TIME-VARYING COEFFICIENT MODELS: A PROPOSAL FOR SELECTING THE COEFFICIENT DRIVER SETS

STEPHEN G. HALL
Leicester University
Bank of Greece
and
NIESR

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

GEORGE S. TAVLAS
Bank of Greece
and
Leicester University

Coefficient drivers are observable variables that feed into time-varying coefficients (TVCs) and explain at least part of their movement. To implement the TVC approach, the drivers are split into two subsets, one of which is correlated with the bias-free coefficient that we want to estimate and the other with the misspecification in the model. This split, however, can appear to be arbitrary. We provide a way of splitting the drivers that takes account of any nonlinearity that may be present in the data, with the aim of removing the arbitrary element in driver selection. We also provide an example of the practical use of our method by applying it to modeling the effect of ratings on sovereign-bond spreads.

Keywords: Time-Varying Coefficient Model, Coefficient Driver, Specification Problem, Correct Interpretation of Coefficients

1. INTRODUCTION

Time-varying-coefficient (TVC) estimation is a way of estimating parameters of a model even when (i) the true functional form is unknown, (ii) there are missing important regressors, and (iii) the included variables contain measurement errors.[1] Several successful applications of this technique have appeared in the recent

1158

literature, including Hall et al. (2008, 2009a, 2009b, 2012, 2013, 2015) and Swamy et al. (2015). However, an important assumption is needed to make the technique operational. This assumption concerns the choice of what are called "coefficient drivers" (formally defined later) and the separation of these drivers into subsets. This separation allows us to derive estimates of bias-free coefficients. Intuitively, coefficient drivers are a set of variables that feed into the TVCs and explain at least part of their movement. As explained in what follows, the set of drivers is split into two subsets, one of which is correlated with the bias-free coefficient that we want to estimate and the other with the misspecification in the model. This split can appear to be somewhat arbitrary [much as in the case of choosing instrumental variables) and has been a source of criticism of the TVC approach—see Hall et al. (2013)]. The problem here is that different driver sets will give different results and there is very little guidance as to which set of drivers should be preferred, much as is true for instrumental variable estimation. The objective of this paper is to put forward a method for producing this split that takes account of the nonlinearity that may be present in the original data. As we argue in the following, this method provides a natural split in the driver set.

The remainder of this paper is divided into four sections. Section 2 presents a summary of the TVC approach and formally defines the concept of coefficient drivers and the need to split the drivers into two sets. Section 3 then proposes a method for determining this split. Section 4 provides an example of the practical use of the technique by applying it to modeling the effect of rating agencies ratings on sovereign bond spreads. Section 5 concludes.

2. TVC ESTIMATION

Here, we summarize the approach to TVC estimation that has been formalized in Swamy et al. (2010). TVC estimation proceeds from an important theorem that was first established by Swamy and Mehta (1975) and that has subsequently been 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 TVC relationship and thus estimate it. Hence, any nonlinear relationship may be stated as

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

where $x_{1t}, \ldots, x_{K-1,t}$ are K-1 observed determinants of y_t. 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 reasonable estimates of the true partial derivatives of the dependent variable with respect to each of the independent variables given the unknown nonlinear functional form. Another theorem that is used to derive (1) should also be stated.

THEOREM.   *A necessary condition for a model to be true is that its coefficients and error term, having the correct functional forms, are unique.*

To explain this theorem, we need to introduce some appropriate terminology. The term "true model" is taken to mean the original real-world relationship underlying (1), which does not permit spurious relations.[2] A compressed real-world relationship is taken to mean the original real-world relationship compressed to have fewer regressors than the original relationship. The included regressors are those that are included in both original and compressed relationships. Omitted regressors are those that are included in the original relationship but not in the compressed relationship. No relevant regressors, including relevant preexisting conditions, are omitted from the original relationship, and hence it does not have an error term. The coefficients and error term of the compressed real-world relationship are unique if they are invariant under the changes made in the real-world relationships between each omitted regressor and the included regressors such that the equality signs in these relationships are unchanged; the unique coefficients on nonconstant regressors have the form of the sums of partial derivatives of the true value of the dependent variable with respect to the true values of included regressors and the corresponding omitted-regressors bias; the unique error term has the form of a function (with the correct functional form) of certain "sufficient sets" of omitted regressors; the coefficients cannot be unique unless they have the correct functional forms. These explanations of the theorem and its proof are given in Swamy et al. (2014). If we allow for the fact that we do *not* know the full set of independent variables and that some, or perhaps all, of them may be measured with error, then the TVCs become biased (for the usual reasons). What we would like to have is some way to decompose the full set of biased TVCs into two parts—the biased component and the remaining part; the latter would be a bias-free true component. Although this is asking a great deal of an estimation technique, it is precisely what TVC estimation aims to provide [see Swamy et al. (2010)]. This technique builds from the Swamy and Mehta theorem, mentioned earlier, to produce such a decomposition.[3]

Swamy et al. (2010) show what happens to the TVCs as other forms of misspecification are added to the model. If we compress a real-world relationship, then the true TVCs get contaminated by a term that involves the relationship between the omitted and included regressors. If we also allow for measurement error, then the TVCs become further contaminated by a term that allows for the relationship between the included regressors and their measurement errors. Thus, as one might expect, the estimated TVCs are no longer the true partial derivatives of the nonlinear function. Instead, they are biased because of the effects of omitted regressors and measurement error. There is no estimation so far. There are exact mathematical proofs for our statements up to this point.

