Macroeconomic Dynamics, **19**, 2015, 1633–1646. Printed in the United States of America.
doi:10.1017/S1365100513000928

A NOTE ON GENERALIZING THE CONCEPT OF COINTEGRATION

STEPHEN G. HALL
Leicester University
Bank of Greece
and
Pretoria University

P.A.V.B. SWAMY
Federal Reserve Board (Retired)

GEORGE S. TAVLAS
Bank of Greece

Building on the time-varying-coefficient (TVC) model, we propose a generalization of the concept of cointegration, allowing for the possibility that a set of variables measured with error entails a nonlinear relationship with unknown functional form. Both the dependent and explanatory variables of this relationship may be nonstationary (not necessarily of unit-root type), but there exists a nonlinear combination of all these explanatory variables that completely explains all the variation in the dependent variable. The TVC model allows us to test for the presence of this generalized cointegration in the absence of knowledge of the true nonlinear functional form and the full set of explanatory variables. We present the basic stages of the technique and discuss in detail how the issues of nonstationarity and cointegration affect each stage of the TVC estimation procedure.

Keywords: Generalized Cointegration, Time-Varying-Coefficient Model

1. INTRODUCTION

As a concept, cointegration is fundamental to empirical work in macroeconomics, as it is at the heart of understanding dynamic structures.[1] The link between cointegration and causality, which is emphasized by the Granger representation theorem, makes this very clear. The conventional definition of cointegration, however, will identify an economic structure if that structure happens to be linear, but will fail to work adequately if the true structure is nonlinear. Clearly, most macroeconomic theory gives rise to nonlinear structures and so, in general, conventional cointegration is not applicable. In this paper, we relate the recent literature dealing with the time-varying-coefficient (TVC) model to the concept of cointegration.[2]

We are grateful to a referee for comments on an earlier draft of this paper. The views expressed in this paper are the authors' own and do not necessarily represent those of their respective institutions. Address correspondence to: George S. Tavlas, Monetary Policy Council, Bank of Greece, 21 El. Venizelos Ave., 102 50 Athens, Greece; e-mail: gtavlas@bankofgreece.gr.

© 2014 Cambridge University Press 1365-1005/14

https://doi.org/10.1017/S1365100513000928 Published online by Cambridge University Press

Specifically, we develop a more general definition of cointegration than has been previously provided in the literature.[3] The extension to the literature in what follows makes the link between cointegration and causality even more apparent because it places emphasis on the need for identifying true economic structures.

Although developments in cointegration have been a focus of time-series econometrics for about 20 years, these developments have occurred largely within a linear framework. Although there have been various extensions to a nonlinear framework, these extensions have generally been limited to specific nonlinear functional forms.[4] The reason for this situation is as follows: In light of its standard definition, given in Engle and Granger (1987), cointegration becomes much easier to implement when the functional form of the relationship is assumed to be linear. Therefore, although it was found to be relatively straightforward to ask whether a linear functional form linked two or more variables together to produce a cointegrating combination, it was not obvious how to answer 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? Of course, the spirit of this question is precisely what was being asked in the cited paper by Engle and Granger, as well as in other work on cointegration [e.g., Cuthbertson et al. (1990); Enders (2009)]. However, there had been no way to make this general question tractable. Consequently, a much more limited linear framework is typically adopted. In this paper, we 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 remainder of the paper is divided into three sections. In Section 2, we introduce the concept of generalized cointegration. In Section 3, we present unusual interpretations of the coefficients of a nonlinear relationship and its underlying assumptions. In Section 4 we conclude.

2. GENERALIZED COINTEGRATION AND THE DEFINITION OF INTEGRATION

The idea underlying cointegration is that if there is a stable structural relationship linking a dependent variable with a group of explanatory variables, then, regardless of the time-series properties of these variables, there should be a combination (function) of the explanatory variables that gives a plausible explanation of the dependent variable. This combination is usually expressed within a linear regression framework in terms of integrated variables. A unit-root nonstationary series is integrated of order d, denoted $I(d)$, if it becomes stationary after being first-differenced d times. In a special linear case, if y_t is a vector of n variables, all integrated of the same order d, then $\Delta^d y_t = C(L)\varepsilon_t$, where there are no linearly deterministic components; $\{\varepsilon_t\}$, $t = -\infty, +\infty$, is a sequence of zero-mean, uncorrelated n-vectors with the same finite constant covariance matrix; $C(L)$ is an $n \times n$ invertible matrix of the polynomials in the lag operator L; and $(1 - L)^d y_t = \Delta^d y_t$ is the dth difference of y_t. Cointegration is said to occur if

https://doi.org/10.1017/S1365100513000928 Published online by Cambridge University Press

two constant n-vectors, say α and β, exist such that the error term of the model $\beta'\Delta^d y_t = \beta' C(L)\varepsilon_t$ of $\Delta^d y_t$ is stationary and the error term of the model $\alpha' y_t$ of y_t is integrated of order $d - b$ ($I(d - b)$), $d \geq b > 0$. The difference $d - b$ may not be zero unless $d = 1$. It is known that both α and β are not unique; for a survey of cointegration, see Dolado et al. (2001). The TVC models discussed in this paper have the models with unique coefficients and error terms as their bases. Our concept of uniqueness is given in note 2.

