Two Applications of the Random Coefficient Procedure: Correcting for Misspecifications in a Small Area Level Model and Resolving Simpson's Paradox[†]

P.A.V.B. Swamy[a], J.S. Mehta[b], G.S. Tavlas[c] and S.G. Hall[d]

Abstract

We apply a random-coefficient framework to deal with two problems frequently encountered in applied work. First, we use a real-world relationship to derive from it a sub-relationship among fewer variables without introducing a single specification error to correct misspecifications in a small area level model. Second, we use this framework to resolve Simpson's paradox. We show that this paradox does not arise if a statistical relationship between a pair of variables is derived from the corresponding real-world relationship involving all relevant variables including the original pair without introducing a single specification error.

Key words: Specification error, False model, Unique coefficient, Unique error term

---

[†] The views expressed in this paper are the authors' own and do not necessarily represent those of their respective institutions.
institutions.

[a] Federal Reserve Board (Retired), 6333 Brocketts Crossing, Kingstowne, VA 22315, e-mail: swamyparavastu@hotmail.com

[b] Department of Mathematics, Temple University, Philadelphia, PA 19122, e-mail: mehta1007@comcast.net

[c] Economic Research Department, Bank of Greece, 21 El. Venizelos Ave. 102 50, Athens, Greece, e-mail: gtavlas@bankofgreece.gr

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

1. Introduction

Empirical models are subject to several types of specification errors, including those that arise from incorrect functional forms, omission of relevant regressors, measurement errors, and incorrect relationships between included and excluded regressors. In this paper we show how to correct such specification errors within a random-coefficient framework. Specifically, we apply this framework to Fay and Herriot's (FH) (1979) small area level model and use our approach to resolve Simpson's paradox.

Small area level model. In sample surveys, it can be the case that for sub-groups, the data available for those groups -- or domains -- are sparse and, as a result, direct estimates of the relevant parameters based on these data are not reliable.[1] Further problems arise if these data are also corrupted by nonsampling errors, such as nonresponse and measurement errors. Any sample estimator can be expressed as its estimand plus the sum of its sampling and nonsampling errors. A statistical model of this estimand is called "a linking model." If this model is not seriously misspecified, it can be used to improve the precision of the sample estimator. A necessary condition (discussed below) for a linking model to be free from all specification errors has been provided by Swamy, Mehta, Tavlas and Hall (2014). If this necessary condition is not satisfied, the empirical model is misspecified or false because it cannot have unique coefficients and error terms. The intuition underlying this result is that misspecified models will have incorrect functional forms and arbitrary coefficients and error terms. We use the Fay and Herriot (FH) (1979) model to illustrate our methods of correcting linking models for their specification errors.

Simpson's paradox. Simpson's paradox is the (paradoxical) observation in which two variables appear to have, say, a positive relationship towards one another when, in fact, that

[1] A domain estimator is referred to as "direct" if it is based only on the domain-specific sample data.

relationship is reversed or disappears after a third variable is brought into the analysis.[2] This paper presents a method of resolving Simpson's paradox by correcting the relationship for its specification errors.

The remainder of this paper is divided into three sections. Section 2 first presents a theorem giving a necessary condition for a model to be free from all specification errors. Then, the section first corrects FH's model to satisfy this condition and then shows that the same corrections resolve Simpson's paradox. Section 3 concludes.

2. False Linking Models, their Corrections, and Resolution of Simpson's Paradox

2.1 False linking models

In this section we explain a necessary condition for a model not to be false. By "false", we mean that the model is not free from all specification errors. Under the heading: 'specification errors,' we include (i) incorrect functional forms, (ii) arbitrary error terms, (iii) incorrect relationships between included and excluded regressors (or covariates), (iv) omission of relevant regressors, and (v) imperfect measurements.

Swamy, Mehta, Tavlas, Hall (2014) proved the following:

Theorem: A necessary condition for a (mathematical or statistical) model to be completely free from all specification errors is that its coefficients and error term, having the correct functional forms, are unique (henceforth, SMTH's theorem; see Swamy, Mehta, Tavlas and Hall 2014)[3]. We call a model that does not contain these specification errors "the original real-world

[2] For recent discussions of Simpson's paradox, see Christensen (2014), Keli Liu and Xiao-Li Meng (2014), and Pearl (2000, 2014).

[3] Swamy, Mehta, Tavlas and Hall did not formalize the conditions for this theorem, although the essential elements of the theorem were stated in that paper. In this paper, we formalize the theorem.

relationship"[4]. That is, by controlling for all relevant pre-existing conditions, we can help ensure that this relationship does not lead to false relationships among different sub-sets of its variables. [5] Specifically, by not omitting any of relevant regressors or any of relevant pre-existing conditions from the original real-world relationship, that relationship does not need any error term to represent omitted variables. A "shorter real-world relationship" is the original real-world relationship containing fewer regressors than the original real-world relationship and which has the error term. The included regressors are those that are included in both original and shorter real-world relationships. Omitted regressors are those that are included in the original but not in the shorter real-world relationship. We apply these definitions below.

The coefficients and error term of shorter real-world relationship are unique in the sense that they are invariant under those changes that keep the equality signs in the real-world relationships between each omitted regressor and the included regressors unchanged; each of the unique coefficients on nonconstant regressors has the form of the partial derivative of the true value of the dependent variable with respect to the true value of an included regressor plus the corresponding omitted-regressors bias; the unique error term has the form of a function (having the correct functional form) of certain 'sufficient sets' of omitted regressors, a concept due to Pratt and Schlaifer (1988, p. 34); the coefficients cannot be unique unless they have the correct functional forms.

