University of Leicester

DEPARTMENT OF ECONOMICS

ESTIMATION OF PARAMETERS IN THE PRESENCE OF MODEL MISSPECIFICATION AND MEASUREMENT ERROR

P.A.V.B. Swamy, U.S. Bureau of Labor Statistics

George S. Tavlas, Bank of Greece

Stephen G. Hall, University of Leicester, UK and Bank of Greece

George Hondroyiannis, Bank of Greece and Harokopio University

Working Paper No. 08/27

August 2008

Estimation of Parameters in the Presence of Model misspecification and Measurement Error

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

And

George Hondroyiannis

Abstract Misspecifications of econometric models can lead to biased coefficients and error terms, which in turn can lead to incorrect inference and incorrect models. There are specific techniques such as instrumental variables which attempt to deal with some individual forms of model misspecification. However these can typically only address one problem at a time. This paper proposes a general method for estimating underlying parameters in the presence of a range of unknown model misspecifications. It is argued that this method can consistently estimate the direct effect of an independent variable on a dependent variable with all of its other determinants held constant even in the presence of a misspecified functional form, measurement error and omitted variables.

.

Keywords Misspecified model · Correct interpretation of coefficients · Appropriate assumption · 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.

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

G. S. Tavlas (Corresponding author)
Economic Research Department,
Bank of Greece, 21 El. Venizelos
Ave. 102 50. Athens, Greece
E-Mail: GTavlas@bankofgreece.gr

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

George Hondroyiannis
Bank of Greece and Harokopio
University, 21 E. Venizelos Ave.
102 50 Athens, E-Mail:
ghondroyiannis@bankofgreece.gr

1 Introduction

Most econometric relationships are subject to specification errors arising from the following three problems: (i) the true functional forms of economic relationships are usually unknown, (ii) econometric models cannot be specified without omitting some relevant explanatory variables, and (iii) data on economic variables contain measurement errors. Consequently, misspecification of models is difficult to avoid. There are specific techniques which attempt to deal with these problems, usually one at a time. Instrumental variables are an obvious example of a technique designed to deal with measurement error. But this technique cannot deal with a misspecified functional form or omitted variables. Similarly the non-parametric estimators such as neural networks or nearest neighbor estimation are designed to deal with an unknown functional form. These techniques can not however cope with measurement error and they also typically require very large data sets. This paper sets out a new approach to estimation which can deal with all three of these problems at the same time and which is practical in relatively small samples.

More specifically, this paper shows how misinterpretations of model coefficients and error terms in the presence of model misspecifications can be avoided using a coefficient-decomposition approach. As we discuss, a key aspect of this approach involves the use of what we term "coefficient drivers". Intuitively, coefficient drivers may be thought of as variables which, though not part of the explanatory variables in a relationship, serve two important purposes. First, they deal with the correlations between the included explanatory variables and *their coefficients*.[1] In other words, even though it can be shown that the included explanatory variables are not unconditionally independent of their

[1] A formal definition of coefficient drivers is provided in Swamy and Tavlas (2006).

coefficients, they can be conditionally independent of their coefficients given the coefficient drivers. Second, the coefficient drivers allow us to decompose the coefficient on a regressor into three components such that one of these components representing the specification-bias-free component is separately identified from the other two components representing specification biases. In one sense the coefficient drivers may be seen as a dual (and a generalization) of instrumental variables. A good instrument is correlated with the explanatory variable measured with error while being uncorrelated with the model's error term. A good coefficient driver is correlated with the parts of coefficients which arise from the econometric misspecification and therefore provides information that allows us to correct the biases which arise in the coefficients. An issue that arises, however, is the following: How do we select an appropriate set of coefficient drivers? This is much like the problem of how we select an appropriate set of instruments and we will discuss this specifically below.

The remainder of this paper is divided into three sections. Section 2 presents the new way of interpreting the coefficients of misspecified econometric models and the assumptions that are consistent with those interpretations. Such assumptions may require the specification of a time-varying coefficient (TVC) model without the implication that these TVC's are always true. The identifiability conditions for TVC models and the methods of consistently estimating their unknown quantities are presented in Section 3. Unresolved problems will be faced if the TVC's are assumed to follow a random walk process, as shown in Section 3. This section also provides a Bayesian method of estimating TVC models satisfying the identifiability conditions. Section 4 concludes.

2 The Interpretations of Model Coefficients and Appropriate Assumptions

Conventional econometrics is to a large extent the study of individual causes of biased parameters, omitted variable bias, measurement error bias, a misspecified functional form etc. These problems are usually dealt with one at a time in a text book context, but of course practical work is plagued by all these problems at once. In this section we outline the basic problem of interpreting coefficients when these problems are present and our proposed procedure for dealing with these problems simultaneously.

