

Article

# Endogeneity, Time-Varying Coefficients, and Incorrect vs. Correct Ways of Specifying the Error Terms of Econometric Models

P.A.V.B. Swamy<sup>1</sup>, Jatinder S. Mehta<sup>2,\*</sup> and I-Lok Chang<sup>3</sup>

<sup>1</sup> Federal Reserve Board (Retired), Washington, DC 20551, USA; swamyparavastu@hotmail.com

<sup>2</sup> Department of Mathematics (Retired), Temple University, Philadelphia, PA 19122, USA

<sup>3</sup> Department of Mathematics (Retired), American University, Washington, DC 20016, USA; ilchang@american.edu

\* Correspondence: mehta1007@comcast.net; Tel.: +1-215-635-3173

Academic Editors: Gilles Dufrénot, Fredj Jawadi and Alexander Mihailov

Received: 5 September 2016; Accepted: 8 December 2016; Published: 3 February 2017

**Abstract:** Using the net effect of all relevant regressors omitted from a model to form its error term is incorrect because the coefficients and error term of such a model are non-unique. Non-unique coefficients cannot possess consistent estimators. Uniqueness can be achieved if; instead; one uses certain “sufficient sets” of (relevant) regressors omitted from each model to represent the error term. In this case; the unique coefficient on any non-constant regressor takes the form of the sum of a bias-free component and omitted-regressor biases. Measurement-error bias can also be incorporated into this sum. We show that if our procedures are followed; accurate estimation of bias-free components is possible.

**Keywords:** endogenous variable; exogenous variable; time-varying coefficient; unique coefficient and error term; accurate estimation of bias-free component

**JEL Classification:** C13; C51

---

## 1. Introduction

The quality of econometric practice is reflected in the assumptions made to build a model. We compare two kinds of practice: one labeled “conventional” and the other labeled “new”. In *conventional* practice, the structural form of each equation in a complete model of linear simultaneous equations has (i) one of the jointly dependent or endogenous variables as its dependent variable; (ii) some relevant endogenous and exogenous variables with relevant predetermined variables as its included regressors; and (iii) relevant but omitted regressors constituting the structural disturbance. However, as shown by Pratt and Schlaifer (1984, 1988) [1,2] (hereafter PS), problems in estimation arise because the error term of an equation made up of relevant regressors omitted from the equation is non-unique. As a consequence, the coefficients of such equations cannot be unique, however estimated. Conventional practice also uses non-linear regression models which have problems of their own.

The new proposed practice remedies shortcomings, arising from non-uniqueness of the coefficients and error terms of models, uniqueness being a property that holds jointly for both coefficients and error term of each equation. For the purpose of our discussion, we shall employ a definition of uniqueness based on earlier work of ours. To anticipate, we will show that the error term of an equation made up of certain “sufficient sets”, to be defined later, of relevant regressors omitted from the equation and its coefficients are unique, where the unique coefficient on a non-constant regressor takes the form of the

sum of a bias-free component and an omitted-regressors bias component. If necessary, one may add a measurement-error bias component to this sum.<sup>1</sup>

The main purpose of this paper is to illuminate the differences between conventional practices and our new proposed methodology. We begin by demonstrating exactly how conventional practice gives rise to (i) non-unique coefficients that cannot be consistently estimated and (ii) a conflict between non-uniqueness of the coefficients and error term of an equation and the exogeneity of some or all of its regressors. We follow this discussion by introducing our new practice, which, as promised, employs models with unique coefficients and error terms, as laid out in a series of papers by Swamy and his associates, a recent contribution to which is Swamy et al. (2016) [3]. Although our earlier work has dwelled on aspects of the new methodology, there is a substantial amount of material in this paper that is novel. For example, Theorem 1 and Corollaries 1 and 2 proved in this paper are new. Further, in contrast to our earlier work on models with time-varying coefficients, this paper introduces the idea of included endogenous regressors with time-varying coefficients, leading to insights that to our knowledge have not heretofore been discussed in the literature. As a consequence, there is little overlap between this and our earlier work.

To be specific about the problems emanating from non-unique coefficients and error terms of models, we choose a linear regression of the earnings of an individual on a set of regressors including a dummy variable that takes the value 1 if the individual attended a college and the value zero otherwise. If the error term of this regression represents the net effect on the earnings of relevant regressors omitted from the regression, then both of its coefficients and error term are non-unique, as we show in this paper. These non-unique coefficients are not consistently estimable. Furthermore, since causality designates a property of the real-world and the above regression with non-unique coefficients and error term cannot be a real-world relationship, the coefficient on the dummy variable in the regression is not the causal effect of attendance at a college. We show in this paper how this causal effect can be measured when education is endogenous.

The remainder of this paper is arranged as follows: Section 2 has three parts. The first part discusses problems with linear simultaneous equations models employed in conventional practice. The second part derives a new simultaneous-equations system in which the functional form of every non-identity is linear-in-variables and nonlinear-in-coefficients. From this non-linear system we then derive another system of equations that is of considerable importance to estimation in that none of the new equations will contain exogenous variables but instead will feature a unique error term and unique coefficients. The coefficient on each non-constant regressor of the final model featuring observable variables contains a bias-free component plus omitted-regressor and measurement-error bias components. We parameterize the system by making each of its coefficients a linear function of appropriate coefficient drivers plus a random error. The third part of Section 2 shows that the parameterized model can yield accurate estimates of the bias-free components of the final model's coefficients. Section 3 concludes.

## 2. Simultaneous Equations Model

### 2.1. Conventional Practice

#### 2.1.1. Constant Coefficients

Typically, the econometric literature distinguishes between jointly dependent (or endogenous), exogenous, and predetermined variables.<sup>2</sup>

---

<sup>1</sup> The concept of "sufficient sets" of omitted regressors is due to PS (1988) [2] (p. 34). The term "bias-free component" means the component free of omitted-regressor and measurement-error biases.

<sup>2</sup> See Greene (2012) [4] (pp. 317,318). We will have an occasion below to discuss the inaccuracy of exogeneity assumption.

The linear structural form of a model is then usually written as

$$Y\Gamma + XB = U \quad (1)$$

where  $Y$  is a  $T \times M$  matrix containing  $T$  observations on  $M$  jointly dependent or endogenous variables,  $\Gamma$  is an  $M \times M$  matrix of constant coefficients,  $X$  is a  $T \times K$  matrix of  $T$  observations on  $K$  exogenous and predetermined variables,  $B$  is a  $K \times M$  matrix of constant coefficients, and  $U$  is a  $T \times M$  matrix of structural disturbances that are assumed to be serially independent. The above notation implies that model (1) is a system of  $M$  linear simultaneous equations. Suppose that the identities have already been removed from (1). The linearity assumption in Equation (1) and implied constancy of the elements of  $\Gamma$  and  $B$  are unduly restrictive and are relaxed in the next section.<sup>3</sup>

**Interpretation of  $U$ :** Greene (2012) [4] (p. 13) pointed out that the disturbance of each equation in (1) captures the net effect on the dependent variable of those regressors that are omitted from the equation.

This is the usual interpretation. We investigate the consequences of adopting this interpretation in this section.<sup>4</sup>

**The findings of PS (1988) [2] (p. 34):** The condition that the included regressors be independent of “the” omitted regressors themselves is meaningless unless the definite article is deleted and can then be satisfied only for certain “sufficient sets” of omitted regressors, some if not all of which must be defined in a way that makes them unobservable as well as unobserved. We prove below that certain problems cannot be avoided unless one takes these findings of PS seriously.

Regardless of the endogeneity and exogeneity of omitted influences on the dependent variables, each row of  $U$  is assumed to be randomly drawn from an  $M$ -variate distribution. Furthermore, conventional practice imposes three additional assumptions which we now state.