What SMTH's theorem establishes is the following: while the models with nonunique coefficients and error terms are false, a sufficient condition for the truth of a model with unique

[4] We call any relationship which is not misspecified "a real-world relationship" (see Basmann 1988). See also Swamy and Hall (2012). The SMTH theorem holds as a direct consequence of writing a real-world relationship in two forms, one with all relevant arguments including all relevant pre-existing conditions and another with fewer regressors. What these two forms are can be seen from Section 2.3.

[5] Skyrms (1988, p. 59) proved this statement.

coefficients and error term is not known. Using the well-known dictum (attributed to George Box) that "All models are wrong, but some are useful", SMTH's theorem modifies this statement by reducing the set of false models from "the set of all models" to "the set of models with nonunique coefficients and error terms." This reduction is very useful. Intuitively, SMTH's theorem makes a lot of sense. If the error term of a model is nonunique, then the correlations between the error term and the included regressors can be made to appear and disappear at the whim of an arbitrary choice between two equivalent forms of the model (see Swamy, Mehta, Tavlas and Hall, 2014, pp. 217-218). Because of this problem, consistent estimation of the parameters of false models is impossible.

This difficulty is also presented by White's (1980, 1982) assumption that the nonunique error term and the regressors of his nonlinear model are uncorrelated with each other. If the coefficient on a regressor of a model is nonunique, then the effect of the regressor on the dependent variable is not consistently estimable (see Swamy, Mehta, Tavlas and Hall, 2014, pp. 217-218). From this discussion, we can conclude that the models with nonunique coefficients and error terms are not useful.

2.2 Fay and Herriot's model[6]

Let $\hat{\theta}_i = \log \hat{\bar{Y}}_i$ where $\hat{\bar{Y}}_i$ is a direct sample estimator of $\bar{Y}_i$ which is the per capita income (PCI) for the ith small area. FH (1979) explained how they computed $\hat{\bar{Y}}_i$. They write $\hat{\theta}_i = \theta_i + e_i$ where $\theta_i = \log \bar{Y}_i$ and e_i is the sampling error. Suppose that there are *M* areas in the population

[6] The U.S. Bureau of the Census uses FH's model (shown below as equation (2)) to periodically update the estimates of per capita income for areas with population less than 1,000. The U.S. Treasury Department uses these estimates to determine allocations of funds to the local governments within different states under the General Revenue Sharing Program.

and only m ($< M$) areas are selected in the sample. It is assumed that for $i = 1, \ldots, m$, $E_p(e_i \mid \theta_i) = 0$ and $V_p(e_i \mid \theta_i) = \psi_i$.

For $i = 1, \ldots, M$, FH assume a model of the form

$$\theta_i = x_i'\beta + b_i v_i, \tag{1}$$

where $x_i = (x_{1i}, \ldots, x_{pi})'$ is the column vector of p regressors, $\beta = (\beta_1, \ldots, \beta_p)'$ is the column vector of p coefficients, the b_i's are known positive constants, v_i is the model error, FH showed how they obtained these b_i's. They assume that $E_m(v_i) = 0$ and $V_m(v_i) = \sigma_v^2$ (≥ 0) are the model expectation and variance respectively. The regressors in (1) are: $x_1 = 1$, x_2 = log(county PCI), x_3 = log(value of owner-occupied housing for the place), x_4 = log(value of owner-occupied housing for the county), x_5 = log(adjusted gross income per exemption from the 1969 tax returns for the place), x_6 = log(adjusted gross income per exemption from the 1969 tax returns for the county), $p = 6$. FH assume that the sample areas obey the population model in (1). Equation (1) is FH's linking model. Substituting the sum of two terms on the right-hand side of equation (1) for θ_i in $\hat{\theta}_i = \theta_i + e_i$ gives the basic area level model

$$\hat{\theta}_i = x_i'\beta + b_i v_i + e_i \tag{2}$$

It is customary to assume that the sampling variances $V_p(e_i \mid \theta_i) = \psi_i$ are known because they are not identifiable. This means that model (2) is also not identifiable.

The possible misspecifications of (1) are: (i) its log-linear functional form is incorrect, (ii) its coefficients and error term are not unique, (iii) the measurements on its regressors are not perfect, (iv) $\hat{\theta}_i$ is corrupted by nonsampling errors, and (v) the selection of the covariates of (1) is not based on appropriate economic theories.

2.3 Corrections to FH's model

To correct (1) for its misspecifications, we need the following real-world relationship:

$$\bar{Y}_i^* = f_i(x_{1i}^*, x_{gi}^*, g = 2,...,G, x_{\ell i}^*, \ell = G+1,...,L) = f_i(.) \tag{3}$$

where the variables with an asterisk are the true values containing no measurement errors, the function $f_i(.)$ has unknown functional form and has three types of arguments: (i) x_{1i}^*, an included determinant of $\bar{Y}_i^*$, (ii) the $x_{gi}^*, g = 2,...,G,$ are labeled, "omitted determinants" of $\bar{Y}_i^*$, and the $x_{\ell i}^*, \ell = G+1,...,L$, represent all relevant pre-existing conditions.[7] Note that the variable x_{1i}^* in equation (3) is the true value of $e^{x_{2i}}$ in (1).[8] We do not enter the regressors x_{3i}, x_{4i}, x_{5i}, and x_{6i} of (1) into $f_i(.)$ as its arguments because we are not sure that they are the determinants of $\bar{Y}_i^*$. We use them for some other purpose below.

Assumption I (Normalization rule): The coefficient of the dependent variable $\bar{Y}_i^*$ of equation (3) is set equal to 1 for all *i*.

Equation (3) is called "the original real-world relationship." We justify this label by using the following argument: Since the true functional form of (3) is unknown, the idea that is used here is: Not specifying a particular functional form for $f_i(.)$ amounts to not misspecifying its unknown true functional form. Using this idea we do not specify the functional form of (3).