When studying the relation of a dependent variable, denoted by y_t^*, 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^* a number of problems may arise. If there is a correlation between y_t^* and x_{1t}^*, …, $x_{K-1,t}^*$ and a third set of variables, a phenomenon known as spurious correlation may arise (see Lehmann and Casella 1998, p. 107). As a first step in avoiding spurious correlations, economic theories may suggest mechanisms through which x_{1t}^*, …, $x_{K-1,t}^*$ could influence y_t^*. Unfortunately, economic theories may suggest direct relationship but it often has very little to say about the true functional form of this relationship. Any specific functional form may be incorrect and may lead to specification errors resulting from functional-form biases. Another problem that can arise in investigating the relationship between the dependent variable and its determinants is that x_{1t}^*, …, $x_{K-1,t}^*$ may not exhaust the complete list of the determinants 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. 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 errors-in-variables problems. The purpose of this paper is to propose the correct interpretations and the appropriate method of estimation of the coefficients of the relationship between y_t^* and x_{1t}^*, …, $x_{K-1,t}^*$ in the presence of the foregoing problems.

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: $y_t = y_t^* + \mathrm{v}_{0t}$, $x_{jt} = x_{jt}^* + \mathrm{v}_{jt}$, $j = 1, \ldots, K\text{-}1$, $t = 1, \ldots, T$, 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 functional form we are estimating may be misspecified and there may be some important variables missing from x_{1t}, …, $x_{K-1,t}$ we need a model which 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 unbiased estimates of the effect on a dependent variable of changing one independent variable holding all others constant. That is to say we aim to find an unbiased estimate of the partial derivative of y_t^* with respect to any x_{jt}^*. This of course is the interpretation which is usually placed on the coefficients of a standard econometric model but this interpretation depends crucially on the assumption that the conventional model gives unbiased coefficients which is of course not the case in the presence of model misspecification.

One way to proceed is to specify a set of time-varying coefficients which provide a complete explanation of the dependent variable y. Consider the relationship

$$y_t = \gamma_{0t} + \gamma_{1t} x_{1t} + \cdots + \gamma_{K-1,t} x_{K-1,t} . \tag{1}$$

which we call "the time-varying coefficient (TVC) model". As this model provides a complete explanation of y, all the misspecification 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 non-linear the time-varying coefficients may be thought of as the partial derivatives of the true non-linear structure and so they are able to capture any possible function. These coefficients will also capture the effects of measurement error and omitted variables. The trick then is to find a way of decomposing these coefficients into the biased and the bias-free components. Equation (1) is called the observation equation and its coefficients are called the state variables if it is embedded in a state-space model (see Durbin and Koopman 2001, p. 38).

For empirical implementation, model (1) has to be embedded in a stochastic framework. To do so, we need to answer the question: What are the correct stochastic assumptions about the TVC's of (1)? We believe that the correct answer is: 'the correct interpretation of the TVC's and the assumptions about them must be based on an understanding of the model misspecification which comes from any (i) omitted variables, (ii) measurement errors, and (iii) misspecification of the functional form'. We expand on this argument in what follows.

Notation and Assumptions Let m_t denote the total number of the determinants of y_t^*. The exact value of m_t cannot be known 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 determinants of y_t^* that are excluded from equation (1). Let x_{gt}^*, $g = K, \ldots, m_t$, denote these excluded determinants. Let α_{0t}^* denote the intercept and let both α_{jt}^*, $j = 1, \ldots, K-1$, and α_{gt}^*, $g = K, \ldots, m_t$, 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. These time profiles are unknown, since the true functional form is unknown. Note that an equation that is linear in variables accurately represents a non-linear equation, provided the coefficients of the former equation are time-varying with time profiles determined by the true functional form of the latter equation. This type of representation of a non-linear equation is convenient, particularly when the true functional form of the non-linear equation is unknown. Such a representation is not subject to the criticism of misspecified functional form. For $g = K, \ldots, m_t$, let λ_{0gt}^* denote the intercept and let λ_{jgt}^*, $j = 1, \ldots, K-1$, denote the other coefficients of the regression of x_{gt}^* on $x_{1t}^*, \ldots, x_{K-1,t}^*$. The true functional forms of these regressions determine the time profiles of λ^*s.

The following theorem gives the correct interpretations of the coefficients of equation (1):

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

$$\gamma_{0t} = \alpha_{0t}^* + \sum_{g=K}^{m_t} \alpha_{gt}^* \lambda_{0gt}^* + \mathrm{v}_{0t}, \tag{2}$$

and the coefficients of (1) other than the intercept satisfy the equations,

$$\gamma_{jt} = \alpha_{jt}^* + \sum_{g=K}^{m_t} \alpha_{gt}^* \lambda_{jgt}^* - \left(\alpha_{jt}^* + \sum_{g=K}^{m_t} \alpha_{gt}^* \lambda_{jgt}^* \right) \left(\frac{\mathrm{v}_{jt}}{\mathrm{x}_{jt}} \right) \quad (j = 1, \ldots, K\text{-}1). \tag{3}$$

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

Thus, we may interpret the TVC's in terms of the underlying correct coefficients, the observed explanatory variables and their measurement errors. It should be noted that by assuming that the λ^*s in equations (2) and (3) are possibly nonzero we do not require that the determinants of y_t^* included in (1) be independent of the determinants of y_t^* excluded from (1). Pratt and Schlaifer (1988, p. 34) show that this condition is "meaningless". By the same logic, the usual exogeneity assumption of independence between a regressor and the disturbances of an econometric model is "meaningless" if the disturbances are assumed to represent the net effect on the dependent variable of the determinants of the dependent variable excluded from the model. The real culprit appears to be the interpretation that the disturbances of an econometric model represent the net effect on the dependent variable of the unidentified determinants of the dependent variable excluded from the model. Elsewhere, Pratt and Schlaifer (1984, p. 14) point out that although the regressors of (1) cannot be uncorrelated with every *determinant* of y_t^* excluded from (1), they can be uncorrelated with certain *remainder* of every such determinant. The intercept λ_{0gt}^* in (2) is indeed such a *remainder* of x_{gt}^* after the effects of x_{1t}^*, …, $x_{K-1,t}^*$ on x_{gt}^* have been subtracted out.

