

New Improved Small Area Models

P. A. V. B. Swamy¹, Tamara S. Zimmerman², Jatinder S. Mehta³ and Edwin L. Robison^{4*}

Abstract

This paper provides the conditions under which the precisions of direct domain estimators produced within the standard design-based mode of inference used in survey analysis are improved by combining the direct estimators with matched statistical models.

1. Introduction

Estimates for small domains are produced in response to policy makers' demands. Domains may be defined in terms of geographic areas, socio-demographic areas, industry groups, or other subpopulations. An estimator for a domain in a time period is referred to as "direct" if it is based on the domain- and time-specific sample data. A domain is considered as small if the domain sample size is not large enough to yield direct estimates of adequate precision. Reliable estimates for small domains can only be produced by moving away from the design-based estimation of conventional direct estimates to indirect model-dependent estimates. The essence of model-dependent approach is the use of auxiliary data available at the small area level. These data are used to construct predictor variables for use in a regression model that can be used to predict the variable of interest for all small areas. The problem here is that if this model is seriously misspecified, then inferences based on the model can be much worse than design-based inferences. The effectiveness of indirect model-dependent estimation depends on the availability of methods that can limit the effects of model misspecifications. Such methods are presented in this paper.

We consider an area level model to supplement small domain direct estimates with area-specific auxiliary data. This model has two equations: the first says that for each domain in a time period, a direct survey estimator is the sum of the true value and a design-induced error. The second relates the true value to area-specific auxiliary variables. Following J.N.K. Rao (2003, p. 77), we call the first equation "the sampling model" and the second equation "the linking model." To analyze the specification problems associated with the linking model, we begin with the conceptual idea that there is the "true" model linking the "true" variables involving the "true" coefficients. We then show that the linking model is an exact representation of the "true" model if its coefficients are correctly interpreted. In the absence of the conditions implied by the correct interpretations, the linking model is misspecified. In this paper, we prove that not all linking models that yield design-consistent estimators limit the effects of model misspecifications in small samples. To do so, we present two linking models, one of which is misspecified and the other of which is correctly specified. Both these models yield design-consistent estimators, but the predictions from the misspecified model are affected by the misspecifications in small samples.

In the next section, we exploit the connection between a linking model and the underlying "true" model. By so doing, we are able to derive explicit algebraic expressions for the effects of model misspecifications and to develop a method of analyzing such effects. The implications of our discussion are summarized in the final section.

2. Basic Area Level Model

¹Bureau of Labor Statistics, U.S. Department of Labor, 2 Massachusetts Avenue, NE, Washington, DC 20212/ paravastu_s@bls.gov

²Bureau of Labor Statistics, U.S. Department of Labor, 2 Massachusetts Avenue, NE, Washington, DC 20212/
Zimmerman_Tamara@bls.gov

³Temple University, Philadelphia, PA 19122/ mehta1007@comcast.net

⁴Bureau of Labor Statistics, U.S. Department of Labor, 2 Massachusetts Avenue, NE, Washington, DC 20212/ Robison_E@bls.gov

* Any opinions expressed in this paper are those of the authors and do not constitute policy of the Bureau of Labor Statistics or the Bureau of the Census. We thank Professor Roderick Little for giving us a chance to discuss with him some of the issues raised in this paper.

We assume throughout that the available observations on variables are the sums of “true” values and sampling and/or nonsampling errors. In what follows, symbols with an asterisk denote “true” values and symbols with a hat or without either a hat or an asterisk denote observable variables measured with error. Let \hat{y}_{it} be a direct survey estimator of y_{it}^* , which, for domain i and time t , denotes the true value of a population characteristic.