To make TVC estimation fully operational, we need to make two key parametric assumptions. First, we assume that the time-varying coefficients themselves are determined by a set of stochastic linear equations, which makes them a function of

a set of variables we call driver (or coefficient-driver) variables. This is a relatively uncontroversial assumption. The only restrictive assumption is that the relationship is linear; this may be viewed as a good first approximation. It would be possible to generalize this assumption to any other nonlinear parametric function without great difficulty. 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 the time-varying coefficients by removing the effect of the set of coefficient drivers that are correlated with the misspecification. This procedure, then, yields a reasonable set of estimates of the true partial derivatives of the unknown compressed real-world nonlinear relationship, which may then be tested by constructing t tests in the usual way. An important difference between coefficient drivers and instrumental variables is that for a valid instrument we require variables that are uncorrelated with the misspecification. Such variables often prove hard to find. For a valid driver, we need variables that are correlated with the misspecification. We would argue that this is much easier to achieve than finding instruments that are uncorrelated with the misspecification.

To formalize the idea of the coefficient drivers, we assume that each of the TVCs in (1) is generated in the following way.

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

$$\gamma_{jt} = \pi_{j0}z_{0t} + \sum_{d=1}^{p-1} \pi_{jd}z_{dt} + \varepsilon_{jt} (j = 0, 1, \ldots, K-1), \qquad (2)$$

where the π s are fixed parameters, the z_{dt} are what we call the observed coefficient drivers, and $z_{0t} = 1$. Different coefficients of (1) can be functions of different sets of coefficient drivers.

Assumption 2. For all t, $E[\varepsilon_t = (\varepsilon_{0t}, \varepsilon_{1t}, \ldots, \varepsilon_{K-1,t})' | z_t = (1, z_{1t}, \ldots, z_{p-1,t})]$ $= 0$. The regressors and the coefficients of (1) are conditionally independent of each other given the coefficient drivers.[4]

These coefficient drivers alone are merely a set of variables that, to a reasonable extent, jointly explain the movement in γ_{jt}. It should be noted that there are two sources of variation in y_t. The regressors and the coefficients of (1) jointly explain the movement in y_t. Under our method, the coefficient drivers included in equation (2) have two uses. Insertion of equation (2) into equation (1) parameterizes the latter equation. This is the first use of the coefficient drivers. Here, the issue of identification of the parameterized model (1) is important.[5] The other important use of the drivers allows us to separate the bias and bias-free components of the coefficients.

Assumption 3. The set of coefficient drivers and the constant term in equation (2) divide into three different subsets, A_{1j}, A_{2j}, and A_{3j} such that the first set is correlated with any variation in the true parameter that is due to the underlying relationship being nonlinear, the second set is correlated with bias in the parameter coming from any omitted regressors, and the final set is correlated with bias coming from measurement error.

This assumption allows us to identify separately the bias-free, omitted-variables, and measurement-error bias components of the coefficients of equation (1). Assumption 3 is the key to making our procedure operational; it is the assumption that we can associate the various forms of specification biases with sets A_{2j} and A_{3j}, which means that set A_{1j} simply explains the time variation in the coefficients caused by the nonlinearity in the true functional form of the compressed real-world relationship underlying (1). If this relationship is linear, then all that will be required for set A_{1j} is to contain the constant z_{0t} of (2). If this relationship is nonlinear, then the bias-free components should be time-varying and the set of drivers belonging to A_{1j} will explain the time variation in these components. There are essentially two sets of variables here—the A_{1j}, which is associated with the true nonlinearity in (1), and the A_{2j} and A_{3j}, which are associated with the misspecification. For ease of notation, hereafter we will refer to A_{1j} as S_1, and we will refer to the joint set of A_{2j} and A_{3j} as S_2.

3. A SUGGESTION FOR THE CHOICE OF COEFFICIENT DRIVERS

Clearly, Assumptions 1–3 are crucial for the successful implementation of the TVC approach. As noted earlier, the split of coefficient drivers stemming from these assumptions has been a problematic part of the TVC-estimation procedure. There are, however, certain requirements that can help in selecting both the sets of variables that make a good driver set and the split into two subsets inherent under Assumption 3.

3.1. Selecting the Complete Driver Set

Consider, first, the broad requirements that a complete set of drivers should fulfill; these relate to predictive power and relevance. To explain, we again present equation (2),

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

where $z_{0t} \equiv 1$. For this set of drivers to be a good set, the drivers must explain most of the variation in γ_{jt}. Hence, we can define an analog of the conventional

R^2 for the estimated counterparts to these equations as follows:

$$R^2 = 1 - \frac{SS\varepsilon_{jt}}{SS\gamma_{jt}}, \tag{3}$$

where $SS\varepsilon_{jt}$ and $SS\gamma_{jt}$ are the sum of squared residuals and the total variation of the dependent variable, respectively. A main difficulty with (3) is that if $E(\varepsilon_{jt}\varepsilon_{j't'}) \neq 0$ for $j \neq j'$ or $t \neq t'$ and we use a fully efficient estimator such as GLS, then (3) will have a range from $-\infty$ to 1 and it will be difficult to interpret it as the proportion of variation in γ_{jt} explained by equation (2) [see Judge et al. (1985, pp. 31–32)]. To deal with this problem, let $\varepsilon_t = (\varepsilon_{0t}, \varepsilon_{1t}, ..., \varepsilon_{K-1,t})'$. Then assume that, given $z_t = (z_{0t}, z_{1t}, ..., z_{p-1,t})'$, $\varepsilon_t = \Phi\varepsilon_{t-1} + u_t$, where $E(u_t|z_t) = 0$ and $E(u_t u_{t'}'|z_t) = \{^{\sigma_u^2 \Delta_u \text{ if } t=t'}_{0 \text{ if } t\neq t'}$, Δ_u being a non-negative matrix. With these assumptions, the covariance matrix of $\varepsilon = (\varepsilon_1', ..., \varepsilon_T')'$ can be represented as $E(\varepsilon\varepsilon'|z_1, ..., z_T) = \sigma_u^2 \Sigma_u$. The elements of this covariance matrix are displayed in Chang et al. (1992). A goodness of fit measure for (2) that resembles Judge et al.'s (1985, p. 32) goodness measure is