In the preceding discussion, we used a unit-root nonstationary process to describe the conventional linear cointegration model. We now turn to the nonlinear case. To deal with a potentially nonlinear data-generating model, which is assumed to have an unknown functional form, we need a more general definition of nonstationarity and cointegration than is typically assumed. Consider a variable that is integrated of order d. When $d = 0$, such a variable is (weakly or strongly) stationary, and when $d = 1$, it is unit-root nonstationary. Yet it is straightforward to demonstrate that there are also nonstationary variables that are not unit-root nonstationary. In this connection, consider an example provided by Cramér (1946), who showed that, for any general nonstationary process $\{x_t\}$, there is a uniquely determined decomposition $x_t = x_t^* + \varepsilon_t$, where x_t^* and ε_s are uncorrelated $\forall t$, x_t^* is deterministic, and ε_t is purely nondeterministic. The last may be represented as

$$\varepsilon_t = \sum_{j=0}^{\infty} c_{jt} a_{t-j}, \tag{1}$$

where the c_{jt} are time-dependent such that $\sum_{j=0}^{\infty} c_{jt}^2 < \infty$ for all t, and $\{a_t\}$ is a sequence of uncorrelated variables. As this definition makes clear, the time-dependent coefficients (c_{jt}) are associated with nonstationary processes. Furthermore, as shown by Swamy et al. (2003), model (1) of x_t can be transformed into an autoregressive model with time-dependent coefficients. Thus, a simple nonstationary process may be expressed as[5]

$$x_t = \gamma_{0t} + \gamma_{1t} x_{t-1}, \tag{2}$$

where γ_{0t} and γ_{1t} are time-dependent and x_t is dependent on x_{t-1}.[6] Thus, equation (2) is linear in variables and nonlinear in coefficients and its first difference may be expressed as

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

which typically is neither stationary nor unit-root nonstationary because the last term in (3) contains the level of x_{t-2}. Hence, x_t in (2) is non-unit-root nonstationary and is not integrated.

The upshot of this discussion is that the dependent variables of nonlinear autoregressive relationships are not integrated [see Swamy et al. (2003) and

https://doi.org/10.1017/S1365100513000928 Published online by Cambridge University Press

Berenguer-Rico and Gonzalo (2012)]; the same is true of the dependent variable of a nonlinear relationship of the form $y_t = f_t(x_{1t}, \ldots, x_{L_t,t})$, although one or more of its regressors $x_{1t}, \ldots, x_{L_t,t}$ may be integrated unless all of its regressors follow nonlinear relationships of the form $x_{gt} = \psi_{gt}(x_{1t}, \ldots, x_{g-1,t}, x_{g+1,t}, \ldots, x_{L_t,t})$ for g = 1, …, L_t [see Swamy et al. (2010)]. Also, Δx_t in (3) does not possess a finite unconditional mean if x_t and/or the coefficients of (2) follow random-walk processes. Furthermore, it can be seen from (3) that, every time equation (2) is differenced, additional terms enter into the resulting expression, giving a nonparsimonious representation unless equation (2) is linear or its intercept (excluding its error term component) and slope are constant, which will not generally be the case.

Economic theory makes it clear that most economic relationships are nonlinear. In addition, many economic variables are not, in theory, integrated variables (e.g., any series that exhibits long-term growth is not integrated, although its logarithm is integrated; any series that exhibits a break in its growth rate is not integrated, etc.). Thus, although the notion of cointegration is an extremely general one, the specific implementation of it in the standard way is very limited.

In sum, although there are a number of alternative definitions of cointegration, there is no simple formal definition that captures the essence of cointegration in a fully general way.

One recent important generalization of cointegration is the asymptotic nonparametric estimation of a model such as $Y_t = f(X_t) + W_t$, set out in Karlsen et al. (2007), who assume that $f(X_t)$ is some nonlinear function of a nonstationary process X_t and that the error process $\{W_t\}$ is stationary.[7] Those authors use a nonparametric kernel estimator, in that $f(X_t)$ is treated as an unknown function. However, this approach considers only part of the problem that we attempt to tackle here, as their work assumes that X is a single variable, or at least that if X is a vector, then the complete set of X variables is included. In the approach we adopt in the following, our definition of generalized cointegration and our implementation of the concept allow the researcher to observe only a subset of the complete X vector, whereas, at the same time, it permits this subvector to be observed with error. This approach requires a rather different definition of cointegration, as we now explain.