2.1 Sampling Model

We assume that \hat{y}_{it} is design-unbiased (or p-unbiased) for y_{it}^* , in symbols, $E_p(\hat{y}_{it}) = y_{it}^*$. The design variance of \hat{y}_{it} is denoted by $V_p(\hat{y}_{it})$. We assume further that

$$\hat{y}_{it} = y_{it}^* + e_{it}, \quad i = 1, 2, \dots, n, \quad t = 1, 2, \dots, T_i \quad (1)$$

where the sampling errors e_{it} are independent with

$$E_p(e_{it} | y_{it}^*) = 0, \quad V_p(e_{it} | y_{it}^*) = \psi_{it}. \quad (2)$$

These assumptions may be quite restrictive in some applications. For example, some sample designs imply a nonzero autocorrelation structure for the sampling errors. Also, the absolute size of the sampling error changes over time because of redesigns, sample size changes, and variation in the levels of certain population characteristics. When modeling the sampling error, it is important to account for these types of sampling design features. We make use of the procedures suggested in the literature to estimate the variance ψ_{it} and the parameters of the stochastic process e_{it} may follow (see J.N.K. Rao (2003, p. 166)). If the direct survey estimates given by \hat{y}_{it} contain nonsampling error, then we show below how we might quantify this error.

2.2 Derivation and Properties of a Linking Model

In superpopulation modeling, the population values, y_{it}^* , are assumed to be a random sample from an infinite “superpopulation” and are assigned a probability distribution. To derive this distribution, we adopt the following definition:

Definition 1 (Swamy and Tavlas, 2001) *Any variable or value that is not mismeasured is true and any economic relationship with the correct functional form, without any omitted explanatory variables and without mismeasured variables, is true.*

Accordingly, if the “true” relationship among y_{it}^* and a set of observable area-specific auxiliary variables, denoted by $x_{1it}, \dots, x_{K-1,it}$, exists, then it is expressible in the form:

$$y_{it}^* = \alpha_{0it} + \sum_{j=1}^{K-1} \alpha_{jit} x_{jit}^* + \sum_{g=K}^{m_{it}} \alpha_{git} x_{git}^* \quad (3)$$

where all the determinants (both observed and unobserved) of y_{it}^* are included on the right-hand side, even though we may know nothing about some of the unobserved determinants. In other words, there are no excluded explanatory variables in equation (3). To avoid the possibility of excluding from (3) any determinants of y_{it}^* at any time for any domain, we assume that the number of the determinants of y_{it}^* may change across i at a point in time and through time. Hence m_{it} is domain-specific and time-dependent.

Freedman and Navidi (1986, p. 7) criticized the modeling assumptions that had no foundation in theory or in fact. To make sure that (3) has such a foundation, we use only the determinants of y_{it}^* suggested by relevant economic theories. Equation (3) avoids all restrictions *not* implied by these theories. In this connection, it can be thought of as the relationship implied by economic theories, in which case, we can assume that a mechanism exists through which the right-hand side variables in (3) exactly determine y_{it}^* . If we exclude from (3) the variables, say x_{git}^* , $g = K, \dots, m_{it}$, then an explanation of the relationship between y_{it}^* and x_{jit}^* , $j = 1, \dots, K-1$, can be found in the dependence of both y_{it}^* and x_{jit}^* , $j = 1, \dots, K-1$, on some of excluded variables, x_{git}^* , $g = K, \dots, m_{it}$, a phenomenon known as *spurious correlation* (see Lehmann and Casella (1998, p. 107)). Thus, by virtue of it's including all relevant explanatory variables in the right way, the formulation in (3) avoids all possible spurious correlations.

Although the relationship in (3) corresponds to that suggested by economic theories, typically, its correct functional form is unknown. Consequently, any specific assumption concerning its functional form may be incorrect.

Solution to the Unknown-Functional-Form Problem Equation (3) is linear if $\alpha_{jit}, j = 1, \dots, K-1$, and $\alpha_{git}, g = K, \dots, m_{it}$, are constants; otherwise, the equation is nonlinear. Different time profiles of its coefficients assign different functional forms to (3). Restrictions on the pattern of variation in (e.g., the constancy of) its coefficients may force (3) to have an incorrect functional form. Any finite class of functional forms may not cover the “true” functional form of (3) as a special case. In our state of ignorance about the “true” functional form of (3), permitting all of its coefficients to differ among domains both at a point in time and through time gives an infinite class of functional forms that encompasses its “true” functional form.