[7] The reason why we use these specific labels here will be clear as we proceed.

[8] Given that incomes are usually underreported, it is safe to assume that the counties' per capita incomes are measured with error, i.e., $x_{1i} = x_{1i}^* + \varsigma_{1i}^*$ where ς_{1i}^* is a measurement error.

Another problem is that we do not have the complete list of the determinants of $\bar{Y}_i^*$. To stop this ignorance of ours from making us omit any of the relevant arguments of $f_i(.)$, we neither define the regressors $x_{gi}^*, g = 2,...,G,$ nor specify their number *G*. Finally, we do not know what the relevant pre-existing conditions are and what their number is. So we neither define the variables $x_{\ell i}^*, \ell = G+1,...,L$, nor specify their number *L*. In (3), there is no need to include an error term representing omitted determinant of $\bar{Y}_i^*$ and omitted pre-existing conditions because there are no such omitted variables. The function $f_i(.)$ is not stochastic. Basmann (1988, p. 73) pointed out that "causal relations and orderings are unique in the real world and that they remain invariant … [under] mere changes in the language we use to describe them." It should be noted that equation (3) has this property of causal invariance. All these properties of (3) assure that it is a real-world relationship.

Assumption II: The function $f_i(.)$ is differentiable with respect to its arguments.

The form of (3) is not convenient for further analysis. Therefore, without misspecifying its functional form, (3) is written as

$$\bar{Y}_i^* = \frac{\partial \bar{Y}_i^*}{\partial x_{1i}^*} x_{1i}^* + \sum_{j=2}^{L} \frac{\partial \bar{Y}_i^*}{\partial x_{ji}^*} x_{ji}^* + (\bar{Y}_i^* - \frac{\partial \bar{Y}_i^*}{\partial x_{1i}^*} x_{1i}^* - \sum_{j=2}^{L} \frac{\partial \bar{Y}_i^*}{\partial x_{ji}^*} x_{ji}^*) \tag{4a}$$

$$= \alpha_{1i}^* x_{1i}^* + \sum_{j=2}^{L} \alpha_{ji}^* x_{ji}^* + \alpha_{0i}^* \tag{4b}$$

where for j = 1, 2, …, L, $\alpha_{ji}^* = \frac{\partial \bar{Y}_i^*}{\partial x_{ji}^*}$ and $\alpha_{0i}^* = (\bar{Y}_i^* - \sum_{j=1}^{L} \frac{\partial \bar{Y}_i^*}{\partial x_{ji}^*} x_{ji}^*)$ is the intercept. Note that in taking the partial derivatives $\frac{\partial \bar{Y}_i^*}{\partial x_{gi}^*}$, g = 1, …, G, the values of all the determinants of $\bar{Y}_i^*$ including all relevant pre-existing conditions ($x_{\ell i}^*, \ell = G+1,...,L$) but not including x_{gi}^* are held constant. As a consequence, $\frac{\partial \bar{Y}_i^*}{\partial x_{gi}^*}$ = 0 if the relation of $\bar{Y}_i^*$ to x_{gi}^* is false and is equal to a nonzero function otherwise, as shown by Skyrms (1988, p. 59).[9] This explains why we enter $x_{\ell i}^*, \ell = G+1,...,L$, into $f_i(.)$ of (3) as its arguments.

The coefficients of equation (4b) or the partial derivatives in equation (4a) are constants when the function $f_i(.)$ in (3) is linear. In the case where this function is nonlinear, the coefficients of (4b) are variables and are functionally related to its regressors.

In (1), FH follow the usual practice of treating the net effect of omitted regressors on θ_i as the error term. If we follow this practice, then the coefficients (α_{0i}^*, α_{1i}^*) and the error term ($\sum_{j=2}^{L} \frac{\partial \bar{Y}_i^*}{\partial x_{ji}^*} x_{ji}^*$) become nonunique, in which case model (4b) is false by SMTH's theorem (see Swamy, Mehta, Tavlas and Hall 2014). An explanation of this nonuniqueness is that the error term ($\sum_{j=2}^{L} \frac{\partial \bar{Y}_i^*}{\partial x_{ji}^*} x_{ji}^*$) is a function of omitted regressors (x_{ji}^*, j = 2, …, L) which are not unique. Pratt and Schlaifer (1988, p. 34) pointed out that the assumption that the included regressor (i.e., x_{1i}^* in (4b)) is independent of 'the' omitted regressors (i.e., x_{ji}^*, j = 2, …, L, in (4b)) 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

[9] Skyrms (1988, p. 59) pointed out that if any partial derivative, say $\frac{\partial \bar{Y}_i^*}{\partial x_{ji}^*}$, in (4a) becomes zero when the values of all relevant pre-existing conditions are held constant, then this means that the sub-relationship between $\bar{Y}_i^*$ and x_{ji}^* is false. In taking these partial derivatives we do hold the values of all relevant pre-existing conditions constant.

makes them unobservable as well as unobserved."[10] We derive such sufficient sets below. The error term v_i of FH's model in (1) is not a function of 'sufficient sets' of omitted regressors. Therefore, by Pratt and Schlaifer's logic, both FH's assumption that $E_m(v_i|x_i)$ = 0 and White's (1980, 1982) assumption that an arbitrary error term of his nonlinear model is uncorrelated with its regressors are equally "meaningless". To avoid such "meaningless" assumptions, we will replace FH's v_i by a function of certain 'sufficient sets' of the omitted regressors $x^*_{gi}, g = 2,...,G$, derived below. We will also ensure that this function has the correct functional form.