$$R_\gamma^2 = 1 - \frac{\varepsilon'\Sigma_u^{-1}\varepsilon}{(\gamma - \iota_{TK}\bar{\gamma}_w^*)'\Sigma_u^{-1}(\gamma - \iota_{TK}\bar{\gamma}_w^*)}, \tag{4}$$

where (2) is written in matrix form as $\gamma_t = \Pi z_t + \varepsilon_t$, $\gamma_t = (\gamma_{0t}, \gamma_{1t}, ..., \gamma_{K-1,t})'$ is $K \times 1$, Π is a $K \times p$ matrix of the π's in (2), z_t is $p \times 1$, ε_t is $K \times 1$, $\gamma = (I_T \otimes \Pi)z + \varepsilon$, γ is a $TK \times 1$ vector of γ's in (2), $z = (z_1', ..., z_T')'$, $\varepsilon = (\varepsilon_1', ..., \varepsilon_T')'$, ι_{TK} is a $TK \times 1$ vector of unit elements, and $\bar{\gamma}_w^* = \frac{\iota_{TK}'\Sigma_u^{-1}\gamma}{\iota_{TK}'\Sigma_u^{-1}\iota_{TK}}$ is a weighted mean of the dependent variable of (2). Like Judge et al.'s (1985, p. 32, (2.3.16)) newly defined R^2, R_γ^2 in (4) also lies between 0 and 1 and its value can be interpreted as the proportion of weighted variation in γ explained by regression (2). However, the methods of estimation of (4) and Judge et al.'s newly defined R^2 are not the same.

We will now show how to compute measure (4). Substitute the right-hand side of (2) for γ_{jt} in (1). Doing this for all j gives the fixed coefficient version of (1). The unknown coefficients (π's) and the unobserved errors (ε's) in (2) and the unknown Σ_u in (4) can be estimated for all j and t, using an iteratively rescaled generalized least squares (IRSGLS) method [see Chang et al. (1992, 2000)]. Inserting these estimates into (2) gives the estimates of γ in (2) for all j. These estimates can be used in (4) to evaluate R_γ^2. Here, we require R_γ^2 to be as close to 1 as possible[6] so that the drivers explain a large proportion of variation in the TVC. This result could, of course, be achieved simply by having a very large number of drivers. Therefore, we also require the drivers to be relevant in the sense that the π_{jd}'s are significantly different from zero. Estimation of the full TVC model produces a covariance matrix for the estimated π_{jd}, so that conventional t statistics and probability levels may be produced in the standard way.

These two conditions are closely analogous to the idea of relevance in instrumental variable estimation, where the instruments must be highly correlated with the variables being instrumented. If the R_γ^2 in (4) is low, then we can infer that we have a weak set of coefficient drivers. There is, however, no requirement for the drivers to be independent of the coefficient γ_{jt}, as there is for instruments to be independent of the error term under an IV estimation procedure. Finally, we also do not require the coefficient drivers to be independent of the errors in the model. In fact, we need to pick drivers that are highly correlated with the errors, and hence the misspecification.

Estimating the components of the coefficients of (1) is our next task.

3.2. Splitting the Driver Set

The more difficult issue is how to perform the split in the coefficient drivers into the two sets outlined under Assumption 3—that is, S_1, which contains the variables correlated with the bias-free coefficient, and S_2, which is the set of drivers correlated with the misspecification. The suggestion being made here is that certain drivers should be chosen to explicitly capture any nonlinearities that may exist in the compressed real-world relationship underlying (1). We will discuss later exactly how this should be done. All other driver variables would then be assumed to be associated with misspecification and should therefore be removed when obtaining the bias-free component. The following examples should make this clear. Let us assume that the original real-world relationship underlying (1) is given by

$$y_t^* = f(x_{1t}^*, ..., x_{m_t t}^*). \tag{5}$$

This function meets all relevant preexisting conditions, as some of its arguments are such that their values are automatically held constant when the partial derivatives of y_t^* with respect to the included regressors other than the preexisting conditions are taken.[7] The true functional form of model (5) is unknown and all of its variables are unknown and unobserved. We are interested in estimating

$$\frac{\partial y_t^*}{\partial x_{jt}^*} \text{ for } j = 1, ..., K - 1, \tag{6}$$

where the values of all the arguments of $f(x_{1t}^*, ..., x_{m_t t}^*)$ other than x_{jt}^* are held constant.

To understand how the split of drivers may be accurately done, consider the following.

Example 1

If (5) is linear, then the S_1 set consists of just the constant $z_{0t} \equiv 1$, and all other drivers explain the biases that stem from missing regressors and measurement error.

Example 2

Suppose that (5) is a polynomial, such as a quadratic form. Consider, for simplicity, the case of only two explanatory variables. Then the original real-world relationship becomes[8]

$$y_t^* = \beta_0 + \beta_1 x_{1t}^* + \beta_2 x_{1t}^{*2} + \beta_3 x_{2t}^* + \beta_4 x_{2t}^{*2} \tag{7}$$

We are interested in treating x_{2t}^* as an omitted regressor and also in estimating

$$\frac{\partial y_t^*}{\partial x_{1t}^*} = \beta_1 + 2\beta_2 x_{1t}^*. \tag{8}$$