We make use of this solution by allowing all the coefficients of (3) to vary freely. The coefficients of (3) with the “true” time profiles are called ‘the “true” coefficients.’ They are denoted by $\alpha_{jit}^*, j = 0, 1, \dots, K-1$, and $\alpha_{git}^*, g = K, \dots, m_{it}$, the existence of which is assumed here. This assumption is equivalent to the assumption that the “true” functional form of (3) exists.

Therefore, the “true” model that is a member of the class in (3) is

$$y_{it}^* = \alpha_{0it}^* + \sum_{j=1}^{K-1} \alpha_{jit}^* x_{jit}^* + \sum_{g=K}^{m_{it}} \alpha_{git}^* x_{git}^* . \quad (4)$$

This equation satisfies Definition 1. Equation (4) differs from (3) in that the former equation includes the “true” values of both variables and coefficients, whereas the class of models represented by (3) includes only the “true” values of variables.

The values of y_{it}^* that are defined for the unrealized values of $x_{jit}^*, j = 1, \dots, K-1$, and $x_{git}^*, g = K, \dots, m_{it}$, will not be realized and Pratt and Schlaifer (1988, p. 28) call them *potential* values. The only values of y_{it}^* that are determined by the realized values of $x_{jit}^*, j = 1, \dots, K-1$, and $x_{git}^*, g = K, \dots, m_{it}$ will be realized. It is the existence of these potential values that guarantees the existence of the “true” model. The “true” model cannot be a real-world relation unless it exists. Basmann (1988, p. 99) shows that causation is a real-world relation between events. Consequently, if the potential values of y_{it}^* do not exist, any version of (4) fitted to observations is a pure statistical artifact.

Unfortunately, we cannot estimate the “true” model because it has certain unknown determinants. For example, data on some of the determinants of y_{it}^* are not available and even the data we have on the other determinants of y_{it}^* contain errors. Suppose that data on $x_{jit}^*, j = 1, \dots, K-1$, are available and data on $x_{git}^*, g = K, \dots, m_{it}$, are not available. The observed measurements, $x_{jit} = x_{jit}^* + v_{jit}, j = 1, \dots, K-1$, are the sums of “true” values, x_{jit}^* , and measurement errors, v_{jit} . We now show that excluding $x_{git}^*, g = K, \dots, m_{it}$, from the “true” model or mismeasuring $x_{jit}^*, j = 1, \dots, K-1$, introduces biases into $\alpha_{jit}^*, j = 0, 1, \dots, K-1$.

A model involving only the observable counterparts of the first $K-1$ explanatory variables of the “true” model can be written as

$$y_{it}^* = \gamma_{0it} + \sum_{j=1}^{K-1} \gamma_{jit} x_{jit} \quad (5)$$

We call this model “the time-varying coefficient (TVC) model.” Substituting the right-hand side of this equation for y_{it}^* in (1) gives a model with \hat{y}_{it} as its dependent variable and the design-induced errors, e_{it} , as its errors.

A Classification of the Explanatory Variables of the “True” Model The explanatory variables, $x_{jit}, j = 1, \dots, K-1$, are called the included explanatory variables because they are included in the TVC model. The variables, $x_{git}^*, g = K, \dots, m_{it}$, are called excluded variables because they are excluded from the TVC model.

Connections between the Included and Excluded Explanatory Variables Pratt and Schlaifer (1988, p. 34) show that the condition that the included explanatory variables be (mean) independent of ‘the’ excluded variables themselves “is

meaningless unless the definite article is deleted and can then be satisfied only for certain ‘sufficient sets’ of excluded variables some if not all of which must be defined in a way that makes them unobservable as well as unobserved.” (p. 34) (see also Pratt and Schlaifer (1984, pp. 11-13)). From this result it follows that a meaningful assumption is

$$x_{git}^* = \lambda_{0git}^* + \sum_{j=1}^{K-1} \lambda_{jgit}^* x_{jit}^*, g = K, \dots, m_{it} \quad (6)$$

where the coefficients with the correct time profiles are denoted by $\lambda_{jgit}^*, j = 0, 1, \dots, K-1$.