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 (1) on the determinants of y_t^*

included in (1) are non-linear. Note that the last term on the right-hand side of equations in (3) implies that the regressors of (1) are correlated with their own coefficients.[2]

Theorem 2 *For* $j = 1, \ldots, K\text{-}1$, *the component* α_{jt}^{*} *of* γ_{jt} *in (3) is the direct or bias-free effect of* x_{jt}^{*} *on* y_{t}^{*} *with all the other determinants of* y_{t}^{*} *held constant and is unique.*

Proof It can be seen from equation (3) 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 (= $-\alpha_{jt}^{*} + \sum_{g=K}^{m_t} \alpha_{gt}^{*}\lambda_{jgt}^{*} \times$ v_{jt} / x_{jt}), 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 (1) and the v_{jt}. However, the γ_{jt} are unique when their correct interpretations given by (2) and (3) are adopted (see Swamy and Tavlas 2007, p. 300). Note that α_{jt}^{*} is the coefficient of x_{jt}^{*} in the correctly specified relation of y_{t}^{*} to all of its determinants. Hence α_{jt}^{*} represents the direct, or bias-free, effect of x_{jt}^{*} on y_{t}^{*} with all the other determinants of y_{t}^{*} held constant. The direct effect is 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). □

The direct effect α_{jt}^{*} is constant if the relationship between y_{t}^{*} and all of its determinants are linear; alternatively, it is variable if the relationship is non-linear. We

---

[2] These correlations are typically ignored in the analyses of state-space models. Thus, inexpressive conditions and restrictive functional forms are avoided in arriving at equations (2) and (3) so that Theorem 1 can easily hold; for further discussion and interpretation of the terms in (2) and (3), see Swamy and Tavlas (2001, 2007) and Hondroyiannis, Swamy and Tavlas (2007).

often have information from theory as to the right sign of α^*_{jt}. Any observed correlation between y_t and x_{jt} is spurious if α^*_{jt} = 0 (see Swamy, Tavlas and Mehta 2007).[3]

So to put the previous formal arguments into words; if the true model has some unknown non-linear functional form this model may be represented as a linear model with time-varying coefficients. If we then add omitted-variable and measurement-error effects we have shown that the time-varying coefficients in the estimated model are a function of the time-varying coefficients from the true model plus components reflecting the omitted-variables and the measurement-error effects. This argument is a matter of pure logic and must always be true and so it gives us an unambiguous way of thinking about and interpreting coefficients. The next issue is how to make some identifying assumptions which will allow us to separate these components.

As noted above we believe that empirical researchers are interested in the direct effects α^*s, not in the omitted-variable and measurement-error biases. That is, they are not interested in the γ_{jt} which are contaminated by omitted-variable and measurement-error biases. To obtain accurate estimates of the α^*_{jt} using the observations in (1), we need to first decompose each γ_{jt} with j > 0 into its components in (3). Our method of identifying these components and performing the decomposition is based on the following assumptions that are consistent with the correct interpretations of γs:

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

[3] Granger and Newbold's (1974) definition of spurious regressions does not apply to non-linear equations and equation (1) can be non-linear.

$$\gamma_{jt} = \pi_{j0} + \sum_{d=1}^{p-1} \pi_{jd} z_{dt} + \varepsilon_{jt} \quad (j = 0, 1, \ldots, K\text{-}1), \tag{4}$$

where the πs *are fixed parameters, the* z_{dt} *are what are called the coefficient drivers, and different coefficients of (1) can be functions of different sets of coefficient drivers.*

Assumption 2 *For* $j = 1, \ldots, K$-1, *the set of* p-1 *coefficient drivers and the constant term in (4) divides into two disjoint subsets* S_1 *and* S_2 *so that* $\pi_{j0} + \sum_{d \in S_1} \pi_{jd} z_{dt}$ *has the same pattern of time variation as* α_{jt}^* *and* $\sum_{d \in S_2} \pi_{jd} z_{dt} + \varepsilon_{jt}$ *has the same pattern of time variation as the sum of the last two terms on the right-hand side of equation (3) over the relevant estimation and forecasting periods.*

This assumption is like the dual of the instrumental variable assumption. Here we are assuming that the drivers in the set S_2 are correlated with the misspecification in the model

Assumption 3 *The K-vector* $\varepsilon_t = (\varepsilon_{0t}, \varepsilon_{1t}, \ldots, \varepsilon_{K-1,t})'$ *of errors in (4) follows the stochastic equation,*

$$\varepsilon_t = \Phi \varepsilon_{t-1} + u_t, \tag{5}$$

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

$$E(u_t u_{t'}' \mid z_{1t}, \ldots, z_{p-1,t}) = \begin{matrix} \sigma_u^2 \Delta_u & \text{if} & t = t' \\ 0 & \text{if} & t \neq t' \end{matrix}, \tag{6}$$

where Δ_u *may not be diagonal.*

This assumption considerably generalizes (4). If we assumed that the errors in (4) were 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.