If the true functional form of equation (6) were unknown to us, we would need to rely on the TVC model $y_t^* = \frac{\partial y_t^*}{\partial x_{1t}^*} x_{1t}^* + \frac{\partial y_t^*}{\partial x_{2t}^*} x_{2t}^* + (y_t^* - \frac{\partial y_t^*}{\partial x_{1t}^*} x_{1t}^* - \frac{\partial y_t^*}{\partial x_{2t}^*} x_{2t}^*) = \alpha_{1t}^* x_{1t}^* + \alpha_{2t}^* x_{2t}^* + \alpha_{0t}^*$, where $\alpha_{1t}^* = \frac{\partial y_t^*}{\partial x_{1t}^*} = \beta_1 + 2\beta_2 x_{1t}^*$, $\alpha_{2t}^* = \frac{\partial y_t^*}{\partial x_{2t}^*} = \beta_3 + 2\beta_4 x_{2t}^*$, and $\alpha_{0t}^* = (y_t^* - \frac{\partial y_t^*}{\partial x_{1t}^*} x_{1t}^* - \frac{\partial y_t^*}{\partial x_{2t}^*} x_{2t}^*)$. Substituting in this TVC model the right-hand side of the second equality sign in the equation $x_{2t}^* = \frac{\partial x_{2t}^*}{\partial x_{1t}^*} x_{1t}^* + (x_{2t}^* - \frac{\partial x_{2t}^*}{\partial x_{1t}^*} x_{1t}^*) = \lambda_{1t}^* x_{1t}^* + \lambda_{0t}^*$ gives $y_t^* = (\alpha_{0t}^* + \alpha_{2t}^* \lambda_{0t}^*) + (\alpha_{1t}^* + \alpha_{2t}^* \lambda_{1t}^*) x_{1t}^*$, where $\alpha_{2t}^* \lambda_{0t}^*$ is the function with the correct functional form of the "sufficient set" λ_{0t}^* of the omitted regressor x_{2t}^*; it is unique and is treated as the error term. The coefficient $(\alpha_{1t}^* + \alpha_{2t}^* \lambda_{1t}^*)$ of x_{1t}^* is unique, being the sum of the partial derivative α_{1t}^* in (8), and the omitted-regressor bias $\alpha_{2t}^* \lambda_{1t}^*$; α_{1t}^* is the bias-free component of the coefficient of x_{1t}^*. $y_t^* = (\alpha_{0t}^* + \alpha_{2t}^* \lambda_{0t}^*) + (\alpha_{1t}^* + \alpha_{2t}^* \lambda_{1t}^*) x_{1t}^*$ is the compressed real-world relationship, which treats x_{2t}^* as an omitted regressor. The reason that we call both this compressed equation and the original equation in (7) real-world relationships is that no approximation is followed in going from the original to the compressed form. Place the measurement errors, $y_t = y_t^* + v_{0t}^*$ and $x_{1t} = x_{1t}^* + v_{1t}^*$, at the appropriate places in the compressed real-world relationship. Doing so gives $y_t = (\alpha_{0t}^* + \alpha_{2t}^* \lambda_{0t}^* + v_{0t}^*) + (\alpha_{1t}^* + \alpha_{2t}^* \lambda_{1t}^*)(1 - \frac{v_{1t}^*}{x_{1t}}) x_{1t}$, where the intercept and the coefficient of x_{1t} are the same as γ_{0t} and γ_{1t}, respectively, in the following. Clearly because of the treatment given to x$_2$, OLS applied to the compressed equation would give biased estimates of the parameter. The TVC model to be estimated is then given by

$$y_t = \gamma_{0t} + \gamma_{1t} x_{1t}. \tag{9}$$

Now, if we include an explicit driver to capture the nonlinearity of (7), we will get the equations

$$\gamma_{0t} = \pi_{00} z_{0t} + \pi_{01} z_{1t} + \pi_{02} z_{2t} + \pi_{03} z_{3t} + \pi_{04} z_{4t} + \pi_{05} z_{5t} + \pi_{06} z_{6t} + \varepsilon_{0t}, \tag{10}$$

$$\gamma_{1t} = \pi_{10} z_{0t} + \pi_{11} z_{1t} + \pi_{12} z_{2t} + \pi_{13} z_{3t} + \pi_{14} z_{4t} + \pi_{15} z_{5t} + \pi_{16} z_{6t} + \varepsilon_{1t}, \tag{11}$$

where we need to choose z_{1t}, z_{2t}, z_{3t}, z_{4t}, z_{5t}, and z_{6t} such that they are highly correlated with α_{0t}^*, $\alpha_{2t}^* \lambda_{0t}^*$, v_{0t}^*, α_{1t}^*, $\alpha_{2t}^* \lambda_{1t}^*$, and $(\alpha_{1t}^* + \alpha_{2t}^* \lambda_{1t}^*)(-\frac{v_{1t}^*}{x_{1t}})$, respectively, and set $\pi_{04} = \pi_{05} = \pi_{06} = 0$ and $\pi_{11} = \pi_{12} = \pi_{13} = 0$. Note that each of the

z's in equations (10) and (11) need not be a single coefficient driver. For example, for $j = 0, 1, h = 1, \ldots, 6$, $\pi_{jh}z_{ht}$ can be equal to $\sum_g \pi_{jhg}z_{htg}$. Now the question is, how do we find such z's when the true functional form of (7) is unknown? Had we known the true functional form of (7) we would have understood that α_{1t}^* is a linear function of x_{1t}^*, α_{2t}^* is a linear function of x_{2t}^*, $\alpha_{2t}^*\lambda_{1t}^*$ is a nonlinear function of x_{2t}^* and $\frac{\partial x_{2t}^*}{\partial x_{1t}^*}$, and $(\alpha_{1t}^* + \alpha_{2t}^*\lambda_{1t}^*)(-\frac{v_{1t}^*}{x_{1t}})$ is a nonlinear function of all the variables: $x_{1t}^*, x_{2t}^*, \frac{\partial x_{2t}^*}{\partial x_{1t}^*}$, and the measurement error in x_{1t}. Suppose that we could somehow guess this and based on this guess we set $z_{4t} = x_{1t}$. Let us substitute the right-hand sides of equations (10) and (11) for γ_{0t} and γ_{1t} in equation (9), respectively. Then

$$y_t = \pi_{00}z_{0t} + \pi_{01}z_{1t} + \pi_{02}z_{2t} + \pi_{03}z_{3t} + (\pi_{10}z_{0t} + \pi_{14}z_{4t} + \pi_{15}z_{5t} + \pi_{16}z_{6t})x_{1t}$$
$$+ \varepsilon_{0t} + \varepsilon_{1t}x_{1t}. \tag{12}$$