Mapping between the Coefficients of the “True” Model and the Coefficients of the TVC Model Substituting $x_{jit} - v_{jit}$ for x_{jit}^* in the equation that results when the right-hand side of equation (6) is substituted for x_{git}^* in the “true” model, we obtain

$$\gamma_{0it} = \alpha_{0it}^* + \sum_{g=K}^{m_{it}} \alpha_{git}^* \lambda_{0git}^* \text{ and } \gamma_{jit} = (\alpha_{jit}^* + \sum_{g=K}^{m_{it}} \alpha_{git}^* \lambda_{jgit}^*) (1 - \frac{v_{jit}}{x_{jit}}) \quad (j = 1, \dots, K-1) \quad (7)$$

Correct Interpretations of the Coefficients of the TVC Model The first equation in (7) can be interpreted as implying that the intercept, γ_{0it} , of the TVC model is the sum of (i) the intercept (α_{0it}^*) of the “true” model, (ii) the joint effect ($\sum_{g=K}^{m_{it}} \alpha_{git}^* \lambda_{0git}^*$) on the dependent variable (y_{it}^*) of the portions of the “true” values, $x_{git}^*, g = K, \dots, m_{it}$, of excluded variables remaining after the effects of the “true” values, $x_{jit}^*, j = 1, \dots, K-1$, of the included explanatory variables have been removed. The last $K-1$ equations in (7) can be interpreted as implying that for $j = 1, \dots, K-1$, γ_{jit} is the sum of (i) the coefficient α_{jit}^* on x_{jit}^* of the “true” model, (ii) a term ($\sum_{g=K}^{m_{it}} \alpha_{git}^* \lambda_{jgit}^*$) capturing omitted-variables bias due to excluded variables, and (iii) a measurement-error bias, $-(\alpha_{jit}^* + \sum_{g=K}^{m_{it}} \alpha_{git}^* \lambda_{jgit}^*) (v_{jit} / x_{jit})$, due to mismeasuring the included explanatory variable x_{jit} (see Chang, Swamy, Hallahan and Tavlás (2000), Swamy and Tavlás (2001) and Swamy, Chang, Mehta and Tavlás (2003)).

The omitted-variable biases are zero if the included explanatory variables are uncorrelated with every excluded variable and the measurement-error biases are zero if the included explanatory variables are measured without error. These conditions are rarely, if ever, satisfied. Thus, Freedman and Navidi’s (1986, pp. 6 and 7) point about omitted variables and measurement errors causing serious bias in the estimates of y_{it}^* is correct. The coefficients of the TVC model are called “the biased coefficients” because they contain omitted-variable and measurement-error biases. The coefficients of the “true” model are called “the bias-free coefficients” since they are not subject to any biases. The component α_{jit}^* is called “the bias-free component” of γ_{jit} . The relationship between \hat{y}_{it} and x_{jit} is spurious if the bias-free component of γ_{jit} is zero for all t . If an estimate of the coefficient on x_{jit} in the regression of \hat{y}_{it} on $x_{1it}, \dots, x_{K-1,it}$ has a wrong sign, then omitted-variables and measurement-error bias contained in the coefficient and the incorrect functional-form of the regression must have caused this wrong sign. The significance (or insignificance) of an estimate of γ_{jit} is not a good indicator of the significance (or insignificance) of the implied estimate of α_{jit}^* . To avoid this difficulty, we suggest below a method of decomposing an efficient estimator of γ_{jit} into the estimators of its components in (7).

An Important Result Under the decomposition of its coefficients in (7), the TVC model is an exact representation of the “true” model because the right-hand sides of equations (4) and (5) are exactly equal to each other when the coefficients of (5) satisfy the equations in (7).