Assumption III: Every one of the omitted regressors (x^*_{ji}, j = 2, …, L) is differentiable with respect to the included regressor (x^*_{1i}).

The fully corrected FH's model with unique coefficients and error term can be obtained by using the following real-world relationships between omitted and included regressors. For j = 2, …, L:

$$x^*_{ji} = \frac{\partial x^*_{ji}}{\partial x^*_{1i}} x^*_{1i} + \left(x^*_{ji} - \frac{\partial x^*_{ji}}{\partial x^*_{1i}} x^*_{1i} \right) \tag{5a}$$

$$= \lambda^*_{1i} x^*_{1i} + \lambda^*_{0i} \tag{5b}$$

where λ^*_{1i} = $\frac{\partial x^*_{ji}}{\partial x^*_{1i}}$ and λ^*_{0i} = $\left(x^*_{ji} - \frac{\partial x^*_{ji}}{\partial x^*_{1i}} x^*_{1i} \right)$. Following Pratt and Schlaifer (1984, 1988), we

[10] Here Pratt and Schlaifer (1988) are not talking about the covariate selection problem. By assuming that (3) is a real-world relationship we have bypassed this problem. It can be solved by using the "ignorability" condition: In (4b), the set X^*_{ji}, j = 2, …, L, is an admissible set of covariates if, given X^*_{ji}, j = 2, …, L, the value that θ_i would take had X^*_{1i} been x^*_{1i} is independent of X^*_{1i} (see Pearl 2000, p. 79). Since this condition is impossible to verify, we cannot determine whether the covariates of (1) are admissible or not. Because of this difficulty econometricians use economic theories to select the regressors for their models.

interpret λ_{0i}^* as the portion of the omitted regressor x_{ji}^* remaining after subtracting the effect ($\lambda_{1i}^* x_{1i}^*$) of the included regressor (x_{1i}^*) on x_{ji}^* from x_{ji}^*.

Substituting the right-hand side of equation (5b) for x_{ji}^* in (4b) gives

$$\bar{Y}_i^* = \alpha_{0i}^* + \sum_{j=2}^{L} \alpha_{ji}^* \lambda_{0i}^* + (\alpha_{1i}^* + \sum_{j=2}^{L} \alpha_{ji}^* \lambda_{1i}^*) x_{1i}^* \tag{6}$$

This is a corrected version of FH's model. Note that we go from (3) to (6) without resorting to any approximation or committing any misspecification. For this reason, equation (6) is called "the original real-world relationship shortened to have only one regressor" x_{1i}^* (or, simply, "the shortened real-world relationship"). The regressor x_{1i}^* is called "the included regressor" because it is included in both (3) and (6); the regressors x_{ji}^*, $j = 2, \ldots, L$, are called "omitted regressors" because they are included in (3) but not in (6).

The coefficients of (6) have the correct functional forms. Furthermore, in conjunction with the included regressor (x_{1i}^*) the portions (λ_{0i}^*'s) of omitted regressors (x_{ji}^*'s) are sufficient to determine the value of $\bar{Y}_i^*$ in (6) exactly. This justifies Pratt and Schlaifer's label "sufficient sets of omitted regressors" for the λ_{0i}^*'s. According to them, the second term ($\sum_{j=2}^{L} \alpha_{ji}^* \lambda_{0i}^*$) on the right-hand side of equation (6) can be treated as the error term. One can easily see how different this error term is from FH's error term $b_i v_i$ in (1).

Swamy, Mehta, Tavlas and Hall (2014, Appendix 1, pp. 217-219) proved that the error term ($\sum_{j=2}^{L} \alpha_{ji}^* \lambda_{0i}^*$) and the coefficient $(\alpha_{1i}^* + \sum_{j=2}^{L} \alpha_{ji}^* \lambda_{1i}^*)$ of x_{1i}^* in (6) are unique in the sense that they are invariant under those changes that keep the equality signs of relationships (5b) unchanged.

The term $\sum_{j=2}^{L}\alpha_{ji}^{*}\lambda_{1i}^{*}$ is the correct measure of omitted-regressors bias and the term α_{1i}^{*} is the bias-free component of the coefficient of x_{1i}^{*} in (6).

It should be noted that equations (4b) and (6) are two different forms of the same real-world relationship in (3). If we are only interested in predicting $\bar{Y}_i^{*}$, then there is no need to eliminate omitted-regressor bias from the coefficient of x_{1i}^{*} because the right-hand side of the equality sign in (6) is exactly equal to its left-hand side.

2.4 Measurement errors

Now, we assume that the dependent variable and the regressor $e^{x_{2i}}$ of (1) contain measurement errors. We insert these errors at the appropriate places in (6) to express model (6) in terms of observed variables. Let x_{1i} (= $e^{x_{2i}}$) in (1) be equal to x_{1i}^{*} in (3) plus a measurement error, denoted by ς_{1i}^{*}. Algebraically, $x_{1i} = x_{1i}^{*} + \varsigma_{1i}^{*}$.[11]

Assumption IV: The variable x_{1i} is continuous.