**Assumption 1.**

$$(i) E(U|X) = 0; (ii) E(Y|X) \text{ is finite; and } (iii) E[(1/T) U'U|X] = \Sigma \quad (2)$$

The conditional expectation  $E(Y|X)$  does not always exist, but sufficient conditions for its existence are given in Rao (1973) [5] (p. 97), although not all economists and statisticians interpret the condition  $E(U|X) = 0$  in the same way. For example, Greene (2012) [4] (p. 223) interpreted it to mean that  $X$  is exogenous in model (1) in the sense that  $X$  is determined outside of the model. Engle et al. (1983) [6] listed four distinct concepts of exogeneity corresponding to different notions of what is “determined outside the model under consideration” according to the purposes of the inferences being conducted. In Friedman and Schwartz’s (1991) [7] (pp. 41–42) view, it may be appropriate to regard a variable as exogenous for some purposes and as endogenous for others. In this respect, the assumptions of one of Lehmann and Casella’s (1988) [8] (Theorem 4.12, p. 184) theorems are the same as our Assumption 1. Finally, PS (1988) [2] (p. 34) showed that the stronger version of  $E(U|X) = 0$ , i.e.,  $X$  independent of  $U$ , is meaningless if the error term of each equation in (1) is made up of relevant regressors omitted from

<sup>3</sup> The constancy assumption about the coefficients of (1) may mean that this equation system is not the correct specification of the model of  $Y$ . Here we do not want to use the term “true specification”. Econometricians generally disapprove of the use of the word “true model.” Note that we do not use the econometrician’s term “data-generating process” because it is not informative about omitted-regressors unrepresented by any data in our analysis, preferring instead the term “correct” to “true”.

<sup>4</sup> Some economists and statisticians believe that if model (1) were correctly specified, then the rows of  $U$  would be identically and independently distributed (i.i.d.), being free of omitted influences. First of all, one cannot prove that any model is “correctly specified,” and second, the i.i.d. assumption about the rows of  $U$  does not mean that each row of  $U$  is free of omitted influences.

the equation. We show below that Assumption 1(i) does not hold if the coefficients and error term of each equation in (1) are non-unique.

Next, consider the important issue of identification. For this, define

**Reduced form:** When  $\Gamma^{-1}$  exists,

$$Y = X(-B\Gamma^{-1}) + U\Gamma^{-1} = X\Pi + V \quad (3)$$

where  $\Pi = -B\Gamma^{-1}$  and  $V = U\Gamma^{-1}$ .

**Normalization rule:** There will be at least one “1” in each column of  $\Gamma$ .

**Identification:** To achieve identification in all  $M$  equations of model (1), one imposes the above normalization rule and certain exclusion and other restrictions on the elements of  $\Gamma$  and  $B$ , such that only the identity matrix of order  $M$  is the admissible value for an  $M \times M$  nonsingular matrix  $P$  in  $-BPP^{-1}\Gamma^{-1} = \Pi$ . The econometric literature has evolved a necessary order and a sufficient rank condition for obtaining unique solutions for the unknown coefficients of the equations in (1) using equations  $\Pi\Gamma = -B$  where, under Assumption 1(i), the conditional mean of  $Y$  given  $X$  is<sup>5</sup>

$$E(Y|X) = X\Pi. \quad (4)$$

But this conditional mean may not exist if Rao's (1973) [5] (p. 97) conditions for its existence do not apply. In order to trace through the effects of autonomous changes in the variables in (1), it is necessary to work through the reduced form, where by convention, the change in  $Y$  induced by a change in  $X$  has the interpretation of a partial derivative, since  $X$  is determined outside model (1).<sup>6</sup> However, in the case of endogenous variables, the ratio of a change in one of them to a change in another cannot have a partial derivative interpretation and is therefore meaningless without first determining what caused the change in the denominator (see Greene (2012) [4] (p. 320)).

To demonstrate cases where Assumption 1(i) is false, we consider the following  $j$ th equation of (1):

$$y_j = Y_j\gamma_j + X_j\beta_j + u_j \quad (j = 1, \dots, M) \quad (5)$$

where  $y_j$  is a  $T \times 1$  vector of observations on the dependent variable of the  $j$ th equation,  $Y_j$  is a  $T \times (M_j - 1)$  matrix consisting of  $T$  observations on a set of  $M_j - 1$  included endogenous regressors that appear on the right-hand side of the  $j$ th equation,  $\gamma_j$  is a column vector of  $(M_j - 1)$  coefficients on the included endogenous regressors,  $X_j$  is a  $T \times K_j$  matrix consisting of  $T$  observations on  $K_j$  included exogenous regressors,  $\beta_j$  is a  $K_j \times 1$  vector of coefficients on the included exogenous regressors, and  $u_j$  is a  $T \times 1$  vector of disturbances.

**Specific Example:** An economic example of Equation (5) is

$$\text{Earnings-Education (EE) Relationship : } \text{earnings}_i = x_i'\beta_i + \delta C_i + u_i \quad (6)$$

where  $i$  indexes individuals, the non-constant elements of  $x_i$  are defined in Krueger and Dale (1999) [9],  $C_i$  is a dummy variable taking the value 1 if individual  $i$  attended a college and taking the value 0 if individual  $i$  did not attend a college, and  $u_i$  is the error term.

<sup>5</sup> These order and rank conditions do not hold if the coefficients and error term of each equation in (1) are non-unique, as shown below.

<sup>6</sup> It is shown below that the exogeneity of  $X$  does not hold; so analyses based on the reduced form in (4) cannot be carried out if the coefficients and error term of each equation in (1) are non-unique.

Greene (2012) [4] (p. 890) showed that the coefficient  $\delta$  does not measure the causal effect of a college education if individuals who choose to go to college would have relatively high earnings whether or not they had gone to college. He further pointed out (see [4] (p. 252)) that (i)  $C_i$  cannot vary autonomously outside the model of the EE relationship; and (ii) variations in  $C_i$  are determined partly by the same hidden influences that determine lifetime earnings. Statements (i) and (ii) mean that  $C_i$  is an endogenous regressor. For this reason, measurement of the effect  $\delta$  of a college education cannot be done with multiple linear regressions, as shown by Greene (2012) [4] (p. 252). Causal implications can only be drawn from the EE relationship in (6), if it is a real-world (or misspecifications-free) relationship (see Swamy et al. (2016) [10]). We show below that (6) is not a misspecifications-free relationship. Thus, the EE relationship in (6) nicely illustrates the problems of interpretation that can arise with (5). We will refer to the EE relationship several times below.

### 2.1.2. Conflict between the Exogeneity Assumption about Certain Regressors in a Model and Non-Uniqueness of Its Coefficients and Error Term

Conventional practice always obeys this assumption:

**Assumption 2.** *Omitted relevant regressors constituting the error term of an econometric model do not introduce omitted-regressor biases into the coefficients of the included regressors.*

Confusion may arise if we do not point out here that Theil's specification-error analysis reproduced in Greene (2012) [4] (p. 56) and other econometric textbooks also involves terms such as "omitted regressors" and "omitted-regressor biases," but that their meanings are different from those used in Assumption 2. For Theil, omitted regressors are those relevant regressors that get *removed* from model (5) when some columns of  $X_j$  are deleted; and the omitted-regressor biases are those biases that get *introduced* into the least squares estimators of some of the elements of  $\Pi$  as a result of this deletion. These omitted regressors and omitted-regressor biases are different from omitted regressors constituting  $u_j$  and the biases they introduce, respectively. A less confusing definition of uniqueness is the following:

**Definition (Uniqueness):** *The coefficients and error term of a model are said to be unique if they are invariant under the addition and subtraction of the product of the coefficient of any omitted relevant regressor and any included regressor on the right-hand side of the model.*

Note that the coefficients and error term of any model are non-unique if they are not unique.

Now we use the preceding definition to show that the coefficients and error term of Equation (5) are not unique, which, for the  $t$ th element of  $y_j$  is

$$y_{tj} = \gamma'_j y_{t,-j} + \beta'_j x_{tj} + \omega'_j w_{tj} \quad (7)$$

where  $y_{tj}$  is the  $j$ th element of  $y_t = (y_{t1}, y_{t2}, \dots, y_{tM})'$ ,  $y_{t,-j}$  is the transpose of the  $t$ th row  $(y_{t1}, \dots, y_{t,j-1}, y_{t,j+1}, \dots, y_{t,M_j})$  of  $Y_j$ ,  $\gamma'_j$  is the transpose of  $\gamma_j = (\gamma_{1j}, \dots, \gamma_{j-1,j}, \gamma_{j+1,j}, \dots, \gamma_{M_j,j})'$ ,  $x_{tj}$  is the transpose of the  $t$ th row  $(x_{t1}, \dots, x_{tK_j})$  of  $X_j$ ,  $\beta'_j$  is the transpose of the column vector  $\beta_j = (\beta_{1j}, \dots, \beta_{K_j,j})'$ ,  $w_{tj} = (w_{t1}, \dots, w_{tL_j})'$  is the column vector of (unknown) observations at time  $t$  on omitted regressors constituting  $u_j$ . To forestall omission of any relevant element of  $w_{tj}$ , we further assume that the value of  $L_j$  is unknown,  $\omega'_j = (\omega_{1j}, \dots, \omega_{L_j,j})$  is a row vector of the coefficients of omitted regressors,  $w_{tj}$ ,  $u_{tj} = \omega'_j w_{tj}$  is the  $t$ th element of  $u_j$  appearing in (5).