Assumption 4 *The regressor* x_{jt} *of (1) is conditionally independent of its coefficient* γ_{jt} *given the coefficient drivers in (4) for all j and t.*

Assumptions 1 and 2 answer the question of parameterization: which features of equation (1) ought to be treated as constant parameters?[4] Correct coefficient drivers are those that satisfy Assumptions 2 and 4.[5] Assumption 4 is weaker than the assumption that x_{jt} is unconditionally independent of γ_{jt}, which is false because the third term on the right-hand side of equation (3) introduces a non-zero correlation between x_{jt} and γ_{jt}.

A vector formulation of model (1) is

$$y_t = x_t'\gamma_t, \tag{7}$$

where $x_t = (1, x_{1t}, ..., x_{K-1,t})'$ and $\gamma_t = (\gamma_{0t}, \gamma_{1t}, ..., \gamma_{K-1,t})'$. A matrix formulation of (4) is

[4] Assumption 1 was considered in Swamy and Tinsley (1980) who adopted a more general Assumption 3.
[5] Extensions of model (1) and Assumptions 1-4 to situations where panel data, i.e., multiple observations on each of many observational units, are available are made in Swamy, Yaghi, Mehta and Chang (2007) and Swamy, Mehta, Chang and Zimmerman (2008). Greene (2008, p. 184) points out that in such situations, relating the means of the random coefficients of a model to a set of observable, person specific variables makes the model extremely versatile. Whether this is true or not, we do need correct coefficient drivers to estimate the direct or bias-free effects α_{jt}^*.

$$\gamma_t = \Pi z_t + \varepsilon_t , \tag{8}$$

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 (8) into (7) gives

$$y_t = (z_t' \otimes x_t')\pi^{Long} + x_t'\varepsilon_t , \tag{9}$$

where $\otimes$ denotes a Kronecker product, and π^{Long} is a *Kp*-vector, denoting a column stack of Π. The observations in (1) can be displayed in a matrix form as

$$y = X_z \pi^{Long} + D_x \varepsilon , \tag{10}$$

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). □

If correct coefficient drivers are not included in equation (4), then Assumption 4 and Theorem 3 do not hold, since the regressors of (1) are correlated with their coefficients γ_{jt}. 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. The magnitude of ε_{jt} gets reduced as the number of correct coefficient drivers in (4) increases. The larger the number of correct coefficient drivers in (4), the smaller the magnitude of ε_{jt} and hence the smaller the variance of γ_{jt}. Including many correct coefficient drivers in (4) may imply that the errors of equation (4) are white-noise variables or the matrix Φ in equation (5) is null. If the error term of (4) is assumed to follow a random walk, then the unconditional variance of γ_{jt} is not finite. Some workers regard the assumption of an infinite variance as unnatural since all observed time series have finite values, as Durbin and Koopman (2001, p. 29) point out.

The fixed-parameter model in (10) performs well in explanation if Assumptions 1-4 are true because under these assumptions, this model gives an accurate estimate of the direct effect of each of x^*_{jt}, j = 1, …, K-1, on y^*_t with all the other determinants of y^*_t held constant. Without the coefficient drivers that satisfy Assumptions 2 and 4, the explanatory power of model (1) is zero because the observations in (1) intermix α^*_{jt} and the other components of γ_{jt} with no prospect of separation. Model (10) has the correct functional form and hence performs well in prediction if the coefficient drivers included in (4) satisfy Assumptions 1-4 (see Swamy, Yaghi, Mehta and Chang 2007).[6] The stochastic coefficient approach based on the TVC model in (1) and Assumptions 1-4 leads to the improved fixed-coefficient model in (10) without misinterpreting the coefficients of the TVC model, without using meaningless conditions, and without misspecifying the time profiles of the coefficients of the TVC model.

[6] This result provides the conditions under which TVC approaches dominate the other approaches to model building including state-space approaches. Thus, model (10) can satisfy Zellner's (1988) definition of a good model: "… logically consistent and sufficient mathematical economic theorems do not qualify to be termed laws unless it can be shown that they actually explain a wide range of past data and experience and yield good predictions over a broad range of data and experience". (p. 9)

3 Identification and Consistent Estimation of Time-Varying Coefficient Models

3.1 Identification

Lehmann and Casella (1998, p. 24) show that unidentifiable parameters are statistically meaningless. To show that the parameters of model (10) are statistically meaningful, we need to demonstrate that the identifiability conditions for these parameters are satisfied.

The fixed coefficient vector π^{Long} in (10) is identified if X_z has full column rank. A necessary condition for X_z to have full column rank is that $T > Kp$. That has been the case for the received applications. 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 are used in (4) to reduce the unidentifiable portions (the ε_{jt}) of the coefficients of (1). However, $D_x\varepsilon$ being equal to y - $X_z\pi^{Long}$ with identifiable π^{Long} is identifiable and 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 (5) and (6). Under Assumptions 1-4, the BLUP of $D_x\varepsilon$ exists (see Swamy, Yaghi, Mehta and Chang 2007, p. 3387).