Inserting x_{1i} - ς_{1i}^{*} for x_{1i}^{*} in (6) gives

$$\hat{\bar{Y}}_i = \gamma_{0i} + \gamma_{1i}x_{1i} \tag{7a}$$

where $\hat{\bar{Y}}_i$ = $\bar{Y}_i^{*}$ + ς_i^{*} + u_i, the ς_i^{*}'s are the sampling errors, the u_i's are nonsampling errors, γ_{0i} = α_{0i}^{*} + $\sum_{j=2}^{L}\alpha_{ji}^{*}\lambda_{0i}^{*}$ + ς_i^{*} + u_i, γ_{1i} = $(\alpha_{1i}^{*} + \sum_{j=2}^{L}\alpha_{ji}^{*}\lambda_{1i}^{*})(1 - \frac{\varsigma_{1i}^{*}}{x_{1i}})$, and $(\alpha_{1i}^{*} + \sum_{j=2}^{L}\alpha_{ji}^{*}\lambda_{1i}^{*})(-\frac{\varsigma_{1i}^{*}}{x_{1i}})$ is the

[11] This x_{1i} should not be confused with x_1 in (1).

right measure of measurement-error bias of γ_{1i}. We call equation (7a) "the fully corrected FH model with measurement errors".[12]

Now it can be seen that the three components of the intercept of (7a) are: (i) the intercept of (4b), (ii) the error term of the shorter real-world relationship in (6), (iii) the sum of sampling and nonsampling errors in the dependent variable of (7a). The components of the coefficient on the nonconstant regressor of (7a) are: (i) the bias-free partial derivative α_{1i}^{*}, (ii) the omitted-regressor bias $(\sum_{j=2}^{L}\alpha_{ji}^{*}\lambda_{1i}^{*})$, and (iii) the measurement-error bias $(\alpha_{1i}^{*}+\sum_{j=2}^{L}\alpha_{ji}^{*}\lambda_{1i}^{*})(-\frac{\varsigma_{1i}^{*}}{x_{1i}})$. Now, we need to estimate the coefficients of (7a) without distorting these interpretations.[13]

To do so, we need to parameterize (7a) to facilitate its estimation. Let z_i = (1, $z_{1i} = e^{x_{3i}},\ldots,z_{4i}=e^{x_{6i}})'$ be a 5 × 1 vector. The elements of this vector are the same as a subvector of the regressors of (1). We write γ_{0i} + $\gamma_{1i}x_{1i}$ in a matrix form as $x_i'\gamma_i$ where x_i = (1, $x_{1i})'$ and γ_i = $(\gamma_{0i},\gamma_{1i})'$. Let $\Pi = \begin{pmatrix}\pi_{00},\pi_{01},\ldots,\pi_{04}\\ \pi_{10},\pi_{11},\ldots,\pi_{14}\end{pmatrix}$ be a 2 × 5 matrix. Using these definitions we write

$$\hat{\bar{Y}}_i = x_i'\Pi z_i + x_i'\varepsilon_i \quad (i = 1, \ldots, m) \tag{7b}$$

where $\gamma_i = \Pi z_i + \varepsilon_i$, $\varepsilon_i = (\varepsilon_{0i},\varepsilon_{1i})'$.

---

[12] The cases where x_{1i} takes the value 0 with positive probability can be handled by adding the term $(\alpha_{1i}^{*}+\sum_{j=2}^{L}\alpha_{ji}^{*}\lambda_{1i}^{*})(-\varsigma_{1i}^{*})$ to γ_{0i} after rewriting the term $(\alpha_{1i}^{*}+\sum_{j=2}^{L}\alpha_{ji}^{*}\lambda_{1i}^{*})(1-\frac{\varsigma_{1i}^{*}}{x_{1i}})$ as $(\alpha_{1i}^{*}+\sum_{j=2}^{L}\alpha_{ji}^{*}\lambda_{1i}^{*})(x_{1i}-\varsigma_{1i}^{*})$ (see Swamy, Mehta, Tavlas and Hall 2014, pp. 198-199).

[13] The coefficients and error terms of the observation equations of state space models are not unique and do not have the same interpretations as the coefficients of (7a). For this reason, the maximum likelihood method of estimating the state space models given in Durbin and Koopman (2001, pp. 30-32) and Kim and Nelson (1999) cannot be used to estimate (7a).

It is assumed that for i, i' = 1, …, m, $E(\varepsilon_i \mid z_i, x_i)$ = 0 and $E(\varepsilon_i \varepsilon_{i'}' \mid z_i, x_i)$ = $\{^{\sigma_\varepsilon^2 \Delta_\varepsilon \text{if } i=i'}_{0 \; i \neq i'}$ where Δ_ε is a 2 × 2 nonnegatuve definite matrix.

The coefficient vector of equation (7a) is γ_i. The elements of the vector z_i in the model γ_i = Πz_i + ε_i are aptly called the coefficient drivers.[14] These drivers explain the variation in the coefficient vector γ_i. In conjunction with this vector, the vector x_i of the regressors of (7a) explains the variation in the dependent variable $\hat{\bar{Y}}_i$. Thus, this dependent variable has two sources of variation whereas the dependent variable of FH's model in (1) has only one source of variation in the regressor vector of (1). The error term ε_i of γ_i = Πz_i + ε_i is part of γ_i. Part of the first element of γ_i is the error term of (6).

The function $e^{x_{2i}}$ of one of the regressors of FH's model appears as the regressor x_{1i} of equation (7a) and the antilog of each of the remaining regressors of FH's model appears as a coefficient driver in γ_i = Πz_i + ε_i. This use introduces the interactions between the antilog of x_{2i} and the antilog of each of the other regressors of FH's model into (7b). Thus, (7b) is richer than FH's model. The reason why we use the antilog of FH's regressor x_{2i} in (1) as the regressor of (7a) and use the antilog of each of FH's other regressors (x_{3i}, x_{4i}, x_{5i}, and x_{6i}) in (1) as the coefficient drivers is that $\hat{\bar{Y}}_i$ is part of the antilog of x_{2i} but not of the antilog of each of FH's other regressors. The motive for introducing the equation γ_i = Πz_i + ε_i is the aim that different combinations of the coefficient drivers would give accurate estimates of the components of the

[14] We only choose observable variables as the coefficient drivers.

coefficients of (7a). If this aim turns out to be false, then FH's regressors of their model can be judged to be inappropriate.