In this equation, the sum $\pi_{00}z_{0t} + (\pi_{10}z_{0t} + \pi_{14}z_{4t})x_{1t}$ gives a quadratic function in x_{1t} if $z_{4t} = x_{1t}$. This quadratic function can represent the sum of the first three terms on the right-hand side of (7) well if data on x_{1t} contain negligible magnitudes of measurement error. The coefficients of equation (12) can be estimated consistently by applying an IRSGLS method to (12). With the choice $z_{4t} = x_{1t}$, the estimates of the coefficients of (12) may give better estimates of the components of the coefficients of equation (9) than other choices. If this is true, then an estimate of α_{1t}^* is a kernel density estimate given by $\hat{\pi}_{10} + \hat{\pi}_{14}x_{1t}$, $t = 1, \ldots, T$, where $\hat{\pi}_{10}$ and $\hat{\pi}_{14}$ are the IRSGLS estimates of π_{10} and π_{14}, respectively. If the central tendencies of this kernel density estimates are unreasonable, then the estimates of the other terms of (11) can be added to the estimate $\hat{\pi}_{10} + \hat{\pi}_{14}z_{4t}(= x_{1t})$. The large sample properties of kernel density estimates are given in Lehmann (1999). It is a good idea to study the robustness of these kernel density estimates to changes in the set of coefficient drivers. Because we would not know whether the true model was quadratic, we could include higher-order polynomial terms and test their significance in the usual way to see how many polynomial terms would be needed. If the nonlinearity was not, in fact, a polynomial, then there are two possible courses of action.

1. We could include a number of polynomial terms and think of this as a Taylor series approximation to the true unknown form.
2. We could try a range of specific nonlinear forms, again testing one form against another.

With regard to the second option, by using this option the standard TVC model is able to nest a number of popular nonlinear models within a single framework, which also allows for measurement error and missing variables. This procedure is very different from other standard procedures. For example, a popular non-linear model is the smooth transition autoregressive model (STAR). This allows a parameter to move smoothly between two values according to a function that responds to some threshold variable. If we were estimating a TVC such as (9) but

believed the true nonlinearity followed a STAR form, then we could specify the driver equations as

$$\beta_{0t} = \pi_{00} + \sum_{i=1}^{p} \pi_{01+i} z_{it} + \varepsilon_{0t}, \tag{13}$$

$$\beta_{1t} = \pi_{10} + \pi_{11} G(z_t, \zeta, c) + \pi_{12}(1 - G(z_t, \zeta, c)) + \sum_{i=1}^{p} \pi_{12+i} z_{it} + \varepsilon_{1t}, \tag{14}$$

where $G(z_t, \zeta, c)$ is the transition function—typically a logistic function for the LSTAR model, an exponential function for the ESTAR model, or a second-order logistic function, z is the transition variable, and ζ and c are parameters [Ahmad and Lo (2014)]. The model given by (14) is more general than the standard STAR model, as it includes the drivers associated with measurement error and missing variables and will hence correct for these misspecifications. Model (14) also has a stochastic error term and thus becomes a stochastic STAR model. The split into the two subsets is again obvious, as the two terms capturing the STAR effect are clearly the set that is appropriate for S_1.

Another interesting nonlinearity would be to use a set of combinations of simple nonlinear functions such as the log or exponential function. In this case the TVC model would begin to encompass a neural net and the universal approximation theorems of Hal White would suggest that with sufficient complexity the model could then approximate any unknown functional form to any degree of accuracy.

Under certain conditions, the IRSGLS estimators of the π's in (2) are consistent and asymptotically efficient [see Swamy et al. (2010)].[9] The distributional theory underlying this estimation technique and the method for constructing inference are given in Swamy et al. (2010). However, this software is not used widely, and so here we point out a way in which these models may be estimated using standard software such as EVIEWS. The preceding models are written in a form that exactly corresponds to the state space form. Although the usual interpretation of state space models is rather different from the TVC procedure being discussed here, the mathematics of the Kalman filter goes through exactly to yield minimum least-squares estimates [Harvey (1989)]. If the errors of the driver equations are not normal with a constant variance, then the Kalman filter will not yield maximum likelihood estimates; they may, however, be interpreted as quasi-maximum likelihood estimates [White (1980)], and they will be consistent, although not generally as efficient as the IRSGLS technique mentioned earlier. (This means that all of these models may be estimated using the Kalman filter, which can estimate all of the parameters of the model by maximizing the quasi-likelihood function. The Kalman filter and the state space form must be linear in the state variables, but they can easily handle nonlinearities in the other variables; all of the preceding variants may be estimated in standard software such as EVIEWS. An Appendix to this paper provides the EVIEWS code that was used to estimate the example following.

4. AN APPLICATION

In this section, we investigate the effects of rating-agency decisions on the sovereign bond spread between Greece and Germany. The underlying hypothesis is that this relationship is highly nonlinear—for example, a decline in ratings by one notch, from, say, AA to A, will have a relatively small effect on spreads, whereas a decline from, say, B-minus to CCC will have a much larger effect on spreads. Thus, as the rating goes down, the effect on spreads becomes proportionally larger. We will be undertaking quasi-maximum likelihood estimation to illustrate the application of these techniques in standard software (EVIEWS).

The intuition underlying this mechanism is as follows. Consider a world that includes two rating agencies, X and Y. In assigning ratings to a particular sovereign, both agencies have access to essentially identical information sets composed of the (present and projected) fundamentals, including spreads, competitiveness, real growth, inflation, fiscal and external positions, and, perhaps, noneconomic variables such as measures of political stability. Suppose that, based on its assessment of the information set of a particular country, rating agency X moves to downgrade the sovereign debt of the country in question. The announcement of the downgrade will very likely trigger a rise in the sovereign's interest rate. In addition, under the ECB's collateral framework, haircuts on sovereigns rise if ratings fall to a specified (triple-B) level and are ineligible as collateral below single-B minus. For these reasons, the very action by rating agency X changes the information set of rating agency Y, because that information set now includes X's downgrade, the resulting higher interest rates, and possibly higher haircuts on collateral, lower projected growth (because of the rise in interest rates), and less sustainable fiscal balances for the country in question. Consequently, rating agency Y, which may have been content with the rating it had assigned to the sovereign in question prior to X's downgrade, may move to downgrade the sovereign's rating based on the changed information set. In this way, X's original action can precipitate a downgrade by Y, triggering self-perpetuating feedback loops between ratings and spreads.