The elements of  $w_{tj}$  in (7), labeled "omitted regressors", are not used as the included regressors but are used to form the error term  $u_{tj}$  of (5). This is what we mean whenever we say that the elements of  $w_{tj}$  are omitted regressors constituting the error term  $u_{tj}$ . PS (1984) [1] (p. 13) pointed out that Equation (7) can be treated as a linear deterministic equation, even though econometricians treat  $u_j$  in

(5) as random. Given that econometricians’ treatment is arbitrary, PS’s treatment is entirely appropriate. Therefore, we shall use only mathematical methods to analyze (7).

PS (1984) [1] (p. 13) proved that  $\gamma_j$ ,  $\beta_j$ ,  $\omega_j$ , and  $w_{tj}$  in (7) are not unique without the help of our definition of uniqueness. Nevertheless, because little attention has been given to this important result by mainstream econometricians, it is useful to restate it here as Theorem 1 and to prove it by employing our definition of uniqueness.

**Theorem 1.** *If relevant regressors omitted from each of several simultaneous-econometric equations form its error term, then its coefficients and error term are non-unique, and such coefficients are not consistently estimable.*

**Proof.** The omitted regressors  $w_{tj}$  in (7) are not unique because  $\omega_j'w_{tj}$  does not change if it is written as  $(\omega_j'P)(P^{-1}w_{tj})$  for any  $L_j \times L_j$  nonsingular matrix  $P \neq I$ , the  $L_j \times L_j$  identity matrix. Hence the error term  $u_{tj}$  is not unique. Lehmann and Casella (1998) [8] (p. 57) proved that a parameter that is unidentifiable cannot be estimated consistently. Therefore, we should first check whether the coefficients of (5) with random  $u_j$  are identifiable. According to econometric textbooks, a necessary condition for the coefficients of (5) to be identifiable is that the number of exogenous variables omitted from (5) but included in other equations of model (1) must be at least as large as  $M_j - 1$ . This condition is inappropriate if Assumption 1(i) is false. (We show below that this assumption is indeed false when the coefficients and error term of (5) are not unique). To prove non-uniqueness, rewrite (7) as

$$y_{tj} = \sum_{\substack{h=1 \\ h \neq j}}^{M_j} \gamma_{hj}y_{th} + \sum_{k=1}^{K_j} \beta_{kj}x_{tk} + \sum_{\ell=1}^{L_j} \omega_{\ell j}w_{t\ell} \tag{8}$$

Let  $k'$  be one of the values the subscript  $k$  takes and let  $\lambda$  be one of the values the subscript  $\ell$  takes. The term  $\omega_{\lambda j}x_{tk'}$  is the product of an element of  $\omega_j$  and an element of  $x_{tj}$ . To apply the above definition of uniqueness, we add and subtract this product on the right-hand side of Equation (8). Doing so gives

$$y_{tj} = \sum_{\substack{h=1 \\ h \neq j}}^{M_j} \gamma_{hj}y_{th} + \sum_{\substack{k=1 \\ k \neq k'}}^{K_j} \beta_{kj}x_{tk} + (\beta_{k'j} + \omega_{\lambda j})x_{tk'} + \sum_{\substack{\ell=1 \\ \ell \neq \lambda}}^{L_j} \omega_{\ell j}w_{t\ell} + \omega_{\lambda j}(w_{t\lambda} - x_{tk'}) \tag{9}$$

Thus, going from (8) to Equation (9) makes one of the coefficients of Equation (8) to change from  $\beta_{k'j}$  to  $(\beta_{k'j} + \omega_{\lambda j})$  and makes one of the terms of the sum  $\sum_{\ell=1}^{L_j} \omega_{\ell j}w_{t\ell}$  in (8) to change from  $\omega_{\lambda j}w_{t\lambda}$  to  $\omega_{\lambda j}(w_{t\lambda} - x_{tk'})$ .<sup>7</sup> Even when  $x_{tk'}$  is not associated with  $w_{t\lambda}$  in (8),  $x_{tk'}$  is associated with  $(w_{t\lambda} - x_{tk'})$  in (9). Since the coefficients and omitted variables  $w_{tj}$  in (8) are unknown, we cannot prove that the values  $(\beta_{k'j} + \omega_{\lambda j})$  and  $(w_{t\lambda} - x_{tk'})$  in (9) are inadmissible. Therefore, we can validly state that the coefficient  $\beta_{k'j}$  and the term  $\omega_{\lambda j}w_{t\lambda}$  of  $\sum_{\ell=1}^{L_j} \omega_{\ell j}w_{t\ell}$  taking two different values in (8) and (9) are not unique. Similarly, assuming that  $K_j < L_j$ , we can show that the coefficients and error term of (5) are not unique and also show that all the regressors of (5) assumed to be exogenous are associated with any  $K_j$  terms of  $\sum_{\ell=1}^{L_j} \omega_{\ell j}w_{t\ell}$ . This means that when the coefficients and error term of (5) are not unique, the exogeneity assumption about  $X_j$  stated in Assumption 1(i) can be made true and false at the whim of an arbitrary choice between two observationally equivalent models in (8) and (9). Since the  $j$ th

<sup>7</sup> Equations (8) and (9) are treated as deterministic.

equation can be any one of the equations in (1), what we have proved about (5) is also true of other equations in (1). If the unknown coefficients and error term of every equation in (1) are not unique, then Assumption 1(i) and the so-called necessary order condition for identification in every equation of model (1) do not hold. Hence, the unknown coefficients of model (1) are not identified and are therefore not consistently estimable. Q.E.D.

Theorem 1 essentially warns against interpreting the disturbance in (5) as capturing the net effect of omitted regressors on the dependent variable because under such an interpretation, the coefficients and error term of (5) are non-unique; and non-unique coefficients are not consistently estimable. It is in this sense that there is a conflict—alluded to in the introduction—between non-uniqueness of the coefficients and error term of (5) and the exogeneity of some or all of its regressors. We have shown here that if one follows conventional practice, employing a linear simultaneous equations model with non-unique coefficients and error term, then the assumption that any of its regressors are exogenous is false. In this case, it is futile to impose restrictions on the model that ostensibly “identify” it.<sup>8</sup>

**Corollary 1.** *The least squares estimators of the non-unique coefficients of a reduced form with non-unique error terms are biased and inconsistent.*

**Proof.** Whenever Assumption 1(i) is false,  $E(\mathbf{U} | \mathbf{X}) \neq 0$  which proves the corollary.

**Corollary 2.** *None of the regressors of any linear simultaneous equation with non-unique coefficients and error term can be exogenous in the sense of Assumption 1(i).*

Therefore, Theorem 1 and Corollaries 1 and 2 are in complete alignment with results given in PS (1984, 1988) [1,2]. Note that Lehmann and Casella (1998) [8] claim to have proved (see their Theorem 4.12, p. 184) that under certain assumptions, the least squares estimators of the coefficients of a general linear model are uniform minimum variance and unbiased among all linear estimators. However, their conclusion conflicts with PS (1984) [1] in that they neither (i) take account of the real-world sources of the error term in the general linear model nor (ii) offer any examination of possible non-uniqueness of its coefficients and error terms.<sup>9</sup> The consistency proofs of limited and full information estimators given, e.g., in Greene (2012) [4] (p. 326–336), are based on Assumption 1(i) which is not satisfied when the coefficients and error terms of the  $M$  equations in (1) are not unique, as shown by PS (1984, 1988) [1,2].

Referring to the EE example in (6), it follows from Theorem 1 that its coefficients  $\beta$  and  $\delta$  are not unique and therefore not consistently estimable, and that the non-constant regressors in  $\mathbf{x}_i$  cannot be exogenous if  $u_i$  is made up of relevant regressors omitted from the EE relationship. Causality is the property of real-world relationships which will have the unique coefficients and error terms. The linear functional form of the EE relationship in (6) can mean that its functional form is misspecified. However, misspecified models cannot be real-world relationships and hence cannot be causal. All these statements suggest that  $\delta$  cannot be the causal effect of attending any college.

<sup>8</sup> There is a connection between Theorem 1 and a related theorem in Swamy et al. (2015) [11] that derives uniqueness of the coefficients and error term of a model as a necessary condition for its correct specification.