To generalize the notion of cointegration, we propose the following definition, which allows for nonlinearity and omitted regressors. The key to our definition is that we assert that the existence of cointegration implies a structural economic relationship. By this, we mean that a (possibly) nonlinear relationship exists between a dependent variable and a set of variables that includes all relevant preexisting conditions, besides all the determinants of the dependent variable. As shown in the following, this relationship (i) involves certain regressors for which there are no data and, hence, is reduced to another relationship in which the intercept contains three components, including the function (with the correct functional form) of certain "sufficient sets" of omitted regressors treated as the error term, and (ii) the coefficient on each included regressor contains three components including two (specification) bias terms and one bias-free term. This bias-free component of

https://doi.org/10.1017/S1365100513000928 Published online by Cambridge University Press

the coefficient on an included explanatory variable is the partial derivative of the dependent variable with respect to the explanatory variable, holding constant the values of all relevant preexisting conditions and the determinants of the dependent variable other than the explanatory variable. As also shown in the following, these coefficients, including the intercept, are expressed as linear functions of certain coefficient drivers and error terms that can be stationary. Thus, in this framework, there are several error terms. We may think of this either as a full dynamic model, in which case the error terms should be white noise, or as a long-run relationship involving only the nonstationary variables, in which case the error terms would normally be stationary ARMA processes and would capture the relevant dynamics.

DEFINITION

With this background, the variables y_t and x_t are cointegrated in a general sense if y and x are nonstationary and the "true" bias-free component of the time-varying coefficient of x_t (that is, its coefficient without specification biases) in the relation of y_t to x_t is nonzero.[8]

To explain, consider the following (real world) structural general relationship between y, x, and a set of other variables w, all of which are assumed to be nonstationary:

$$y_t = f_t(x_t, w_t), \tag{4}$$

where w_t includes all relevant preexisting conditions, besides all the determinants of y_t other than x_t. Therefore, under our definition of generalized cointegration, y and x, both of which are measured without errors, are cointegrated if

$$\frac{\partial y_t}{\partial x_t} \neq 0, \tag{5}$$

where the values of all the elements of w_t are held constant. Under this definition of generalized cointegration[9], cointegration is clearly defined as a property of the real world—not of any particular statistical model. This definition allows y and x to have different forms of nonstationarity, as w (which, of course, may be a vector) will allow us to reduce any spurious correlation to zero by letting us control for all relevant preexisting conditions while maintaining balance in the overall equation.[10]

The preceding formulation is very much in keeping with the original idea of cointegration. That is, cointegration should arise only if there is a (possibly nonlinear) stable structural relationship holding a set of variables together. If there is such a relationship, it implies that the true effect of x on y will be nonzero. Thus, if the following equation holds,

$$\frac{\partial f_t(x_t, w_t)}{\partial x_t} = 0, \tag{6}$$

it implies that there is no structural relationship between the variables (y_t,x_t,w_t), so that any observed correlation between the two variables (y_t,x_t) is spurious.

https://doi.org/10.1017/S1365100513000928 Published online by Cambridge University Press

Alternatively, if we run a standard regression between x and y, we may falsely obtain a significant coefficient. To make this definition of cointegration operational, we need an estimation technique that will provide bias-free estimates of parameters for which the true functional form is unknown and where, in addition, there may be omitted regressors.

3. UNUSUAL INTERPRETATIONS OF THE COEFFICIENTS OF A NONLINEAR RELATIONSHIP AND ITS UNDERLYING ASSUMPTIONS

3.1. Interpretations of Model Coefficients and Appropriate Assumptions

In this section, we will begin by giving a largely intuitive account of our estimation strategy, which makes the idea of generalized cointegration operational.[11] TVC estimation proceeds from an important theorem that was first established by Swamy and Mehta (1975) and that was subsequently confirmed by Granger (2008). This theorem states that any nonlinear functional form can be exactly represented by a model that is linear in variables but that has time-varying coefficients. The implication of this result is that, even if we do not know the correct functional form of a relationship, we can always represent this relationship as a time-varying coefficient relationship and thus estimate it. Hence, any nonlinear real-world relationship may be stated as