Thus, Assumptions 1-4 make all the fixed parameters of model (10) statistically meaningful. Equation (4) which establishes a link between the coefficients of (1) and the coefficients and errors of (10), shows that if the coefficients and $D_x\varepsilon$ of (10) are statistically meaningful, so are the coefficients of (1). We next show that these identification results do not hold if Assumption 1 is changed to

Assumption $1'$ *For all t, each coefficient of (1) is modeled by a random walk of the form*

$$\gamma_{jt} = \gamma_{j,t-1} + \zeta_{jt}, \tag{11}$$

where ζ_{jt}, j = 0, 1, …, K-1, *are mutually uncorrelated, each of which is serially uncorrelated and is distributed with mean zero and constant variance.*

Suppose that Assumption 4 is also changed to

Assumption $4'$ *The regressor* x_{jt} *of (1) is unconditionally independent of its coefficient* γ_{jt} *for all j and t.*

We have already shown that Assumption $1'$ implying infinite unconditional variance for γ_{jt} may be regarded as unnatural and that Assumption $4'$ is false in the setting of the γ_{jt} satisfying equations (2) and (3). We now show the consequences of these assumptions. Under Assumption $1'$, the coefficients of (1) and their components shown in (2) and (3) are not identified on the basis of the observations in (1). If so, the coefficients of (1) are statistically meaningless because it is not possible to obtain an accurate estimate of the direct effect of each of x^*_{jt}, j = 1, …, K-1, on y^*_t with all the other determinants of y^*_t held constant. In the studies that make Assumptions $1'$ and $4'$, the time path of γ_{jt} is estimated using the Kalman filter with an arbitrary estimate of the unidentified initial value γ_{j1}. The formula for the Kalman filter (see Durbin and Koopman 2001, p. 12) of γ_{jt} derived under Assumptions $1'$ and $4'$ is inappropriate, since Assumption $4'$ is false. Furthermore, the Kalman filters of the γ_{jt} derived under Assumptions $1'$ and $4'$ are unconditionally inadmissible relative to quadratic loss functions because they do not possess finite unconditional means. Under Assumptions $1'$ and $4'$, the Kalman filters of γs do not provide the predictions of y_t with good conditional and unconditional properties. Brown's (1990) argument favoring such predictors is as follows:

> "Ordinary notions of consistency demand use of procedures which are valid and admissible both conditionally and unconditionally. (Numerically minor deviations from this goal may be satisfactory and justifiable on the grounds of convenience. The preceding statement also requires the qualification that the problem be correctly modelled, otherwise it may be desirable to adopt robust but formally inadmissible procedures to reflect realistic possibilities that have been omitted from the formal model.)
>
> … It seems to me the conclusion is that none of [(objectively or subjectively specified) formal Bayes estimators or empirical or robust Bayes methods] … should be applied conditionally without also taking into account the unconditional, frequentist structure of the situation." (p. 491)

Under Assumptions $1'$ and $4'$, model (1) does not satisfy Brown's qualification that it needs to be correctly specified because it has statistically meaningless coefficients.

3.2 Estimation under Assumptions $1'$ and $4'$

Methods for calculating the loglikelihood function, and the maximization of it with respect to the variances of ζ s in (11), require the joint density of y_1, …, y_T. Suppose that for *j* = 0, 1, …, *K*-1, ζ_{j1}, ζ_{j2}, … are independent according to a normal distribution with mean zero and variance σ_j^2. Let $\dot{y}_1$, …, $\dot{y}_T$ denote the values taken by the random variables $\dot{y}_1$, …, $\dot{y}_T$, respectively. Then if Assumption $4'$ were true, the conditional distribution of $\dot{y}_t$ given x_{1t}, …, $x_{K-1,t}$, $\dot{y}_1$, …, $\dot{y}_{t-1}$ would be normal with certain mean and variance implied by equation (1) and Assumption $1'$ (see Durbin and Koopman 2001, p.14). Let $p(\dot{y}_t \mid \dot{y}_1,...,\dot{y}_{t-1}, x_{1t},..., x_{K-1,t})$ be the probability density function (pdf) of this distribution. Then the joint density of $\dot{y}_1$, …, $\dot{y}_T$ can be written as

$$p(\dot{y}_1,...,\dot{y}_T \mid x_{11},...,x_{K-1,T}) = p(\dot{y}_1 \mid x_{11},...,x_{K-1,1})\prod_{t=2}^{T} p(\dot{y}_t \mid \dot{y}_1,...,\dot{y}_{t-1}, x_{1t},...,x_{K-1,t}), \qquad (12)$$

where it is assumed that

$$p(\dot{y}_1 \mid x_{11},...,x_{K-1,1}) \text{ with finite unconditional variance exists.} \qquad (13)$$

The joint pdf in (12) is the result of a contradiction because assumption (13) contradicts the assumption that model (11) holds for all *t*. This is because, according to the latter assumption, y_t does not possess finite unconditional mean for all *t*, and, according to the former assumption, y_t possesses finite unconditional mean for *t* = 1. Time *t* = 1 is not unique. Both assumption (13) and model (11) for *t* > 1 cannot hold if time t = 1 is not unique. Wrong initialization of the Kalman filter of γ_{jt} (see Durbin and Koopman 2001, pp. 17-30) using a wrong distribution of the initial value γ_{j1} can lead to wrong time-paths of the coefficients of (1).