One question that remains to be answered is that of parameterization: which features of the TVC model ought to be treated as constant parameters? To answer this question, we need the following implications:

Implications of the Correct Interpretations of the Coefficients of the TVC Model (i) Variations in its components lead to variations in γ_{jit} . It can be seen from (7) that the real-world sources of variations in its components are: (a) the nonlinearities

of the “true” model resulting in variations in the α^* ’s, (b) the nonlinearities of the relationships among the “true” values of excluded and included explanatory variables resulting in variations in the λ^* ’s, (c) variation in the ratio of v_{jit} to x_{jit} , and (d) changes in m_{it} . (ii) The measurement-error bias component of γ_{jit} is a function of both x_{jit} and v_{jit} , implying that the included explanatory variable x_{jit} is correlated with its own coefficient γ_{jit} (i.e., in the TVC model, the included explanatory variables are correlated with their own coefficients). (iii) For $j = 0, 1, \dots, K-1$, the γ_{jit} ’s cannot be uncorrelated with each other because the coefficients, α_{git}^* , $g = K, \dots, m_{it}$, on excluded variables are their common determinants.

In our work, these implications are the prime considerations guiding the selection of parameters. We decompose each coefficient of the TVC model into its components shown in (7) by assuming that γ_{jit} is linearly related to some observable variables, denoted by z_{dit} , $d = 1, \dots, p-1$, plus a linear combination of stochastic errors.

$$\gamma_{jit} = \pi_{j0} + \sum_{d=1}^{p-1} \pi_{jd} z_{dit} + \mu_{ji} + \sum_{h=0}^{q-1} l_{jh} \varepsilon_{hit} \quad (8)$$

where none of the z_{dit} is equal to 1 for all i and t , the l_{jh} ’s are known positive constants, and the μ_{ji} ’s and ε_{hit} ’s are random variables. We give the z ’s the new name, “the coefficient drivers.” The TVC model is called “the stochastic coefficients (SC) model” if its coefficients follow assumption (8). This assumption was made previously in Swamy and Mehta (1975), Swamy and Tinsley (1980), and Swamy, Chang, Mehta, and Tavlás (2003). We further assume that for $j > 0$, the sum of p_1 ($< p$) specific terms in $\pi_{j0} + \sum_{d=1}^{p-1} \pi_{jd} z_{dit}$ is equal to the bias-free component, α_{jit}^* , of γ_{jit} and the sum of the remaining $p - p_1$ terms and $\mu_{ji} + \sum_{h=0}^{q-1} l_{jh} \varepsilon_{hit}$ is equal to the sum of omitted-variables and measurement-error bias components of γ_{jit} (see Swamy, Tavlás and Chang (2005)). From this assumption it follows that only those bias-free coefficients of the “true” model that are also the components of the coefficients of the TVC model are identifiable--subject to the restrictions implied by (8)--on the basis of the available data, whereas the bias-free coefficients on excluded variables are not identifiable. The only way in which the bias-free coefficient on an excluded variable is identifiable is through converting the excluded variable into an included variable. Assumption (8) does not contradict the implications of the correct interpretations of the γ_{jit} if (i) the function, $\pi_{j0} + \sum_{d=1}^{p-1} \pi_{jd} z_{dit}$, completely accounts for the correlation between x_{jit} and γ_{jit} so that the remainder, $\mu_{ji} + \sum_{h=0}^{q-1} l_{jh} \varepsilon_{hit}$, obtained by subtracting the function from γ_{jit} is independent of x_{jit} and (ii) the right-hand side of equation (8) is expressible as the sum of two sums, one of which is equal to the bias-free component of γ_{jit} and the other of which is equal to the sum of omitted-variables and measurement-error bias components of γ_{jit} . The satisfaction of these conditions should underpin the selection of coefficient drivers.

Linking Model Substituting the right-hand side of equation (8) for γ_{jit} in the TVC model gives the reduced-form model:

$$y_{it}^* = \pi_{00} + \sum_{d=1}^{p-1} \pi_{0d} z_{dit} + \sum_{j=1}^{K-1} (\pi_{j0} + \sum_{d=1}^{p-1} \pi_{jd} z_{dit}) x_{jit} + \mu_{0i} + \sum_{h=0}^{q-1} l_{0h} \varepsilon_{hit} + \sum_{j=1}^{K-1} (\mu_{ji} + \sum_{h=0}^{q-1} l_{jh} \varepsilon_{hit}) x_{jit} \quad (9)$$

This is our assumed linking model. It is a nonlinear regression model with heteroscedastic (and possibly serially correlated) error terms and coincides with the “true” model in (4) if the decompositions of the coefficients of the TVC model in (8) coincide with those in (7). Substituting the right-hand side of equation (9) for y_{it}^* in (1) gives an area level model, an analysis of which leads to a combination of design-based and model-based weighting. The virtue of this model-based weighting is that it has been derived from the “true” model using very weak parametric assumptions. Estimation of the area level model can be done using an Iteratively Re-Scaled Generalized Least Squares (IRSGLS) method of Swamy, Chang, Mehta and Tavlás (2003).