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

Consequently, this theorem leads to the result that, if we have the complete set of relevant variables with no measurement error, then by estimating a TVC model, we will get consistent estimates of the true partial derivatives of the dependent variable with respect to each of the independent variables, given the unknown, nonlinear functional form. If we then allow for the fact that we do *not* know the full set of independent variables and that some, or all, of them may be measured with error, then the TVC become biased (for the usual reasons).[12] What we would like is to have is some way to decompose the full, biased, time-varying coefficients into two parts, the bias component and the remaining part, which would again be a consistent estimate of the true component. Of course, this is asking a great deal of an estimation technique. However, that is precisely what TVC estimation aims to provide [see Swamy et al. (2010)]. This technique builds from the Swamy and Mehta theorem, mentioned previously, to produce such a decomposition.[13]

Swamy et al. (2010) show exactly what happens to the time-varying coefficients as other forms of misspecification are added to the model. If we omit some relevant variables from the model, then the true partial derivative components of the time-varying coefficients get contaminated by a term that involves the relationship between the omitted and included variables. Also, if we allow for measurement error, then the time-varying coefficient gets further contaminated by a term that allows for the relationship between the exogenous variables and the error terms. Thus, as one might expect, the estimated time-varying coefficient is no longer a

https://doi.org/10.1017/S1365100513000928 Published online by Cambridge University Press

consistent estimate of the true partial derivatives of the nonlinear function, but is now biased because of the effects of omitted variables and measurement error. There are exact mathematical proofs for our statements up to this point.

To make TVC estimation fully operational, we need to make some parametric assumptions. We make two key assumptions. First, we assume that the time-varying coefficients themselves are determined by a set of stochastic linear equations, which makes them functions of a set of variables, which we call driver (or coefficient-driver) variables. This is a relatively uncontroversial assumption. Second, we assume that some of these drivers are correlated with the misspecification in the model and some of them are correlated with the time variation coming from the nonlinear (true) functional form. Having made this assumption, we can then simply remove the bias from each time-varying coefficient by removing the effect of the set of coefficient drivers that are correlated with the misspecification. This procedure, then, yields a consistent set of estimates of the true partial derivatives of the unknown nonlinear function, which may then be tested by constructing t tests in the usual way.

3.2. Identification and Coefficient Drivers

We have argued that generalized cointegration takes place if the bias-free component of the coefficient linking two variables is nonzero. To test whether this situation applies. we are interested in the bias-free components of γ's—not in the omitted-variable and measurement-error biases. To obtain accurate estimates of theα_{jt}^{*}'s using the observations in (7), we need to first decompose each γ_{jt} of (7) into its bias and unbiased components. Our method of identifying these components and performing the decomposition is based on the following assumptions.[14]

Assumption 1 (Auxiliary Information). Each coefficient is linearly related to certain drivers plus a random error,

$$\gamma_{jt} = \pi_{j0} + \sum_{d=1}^{p-1} \pi_{jd} z_{dt} + \varepsilon_{jt} \quad (j = 0, 1, \ldots, K-1), \tag{8}$$

where the πs are fixed parameters and the z_{dt} are what we call the coefficient drivers; different coefficients of (7) can be functions of different sets of coefficient drivers.

The regressors and the coefficients of (7) are conditionally independent of each other given the coefficient drivers.[15] These coefficient drivers are merely a set of variables that, to a reasonable extent, jointly explain the movement in γ_{jt}. If the variation in the coefficients is due to some form of misspecification (say, omitted variables), then any variable that is correlated with the misspecification may act as a driver; for example, the drivers might include lagged explanatory variables. If the variation is due to nonlinearity, then, again, lags in some of the variables in the model will likely be related to the changing coefficient. An important part of our

https://doi.org/10.1017/S1365100513000928 Published online by Cambridge University Press

contention here is that it is relatively straightforward in practice to find variables that are correlated with the misspecification.

The total number of components in each coefficient of (7) is three, as shown in note 12. If the number of nonconstant coefficient drivers we could find is greater than or equal to $3K$, then in equation (8), for each coefficient of (7) there will be at least three appropriate nonconstant coefficient drivers, one constant, and one error term. In this case, there will be at least one nonconstant coefficient driver to estimate each component of every coefficient of (7). To estimate a component accurately, we need to choose at least one nonconstant coefficient driver in such a way that a linear function of the chosen driver or drivers has the same kernel density estimator as the component. Such coefficient drivers exist and Assumption 1 is not unrealistic. Thus, although it is not easy to find such coefficient drivers, it is easy to prove their existence.

Under our method, the coefficient drivers included in equation (8) have two uses. Insertion of equation (8) into equation (7) parameterizes the latter equation. This is the first use of the coefficient drivers. Here, the issue of identification of the parameterized model (7) is important.[16] The other important use of the drivers is to allow us to separate the bias and bias-free components of the coefficients.