Note that the instrumental variables that are highly correlated with x_i and uncorrelated with $x_i'\varepsilon_i$ cannot exist. Therefore, the method of instrumental variables does not apply to model (7b). This result cannot be avoided if one wants to specify a model without introducing any specification errors. Equations (3), (4b), (5b), (6), (7a) and (7b) always end up with the multiplication of the error vector ε_i of $\gamma_i = \Pi z_i + \varepsilon_i$ by the regressor vector x_i of (7a), as in (7b).

Admissible coefficient drivers: The elements of the vector $Z_i = (1, Z_{1i},...Z_{4i})'$ in $\gamma_i = \Pi z_i + \varepsilon_i$ is an admissible set of coefficient drivers if, given Z_i, the value that the vector of the coefficients of (7a) would take had $X_i = (1, X_{1i})'$ been $x_i = (1, x_{1i})'$ is independent of X_i for all *i*.

Assuming that the elements of ε_i are contemporaneously and serially correlated, Swamy, Tavlas, Hall and Hondroyiannis (2010) developed an iteratively rescaled generalized least squares (IRSGLS) method to estimate the parameters and to predict the ε_i's of (7b). They have also shown that under certain conditions, these estimators are consistent. Next, we use these estimates in Lehmann and Casella's (1998, p. 467, Theorem 5.3) way to obtain asymptotically efficient estimates of the parameters of (7b). Inserting these estimates into (7b) gives

$$\hat{\hat{Y}}_i = x_i'\hat{\Pi}z_i + x_i'\hat{\varepsilon}_i \tag{8a}$$

$$\hat{\gamma}_i = \hat{\Pi}z_i + \hat{\varepsilon}_i \tag{8b}$$

where the symbols with hat are IRSGLS or asymptotically efficient estimates. We partition two-equation system (8b) into

$$\hat{\gamma}_{0i} = \hat{\pi}_0' z_i + \hat{\varepsilon}_{0i} \tag{9}$$

$$\hat{\gamma}_{1i} = \hat{\pi}_1' z_i + \hat{\varepsilon}_{1i} \tag{10}$$

where $\hat{\pi}_0'$ and $\hat{\pi}_1'$ are the first and second rows of $\hat{\Pi}$, respectively.

If the regressors of FH's model are appropriate and adequate, then equations (9) and (10) provide useful information about the components of the coefficients γ_{0i} and γ_{1i} of (7a), respectively. To extract the estimates of the coefficients of (6) from the dependent variables of (9) and (10), we need to subtract an accurate estimate of (i) the sum ($\varsigma_i^* + u_i$) of sampling and nonsampling errors from $\hat{\gamma}_{0i}$ ($\gamma_{0i} = \alpha_{0i}^* + \sum_{j=2}^{L} \alpha_{ji}^* \lambda_{0i}^* + \varsigma_i^* + u_i$) and (ii) the measurement-error bias $[(\alpha_{1i}^* + \sum_{j=2}^{L} \alpha_{ji}^* \lambda_{1i}^*)(-\frac{\varsigma_{1i}^*}{x_{1i}})]$ of γ_{1i} $[=(\alpha_{1i}^* + \sum_{j=2}^{L} \alpha_{ji}^* \lambda_{1i}^*)(1-\frac{\varsigma_{1i}^*}{x_{1i}})]$ from $\hat{\gamma}_{1i}$. For this purpose, we exploit the information contained in (9) and (10). The quantities $\hat{\varepsilon}_{0i}$ and $\hat{\varepsilon}_{1i}$ are treated as parts of the estimates of ($e_i + u_i$) and measurement-error bias of γ_{1i}, respectively.

Since the information contained in $\hat{\varepsilon}_{0i}$ and $\hat{\varepsilon}_{1i}$ is already used up, we can now see what portion of the unused information contained in $\hat{\pi}_0' z_i$ and $\hat{\pi}_1' z_i$ of (9) and (10) can be used to obtain accurate estimates of ($e_i + u_i$) and measurement-error bias of γ_{1i}, respectively. With 5 coefficient drivers, we can have 2^5 - 1 = 31 combinations of them. Each of these combinations can be written as

$$\hat{\Pi}_\tau z_{\tau i} \;\; (\tau = 1, \ldots, n, i = 1, \ldots, m) \tag{11}$$

where $z_{\tau i}$ is a truncated vector of coefficient drivers obtained by deleting from z_i one, two, three, or four elements at a time; and $\hat{\Pi}_\tau$ is a submatrix of $\hat{\Pi}$ having all the columns that correspond to the elements in $\mathrm{z}_{\tau i}$. For each τ, we compute kernel density estimates using $\hat{\Pi}_\tau \mathrm{z}_{\tau i}$, i = 1, …, m. The large sample properties of these estimates are stated in Lehmann (1999, pp. 406-419). By comparing the locations, spreads, modes and medians of these kernel estimates we might be able to pick those kernel density estimates of γ_{0i} and γ_{1i} that have negligible magnitudes of ($e_i + u_i$) and measurement-error bias, respectively.

2.5 Simpson's Paradox

To illustrate the paradox, we have taken the following example from Armistead (2014): When a hypothetical quasi-experiment was conducted for both male and female subjects assigned to a medical treatment, the combined recovery rates for males and females from treatment are 50% for the treated subjects and 40% for the controls (untreated subjects). In the subgroups of males and females, the recovery rates for treated males and untreated males (controls) are 60% and 70% and those for treated females and untreated females (controls) are 20% and 30%, respectively. Thus, in the aggregate (combined) group of men and women, treated subjects fared better than untreated subjects (controls). However, in the subgroups of males and females, the finding is reversed: untreated males (controls) are shown to recover better than treated males and untreated females (controls) are shown to recover better than treated females, respectively. What is paradoxical about these results is that the treatment is good for everyone but bad for males and females when they are treated as subgroups. For this type of result to arise, there needs to be different proportions of men and women in the treated and untreated groups.

