. Annals of International Medicine **138**, 283-287 (2003)

Garber, A.M., Skinner, J.: Is American health care uniquely inefficient. Journal of Economic Perspectives **22**, 27-50 ((2008)

Granger C.W.J.: Non-linear models: Where do we go next – Time varying parameter models? Studies in Nonlinear Dynamics and Econometrics **12**, 1-9 (2008).

Greene, W.H.: Econometric analysis, 6th edn. Upper Saddle River, New Jersey: Pearson, Prentice Hall (2008)

Hall R.E., Jones, C.I.: The value of life and the rise in health spending. Quarterly Journal of Economics, **122**, 39-72 (2007)

Jones, C.I.: More life vs more goods: Explaining rising health expenditure, FRBSF Economic Letter **27,** (2005)

Judge, G.G., Griffiths, W.E., Hill, R.C., Lütkepohl, H., Lee, T.: The theory and practice of econometrics, 2nd edn. New York: John Wiley and Sons (1985)

Lehmann, E.L.: Elements of large-sample theory. Berlin Heidelberg New York: Springer (1999) Lehmann, E.L., Casella, G.: Theory of point estimation, 2nd edn. Berlin Heidelberg New York: Springer (1998) Pratt, J.W., Schlaifer, R.: On the nature and discovery of structure. J Am Stat Assoc **79**, 9-21, 29-33 (1984)

Skinner, J.S., Fisher, E.S. Wennberg, J.: The efficiency of madicare in: Analyses in the economics of ageing (ed.) David A. Wise, NBER research, University of Chicago Press, Chicago (2005)

Skinner J.S., Staiger, D.: Technology diffusion and productivity growth in health care, NBER working paper No. 14865 (2009)

Swamy, P.A.V.B., Tinsley, P.A.: Linear prediction and estimation methods for regression models with stationary stochastic coefficients. J Econ **12**, 103-142 (1980)

Swamy, P.A.V.B., Mehta, J.S.: Bayesian and non-Bayesian analysis of switching regressions and a random coefficient regression model. Journal of the American Statistical Association **70**, 593-602 (1975)

Swamy, P.A.V.B., Mehta, J.S., Singamsetti, R.N.: Circumstances in which different criteria of estimation can be applied to estimate policy effects. J Statis Plan Inf **50,** 121-153 (1996)

Swamy, P.A.V.B., Tavlas, G.S.: Random coefficient models, Chap. 19. In: Baltagi, B.H. (ed.) A companion to theoretical econometrics. Malden: Blackwell 2001

Swamy, P.A.V.B., Chang I-L., Mehta J.S., Tavlas, G.S.: Correcting for omitted-variables and measurement-error bias in autoregressive model estimation with panel data. Comput Econ **22**, 225–253 (2003)

Swamy, P.A.V.B., Tavlas, G.S.: The new Keynesian Phillips curve and inflation expectations: Re-specification and interpretation. Econ Theory **31**, 293-306 (2007)

Swamy, P.A.V.B., Tavlas, G.S., Mehta, J.S.: Methods of distinguishing between spurious regressions and causality. J Stat Theory Appl **1**, 83-96 (2007)

Swamy, P.A.V.B., Yaghi, W., Mehta, J.S., Chang, I-L.: Empirical best linear unbiased prediction in misspecified and improved panel data models with an application to gasoline demand. Comput Statist Data Anal **51**, 3381-3392 (2007)

Swamy, P.A.V.B., Tavlas, G. S., Hall, S. G., Hondroyiannis, G.: Estimation of parameters in the presence of model misspecification and measurement error. Studies in Non-linear Dynamics and Econometrics **14**, 1-33 (2010)

Zellner, A.: Causality and econometrics, pp. 9-54. In: Brunner, K., Meltzer, A.H. (eds.) Three aspects of policy and policy making. Amsterdam: North-Holland 1979

Zellner, A.: Causality and causal laws in economics, J Econometrics, Annals, **39,** 7-21 (1988)