We divide the complete set of coefficient drivers in each equation (8) into three sets, the first of which is associated with the time variation in the true coefficient, the second with the omitted-variable bias, and the third with any measurement error. This division allows us to identify separately the bias-free, omitted-variables, and measurement-error bias components of the coefficients of (7).

This division is the key to making our procedure operational; it is the division in which we can associate the various forms of specification biases with the second and third sets, which means that the first set simply explains the time variation in the coefficients, which is caused by the nonlinearity in the true function with unknown functional form. If the true (or data-generating) model is linear, then all that is required for the first set is to contain the constant of (8). If the true model is nonlinear, then the bias-free components should be time-varying and the set of drivers belonging to the first set will explain the time variation in these components.

3.3. Consistent Estimation

Under certain conditions, the iteratively rescaled generalized least-squares estimators of the coefficients in (8) are consistent. With these estimates, Lehmann and Casella's (1998, Theorem 5.3, p. 467) method of solving the likelihood equations gives asymptotically efficient estimators.[17] The distributional theory underlying this estimation technique and the method for conducting inference are given in Swamy et al. (2010). It may seem surprising that the inference is standard rather than dependent on the Dickey–Fuller distribution (or some other nonstandard distribution); the intuitive reason that this comes about is that the distribution of the TVCs is derived from the errors in the coefficient driver equations (8). As long

https://doi.org/10.1017/S1365100513000928 Published online by Cambridge University Press

as these errors are stationary, the distribution of the coefficients of (7) will be of the Cavanagh–Rothenberg (1995, 279–280) type [see Swamy et al. (2010)]. This might be thought to be a challenging requirement, as of course the time-varying coefficients may well be nonstationary, and so, to achieve a stationary error process in the fixed coefficient linear driver, equation (8) might at first seem to require conventional cointegration to exist here. However, this is not the case, as the driver equations may be dynamic and therefore may contain lags of all the variables included in (7). It is possible to show [using Cramér's (1946) decomposition] that sufficient lags in these variables will always ensure a stationary error in (8), and hence inference is standard.

To illustrate, consider the standard case of testing a linear relationship between x and y for cointegration. Dolado et al. (2001, 639–642) give a clear description of these tests. Assume we have $x_t \sim I(1)$, $y_t \sim I(1)$. Then the conventional approach would be to run the regression $y_t = \beta_0 + \beta_1 x_t + \varepsilon_t$ using ordinary least squares and to test whether the resulting residuals are $I(1)$ against the alternative that they are $I(0)$. If, with some adjustments discussed in Dolado et al. (2001), the alternative is accepted, then it is concluded that x and y are cointegrated. We first estimated the regression $y_t = \beta_0 + \beta_1 x_t + \varepsilon_t$ under the null of no cointegration, $\varepsilon_t \sim I(1)$, and then drew the conclusion of cointegration under the alternative of cointegration, $\varepsilon_t \sim I(0)$, and it is this change in the properties of the errors under the null and the alternative that gives rise to the nonstandard distributions. In the generalized cointegration/TVC framework, the problem would be formalized in the following way. We would run the time-varying regression

$$y_t = \beta_{0t} + \beta_{1t} x_t,$$

where the coefficient driver equation would be

$$\beta_{it} = \alpha_{i0} + \alpha_{i1} x_{t-1} + \alpha_{i2} y_{t-1} + v_{it} \quad i = 0, 1.$$

Now, substituting the driver equations into the model yields

$$y_t = \alpha_{00} + \alpha_{01} x_{t-1} + \alpha_{02} y_{t-1} + (\alpha_{10} + \alpha_{11} x_{t-1} + \alpha_{12} y_{t-1}) x_t + v_{0t} + v_{1t} x_t.$$

Under the null of no cointegration, $\beta_{1t} = 0$ for all t, $\beta_{1t} = 0$ for all t, $\alpha_{00} = \alpha_{01} = 0$, and $\alpha_{02} = 1$ and the errors from this regression are stationary if there are no omitted lagged dependent variables. Under the alternative of cointegration, $\beta_{1t} = \beta_1 = \alpha_{10}$, $\alpha_{11} = \alpha_{12} = \alpha_{01} = \alpha_{02} = 0$ and again the error process is stationary if there are no omitted regressors. So under both the null and the alternative, the errors are stationary and standard inference results.