3.3 Estimation under Assumptions 1-4

While instrumental variable estimation can work in the specific case of measurement error alone, we argue that, in the case of the general forms of misspecification we consider, there can be no variables which meet the requirements for valid instruments, so that instrumental variable estimation is not a sensible way forward. Instrumental variables estimation requires a set of instruments that are correlated with the variable that is subject to measurement error, but not correlated with the error process. This is achieved in the standard econometric model of measurement error as it is assumed that the error process in the equation comprises two components, a term involving the measurement error and a random error term (see Greene 2008, p. 326, (12-13)). If this second component is, indeed, random then valid instruments may exist. However, an alternative view on the nature of this error is that it represents all the misspecification in the model, including omitted variables and misspecified functional form; equation 10 makes this clear. In the latter case, valid instruments cannot exist as the error will be a function of all relevant omitted variables. In effect, we argue that, instrumental variables can only work in the complete absence of any other form of misspecification in the model.

Once we assume that there are omitted variables and other forms of misspecification in our model the instrumental variables that are correlated with the regressors of model (10) but not with its error term do not exist because X_z and $D_x\varepsilon$ have the regressors of (1) in common. An extension of this result was made by Pratt and Schlaifer (1988) who had warned that "it must not be assumed that because the value of a lagged included variable x_{t-1} was determined before the value of the current joint effect U_t of excluded variables, x_{t-1} necessarily satisfies the condition for observability – i.e., was independent of U_t. It may well have been influenced by a forecast of an excluded variable represented in U_t, or both x_{t-1} and U_t may have been affected by some third variable – in common parlance, a 'common cause'." (p. 47). What was assumed in the studies covered in Greene (2008, pp. 341-349) is what Pratt and Schlaifer said must not be assumed. The warning of Pratt and Schlaifer and the nonexistence of instrumental variables in model (10) clarify the distinction between the method of instrumental variables, which has become a workhorse technique in the empirical literature, and the use of coefficient drivers. The latter variables, with the same pattern of time variation as the components (3) of the coefficients of (1) over the relevant estimation and forecasting periods can exist. The motive for introducing model (10) is the expectation that such coefficient drivers can be found.

3.4 Practical Estimation

Under Assumptions 1-4, we can use the following strategy to construct a practical estimation method. Model (10) can be estimated by an iteratively rescaled generalized least squares (IRSGLS) method developed in Chang, Swamy, Hallahan and Tavlas (2000).

This criterion leads to a good determination of the coefficients of (1) if Assumptions 1-4 hold with $\Phi = 0$ and small variances for the errors of (4), as the Chebychev inequality shows (see Lehmann 1999, p. 52). The appropriate formulas for computing the standard errors of IRSGLS estimates are given in Swamy, Yaghi, Mehta and Chang (2007). A Monte Carlo study by Yokum, Wildt and Swamy (1998) shows that model (10) performs well in prediction. Further Monte Carlo studies on model (10) are in progress.

Before applying the IRSGLS method to model (10) we need to select the coefficient drivers that satisfy Assumptions 1-4. To operationalize the method of the decomposition of γ_{jt} outlined in Assumption 2, we make γ_{jt} a function of all the variables which we can use with the data available to us and which we believe are correct coefficient drivers from the following reasoning: If the parameters of the "decision rules" embodied in equation

(1) change when economic policies change, then it is sensible to use these policy changes as coefficient drivers in (4). With the relevant policy changes entering into (4) as coefficient drivers, equation (1) is not subject to "the Lucas (1976) critique". Any shift variables that we believe have changed the functional form of equation (1) and lagged changes in the included explanatory variables can also be correct coefficient drivers.[7] We need to satisfy the condition that model (10) provides an adequate approximation, over the relevant range of variation in y_t^* and its determinants, to the relation of y_t^* to all of its determinants.

In developing a procedure for selecting a subset of the included coefficient drivers that accurately estimates the direct effect α_{jt}^* in (3), we use the prior information: α_{jt}^* is unique with known sign so that our procedure is not completely arbitrary. A test statistic that is consistent with this prior information is due to Pratt and Schlaifer (1988, p. 44) who say that "the relevant 'test statistic' for a law as opposed to a regression is not R^2 or F, but the vector of changes in the estimated effects of X on Y that result when 'test concomitants' are included in the relation". We cannot use this test statistic because our method of using coefficient drivers is different from Pratt and Schlaifer's method of using concomitants. They include concomitants as additional regressors in (1), whereas we include the coefficient drivers in (4) to decompose the coefficients of (1) into their components in (3). Hence the following selection procedure is different from Pratt and Schlaifer's test statistic.

[7] Examples of such coefficient drivers are given in Hall, Hondroyiannis, Swamy and Tavlas (2007).

3.5 Decomposing the Coefficients

The final part of the procedure is to decompose the time-varying coefficients into their components representing biases and bias-free effects;,this requires the allocation of the coefficient drivers into the two sets S_1 and S_2. This element of the procedure involves some judgment and we have the following advice to try and make this process as objective as possible.