Generalized Linear Mixed Model All the area level models that are covered in J.N.K. Rao’s (2003) survey are the special cases of the following general linear mixed (GLM) model:

$$y_{it}^* = \beta_0 + \sum_{j=1}^{K-1} \beta_j x_{jit} + b_i v_i + u_{it} \quad (10)$$

where the b_i 's are known positive constants, the v_i 's are domain-specific random effects, and the u_{it} 's are area-by-time specific effects (see J.N.K. Rao (2003, p. 83)). The dependent variable of model (10) becomes observable if (10) is inserted into (1). Such a model involves both design-induced and model errors. Estimation of a GLM model with an observable dependent variable is discussed in Lehmann and Casella (1998, p. 518).

2.3 Comparison of the “True”, TVC, SC, GLM, and Hierarchical Bayes Models

(i) Rewriting (3) in terms of $x_{jit}^*, j = 1, \dots, K-1$, and a function of $x_{jit}^*, j = 1, \dots, K-1$, and $x_{git}^*, g = K, \dots, m_{it}$, leaves the coefficients of the TVC model invariant (see Swamy, Mehta and Singamsetti (1996)). The coefficients of (3), however, do not possess this invariance property. Consequently, in (3), excluded explanatory variables, $x_{git}^*, g = K, \dots, m_{it}$, and the coefficients on the included explanatory variables are not unique, as shown originally by Pratt and Schlaifer (1984, p. 13). This non-uniqueness implies the non-uniqueness of the coefficients and the error terms of the GLM model. The non-unique coefficients of the GLM model do not measure the direct effects of its explanatory variables on its dependent variable (see Pratt and Schlaifer (1984, 1988)). They also cannot account for omitted-variable and measurement-error biases and hence the GLM model is misspecified. The linear functional form of the GLM model may also be incorrect. By contrast, the “true” model is unique if it is a real-world relation that remains invariant against changes in the language we use to describe it (see Basmann (1988, pp. 72-74)). For example, both adding and subtracting a term on the right-hand side of a representation of a real-world relation change only the representation but not the relation itself. The TVC model, but not the GLM model, shares this invariance property with the real-world relations.

(ii) The SC model coincides with the GLM model if for $j = 1, \dots, K-1$, the distribution of γ_{jit} is degenerate at β_j and the distribution of γ_{oit} is the same as that of $\beta_0 + b_i v_i + u_{it}$. Thus, the parametric assumptions underlying the GLM model are much stronger than (8).

(iii) The assumption that the x_{jit} 's in the GLM model are (mean) independent of v_i and u_{it} is, in the terminology of Pratt and Schlaifer, “meaningless” if v_i and u_{it} are used to denote ‘the’ excluded variables. Hence restricted maximum likelihood estimation discussed in Lehmann and Casella (1998, p. 518) may not lead to consistent estimators when it is applied to the GLM model. The area level model in (9) can be consistently estimated using an IRSGLS method if the coefficient drivers in (8) assign the correct functional form to the TVC model and if the included explanatory variables in the TVC model are conditionally independent of the error terms in (8), given the coefficient drivers.