Generalized cointegration does two things. First, it allows for the possibility that we may have important omitted variables. Second, it allows for the possibility that we may have misspecified or not know the true functional form. That is, under generalized cointegration, we are able to estimate bias-free relationships among a set of variables even if we do not know the true, underlying functional form and even if there are missing regressors. Underlying generalized cointegration is

https://doi.org/10.1017/S1365100513000928 Published online by Cambridge University Press

TABLE 1. Coefficient drivers versus instruments

	Coefficient drivers	Instruments
Correlation with error term	No assumption	Zero correlation; if this fails instruments are invalid
Correlation with the misspecification	Should be correlated with the misspecification	Should be uncorrelated with misspecification, but correlated with the variable being instrumented
Correlation between variables	Set of variables correlated with the true coefficient should be uncorrelated with the other sets	No assumption

a new way of thinking about, and testing for, cointegration that emphasizes the properties of the real-world relationship rather than a particular model. If, in the real world, a causal cointegrating vector exists that determines a variable, say, the demand for a particular commodity, then, obviously, if one of the variables (say X) in that relationship changes, demand will also change. This implies that the partial derivative of demand with respect to X is nonzero.

3.4. Coefficient Drivers versus Instruments

How do coefficient drivers differ from instruments? In a practical application the choice of the coefficient drivers and the decision as to how to use them are somewhat arbitrary, in much the same way as the choice of instruments in an instrumental variable estimation. Different drivers can give different answers, as can dividing the drivers into the relevant three sets in different ways. It is worth contrasting the different assumptions regarding drivers and instruments; Table 1 provides a comparison.

For instrumental variables we need variables that are relevant (correlated with the variable being instrumented), but independent of the error process (the misspecification in the model); for good drivers we need variables that are correlated with the misspecification, but that can be split into two sets that identify the bias from the total coefficient. In practice, it is typically much easier to find variables correlated with the misspecification than variables uncorrelated with the misspecification, so this argues in favor of the driver approach.

4. CONCLUSIONS

Building on the TVC model, we proposed a generalization of the standard definition of cointegration that allows for the existence of an unknown structural

https://doi.org/10.1017/S1365100513000928 Published online by Cambridge University Press

nonlinear relationship among a set of nonstationary variables. The idea underlying this definition is straightforward: If a structural relationship exists between two or more variables, then the implication is that there will be a nonzero bias-free effect of any of the independent variables on the dependent variable. Thus, the significance of an estimate of this bias-free effect becomes a simple direct test of generalized cointegration. Furthermore, we can estimate this effect and test its significance without knowing the true functional form of the relationship or the full set of regressors that should enter into it. This definition can be made operational by applying the TVC estimation technique, which provides an estimate of the bias-free effect.

Nonstationarity does not pose any problem for TVC estimation. TVC estimation by construction produces a unique error term that is the correct function of certain "sufficient sets" of omitted variables, whereas standard cointegration aims at a stationary error term, but does not always produce such a stationary error. However, as in other modeling situations, the explicit recognition of nonstationarity does offer advantages—n particular, in the identification of the correct set of coefficient drivers to identify the bias-free component of the time-varying coefficient correctly.

NOTES

1. See, for example, Arouri et al. (2012), and Lee (2013).

2. A recent and expanding literature has been concerned with building on the TVC model of Swamy (1971, 1974). In this connection, Granger (2008) argued that TVC models will be the next major development in econometrics. The so-called correlated random coefficient model is rigorously derived in a long sequence of papers that include Swamy and Tavlas (2001) and Swamy et al. (2014). Here the term "rigorously derived model" is used to convey the idea that the coefficients and error term of the model are unique. The coefficients and error term of a model are said to be unique if they remain invariant under equivalent changes in the relationship between the included and excluded regressors of the model.

3. We use the term "generalized cointegration" despite the fact that integration is an inherently linear concept, as we believe that this conveys the essence of what we are doing here, which is to extend the notion of cointegration to a nonlinear framework.

4. See, for example, Park and Phillips (2001); Kanas (2003); Gonzalo and Pitarakis (2006); Karlsen et al. (2007); Kasparis (2008, 2011); Al-Abri and Goodwin (2009); Wang and Phillips (2009); Choi and Saikkonen (2010); Schienle (2011); Banerjee and Pitarakis (2012) ; and Berenguer-Rico and Gonzalo (2012).

5. It is also possible to represent the process as a function of more than one lag. However, this is the easiest form of the process to handle and is most relevant in demonstrating our point simply. To avoid any misunderstanding here, we point out that between two models that perform equally well in explanation and prediction, the one with fewer unknown parameters is parsimonious. We are not claiming here that equation (2) is a parsimonious model.

6. If the true process is a conventional random walk without drift, then γ_{0t} should be a white noise process, γ_{1t} should be equal to 1 for all t and the model should be linear. Hence, the usual random walk case is a special case of equation (2).