<sup>9</sup> To avoid a possible misunderstanding, we hasten to point out here that Section 2.1 is written not to criticize econometricians and statisticians in general and Lehmann and Casella [8] in particular but merely to point out the implication of a PS’s result about a meaningless assumption typically made in conventional practice for the consistency of regression coefficient estimators. Note that in proving Theorem 1, only Greene’s (2012) [4] (p. 13) interpretation of the error terms of econometric models was required without resort to further potentially arbitrary assumptions.

## 2.2. New Practice

### 2.2.1. Time-Varying Coefficients

Having learned the preceding lesson about the undesirable consequences of non-unique coefficients and error terms of models, we now turn to models with unique coefficients and error terms. In the interest of generality, which characterizes the new practice, we drop Assumption 2 as well as the assumption that the coefficients of model (1) are fixed.

**Assumption 3.** All relevant regressors omitted from each of several simultaneous-econometric equations introduce omitted-regressor biases into the coefficients on the included regressors of the equation.

Now consider (5) with its fixed coefficients changed to time-varying coefficients (TVCs):

$$y_{ij}^* = \alpha_{i0j}^* + \sum_{h=1}^{M_j} \alpha_{thj}^* y_{th}^* + \sum_{k=1}^{K_j} \beta_{tkj}^* x_{tk}^* + \sum_{\ell=1}^{L_j} \omega_{t\ell j}^* w_{t\ell}^* \quad (j = 1, \dots, M) \quad (10)$$

$$h \neq j$$

where all the relevant regressors are explicitly shown, none of  $y_{th}^*$ ,  $x_{tk}^*$ , and  $w_{t\ell}^*$  is equal to 1 for all  $h$ ,  $k$ , and  $\ell$ , respectively, the variables with an asterisk are the true values, and the coefficients are called “time-varying structural coefficients (TVSCs)”. With these coefficients, Equation (10) defines a variety of non-linear functional forms covering the linear form as a special case and the correct functional form of (10) can be any one of those forms. An innovation of Equation (10) is that, in contrast to previous work on time-varying coefficients, we now study a model with endogenous regressors ( $y_{th}^*$ ). In (5), the endogenous regressor matrix  $Y_j$  is correlated with its error term,  $u_j$ . This means that the regressors  $y_{th}^*$  of Equation (10) are associated with the variables ( $w_{t\ell}^*$ ’s). We treat (10) as a deterministic equation. Instead of assuming that the  $x_{tk}^*$ ’s in (10) are exogenous, we assume that they are also associated with “the”  $w_{t\ell}^*$  in (10).<sup>10</sup> This means that we heed the warning by PS (1988, p. 34) regarding the meaninglessness of the assumption that the regressors  $x_{tk}^*$  included in (10) are not associated with “the” regressors ( $w_{t\ell}^*$ ) to be used to form the error term of (10). When all the regressors in the  $M$  equations of (10) are endogenous, the model has more endogenous variables than equations and hence is incomplete. As a remedy, additional  $K$  equations, each with one of the  $K$   $x$ ’s as its dependent variable, should be added to (10) to make it a complete model. The functional form of (10) can be described as linear in variables but nonlinear in coefficients. Since (10), which we treat as deterministic, does not involve measurement errors and explicitly reveals all its (relevant) regressors, we refer to its *non-random* coefficients as “bias-free components”. Equation (10) being very general can cover a misspecifications-free equation as a special case. If this special case occurs, then the coefficient on any regressor of the misspecifications-free equation is the causal effect of the regressor on the dependent variable. This definition of causal effects makes sense because of the misspecifications-free condition.

Importantly, we no longer assume that for  $h = 1, \dots, j - 1, j + 1, \dots, M_j$ , the time-varying coefficient of  $y_{th}^*$  is equal to the partial derivative of  $y_{ij}^*$  with respect to  $y_{th}^*$  because, as noted earlier, the ratio of a change in an endogenous variable to a change in another endogenous variable is meaningless without first determining what caused the change in the denominator variable (see Greene (2012) [4] (p. 320)).

<sup>10</sup> For ease of comparison of the derivation in this section with that in the previous section, we do not change the notation  $x_{tk}^*$  to  $y_{tk}^*$ .

2.2.2. Unique Coefficients and Error Term

**Assumption 4.** In each of several simultaneous-econometric equations, the included regressors act partly as “stand-in” variables for each of its omitted regressors.

Under this assumption, the following theorem is true.

**Theorem 2.** If the error term of a simultaneous-econometric equation with time-varying coefficients is made up of certain “sufficient sets” of relevant regressors omitted from the equation, then the coefficients and error term of the equation are unique.

**Proof.** Assumption 4 implies that for  $\ell = 1, \dots, L_j$ :

$$w_{t\ell}^* = \lambda_{t0\ell}^* + \sum_{\substack{h=1 \\ h \neq j}}^{M_j} \lambda_{th\ell}^* y_{th}^* + \sum_{k=1}^{K_j} \varphi_{tik\ell}^* x_{ik}^* \quad (\ell = 1, \dots, L_j) \tag{11}$$

where each omitted regressor ( $w_{t\ell}^*$ ) constituting the error term of (5) is related to the regressors, the  $y_{th}^*$ 's and  $x_{ik}^*$ 's, included in (10). Previously, PS (1984) [1] (p. 13, (3.2b)) used a linear form of Equation (11). Our new practice uses (11) to split each omitted regressor ( $w_{t\ell}^*$ ) constituting the error term of (10) into “a sufficient ( $\lambda_{t0\ell}^*$ ) piece” and “the effect ( $\sum_{\substack{h=1 \\ h \neq j}}^{M_j} \lambda_{th\ell}^* y_{th}^* + \sum_{k=1}^{K_j} \varphi_{tik\ell}^* x_{ik}^*$ ) of all included regressors (the  $y_{th}^*$ 's and  $x_{ik}^*$ 's) on each omitted regressor ( $w_{t\ell}^*$ ) piece”.

The Equations in (11) are most general in the sense that, in contrast to conventional practice, their functional forms are linear in variables and nonlinear in coefficients. Substituting the right-hand side of Equation (11) for  $w_{t\ell}^*$  in (10) gives the following equations: For  $j = 1, \dots, M$ :

$$y_{tj}^* = \alpha_{t0j}^* + \sum_{\ell=1}^{L_j} \omega_{t\ell j}^* \lambda_{t0\ell}^* + \sum_{\substack{h=1 \\ h \neq j}}^{M_j} (\alpha_{thj}^* + \sum_{\ell=1}^{L_j} \omega_{t\ell j}^* \lambda_{th\ell}^*) y_{th}^* + \sum_{k=1}^{K_j} (\beta_{tikj}^* + \sum_{\ell=1}^{L_j} \omega_{t\ell j}^* \varphi_{tik\ell}^*) x_{ik}^* \tag{12}$$

The deterministic equations in (10) and (11) together give the interdependent system (12) of  $M$  equations. Note, these equations are generalizations of a result PS (1984) [1] (p. 13 (3.3a,b)) obtained previously.

**Noteworthy features of Equation (12):** (i) The pieces, the  $\lambda_{t0\ell}^*$ 's, of omitted regressors ( $w_{t\ell}^*$ 's) in conjunction with the included regressors (the  $y_{th}^*$ 's and  $x_{ik}^*$ 's) are at least sufficient to determine the value of  $y_{tj}^*$ . This is the reason why PS (1988) [2] (p. 34) called the  $\lambda_{t0\ell}^*$ 's “sufficient sets” of omitted regressors”. Equation (11) does not miss any relevant sufficient set as long as  $L_j$  is the correct number of all the terms in the last sum on the right-hand side of (10).

**The error term of (12):**  $\sum_{\ell=1}^{L_j} \omega_{t\ell j}^* \lambda_{t0\ell}^*$

(ii) PS (1988) [2] showed that the function  $\sum_{\ell=1}^{L_j} \omega_{t\ell j}^* \lambda_{t0\ell}^*$  of sufficient sets of omitted regressors can be treated as the error term.

(iii) Swamy et al. (2014) [12] (pp. 199,217–219) proved that the coefficients and error term of (12) are unique.<sup>11</sup> Specifically, the coefficients and error term of a model are non-unique or unique according as the error term is made up of regressors omitted from the model, as in (5), or made up of certain “sufficient sets” of such regressors, as in (12). By construction then, the equations in (12) do not have the defects of (5).