Of course, there are many other things that might affect spreads, such as debt, deficits, relative prices, and politics. Therefore, if we examined a simple relationship between spreads and ratings, omitted variables would cause bias for a standard OLS regression.

The data used are monthly and cover the period 1998m1 to 2012m6. In cases for which the original data are quarterly, the data have been interpolated to a monthly frequency; where appropriate, variables are measured relative to the corresponding variables for Germany. As mentioned, the dependent variable, sp, is the yield spread between the 10-year benchmark government bond yield of Greece and that of Germany. Our explanatory variables are measures of macroeconomic and political fundamentals, as follows:

> pol represents political stability. We use the IFO World Economic Survey Index of Political Stability. A rise in the index implies greater stability.

dgdp is real GDP growth. A relatively high rate of economic growth suggests that a country's existing debt burden will become easier to service over time.

cnewssq is "fiscal news." To capture the news (or surprise) element that has figured strongly in Greece's experience, we use real-time fiscal data. In particular, using the European Commission Spring and Autumn forecasts, we use a series of forecast revisions. For example, the revision in the Spring 2001 forecasts is the 2001 deficit/GDP ratio in the spring compared to the forecast for 2001 made in the autumn of 2000. This procedure generates a series of revisions, which, when cumulated over time, provides a cumulative fiscal news variable.

relp is relative prices. To help capture relative changes in competitiveness, we use Greece's Harmonized Index of Consumer Prices (HICP, all items index) relative to that of Germany.

debtogdp is the debt-to-GDP ratio. A higher debt burden should correspond to a higher risk of default. We include the general government consolidated gross debt-to-GDP ratio (expressed as a percentage), interpolated from a quarterly to a monthly frequency.

rate is the agencies' credit rating for Greek government debt. The ratings of three agencies are used: Fitch, Moody's, and Standard & Poor's. To capture the effects of ratings on spreads, we include ordinal ratings to allow for nonlinearities in the relationship between ratings and spreads. For example, the dummy variable triple-A takes a value of 1 for the period for which the rating was triple-A, and a value of zero otherwise. We date rating changes by identifying the agency that made the first move from one rating to another, on the assumption that the first mover would cause the subsequent reaction. In other words, if rating agency A downgraded Greece from A- to BBB+ in April, say, and subsequently rating agency B downgraded Greece from A to A- in June, then the second downgrade would not register. Our basic TVC model is

$$\mathrm{sp}_t = \alpha_{0t} + \alpha_{1t}\mathrm{rate}. \tag{15}$$

Our coefficient driver equations take the form

$$\alpha_{0t} = \pi_{00} + \pi_{01}\mathrm{pol} + \pi_{02}\mathrm{dgdp} + \pi_{03}\mathrm{cnewssq} + \pi_{04}\mathrm{relp} + \pi_{05}\mathrm{debtogdp} + \varepsilon_{0t}, \tag{16}$$

$$\alpha_{1t} = \pi_{10} + \pi_{11}\mathrm{rate} + \pi_{12}\mathrm{pol} + \pi_{13}\mathrm{dgdp} + \pi_{14}\mathrm{cnewssq} + \pi_{15}\mathrm{relp} + \pi_{16}\mathrm{debtogdp} + \varepsilon_{1t}. \tag{17}$$

In this driver set, rate gives a quadratic effect to the coefficient on ratings, allowing for a strong nonlinearity. We begin by estimating this general model, to obtain the following results (where *t*-stats are in parentheses):

$$\alpha_{0t} = -4.2 - 0.05\mathrm{pol} - 0.9\mathrm{dgdp} + 0.03\mathrm{cnewssq} + 23.27\mathrm{relp} + 0.05\mathrm{debtogdp} + \varepsilon_{1t}$$
$$(0.6)\,(0.4)\qquad(13.4)\quad(2.9)\qquad\qquad(3.2)\qquad\quad(1.6), \tag{16'}$$

$$\alpha_{1t} = -0.6 + 0.07\mathrm{rate} + 0.004\mathrm{pol} + 0.27\mathrm{dgdp} - 0.005\mathrm{cnewssq} - 3.5\mathrm{relp} + 0.0004\mathrm{debtogdp} + \varepsilon_{2t}$$
$$(1.2)\,(17.4)\qquad(0.1)\qquad(0.18)\qquad(3.2)\qquad\qquad(3.35)\qquad(0.12). \tag{17'}$$

As mentioned in Section 3, selection of drivers should be made on the basis of the explanatory power of the driver equations and the significance of the individual
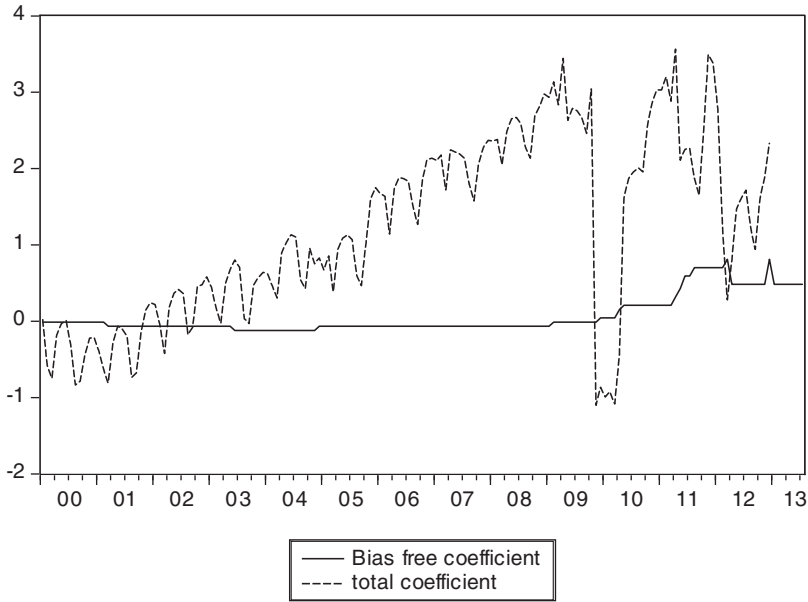
FIGURE 1. Time profiles of α_{1t} in equation (15) and its bias-free component.