7. See, also, the references cited in Karlsen et al. (2007).

8. As discussed in what follows, by "true" we mean the coefficient that links x to y in the real world structural relationship under consideration. The notion that y and x are themselves nonstationary is not crucial to our argument. In fact, in a nonlinear world, it is even possible to think of a variable being

https://doi.org/10.1017/S1365100513000928 Published online by Cambridge University Press

stationary at one point in time and nonstationary at another. However, this assumption does keep the analysis closer to the original spirit of cointegration.

9. To give some specific examples, consider $y_t = \beta x_t + \varepsilon_t$ (which we consider to be the true model), where $x_t \sim I(2)$ and, therefore, $y_t \sim I(2)$. Here the symbol $I(d)$ denotes integrated of order d; a nonstationary series is integrated of order d if it becomes stationary after being first differenced d times [see Greene (2008, p. 740)]. Then y_t and x_t are cointegrated in the conventional sense and in our generalized sense as $dy/dx = \beta$. As another example, consider a true nonlinear relationship $y_t = \beta x_t^2 + v_t$, where $x_t \sim I(1)$ and $v \sim I(0)$; then y will not generally be integrated and y and x will not be cointegrated in the Engle–Granger sense. However, in this nonlinear case, y and x can be cointegrated in our generalized sense under the conditions of Swamy, Tavlas, and Hall (in press).

10. For a standard way to reduce spurious correlations to zero by controlling for all relevant preexisting conditions, see Skyrms (1988, p. 59), and for the definition of a balanced nonlinear relationship based on the concepts of the order of summability of a stochastic process and cosummability of stochastic processes, see Berenguer-Rico and Gonzalo (2012, p. 9). See also Swamy and von zur Muehlen (1988).

11. A formal and detailed description of the approach may be found in Swamy and Tavlas (2001, 2006) and Swamy et al. (2009, 2010, 2014, in press).

12. The intercept γ_{0t} of (7) is the sum of the true intercept, the correct function of certain "sufficient sets" of omitted regressors, and the measurement error in the dependent variable. For $j = 1, \ldots, K-1$, γ_{jt} is the sum of the partial derivative of the true value of y_t with respect to the true value of x_{jt} and the corresponding omitted-regressor and measurement-error biases.

13. Mathematically, model (7) may appear to be the same as the observation equation of a state space model. To show that this similarity is only apparent, we point out that the interpretations of the coefficients of the observation equation in the standard state space representation are quite different from those of the coefficients of (7). Omitted-variable biases, measurement-error biases, and the correct functions of certain "sufficient sets" of excluded regressors are not considered parts of the coefficients of the observation equations of state-space models. Also, the problem of the dependence of the coefficient of a regressor on the regressor does not arise in state space models. These are the major differences between (7) and the observation equations of a standard state-space model. Jawadi (2012) discusses the use of time-varying models in macroeconomics.

14. We say that the condition of identification of the components of coefficients is satisfied if the bias-free component of each coefficient is accurately estimated. This condition is different from Lehmann and Casella's (1998, 24 and 57) condition of identification. The latter condition is used to identify the fixed coefficients of equation (8) in this paper.

15. Given the coefficient drivers, the error terms of (8) are assumed to be conditionally independent of the regressors of (7). The distributional assumptions about the errors in (8) are given in Swamy et al. (2010).

16. To deal with this issue, we use Lehmann and Casella's (1998, 24 and 57) concept of identification.

17. A computer program that implements this technique is available at http://www.le.ac.uk/ec/sh222/soft.htm.

REFERENCES

Al-Abri, A.S. and B.K. Goodwin (2009) Re-examining the exchange rate pass-through into import prices using nonlinear estimation techniques: Threshold cointegration. *International Review of Economics and Finance* 18(1), 142–161.

Arouri, M., D. Jawadi, and D. Nguyen (2012) Modeling nonlinear and heterogeneous links in international monetary markets. *Macroeconomic Dynamics* 16, 232–251.

Banerjee, A. and J. Pitarakis (2012) Functional Cointegration: Definition and Nonparametric Estimation. MPRA Paper 38846, University Library of Munich.

https://doi.org/10.1017/S1365100513000928 Published online by Cambridge University Press

Berenguer-Rico, V. and J. Gonzalo (2012) Summability of Stochastic Processes: A Generalization of Integration and Co-integration Valid for Non-linear Processes. Working paper, Universidad Carlos III de Madrid.

Cavanagh, C.L. and T.J. Rothenberg (1995) Generalized least squares with nonnormal errors. In G.S. Maddala, P.C.B. Phillips, and T.N. Srinivasan (eds.), *Advances in Econometrics and Quantitative Economics*, Chap. 11. Oxford, UK: Blackwell.

Choi, I. and P. Saikkonen (2010) Tests for non-linear cointegration. *Econometric Theory* 26, 682–709.