(iv) If the error term  $\sum_{\ell=1}^{L_j} \omega_{t\ell j}^* \lambda_{t0\ell}^*$  is treated as random, then according to PS (1988) [2] (p. 34), the included regressors of (12) can be assumed to be independent of the error term or, alternatively, Assumption 1(i) can be replaced by  $E(\sum_{\ell=1}^{L_j} \omega_{t\ell j}^* \lambda_{t0\ell}^* \mid \text{the } y_{th}^* \text{'s and } x_{tk}^* \text{'s in (11)}) = 0$ . The equations in (11) ensure that the included regressors (the  $y_{th}^*$ 's and  $x_{tk}^*$ 's) in (12) are independent of the error term  $(\sum_{\ell=1}^{L_j} \omega_{t\ell j}^* \lambda_{t0\ell}^*)$  (see PS (1988) [2] (p. 34)).

(v) Whereas conventional practice treats the function  $\sum_{\ell=1}^{L_j} \omega_{\ell j} w_{t\ell}$  of all omitted regressors (the  $w_{t\ell}$ 's) as the error term of (5), the new practice treats the function  $\sum_{\ell=1}^{L_j} \omega_{t\ell j}^* \lambda_{t0\ell}^*$  of only pieces or “sufficient sets” ( $\lambda_{t0\ell}^*$ 's) of omitted regressors ( $w_{t\ell}$ 's) as the error term.

**Omitted-regressor biases of the coefficients of (12):**  $\sum_{\ell=1}^{L_j} \omega_{t\ell j}^* \lambda_{th\ell}^*$  and  $\sum_{\ell=1}^{L_j} \omega_{t\ell j}^* \varphi_{tk\ell}^*$

(vi) In the new practice, the piece  $\sum_{h=1}^{M_j} \lambda_{th\ell}^* y_{th}^* + \sum_{k=1}^{K_j} \varphi_{tk\ell}^* x_{tk}^*$  of each omitted regressor ( $w_{t\ell}^*$ ) in  $h \neq j$

(11) contributes to omitted-regressor biases of the coefficients of the included regressors (the  $y_{th}^*$ 's and  $x_{tk}^*$ 's) in (12), meaning that Assumption 3 is satisfied.

(vii) The adjectives “biased” and “unbiased” can only be associated with estimators. Since the coefficients of (10) are not estimators, the coefficients of (12) containing omitted-regressor biases cannot be said to be biased. Q.E.D.

**Corollary 3.** (i) Any model with only endogenous regressors and with time-varying coefficients can be expressed as a model with unique coefficients and error term; (ii) All these endogenous regressors can be independent of certain “sufficient sets” of regressors omitted from the model.

**Proof.** Equation (12), featuring unique coefficients and a unique error term, expresses model (10) with time-varying coefficients and without exogenous regressors. Equations (11) assure that all the endogenous regressors of (12) can be independent of the sufficient sets ( $\lambda_{t0\ell}^*$ 's) of omitted relevant regressors (see PS (1988) [2] (p. 34)). Q.E.D.

A failure to accept (12) dooms econometricians to estimating models with non-unique coefficients and error terms, leading to their inconsistent estimation.

**Measurement errors:**  $y_{tj}^* = y_{tj} - v_{tj}^*$ ,  $j = 1, \dots, M$ ,  $y_{th}^* = y_{th} - v_{th}^*$ ,  $h = 1, \dots, j - 1, j + 1, \dots, M_j$ ,  $x_{tk}^* = x_{tk} - v_{tk}^*$ ,  $k = 1, \dots, K$ , where the variables without an asterisk are observed and ( $v_{tj}^*$ ,  $v_{th}^*$ ,  $v_{tk}^*$ ) with different  $j$ ,  $h$ , and  $k$  are measurement errors.

<sup>11</sup> This result arises as a direct consequence of (11).

Inserting measurement errors at the appropriate places in model (12) gives a model that can be expressed in terms of observed variables as

$$y_{tj} = \gamma_{t0j} + \sum_{\substack{h=1 \\ h \neq j}}^{M_j} \gamma_{thj} y_{th} + \sum_{k=1}^{K_j} \eta_{tkj} x_{tk} \quad (j = 1, \dots, M) \tag{13}$$

where  $\gamma_{t0j} = v_{tj}^* + \alpha_{t0j}^* + \sum_{\ell=1}^{L_j} \omega_{t\ell j}^* \lambda_{t0\ell}^*$ ,  $\gamma_{thj} = (\alpha_{thj}^* + \sum_{\ell=1}^{L_j} \omega_{t\ell j}^* \lambda_{th\ell}^*) (1 - \frac{v_{th}^*}{y_{th}})$ , and  $\eta_{tkj} = (\beta_{tkj}^* + \sum_{\ell=1}^{L_j} \omega_{t\ell j}^* \varphi_{tk\ell}^*) (1 - \frac{v_{tk}^*}{x_{tk}})$ .<sup>12</sup>

**Measurement-error biases:** The formulas  $(\alpha_{thj}^* + \sum_{\ell=1}^{L_j} \omega_{t\ell j}^* \lambda_{th\ell}^*) (-\frac{v_{th}^*}{y_{th}})$  and  $(\beta_{tkj}^* + \sum_{\ell=1}^{L_j} \omega_{t\ell j}^* \varphi_{tk\ell}^*) (-\frac{v_{tk}^*}{x_{tk}})$  measure measurement-error biases of  $\gamma_{thj}$  and  $\eta_{tkj}$ , respectively. To simplify the method of estimating the bias-free components ( $\alpha_{thj}^*$ 's and  $\beta_{tkj}^*$ 's), we treat the proportions,  $\frac{v_{th}^*}{y_{th}}$ ,  $h = 1, \dots, j - 1, j + 1, \dots, M_j$ , and  $\frac{v_{tk}^*}{x_{tk}}$ ,  $k = 1, \dots, K_j$ , of measurement errors as unknown deterministic values.

**Components of the coefficients of model (13):** The intercept,  $\gamma_{t0j} = v_{tj}^* + \alpha_{t0j}^* + \sum_{\ell=1}^{L_j} \omega_{t\ell j}^* \lambda_{t0\ell}^* =$  measurement error in the dependent variable ( $y_{tj}$ ) + the intercept of very general equation (10) + the error term of (12); the coefficients of the non-constant regressors,  $\gamma_{thj} = (\alpha_{thj}^* + \sum_{\ell=1}^{L_j} \omega_{t\ell j}^* \lambda_{th\ell}^*) (1 - \frac{v_{th}^*}{y_{th}}) =$  bias-free component + omitted-regressor biases + measurement-error biases, and  $\eta_{tkj} = (\beta_{tkj}^* + \sum_{\ell=1}^{L_j} \omega_{t\ell j}^* \varphi_{tk\ell}^*) (1 - \frac{v_{tk}^*}{x_{tk}}) =$  bias-free component + omitted-regressor biases + measurement-error biases.

The above labeling hopefully helps explain what the components of the coefficients of (13) are and how they arise.

### 2.2.3. Comparison of Conventional and New Practices

In this section, all references to any one of Equations (1)–(9) involve conventional econometric practice, while all references to any one of Equations (10)–(13) relate to our new practice. The normalization rule is the same in both (5) and (10). Swamy et al. (2016) [10] (p. 9) proved that, unlike model (5), model (13) is free of four major specification errors. While conventional practice routinely ignores omitted-regressor and measurement-error biases, the new practice incorporates them into the coefficients of the included regressors of (13). As a consequence, the coefficients and error term of (5) are not unique, and those of (12) are unique. As noted, conventional practice routinely adopts the exogeneity Assumption 1(i); but the presence of non-unique coefficients and error terms in models in (5) renders this assumption invalid. As PS (1988) [2] (p. 34) required, all of certain “sufficient sets” of omitted regressors in model (11) are defined in a way that makes them unobservable as well as unobserved. The included regressors (the  $y_{th}^*$ 's and  $x_{tk}^*$ 's) in (12) can be independent of the error term  $\sum_{\ell=1}^{L_j} \omega_{t\ell j}^* \lambda_{t0\ell}^*$  which is a function of certain “sufficient sets” of omitted regressors, as PS (1988) [2] (p. 34) pointed out. A result due to PS (1988) [2] (p. 34) is that the assumption—routinely made in conventional practice—that the included regressors ( $x_{tk}^*$ 's) are independent of “the” omitted

<sup>12</sup> The  $\gamma$ 's in (13) should not be confused with those in (8).

regressors ( $w_{t\ell}$ 's) in (8) is meaningless. Our new practice, which assumes that Equations (10)–(12) are deterministic, does not rely on such meaningless assumptions.