(iv) **A crucial distinction** The SC model differs from hierarchical Bayes models analyzed in Lehmann and Casella (1998, pp. 253-262) in that the distribution of the coefficients of the former model is part of the likelihood function, whereas the distributions of the coefficients of the latter models are parts of the prior distributions. Two Bayesian statisticians, Pratt and Schlaifer (1988, p. 49), produce a very convincing argument to show that a Bayesian will do much better to search like a non-Bayesian for concomitants that absorb omitted-variable and measurement-error biases. Using (8), we do exactly what Pratt and Schlaifer suggest. The coefficient drivers in (8) are our concomitants. If some of them absorb the bias components of the coefficients of the TVC model, then they should appear as the explanatory variables of the coefficients, as in (8). The GLM model as well as hierarchical Bayes models completely ignores these biases. The prior distributions employed in their Bayesian analyses cannot distinguish between the bias-free and bias components of the coefficients of their TVC versions. Users of hierarchical Bayes models do not do what Pratt and Schlaifer suggest. Hierarchical Bayes models are not the devices capable of providing estimates (and, therefore, predictions) that would empirically be indistinguishably as good as those provided by good approximations to their (classical) TVC versions with unique coefficients.

2.4 Effects of Omitted-Variable and Measurement-Error Biases and Incorrect Functional Forms on GLM Models

A vector formulation of the GLM model without the u_{it} 's but with the sampling errors is

$$\hat{y}_{it} = x_{it}' \beta + b_i v_i + e_{it} \quad (11)$$

where $x_{it} = (1, x_{iit}, \dots, x_{K-1,it})'$, $\beta = (\beta_0, \beta_1, \dots, \beta_{K-1})'$. J.N.K. Rao (2003, p.116) assumes that (i) the v_i 's are identically and independently distributed with mean zero and constant variance, σ_v^2 , (ii) the sampling errors, e_{it} , are

independently distributed with mean zero and known variance, ψ_{it} , and (iii) v_i is independent of e_{it} for all i and t . Suppose that $T_i = 1$. Under these assumptions, the best linear unbiased predictor (BLUP) of the true value, y_{it}^* , from model (11) is

$$\hat{y}_{it}^* = \omega_{it} \hat{y}_{it} + (1 - \omega_{it}) x_{it}' \tilde{\beta} \quad (12)$$

where

$$\omega_{it} = \sigma_v^2 b_i^2 / (\psi_{it} + \sigma_v^2 b_i^2) \quad (13)$$

and

$$\tilde{\beta} = \left[\sum_{i=1}^n x_{it}' x_{it}' / (\psi_{it} + \sigma_v^2 b_i^2) \right]^{-1} \left[\sum_{i=1}^n x_{it}' \hat{y}_{it} / (\psi_{it} + \sigma_v^2 b_i^2) \right]. \quad (14)$$

J.N.K. Rao's (2003, p. 117) elegant method shows that the (average) mean square error (MSE) of (12) is

$$\omega_{it} \psi_{it} + (1 - \omega_{it})^2 x_{it}' \left[\sum_{i=1}^n x_{it}' x_{it}' / (\psi_{it} + \sigma_v^2 b_i^2) \right]^{-1} x_{it} \quad (15)$$

where the second term arises as a direct consequence of using an estimator of β in (12). It can happen that predictor (12) is not more efficient than the direct estimator \hat{y}_{it} because (15) can be larger than ψ_{it} in small samples.

To derive the BLUP of y_{it}^* from the linking model in (9), we make the following assumptions: For $i, i' = 1, \dots, n$, and $t, t' = 1, \dots, T_i$,

(A1) The K -vector $\mu_i = (\mu_{0i}, \mu_{1i}, \dots, \mu_{K-1,i})'$ is distributed with $E_m(\mu_i | z_{1it}, \dots, z_{p-1,it}) = 0$ and

$$E_m(\mu_i \mu_{i'}' | z_{1it}, \dots, z_{p-1,it}) = \begin{cases} \Delta & \text{if } i=i' \\ 0 & \text{if } i \neq i' \end{cases}, \quad (16)$$

where E_m denotes the model expectation and Δ may not be diagonal.