drivers. For equation (16′), the R^2 is 0.80; for equation (17′), the R^2 is 0.84. Thus, both sets of drivers produce a reasonably high degree of explanatory power for the two TVCs. A number of drivers in each equation are insignificant, however, so we simplify the model to exclude insignificant drivers, obtaining the following more parsimonious model (note we do not exclude the constants even if insignificant):

$$\alpha_{0t}= -3.5 + 0.03\text{cnewssq} + 15.6\text{relp} + 0.04\text{debtogdp}+\varepsilon_{1t} \tag{18}$$
$$(2.0)(2.9) \qquad\qquad (3.9) \qquad (2.7).$$

$$\alpha_{1t}= -0.64 + 0.05\text{rate} - 0.005\text{cnewssq} - 2.3\text{relp} + \varepsilon_{2t} \tag{19}$$
$$(1.8) \quad (20.9) \qquad (3.8) \qquad\qquad (4.2)$$

The coefficient of interest is α_{1t}, that is, the effect of ratings on the two-year spread. Effectively, we have taken the variable "rate" from the basic equation for spreads—i.e., equation (15)—and used it as the driver. Then, by substituting back into the basic equation, we obtain a quadratic effect in the basic equation. Figure 1 shows the total value of this TVC, along with the bias-free effect that is given by subtracting the error term and the effect from rate and cnewssq.

Because the scaling of the two coefficients is quite different, Figure 2 shows just the bias-free coefficient. This makes the strong nonlinearity clear as ratings start to rise (that is, deteriorate). After 2008, the effect of ratings on spreads becomes increasingly powerful. We therefore have found a very strong quadratic link between ratings and spreads.
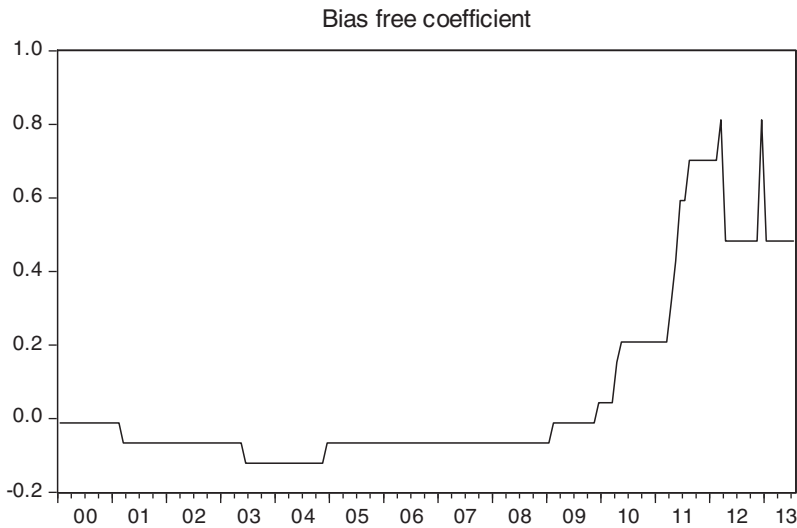
Bias free coefficient

FIGURE 2. Time profile of the bias-free component of α_{1t}.

5. CONCLUSIONS

This paper has proposed a new way of deriving the split between coefficient drivers when deriving the bias-free estimate of a coefficient within the TVC estimation framework. We have argued that, if a model with unique coefficients and error term is linear, then only the constant in the coefficient driver set should be retained. If a model with unique coefficients and error term is nonlinear, however, then an explicit set of drivers that capture the nonlinearity should be chosen. In the absence of any specific information about the precise form of the nonlinearity, this can best be achieved by using a set of polynomials in the explanatory variables. These drivers are then the only ones that should be retained when the bias-free component is derived. We have also argued that two conditions should be applied to the drivers, which we call predictive power and relevance; that is, the drivers should explain a large proportion of the movement in the TVCs and they should be statistically significant. We illustrated this process by estimating a nonlinear relationship between country-risk ratings and sovereign bond spreads for Greece. We showed that there is a highly nonlinear effect here. Finally, the procedure can be implemented using standard software such as EVIEWS. The code for the estimated model is provided in Appendix A.

NOTES

1. The development of a type of variable coefficient estimation is due to Swamy (1970, 1974). See also Swamy and Tavlas (2001, 2007) and Swamy et al. (2010).

2. For a definition of real-world relationships, see Basmann (1988). See also Swamy et al. (2014, p. 196).

3. Mathematically, this model may appear to be a state space one. However, the interpretation of the coefficients is quite different from the standard state space representation. Omitted-variable biases, measurement-error biases, and the correct functions of certain "sufficient sets of excluded variables" are not considered parts of the coefficients of the observation equations of state-space models. This is the major difference between (1) and the observation equations of a standard state-space model.

4. The distributional assumptions about the errors in (2) are given in Swamy et al. (2010).

5. To handle this issue, we use Lehmann and Casella's (1998, pp. 24 and 57) concept of identification.

6. Note that, as (5) is computed using the estimates of its unknown quantities, this cannot be given a formal test interpretation. Instead, we use R_γ^2 as a simple descriptive statistic.