Now consider re-writing the EE example (6) from Section 2.1.1 in the form of (13),

$$\text{earnings} = \gamma_{t0j} + \gamma_{t1j} \text{education} \tag{14}$$

where  $M_j = 1$  and  $K_j = 0$ , the coefficients  $\gamma_{t0j}$  and  $\gamma_{t1j}$  have three components each, as in (13), and the causal effect of education on earnings is the bias-free component of  $\gamma_{t1j}$  times the true value of education, even though education in this equation is treated as endogenous (see Swamy et al. (2016) [10]). Note particularly that Equation (14) has all the good properties of (13) in that it embodies causal implications that (6) cannot have.

### 2.3. Estimation

Our proposed methodology posits as objects of estimation the bias-free components of the coefficients of (13). This task requires accurate separation of the estimates of the bias-free components from those of the corresponding omitted-regressor and measurement-error biases. In this section, we show how this separation can be accomplished. In conventional practice, the structural parameters of model (5) are assumed not to contain any biases and are estimated from a sample drawn from the  $M$ -dimensional distribution of endogenous variables, given  $K$  exogenous variables. This distribution is inherently misspecified, since the so-called  $K$  exogenous variables are not strictly exogenous if the coefficients as well as the error term of (5) are not unique (see Theorem 1).

#### 2.3.1. Parameterization of Model (13)

We assume that for  $h = 0, 1, \dots, j - 1, j + 1, \dots, M_j$ :

$$\gamma_{thj} = \pi_{0hj} + \pi_{1hj}z_{t1j} + \dots + \pi_{phj}z_{tpj} + \varepsilon_{thj} \tag{15}$$

and for  $k = 1, \dots, K_j$ :

$$\eta_{tkj} = \beta_{0kj} + \beta_{1kj}z_{t1j} + \dots + \beta_{pkj}z_{tpj} + \zeta_{tkj} \tag{16}$$

where Equation (15) for  $h = 0$  implies that the second term on the right-hand side of (12) is distributed with nonzero mean, zero restrictions on  $\pi$ 's and  $\beta$ 's can be imposed if they are appropriate, and the  $z$ 's are called "the coefficient drivers" satisfying the following condition:

**Admissibility Condition:** For  $j = 1, \dots, M$ , the vector  $\mathbf{Z}_{tj} = (1, Z_{t1j}, \dots, Z_{tpj})'$  in Equations (15) and (16) is an admissible vector of coefficient drivers if, given  $\mathbf{Z}_{tj}$ , the value that the coefficient vector of (13) would take at time  $t$ , had  $\mathbf{Y}_{t,-j} = (1, Y_{t1}, \dots, Y_{t,j-1}, Y_{t,j+1}, \dots, Y_{tM_j})'$  and  $\mathbf{X}_{tj} = (X_{t1}, \dots, X_{tK_j})'$  been  $\mathbf{y}_{t,-j} = (1, y_{t1}, \dots, y_{t,j-1}, y_{t,j+1}, \dots, y_{tM_j})'$  and  $\mathbf{x}_{tj} = (x_{t1}, \dots, x_{tK_j})'$  is independent of  $\mathbf{Y}_{t,-j}$  and  $\mathbf{X}_{tj}$  for all  $t$ , respectively.<sup>13</sup>

The purpose of Equations (15) and (16) is to decompose the coefficients of (13) into their respective parts, necessary for estimation of bias-free components of the coefficients of (13), as shown below. A further condition is that the ranges of the coefficient drivers in (15) (or (16)) should be the same as that of the dependent variable of (15) (or (16)). It is important to stress here that the bias-free components ( $\alpha_{thj}^*, \beta_{tkj}^*$ ) of the coefficients of (13) will have theoretically correct signs and magnitudes only if one

accounts for omitted-regressor and measurement-error biases  $(\sum_{\ell=1}^{L_j} \omega_{t\ell j}^* \lambda_{th\ell}^*, (\alpha_{thj}^* + \sum_{\ell=1}^{L_j} \omega_{t\ell j}^* \lambda_{th\ell}^*) (-\frac{v_{th}^*}{y_{th}}),$

<sup>13</sup> Pearl (2000) [13] (p. 99) elaborated on this condition.

$\sum_{\ell=1}^{L_j} \omega_{t\ell j}^* \varphi_{tk\ell}^* (\beta_{tkj}^* + \sum_{\ell=1}^{L_j} \omega_{t\ell j}^* \varphi_{tk\ell}^* (-\frac{v_{tk}^*}{x_{tk}}))$  separately. To explain, note that the coefficients of (13) have three components each, where the components of the intercept are different from those of the coefficients on non-constant regressors. The coefficient on each non-constant regressor consists of (i) a bias-free component; (ii) an omitted-regressor bias; and (iii) a measurement-error bias. Of these, only (i) and (ii) are the additive components of  $\gamma_{thj}$  and  $\eta_{tkj}$  for all  $h \neq 0$  and all  $k$ . The theoretical signs of the values of bias-free components may be known a priori from economic theory. However, theory will not generally instruct us about the signs of omitted-regressor and measurement-error biases. Therefore, the signs of the coefficients, being functions of (i), (ii), and (iii), will generally not be known a priori, and estimates of their bias-free components will have correct theoretical signs and magnitudes only, if they are separated accurately from those of the corresponding omitted-regressor and measurement-error biases. For example, estimates of the causal effect of education on earnings in model (14) are accurate with correct sign and magnitude only if the bias components of  $\gamma_{t1j}$  are removed completely from it. Likewise, published estimates of own- and cross-price elasticities of the demand for goods and services or the demand for liquid assets using model (5) are very likely incorrect in sign and magnitude because they are based on Assumption 2, which we have shown to be false, rather than on Assumption 3, which is true.

### 2.3.2. Choice of Dependent Variable and Regressors to be Included in (13) and Choice of Coefficient Drivers to Be Included in (15) and (16)

In our proposed methodology, the coefficients of model (13) are the sources of the error terms of Equations (15) and (16). Note that the error term of (12) is absorbed into  $\gamma_{t0j}$  appearing in (13). Equation (15) for  $h = 0$  implies that  $\gamma_{t0j}$  is random with a nonzero mean. This is a reasonable assumption. The choice of dependent variable and regressors to be included in (13) is entirely dictated by the bias-free components one wants to learn. For example, in the EE relationship in (14), the variable “earnings” is its dependent variable, and the variable “education” is its non-constant regressor because we want to learn about the bias-free component of the coefficient  $\gamma_{t1j}$ . After choosing the dependent variable and a set of non-constant regressors on this basis, we can insert them into (13) and thus complete its specification.

As far as possible the coefficient drivers in (15) (or (16)) should be selected in such a way that some of them are strongly related to (and has the same range and variation as) the bias-free component, and the rest of them are strongly related to the omitted-regressor bias component of the dependent variable of Equation (15) (or (16)).<sup>14</sup> The choice of coefficient drivers in (15) and (16) is best explained in terms of a specific example, for which we again resort to EE relationship (14). Greene (2012) [4] (p. 14) presented various arguments justifying the inclusion of additional variables such as age, age square, number of children, the husband’s age, the husband’s education, family income, etc., as separate regressors in a constant-coefficient version of the relationship between earnings and education. In contrast, Swamy et al. (2016) [3] included these additional variables as *coefficient drivers* and *not as separate regressors (or explanatory variables)*, as is common practice when studying what are theoretically bivariate relationships. In their methodology, Swamy et al. (2016) [3] do not merely include such additional variables but they also study the interactions between them and education as separate regressors in a constant-coefficient version of the EE relationship, an approach that we believe is preferable to Greene’s (2012) [4] (pp. 14, 15, 708) conventions described above. Based on our preferred model (13), we use Greene’s proposed additional explanatory variables not as separate regressors but as coefficient

<sup>14</sup> This procedure is different from that of PS (1988) [2] (p. 49). Their method is to search like a non-Bayesian for concomitants that absorb “proxy effects” for omitted regressors. Section 4.2 of their paper shows how they use the concomitants they found.

drivers in (15) and (16).<sup>15</sup> Given that (13) but not (5) should be estimated, the choice of appropriate coefficient drivers for (15) and (16) is a must. If all econometricians use Equations (13), (15), and (16), then given (13), there can be a consensus about what coefficient drivers one should include in (15) and (16). In any case, no one should use false models like (5).