The whole set or any sub-set of the included coefficient drivers that gives an estimate of α_{jt}^{*} with the wrong sign should be rejected. The largest possible sub-set of the included coefficient drivers that gives an estimate of α_{jt}^{*} with the right sign is acceptable unless this sub-set gives an overestimate of the magnitude of α_{jt}^{*}, in which case the coefficient drivers with insignificant or relatively small magnitudes of π estimates can be eliminated from the sub-set.

This decomposition is based on the intuition that an estimate of α_{jt}^{*} can have the wrong sign and a wrong magnitude if an accurate estimate of the sum of the last two terms on the right-hand side of equation (3) is not subtracted from the corresponding accurate estimate of γ_{jt}.

Each coefficient of the TVC model considered in this paper is the sum of three components. One of the components of the coefficient of a regressor in this model measures the direct effect of a determinant of the dependent variable on the dependent variable with all of its other determinants held constant. Under Assumptions 1-4, the fitting criterion described in this section gives a formal statistical way of determining the sum of the three components and the decomposition of this sum described above gives a formal statistical way of determining which coefficient drivers are used to derive the direct-effect component.

3.6 Bayesian estimation of the direct effects with probably correct coefficient drivers

DeGroot (1982) wrote, "All good Bayesian statisticians reserve a little pinch of probability for the possibility that their model is wrong". Accordingly, we assign a less than 1 prior probability for the possibility that the coefficient drivers included in (4) are correct. Models of the form (10) with the same explanatory variables as in (1) but with different coefficient drivers are considered as separate elements of a model space. Models of this space differ only in the definitions of coefficient drivers. We assume that this space is a finite set. We assume that the prior probabilities assigned to its elements add up to 1. That is, the prior probability, denoted by $P(M_i)$, assigned to the ith model, denoted by M_i, of the space satisfies the condition $\sum_i P(M_i)$ = 1. If the set of coefficient drivers included in M_i leads to the estimates of the direct effects α^*_{jt} with the wrong sign and/or to poor predictions of future values of the dependent variable y_t of (1), then we will set $P(M_i)$ = 0 and adjust the non-zero probabilities assigned to other models in the space so that they add up to 1. Let $p(y \mid X_z, \pi^{Long}, \omega, M_i)$ denote the probability density function (pdf) of *y* under M_i with $P(M_i) \neq 0$, where ω is a vector having the unknown elements of Φ, the distinct but unknown elements of Δ_u, and σ_u^2 in (5) and (6) as its elements.

Note that for notational convenience we suppress the subscript i of X_z, π^{Long}, and ω. We denote the prior pdf for π^{Long} and ω under M_i by $p(\pi^{Long}, \omega \mid M_i)$.

For fixed y and X_z of M_i, viewed as a function of π^{Long} and ω, $p(y \mid X_z, \pi^{Long}, \omega, M_i)$ is called the likelihood of π^{Long} and ω. Note that the parameters π^{Long} and ω do not distort the correct interpretations of the coefficients of (1) because these parameters are defined by Assumptions 1-4 that establish a close connection between the coefficients of (1) and the unknown quantities π^{Long} and ε of (10) and these parameters call a spade a spade. Suppose that we cannot maintain this connection because we have included inappropriate coefficient drivers in (4). Then we will not be estimating (1) but will be estimating something else using the likelihood function $p(y \mid X_z, \pi^{Long}, \omega, M_i)$. In this regard, Pratt and Schlaifer (1988) point out that "If a statistician with a Bayesian computer program treats the likelihood function of **c** as if it were the likelihood function of **b**, as he must if he supplies no proper prior distribution of **b** given **c**, then what his printout will contain will be neither **b**eans nor **c**orn but **s**uccotash." (p. 49). To avoid these mistakes, we are trying to satisfy Assumptions 1-4 that maintain the connection between (1) and (10).

We may now compute the marginal density of the observations (given M_i) by

$$p(y \mid X_z, M_i) = \int_{\pi^{Long}} \int_{\omega} p(y \mid X_z, \pi^{Long}, \omega, M_i) p(\pi^{Long}, \omega \mid M_i) \mathrm{d}\omega \mathrm{d}\pi^{\mathrm{Long}} \tag{14}$$

and average over the models to obtain a density which is unconditional of them:

$$p(y \mid \mathrm{X_z} \text{ of all } M_\mathrm{i}) = \sum_i p(y \mid X_z, M_i) P(M_i) . \tag{15}$$

By the Bayes theorem, the posterior probability of M_i is

$$P(M_i \mid y,\ X_z \text{ of all } M_\mathrm{i}) = \frac{P(M_i)\,p(y \mid X_z, M_i)}{p(y \mid X_z \text{ of all } M_i)}. \tag{16}$$

This posterior probability shows how the prior probability assigned to M_i is revised by the data. Further, the posterior pdf for the (*j*, *d*)th element π_{jd} of π^{Long} in M_i is given by

$$p(\pi_{jd} \mid y,\ X_z, M_i) = \frac{\int_{\pi_{-jd}^{Long}}\int_{\omega} p(y \mid X_z, \pi^{Long}, \omega, M_i)\,p(\pi^{Long}, \omega \mid M_i)\mathrm{d}\omega\,\mathrm{d}\pi_{\text{-jd}}^{\text{Long}}}{\int_{\pi^{Long}}\int_{\omega} p(y \mid X_z, \pi^{Long}, \omega, M_i)\,p(\pi^{Long}, \omega \mid M_i)\mathrm{d}\omega\,\mathrm{d}\pi^{\text{Long}}} \tag{17}$$

where π_{-jd}^{Long} is π^{Long} without the (*j*, *d*)th element. The mean of this distribution gives the posterior mean, denoted by $\mathrm{E}(\pi_{\mathrm{jd}} \mid y, X_z, M_i)$, of π_{jd}.