7. Spurious relationships disappear when we control for all relevant preexisting conditions [see Skyrms (1988, p. 59)].

8. The real-world relationships with more than two explanatory variables are considered in our previous work [see Chang, Swamy, Hallahan, and Tavlas (2000)].

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

REFERENCES

Ahmad, Yamin and Ming Chien Lo (2014) Nonlinear time series models and model selection. In J. Ma and M. Wohar (eds.), *Recent Advances in Estimating Nonlinear Models with Applications in Economics and Finance*, pp. 99–121. New York: Springer.

Basmann, Robert L. (1988) Causality tests and observationally equivalent representations of econometric models. *Journal of Econometrics* 39, 69–104.

Chang, I-Lok, Charles Hallahan, and P.A.V.B. Swamy (1992) Efficient computation of stochastic coefficient models. In H.M. Amman, D.A. Belsley, and L.F. Pau (eds.), *Computational Economics and Econometrics*, pp. 43–53. Boston: Kluwer Academic Publishers.

Chang, I-Lok, P.A.V.B. Swamy, Charles Hallahan, and George S. Tavlas (2000) A computational approach to finding causal economic laws. *Computational Economics* 16, 105–136.

Granger, Clive W.J. (2008) Non-linear models: Where do we go next—Time varying parameter models. *Studies in Nonlinear Dynamics and Econometrics* 12, 1–10.

Hall, Stephen G., George Hondroyiannis, P.A.V.B. Swamy, and George S. Tavlas (2008) A Portfolio Balance Approach to Euro-area Money Demand in a Time-Varying Environment. Paper presented at Bank of England/MMF workshop on money and macromodels.

Hall, Stephen G., George Hondroyiannis, P.A.V.B. Swamy, and George S. Tavlas (2009a) The New Keynesian Phillips curve and lagged inflation: A case of spurious correlation? *Southern Economic Journal* 76, 467–481.

Hall, Stephen G., George Hondroyiannis, P.A.V.B. Swamy, and George S. Tavlas (2009b) Where has all the money gone? Wealth and the demand for money in South Africa. *Journal of African Economies* 18, 84–112.

Hall, Stephen G., Amangeldi Kenjegaliev, P.A.V.B. Swamy, and George S. Tavlas (2013) The forward rate premium puzzle: A case of misspecification? *Studies in Nonlinear Dynamics and Econometrics* 17, 265–279.

Hall, Stephen G., P.A.V.B. Swamy, and George S. Tavlas (2012) Generalized cointegration: A new concept with an application to health expenditure and health outcomes. *Empirical Economics* 42, 303–318.

Hall, Stephen G., P.A.V.B. Swamy, and George S. Tavlas (2015) A note on generalizing the concept of cointegration. *Macroeconomic Dynamics* 19(7), 1633–1646.

Harvey, Andrew C. (1989) *Forecasting Structural Time Series Models and the Kalman Filter*. Cambridge, UK: Cambridge University Press.

Judge, George G., William E. Griffiths, R. Carter Hill, Helmut Lütkepohl, and Tsoung-Chao Lee (1985) *The Theory and Practice of Econometrics*, 2nd ed. New York: Wiley.

Lehmann, Erich L. (1999) *Elements of Large-Sample Theory*. Berlin: Springer.

Lehmann, Erich L. and George Casella (1998) *Theory of Point Estimation*, 2nd ed. Berlin: Springer.

Skyrms, Brian (1988) Probability and causation. *Journal of Econometrics* 39, 53–68.

Swamy, P.A.V.B. (1970) Efficient inference in a random coefficient regression model. *Econometrica* 38, 311–323.

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, Stephen G. Hall, and George S. Tavlas (2015) Microproduction functions with unique coefficients and errors. *Macroeconomic Dynamics* 19, 311–333.

Swamy, P.A.V.B. and Jatinder 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., Jatinder S. Mehta, George S. Tavlas, and Stephen G. Hall (2014) Small area estimation with correctly specified linking models. In Jun Ma and Mark Wohar (eds.), *Recent Advances in Estimating Nonlinear Models with Applications in Economics and Finance*, pp. 193–228. New York: Springer.

Swamy, P.A.V.B. and George S. Tavlas (2001) Random coefficient models. In Badi H. Baltagi (ed.), *Companion to Theoretical Econometrics*, pp. 410–428. Malden, MA: Blackwell.

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

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

White, Halbert (1980) A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 48, 817–838.

APPENDIX A: EVIEWS CODE FOR THE REPORTED MODEL IN SECTION 4

```
@signal sp_gr = sv1 + sv2*rate_gr
@state sv1 = c(1) +c(6)*pol_gr+c(7)*dgdp_gr+c(8)*cnewssq_gr+c(9)*relp_gr
    +c(10)*debtogdp_gr+sv3(-1)+c(16)*sv4(-1)
@state sv2 = c(2)+c(3)*rate_gr+c(11)*pol_gr+c(12)*dgdp_gr+c(13)*cnewssq_gr
    +c(14)*relp_gr+c(15)*debtogdp_gr +sv5(-1)+c(17)*sv6(-1)


@state sv3 = [var = exp(c(18))]
@state sv4 = sv3(-1)
@state sv5 = [var = exp(c(19))]
@state sv6 = sv5(-1)
```

In this code, SP_gr is the spread between Greek and German ten-year government bond yields; Rate_gr is the rating on country risk, denoted as a variable scaled from 1 to 20, where 1 is equivalent to a AAA rating; pol_gr is an indicator of political unrest in Greece; cnewssq is an indicator of news about Greek GDP derived from ECB forecasts; relp_gr is the relative process between Greece and Germany; and debtogdp_gr is the debt-to-GDP ratio for Greece.

This model is also coded to have a first-order moving average error process on the two time-varying coefficients sv1 and sv2, which is usual in TVC models.

The c(?) parameters are the usual EVIEWS notation for parameters to be estimated by maximum likelihood.