In various disciplines, models that look like our model (13), (15), and (16) are labeled “hierarchical,” “mixed,” “random parameter,” or “random effects”. However, because such models are not derived from (10)–(12), they suffer from the same defects previously enumerated for model (5) and do not therefore possess unique coefficients and error terms.

### 2.3.3. Identification

In his book, Greene (2012) [4] (p. 322) provides two examples, noting the standard definition of observational equivalence that if more than one theory is consistent with the same “data”, then the theories are observationally equivalent and cannot be distinguished on the basis of those data alone. In the first example, observational equivalence arises from extreme multicollinearity among the regressors of a model, a problem he eliminates by using some exclusion restrictions. We may do the same if this problem occurs in (13). In his second example, the problem is that of an under-identified model (see Greene (2012) [4] (p. 322)). The problem of identification arises because the probability limit of the least squares estimator of a coefficient is a mixture of all the parameters in the model where both the dependent variable and the non-constant regressor are measured with error. Greene (2012) [4] (p. 241)) points out that in this case, bringing in outside information may provide identification. Here, we follow this procedure in evaluating estimators (21) and (22) given below.

The models in (13) are identified when the coefficients of different models or different coefficients of the same model are made the functions of different coefficient drivers. This is a counterexample to the conventional demonstration that equations with all endogenous regressors are not identifiable.

### 2.3.4. Vector Formulation of Equations (13), (15) and (16)

We use the following vector notation:  $j = 1, \dots, M$ ,  $\mathbf{y}_j = (y_{1j}, \dots, y_{Tj})'$  is the  $T \times 1$  vector of observations on the dependent variable of (13),  $h = 0, 1, \dots, j - 1, j + 1, \dots, M_j$ ,  $\mathbf{y}_{t,-j} = (1, y_{t1}, \dots, y_{t,j-1}, y_{t,j+1}, \dots, y_{tM_j})'$  is the  $M_j \times 1$  vector,  $\boldsymbol{\gamma}_{tj} = (\gamma_{t0j}, \gamma_{t1j}, \dots, \gamma_{t,j-1,j}, \gamma_{t,j+1,j}, \dots, \gamma_{tM_jj})'$  is the  $M_j \times 1$  vector,  $k = 1, \dots, K_j$ ,  $\mathbf{x}_{tj} = (x_{t1}, \dots, x_{tK_j})'$  is the  $K_j \times 1$  vector, and  $\boldsymbol{\eta}_{tj} = (\eta_{t1j}, \dots, \eta_{tK_jj})'$  is the  $K_j \times 1$  vector. Using these notations, (13) can be written as

$$y_{tj} = \mathbf{y}'_{t,-j} \boldsymbol{\gamma}_{tj} + \mathbf{x}'_{tj} \boldsymbol{\eta}_{tj} \tag{17}$$

Another set of vector and matrix notations we use is  $\mathbf{z}_{tj} = (1, z_{t1j}, \dots, z_{tpj})'$  is the  $(p + 1) \times 1$  vector of coefficient drivers,  $\boldsymbol{\pi}_{hj} = (\pi_{0hj}, \pi_{1hj}, \dots, \pi_{phj})'$  is the  $(p + 1) \times 1$  vector of fixed coefficients,  $\Pi_1$  is the  $M_j \times (p + 1)$  matrix having  $\boldsymbol{\pi}'_{hj}$  as its  $h$ th row vector,  $\boldsymbol{\beta}_{kj} = (\beta_{0kj}, \beta_{1kj}, \dots, \beta_{pkj})'$  is the  $(p + 1) \times 1$  vector of fixed coefficients,  $B$  is the  $K_j \times (p + 1)$  matrix having  $\boldsymbol{\beta}'_{kj}$  as its  $k$ th row,  $\boldsymbol{\varepsilon}_{tj} = (\varepsilon_{t0j}, \varepsilon_{t1j}, \dots, \varepsilon_{t,j-1j}, \varepsilon_{t,j+1j}, \dots, \varepsilon_{tM_jj})'$  is the  $M_j \times 1$  vector of errors in Equation (13),

---

<sup>15</sup> The rationale for these coefficient drivers is: (i) If we do not make the coefficients of the EE relationship functions of age, then the relationship neglects the fact that most people have higher incomes when they are older than when they are young, regardless of their education. Thus, without the coefficient driver “Age” or without the interaction term between education and age, the coefficient will overstate the marginal effect of education on earnings; (ii) It is often observed that income tends to rise less rapidly in the latter earning years than in the early years. To accommodate this possibility, we enter the square of age to the list of coefficient drivers; (iii) In addition, previous empirical work of ours has shown that the husband’s education and family income are strongly related to the bias-free component and that the other coefficient drivers are strongly related to the omitted-regressor bias component of  $\gamma_{t1j}$ .

$\zeta_{tj} = (\zeta_{t1j}, \dots, \zeta_{tKj})'$  is the  $K_j \times 1$  vector of errors in Equation (16);  $\gamma_{tj} = \Pi_1 z_{tj} + \varepsilon_{tj}$ ; and  $\eta_{tj} = B z_{tj} + \zeta_{tj}$ . Using these notations, Equation (17) can be written as

$$\begin{aligned} y_{tj} &= \mathbf{y}'_{t,-j} \Pi_1 z_{tj} + \mathbf{x}'_{tj} B z_{tj} + \mathbf{y}'_{t,-j} \varepsilon_{tj} + \mathbf{x}'_{tj} \zeta_{tj} \\ &= (\mathbf{y}'_{t,-j}, \mathbf{x}'_{tj}) \begin{pmatrix} \Pi_1 \\ B \end{pmatrix} z_{tj} + (\mathbf{y}'_{t,-j}, \mathbf{x}'_{tj}) \begin{pmatrix} \varepsilon_{tj} \\ \zeta_{tj} \end{pmatrix} \\ &= \mathbf{y}'_{xt,-j} \Pi_B z_{tj} + \mathbf{y}'_{xt,-j} \varepsilon_{\zeta tj} \end{aligned} \tag{18}$$

where  $\mathbf{y}'_{xt,-j} = (\mathbf{y}'_{t,-j}, \mathbf{x}'_{tj})$ ,  $\Pi_B = \begin{pmatrix} \Pi_1 \\ B \end{pmatrix}$ , and  $\varepsilon_{\zeta tj} = \begin{pmatrix} \varepsilon_{tj} \\ \zeta_{tj} \end{pmatrix}$ .

Since  $\mathbf{y}_{t,-j}$  and  $\mathbf{x}_{tj}$  are not the sources of the errors in (15) and (16), we can assume the following:

**Assumption 5.** For all  $t$  and  $j$ , given  $z_{tj}$ ,  $\mathbf{y}_{xt,-j}$  is conditionally independent of  $\varepsilon_{\zeta tj}$ .

**Assumption 6.** For all  $t$  and  $j$ , given  $z_{tj}$ , the  $\varepsilon_{\zeta tj}$ ,  $t = 1, \dots, T$ , are serially independent with  $E(\varepsilon_{\zeta tj} \mid z_{tj}) = 0$  and  $E(\varepsilon_{\zeta tj} \varepsilon'_{\zeta tj} \mid z_{tj}) = \sigma^2 \Delta$ .

Under Assumptions 5 and 6, we apply an iteratively rescaled generalized least squares (IRSGLS) method to (18) to obtain the estimators of  $\Pi_B$  and  $\sigma^2 \Delta$ . The second-order properties of these estimators are thoroughly studied by Cavanagh and Rothenberg (1995) [14]. Under certain conditions these IRSGLS estimators of  $\Pi_B$  and  $\sigma^2 \Delta$  are consistent.

The IRSGLS method also gives the empirical best linear unbiased predictors of  $\varepsilon_{tj}$  and  $\zeta_{tj}$ . Inserting the observations on  $z$ 's, the predictions of  $\varepsilon_{thj}$  and  $\zeta_{tkj}$ , and the IRSGLS estimates of  $\pi$ 's and  $\beta$ 's in (15) and (16), respectively, gives the predictions of the coefficients of (13).