In this experiment, Simpson's paradox was observed when gender (a third variable) was introduced into a statistical relationship between a pair of variables (treatment and recovery). Let us now derive a relationship between a pair of variables from a relationship between the same pair of variables involving all relevant third variables without committing a single specification error. We will show below that this derivation does not produce Simpson's paradox. Let us look at our algebra behind this derivation. The question which Simpson's paradox poses is: Do the disaggregated (separate male-female) data invalidate the aggregated (combined male-female) data? To rephrase this question, we treat data on the variables in (4b) as Armistead's (2014) type disaggregated data and data on the variables in (6) as Armistead's type aggregated data. To make a comparison of our work in this paper with Armistead's (2014) analysis, we view (6) as containing aggregated data on $(\bar{Y}_i^*, x_{1i}^*)$ and view (4b) as containing data on $(\bar{Y}_i^*, x_{1i}^*)$ disaggregated by including x_{ji}^*, $j = 2, \ldots, L$. Suppose that (6) is a first-order relationship between $\bar{Y}_i^*$ and x_{1i}^*. The variables x_{ji}^*, $j = 2, \ldots, L$, are used in place of a third variable used in Armistead (2014).

The relevant question is: Whether the third variable, gender, is causal or not.[15] Our answer is: Any third variable is causal if it enters into a real-world relationship of (4b)'s type with nonzero coefficient and is not causal if it enters into (4b)'s type equation with zero coefficient.[16] Sections 2.1-2.4 of this paper show how we carefully examine the third variables x_{ji}^*, $j = 2, \ldots, L$.

Suppose that the correlation between treatment and recovery is not spurious or the relationship between $\bar{Y}_i^*$ and x_{1i}^* is not false. Then the numerical results in Armistead (2014) hold but the result, viz., 50% vs. 40% recovery, contains incorrect functional-form and omitted-regressor biases due to omitting the third variable which is gender. Because of these biases it

[15] This question was raised in the debate published in the February 2014 issue of *The American Statistician*.

[16] Armistead (2014) gives a different answer: "Whether causal or not, third variables can convey critical information about a first-order relationship, study design, and previously unobserved variables. Any conditioning on a non-trivial third variable that produces Simpson's Paradox should be carefully examined before either the aggregated or the disaggregated findings are accepted, regardless of whether the third variable is thought to be causal."

cannot be concluded from the result that the disaggregated data invalidate the aggregated data. Disaggregated equation (4b) does not imply false relationships among different subsets of its variables, since all relevant pre-existing conditions are controlled. It can be seen that in disaggregated equation (4b), the coefficient of x_{1i}^* is α_{1i}^* and in aggregated equation (6), the coefficient of the same x_{1i}^* is $(\alpha_{1i}^* + \sum_{j=2}^{L} \alpha_{ji}^* \lambda_{1i}^*)$. These two coefficients will have the same sign and magnitude if the omitted-regressors bias $\sum_{j=2}^{L} \alpha_{ji}^* \lambda_{1i}^*$ is completely removed from $(\alpha_{1i}^* + \sum_{j=2}^{L} \alpha_{ji}^* \lambda_{1i}^*)$. Thus, disaggregated equation (4b) does not reverse the relationship between the pair of variables $(\bar{Y}_i^*, x_{1i}^*)$ implied by aggregated equation (6). Any sign of $(\alpha_{1i}^* + \sum_{j=2}^{L} \alpha_{ji}^* \lambda_{1i}^*)$ that is the opposite of the sign of α_{1i}^* does not mean that disaggregated data in (4b) reversed the relationship in (6) but mean that the addition of the omitted-regressors bias $\sum_{j=2}^{L} \alpha_{ji}^* \lambda_{1i}^*$ to α_{1i}^* has assigned to the sum $(\alpha_{1i}^* + \sum_{j=2}^{L} \alpha_{ji}^* \lambda_{1i}^*)$ a sign that is the opposite of the sign of α_{1i}^*. Incorrect functional-form and omitted-regressor biases cause wrong signs all the time in econometrics. A comparison of (6) with (4b) is not a comparison of apples and oranges because (6) and (4b) are the two forms of the same real-world relationship in (3).

2.6 Discussion

Equations (4b) and (6) have multiple uses: (i) they show how to eliminate specification errors from any misspecified model. We chose FH's model to illustrate this point. (ii) They can resolve Simpson's paradox. (iii) They show how to derive the formulas for omitted-regressor and measurement-error biases with the correct functional forms. (iv) They show how the unique coefficients and error terms of models look like.

With equations (4b) and (6) we could resolve Simpson's paradox. Causality is the property of such real-world relationships as (3). The derivation of disaggregated relationship (4b) from real-world relationship (3) and that of aggregated relationship (6) from disaggregated relationship (4b) avoid all specification errors as well as Simpson's paradox. Model (6) satisfies the necessary condition of SMTH's theorem.

2.7 Prediction

In actual practice, false models with nonunique coefficients and error terms are used for generating predictions under some arbitrary error distributions. A desirable practice would be to use equation (7b) for prediction. If this is done, then the sources of forecast errors are: (i) $\hat{\gamma}_{0i}$ - γ_{0i}, (ii) ($e_i + u_i$) – estimate of ($e_i + u_i$), (iii) ($\hat{\gamma}_{1i}$ - γ_{1i}), (iv) $(\alpha_{1i}^* + \sum_{j=2}^{L}\alpha_{ji}^*\lambda_{1i}^*)(-\frac{\varsigma_{1i}^*}{x_{1i}})$ – estimate of $[(\alpha_{1i}^* + \sum_{j=2}^{L}\alpha_{ji}^*\lambda_{1i}^*)(-\frac{\varsigma_{1i}^*}{x_{1i}})]$, and (v) measurement errors in the x_{1i}. It is very difficult to eliminate any of these sources. Some of the components of forecast errors coming from these sources can be very large.