A Bayes estimator of the direct effect α_{jt}^* is

$$\mathrm{E}(\alpha_{\mathrm{jt}}^* \mid y, X_z \text{ of all } M_i) = \sum_i [P(M_i \mid y, X_z \text{ of all } M_i)\{\mathrm{E}(\pi_{\mathrm{j0}} \mid y, X_z, M_i) + \sum_{d \in S_{1i}} \mathrm{E}(\pi_{\mathrm{jd}} \mid y, X_z, M_i) z_{dt}\}] \tag{18}$$

where S_{1i} is a subset of the coefficient drivers in the ith model M_i such that π_{j0} + $\sum_{d \in S_{1i}} \pi_{jd} z_{dt}$ has the same pattern of time variation as α_{jt}^* and the subscript *i* of z_{dt} is suppressed.

With the model space we have above, we can consider all possible combinations of coefficient drivers and then think about how we can try to narrow down the set of drivers which should be used in (18). The first step is to eliminate those combinations of drivers which would give rise to an estimate of α_{jt}^* which goes outside some acceptable range (possibly just in terms of sign but possibly also in terms of magnitude). For the M_i's providing only such combinations, we set $P(M_i)$ = 0. The next step might be to ask if there are any drivers which simply do not matter much, that is to say the α^{*}_{jt} estimate is much the same regardless of whether they are included or excluded in all the combinations they appear in. This may or may not reduce the number of possible combinations to consider by quite a lot. After we have excluded the driver combinations that give ‘unacceptable’ estimates of direct effects then with equal priors on the remaining combinations we simply take a weighted average.

Further, we may express the predictive pdf for a future value, denoted by y_F, of y_t as

$$p(y_F \mid y, X_z \text{ of all } M_i) = \sum_i P(M_i \mid y, X_z \text{ of all } M_i)\, p(y_F \mid y, X_z, M_i). \tag{19}$$

4 Conclusions

Any specified econometric model is likely to be misspecified. This paper offers the correct interpretations of the coefficients and the error term of a misspecified model. The assumptions that are consistent with these interpretations are also offered. With such assumptions it is possible to correct for omitted-variables and measurement-error bias in misspecified models. Both Bayesian and non-Bayesian solutions to the problems of unknown functional form, omitted variables, and measurement errors are presented.

References

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

Brown, L.D.: An ancillarity paradox which appears in multiple linear regression (with discussion). Ann Statist **18**, 471-538 (1990)

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

DeGroot, M.H.: Comment on 'Lindley's paradox' by G. Shafer. J Amer Statist Assoc **77**, 336-339 (1982)

Durbin, J., Koopman, S.J.: Time series analysis by state space methods. Oxford: Oxford University Press (2001)

Granger, C.W.J. and Newbold, P.: Spurious regression in econometrics. J Econ **2**, 111-120 (1974)

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

Hall, S.G., Hondroyiannis, G., Swamy, P.A.V.B., Tavlas, G.S.: A portfolio balance approach to euro-area money demand in a time-varying environment. A paper presented at Bank of England/MMF workshop on money and macromodels.

Hondroyiannis, G., Swamy, P.A.V.B., Tavlas, G.S.: A note on the new Keynesian Phillips curve in a time-varying coefficient environment: some European evidence. Macroecon Dyn **12**, (2008) forthcoming

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)

Lucas, R.E., Jr.: Econometric policy evaluations: a critique. Carnegie-Rochester Ser Public Policy **1**, 19-46 (1976)

Pratt, J.W., Schlaifer, R.: On the nature and discovery of structure. J Am Stat Assoc **79**, 9-21, 29-33 (1984)

Pratt, J.W., Schlaifer, R.: On the interpretation and observation of laws. J Econ Ann **39**, 23-52 (1988)

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

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., Tavlas, G.S.: A note on Muth's rational expectations hypothesis: A time-varying coefficient interpretation. Macroecon Dyn **10**, 415-425 (2006)

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., Mehta, J.S., Chang, I-L., Zimmerman, T.S.: An efficient method of

estimating the true value of a population characteristic from its discrepant estimates. Comput Statist Data Anal **52**, (2008) forthcoming
Yokum, J.T., Wildt, A.R., Swamy, P.A.V.B.: Forecasting disjoint data structures using appropriate constant and stochastic coefficient models. J Appl Statist Sci **8**, 29-49 (1998)
Zellner, A.: Causality and econometrics, 9-54. In: Brunner, K., Meltzer, A.H. (eds.) Three aspects of policy and policymaking. Amsterdam: North-Holland 1979
Zellner, A.: Causality and causal laws in economics. J Econ Ann **39**, 7-21 (1988)