### 2.3.5. Estimation of the Bias-Free Components of the Coefficients of (13)

To prevent the differences in the functional forms of  $\gamma_{thj}$  in (13) and (15) and of  $\eta_{tkj}$  in (13) and (16) from introducing inconsistencies into our analysis, we consider

$$\gamma_{thj} = \pi_{0hj} + \pi_{1hj} z_{t1j} + \dots + \pi_{phj} z_{tpj} + \varepsilon_{thj} = (\alpha^*_{thj} + \sum_{\ell=1}^{L_j} \omega^*_{t\ell j} \lambda^*_{th\ell}) (1 - \frac{v^*_{th}}{y_{th}}) \tag{19}$$

and

$$\eta_{tkj} = \beta_{0kj} + \beta_{1kj} z_{t1j} + \dots + \beta_{pkj} z_{tpj} + \zeta_{tkj} = (\beta^*_{tkj} + \sum_{\ell=1}^{L_j} \omega^*_{t\ell j} \varphi^*_{tk\ell}) (1 - \frac{v^*_{tk}}{x_{tk}}) \tag{20}$$

Using Equation (19) gives the estimator ( $\hat{\alpha}^*_{thj}$ ) of the bias-free component of  $\gamma_{thj}$  as

$$\left[ 1 - \frac{\hat{v}^*_{th}}{y_{th}} \right]^{-1} (\hat{\pi}_{0hj} + \sum_{s \in G_1} \hat{\pi}_{shj} z_{tsj}) \tag{21}$$

where  $\frac{\hat{v}^*_{th}}{y_{th}}$  is an assumed value of the proportion  $\frac{v^*_{th}}{y_{th}}$ , the  $\hat{\pi}$ 's are the IRSGLS estimates of the  $\pi$ 's in Equation (19), and  $G_1$  is a subset of the  $z$ 's which we believe is appropriate to estimate  $\alpha^*_{thj}$  in (19).

Similarly, using Equation (20) gives the estimator ( $\hat{\beta}^*_{tkj}$ ) of the bias-free component of  $\eta_{tkj}$  as

$$\left[ 1 - \frac{\hat{v}^*_{tk}}{x_{tk}} \right]^{-1} (\hat{\beta}_{0kj} + \sum_{s \in G_2} \hat{\beta}_{skj} z_{tsj}) \tag{22}$$

where  $\frac{\hat{v}^*_{tk}}{x_{tk}}$  is an assumed value of the proportion ( $\frac{v^*_{tk}}{x_{tk}}$ ), the  $\hat{\beta}$ 's are the IRSGLS estimates of the  $\beta$ 's in Equation (20), and  $G_2$  is a subset of the  $z$ 's which we believe is appropriate to estimate  $\beta^*_{tkj}$  in (20). An application of formulas (21) and (22) is given in Swamy et al. (2016) [3].

These formulas are not pure sample estimators because they involve (prior) non-sample values: the numbers  $G_1$  and  $G_2$ ,  $G_1$  and  $G_2$  coefficient drivers, and the assumed values  $\frac{\hat{v}_{ih}^*}{y_{ih}}$  and  $\frac{\hat{v}_{ik}^*}{x_{ik}}$  of the proportions of measurement errors  $\frac{v_{ih}^*}{y_{ih}}$  and  $\frac{v_{ik}^*}{x_{ik}}$ , respectively. We call them “the (prior) non-sample values” because the sample data on the variables,  $y_{tj}$ ,  $y_{xt,-j}$ , and  $z_{tj}$  do not contain any information on them. Therefore, the accuracy of estimates given by  $\hat{\alpha}_{thj}^*$  and  $\hat{\beta}_{tkj}^*$  depends not only on the accuracy of the sample estimates of  $\pi$ 's and  $\beta$ 's but also on our ability to obtain accurate prior information on the non-sample values.<sup>16</sup> This approach would have been objectionable had econometricians never used any non-sample (prior) information. In conventional practice, the crucial issue is econometricians' ability to deduce the values of structural parameters uniquely from sample information in terms of sample moments coupled with non-sample information such as restrictions on parameter values (see Greene (2012) [4] (p. 326)).

### 3. Conclusions

We distinguish between conventional and new practices in econometrics and show that the latter yield different and, in our view, better results than the former. After defining uniqueness of the coefficients and error terms of models, we show that conventional practices are handicapped by a focus on models with necessarily non-unique coefficients and error terms. We prove further that such coefficients do not possess consistent estimators. In contrast, our new practice employs very general models featuring time-varying and unique coefficients and error terms. By construction, these models are free from four major specification errors cited in the body of the paper. Since certain non-sample (prior) information besides sample information is needed to estimate these models, we show how such non-sample information can be obtained and used. Finally, given the importance of empirical validation of our theory, we plan to offer some applications using real-world data in the near future.

**Acknowledgments:** We thank Fredj Jawadi and four referees for their very helpful comments.

**Author Contributions:** All authors contributed equally to the paper.

**Conflicts of Interest:** The authors declare no conflict of interest.

### References

1. Pratt, J.W.; Schlaifer, R. On the Nature and Discovery of Structure (with discussion). *J. Am. Stat. Assoc.* **1984**, *79*, 9–21. [[CrossRef](#)]
2. Pratt, J.W.; Schlaifer, R. On the Interpretation and Observation of Laws. *J. Econom.* **1988**, *39*, 23–52. [[CrossRef](#)]
3. Swamy, P.A.V.B.; Chang, I.; Mehta, J.S.; Greene, W.H.; Hall, S.G.; Tavlas, G.S. Removing Specification Errors from the Usual Formulation of Binary Choice Models. *Econometrics* **2016**, *4*, 26. [[CrossRef](#)]
4. Greene, W.H. *Econometric Analysis*, 7th ed.; Pearson, Prentice Hall: Upper Saddle River, NJ, USA, 2012.
5. Rao, C.R. *Linear Statistical Inference and Its Applications*, 2nd ed.; John Wiley & Sons: New York, NY, USA, 1973.
6. Engle, R.D.; Hendry, D.; Richard, J. Exogeneity. *Econometrica* **1983**, *51*, 277–304. [[CrossRef](#)]
7. Friedman, M.; Schwartz, A.J. Alternative Approaches to Analyzing Economic Data. *Am. Econ. Rev.* **1991**, *81*, 39–49.
8. Lehmann, E.L.; Casella, G. *Theory of Point Estimation*, 2nd ed.; Springer: New York, NY, USA, 1998.
9. Kreuger, A.; Dale, S. *Estimating the Payoff to Attending a More Selective College*; Working Paper 7322; National Bureau of Economic Research (NBER): Cambridge, MA, USA, 1999.

<sup>16</sup> The non-sample (prior) values in estimators (21) and (22) can change from user to user and the bias-free components  $\alpha_{thj}^*$  and  $\beta_{tkj}^*$  are not always constants. It is very hard to study the large sample properties of such estimators. Bayesian methods also cannot be used to estimate the  $\alpha_{thj}^*$ 's and  $\beta_{tkj}^*$ 's because in any Bayesian analysis, it is the knowledge about fixed and unknown parameters that Bayesians model as random and the  $\alpha_{thj}^*$ 's and  $\beta_{tkj}^*$ 's are unknown but may not be fixed. PS (1988) [2] (p. 49), the Bayesian statisticians, did not really recommend Bayesian analysis of laws but said that “a Bayesian will do much better to search like a non-Bayesian for concomitants that absorb . . . [‘proxy effects’ for excluded variables]”. We would use this sentence with “concomitants” replaced by “coefficient drivers.”

10. Swamy, P.A.V.B.; Hall, H.G.; Tavlas, G.S.; Chang, I.; Gibson, H.D.; Greene, W.H.; Mehta, J.S. A Method for Measuring Treatment Effects on the Treated without Randomization. *Econometrics* **2016**, *4*, 19. [[CrossRef](#)]
11. Swamy, P.A.V.B.; Mehta, J.S.; Tavlas, G.S.; Hall, S.G. Two Applications of the Random Coefficient Procedure: Correcting for Misspecifications in a Small Area Level Model and Resolving Simpson's Paradox. *Econ. Model.* **2015**, *45*, 93–98. [[CrossRef](#)]
12. Swamy, P.A.V.B.; Mehta, J.S.; Tavlas, G.S.; Hall, S.G. Small Area Estimation with Correctly Specified Linking Models. In *Recent Advances in Estimating Nonlinear Models, with Applications in Economics and Finance*; Ma, J., Wohar, M., Eds.; Springer: New York, NY, USA, 2014; pp. 193–228.
13. Pearl, J. *Causality*; Cambridge University Press: Cambridge, UK, 2000.
14. Cavanagh, C.L.; Rothenberg, T.J. Generalized Least Squares with Nonnormal Errors. In *Advances in Econometrics and Quantitative Econometrics*; Maddala, G.S., Phillips, P.C.B., Srinivasan, T.N., Eds.; Blackwell Publishers, Inc.: Cambridge, MA, USA, 1995; pp. 276–290.



© 2017 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).