3. Conclusions

We have dealt with the following issues.

(1) Models with nonunique coefficients and error terms are not free from specification errors. FH's misspecified model or any other misspecified model can be modified so that it has unique coefficients and error term. Such modified models are proven to be free from all relevant specification errors. After these modifications have been made in FH's model,

accurate forecasts of its dependent variable can be generated from it if the sum of sampling and nonsampling errors and the measurement-error bias component can be completely removed from its intercept and slope coefficient, respectively.

(2) We use the random coefficient framework to resolve Simpson's paradox. We show that this paradox does not arise if a statistical relationship between a pair of variables is derived from the corresponding real-world relationship involving all relevant variables including the original pair without introducing a single specification error.

References

Armistead, T.W. (2014), Resurrecting the Third Variable: A Critique of Pearl's Causal Analysis of Simpson's Paradox (with Discussion), *The American Statistician*, 68, 1-7, 30-31.

Basmann, R.L. (1988), Causality Tests and Observationally Equivalent Representations of Econometric Models, *Journal of Econometrics*, 39, 69-104.

Brown, L.D. (1990), An Ancillarity Paradox Which Appears in Multiple Linear Regression, *The Annals of Statistics*, 18, 471-538.

Christensen, R. (2014), Comment, *The American Statistician*, 68, 13-17.

Durbin, J. and S.J. Koopman (2001), *Time Series Analysis By State Space Methods*, Oxford: Oxford University Press.

Fay, R.E. and R.A. Herriot (1979), Estimation of Income from Small Places: An Application of James-Stein Procedures to Census Data, *Journal of the American Statistical Association*, 74, 269-277.

Hall, S.G., P.A.V.B. Swamy and G.S. Tavlas (2012a), Generalized Cointegration: A New Concept with an Application to Health Expenditure and Health Outcomes, *Empirical Economics*, 42, 603-618.

Hall, S.G., P.A.V.B. Swamy and G.S. Tavlas (2012b), Milton Friedman, the Demand for Money and the ECB's Monetary-Policy Strategy, *Federal Reserve Bank of St. Louis Review*, 94, 153-185.

Hall, S.G., A. Kenjegaliev, P.A.V.B. Swamy and G.S. Tavlas (2013a), Measuring Currency Pressures: The Cases of the Japanese Yen, the Chinese Yuan, and the UK Pound, *Journal of The Japanese and International Economies*, 29, 1-20.

Hall, S.G., A. Kenjegaliev, P.A.V.B. Swamy, G.S. Tavlas (2013b), The Forward Rate Premium Puzzle: A Case of Misspecification?, *Studies in Nonlinear Dynamics and Econometrics*, 17, 239-343.

Hall, S.G., G. Hondroyiannis, A. Kenjegaliev, P.A.V.B. Swamy and G.S. Tavlas (2013), Is the Relationship Between Prices and Exchange Rates Homogeneous?, *Journal of International Money and Finance*, 37, 411-438.

Keli Liu and Xiao-Li Meng (2014), Comment: A Fruitful Resolution to Simpson's Paradox via Multiresolution Inference, *The American Statistician*, 68, 17-29.

Kim, C. and C.R. Nelson (1999), *State-space Models with Regime Switching*, Cambridge, Massachusetts: The MIT Press.

Lehmann, E.L. (1999), *Elements of Large Sample Theory*, New York: Springer.

Lehmann, E.L. and G. Casella (1998), *Theory of Point Estimation*, Second Edition, New York: Springer.

Pratt, J.W. and B. Schlaifer (1984), On the Nature and Discovery of Structure, *Journal of the American Statistical Association*,79, 9-21, 29-33.

Pratt, J.W. and B. Schlaifer (1988), On the Interpretation and Observation of Laws, *Journal of Econometrics*, 39, 23-52.

Pearl, J. (2000), Causality: Models, Reasoning, and Inference, Second Edition, Cambridge: Cambridge University Press.

Pearl, J. (2014), Comment: Understanding Simpson's Paradox, *The American Statistician*, 68, 8-13.

Skyrms, B. (1988), Probability and Causation, *Journal of Econometrics*, 39, 53-68.

Swamy, P.A.V.B., and S.G.F. Hall (2012), Measurement of Causal Effects, *Economic Change and Restructuring*, 45, 3-23.

Swamy, P.A.V.B., G.S. Tavlas, S.G.F. Hall, and G. Hondroyiannis (2010), Estimation of Parameters in the Presence of Model Misspecification and Measurement Error, *Studies in Nonlinear Dynamics & Econometrics*, 14,

Swamy, P.A.V.B., J.S. Mehta, G.S. Tavlas, and S.G. Hall (2014), Small Area Estimation with Correctly Specified Linking Models, in J. Ma and M. Wohar (Eds.), *Recent Advances in Estimating Nonlinear Models: With Applications in Economics and Finance*, New York: Springer.

Swamy, P.A.V.B., G.S. Tavlas and S.G. Hall (2014), Microproduction Functions with Unique Coefficients and Errors: Reconsideration and Respecification, Macroeconomic Dynamics, forthcoming.

White, H. (1980), Using Least Squares to Approximate Unknown Regression Functions, *International Economic Review*, 21, 149-170.

White, H. (1982), Maximum Likelihood Estimation of Misspecified Models, *Econometrica*, 50, 1-25.