Cramér, H. (1946) *Mathematical Methods of Statistics.* Princeton, NJ: Princeton University Press.

Cuthbertson, K., S.G. Hall, and M. Taylor (1990) *Applied Econometric Techniques.* Ann Arbor: University of Michigan Press.

Dolado, J.J., J. Gonzalo, and F. Marmol (2001) Cointegration. In B.H. Baltagi (ed.), *A Companion to Theoretical Econometrics,* Chap. 30. Malden, MA: Blackwell.

Enders, W. (2009) *Applied Econometric Techniques*. Hoboken, NJ: Wiley.

Engle, R.F. and C.W.J. Granger (1987) Co-integration and error correction: Representation, estimation, and testing. *Econometrica* 55(2), 251–276.

Gonzalo, J. and J.Y. Pitarakis (2006) Threshold effects in co-integrating regressions. *Oxford Bulletin of Economics and Statistics* 68, 813–833.

Granger, C.W.J. (2008) Nonlinear models: Where do we go next—Time-varying parameter models? *Studies in Nonlinear Dynamics and Econometrics* 12(3), 1–9.

Greene, W.H. (2008) *Econometric Analysis*, 6th ed. Upper Saddle River, NJ: Pearson/Prentice Hall.

Jawadi, F. (2012) Introduction to time-varying modeling with macroeconomic and financial data, *Macroeconomic Dynamics* 16, 167–175.

Kanas, A. (2003) Nonlinear cointegration between stock prices and dividends. *Applied Economics Letters* 10(7), 401–405.

Karlsen, H.A., T. Myklebust, and D. Tjøstheim (2007) Nonparametric estimation in a nonlinear cointegration type model. *Annals of Statistics* 35(1), 252–299.

Kasparis, I. (2008) Detection of functional form misspecification in cointegrating relations. *Econometric Theory* 24, 1373–1403.

Kasparis, I. (2011) Functional form misspecification in regressions with a unit root. *Econometric Theory* 27, 285–311.

Lee, C.C. (2013) Insurance and real output: The key role of banking activities. *Macroeconomic Dynamics* 17, 235–260.

Lehmann, E.L. and G. Casella (1998) *Theory of Point Estimation*, 2nd ed. Berlin: Springer-Verlag.

Park, J.Y. and P.C.B. Phillips (2001) Non-linear regression with integrated time series. *Econometrica* 69, 117–161.

Schienle, M. (2011) Nonparametric Nonstationary Regression with Many Covariates. Humboldt-Universität zu Berlin, Discussion paper 076.

Skyrms, B. (1988) Probability and causation. *Journal of Econometrics* 39(1/2), 53–68.

Swamy, P.A.V.B. (1971) *Statistical Inference in Random Coefficient Regression Models.* New York: Springer-Verlag.

Swamy, P.A.V.B. (1974) Linear models with random coefficients. In P. Zarembka (ed.), *Frontiers in Econometrics,* pp. 143–168. New York: Academic Press.

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

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

Swamy, P.A.V.B., J.S. Mehta, G.S. Tavlas, and S.G. Hall (2014) Small area estimation with correctly specified linking models, Chap. 10, In Jun Ma and Mark E. Wohar (eds.), *Recent Advances in Estimating Nonlinear Models,* Chap. 10, pp. 193–228. New York: Springer-Verlag.

https://doi.org/10.1017/S1365100513000928 Published online by Cambridge University Press

Swamy, P.A.V.B. and G.S. Tavlas (2001) Random coefficient models. In B.H. Baltagi (ed.), *A Companion to Theoretical Econometrics,* Chap. 19. Malden, MA: Blackwell.

Swamy, P.A.V.B. and G.S. Tavlas (2006) A note on Muth's rational expectations hypothesis: A time-varying-coefficient interpretation. *Macroeconomic Dynamics* 10, 415–425.

Swamy, P.A.V.B., G.S. Tavlas, and S.G. Hall (in press) Microproduction functions with unique coefficients and errors: A reconsideration and respecification. *Macroeconomic Dynamics*.

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

Swamy, P.A.V.B., G.S. Tavlas, and G. Hondroyiannis (2009) The New Keynesian Phillips curve in a time-varying coefficient environment: Some European evidence. *Macroeconomic Dynamics* 13, 149–166.

Swamy, P.A.V.B., and P. von zur Muehlen (1988) Further thoughts on testing for causality with econometric models. *Journal of Econometrics* 39, 105–147.

Wang, Q. and P.C.B. Phillips (2009) Structural nonparametric cointegrating regressions. *Econometrica* 77, 1901–1948.

https://doi.org/10.1017/S1365100513000928 Published online by Cambridge University Press