(A2) The q -vector $\varepsilon_{it} = (\varepsilon_{0it}, \varepsilon_{1it}, \dots, \varepsilon_{q-1,it})'$ follows the stochastic equation

$$\varepsilon_{it} = \phi_{it} \varepsilon_{it-1} + a_{it} \quad (17)$$

where ϕ_{it} is a $q \times q$ diagonal matrix whose eigenvalues are less than 1 in absolute value, the q -vector $a_{it} = (a_{0it}, a_{1it}, \dots, a_{q-1,it})'$ is distributed with $E_m(a_{it} | z_{1it}, \dots, z_{p-1,it}) = 0$ and

$$E_m(a_{it} a_{i't'}' | z_{1it}, \dots, z_{p-1,it}) = \begin{cases} \sigma_i^2 \Delta_{ii} & \text{if } i=i' \text{ and } t=t' \\ 0 & \text{if } i \neq i' \text{ and } t \neq t' \end{cases}, \quad (18)$$

where Δ_{ii} may not be diagonal.

(A3) Given $z_{dit}, d = 1, \dots, p-1$, the vectors, μ_i and ε_{it} , are independent and each of them varies independently across i .

(A4) Given $z_{dit}, d = 1, \dots, p-1$, the vectors, μ_i and ε_{it} , are independent of the x_{jit} 's.

(A5) The e_{it} 's are independent of the μ_i 's, ε_{it} 's, and x_{jit} 's.

A vector formulation of the model obtained by substituting the linking model in (9) for y_{it}^* in the sampling model is

$$\hat{y}_{it} = (z_{it}' \otimes x_{it}') \pi^{Long} + x_{it}' (\mu_i + L \varepsilon_{it}) + e_{it}, \quad (19)$$

where x_{it} is as defined in (11), μ_i is as defined in (16), ε_{it} is as defined in (17), $z_{it} = (1, z_{1it}, \dots, z_{p-1,it})'$, $\Pi = [\pi_{jd}]_{0 \leq j \leq K-1, 0 \leq d \leq p-1}$ is a $K \times p$ matrix having π_{jd} as its (j, d) element, π^{Long} is a Kp -vector denoting a column stack of Π , \otimes denotes a Kronecker product, and $L = [l_{jh}]_{0 \leq j \leq K-1, 0 \leq h \leq q-1}$ is a $K \times q$ matrix having l_{jh} as its (j, h) element. We call (19) "the improved area level model."

When $T_i = 1$, the BLUP of the true value, y_{it}^* , from the improved area level model is

$$\hat{y}_{it}^* = (z_{it}' \otimes x_{it}') \hat{\pi}^{Long} + \omega_{it}^* \{\hat{y}_{it} - (z_{it}' \otimes x_{it}') \hat{\pi}^{Long}\} = \omega_{it}^* \hat{y}_{it} + (1 - \omega_{it}^*) (z_{it}' \otimes x_{it}') \hat{\pi}^{Long} \quad (20)$$

where

$$\omega_{it}^* = x_{it}' \{ \Delta + L(\phi_{it} \Gamma_{ii} \phi_{it}' + \sigma_i^2 \Delta_{ii}) L' \} x_{it} / \left[x_{it}' \{ \Delta + L(\phi_{it} \Gamma_{ii} \phi_{it}' + \sigma_i^2 \Delta_{ii}) L' \} x_{it} + \psi_{it} \right] \quad (21)$$

$$\hat{\pi}^{Long} = \left[\sum_{i=1}^n (z_{it} \otimes x_{it}) (z_{it}' \otimes x_{it}') / (x_{it}' \{ \Delta + L(\phi_{it} \Gamma_{ii} \phi_{it}' + \sigma_i^2 \Delta_{ii}) L' \} x_{it} + \psi_{it}) \right]^{-1}$$

$$\times \left[\sum_{i=1}^n (z_{it} \otimes x_{it}) \hat{y}_{it} / (x'_{it} \{ \Delta + L(\phi_{it} \Gamma_{it} \phi'_{it} + \sigma_i^2 \Delta_{it}) L' \} x_{it} + \psi_{it}) \right] \quad (22)$$

with $E \varepsilon_{it} \varepsilon'_{it} = \Gamma_{it} = \phi_{it} \Gamma_{it} \phi'_{it} + \sigma_i^2 \Delta_{it}$.

What are the properties of predictor (12) when in fact the improved area level model in (19) is appropriate? Predictor (12) can be still design-consistent. The direct survey estimator, \hat{y}_{it} , in (1) is design-consistent if its design bias goes to zero and if its design variance, ψ_{it} , tends to zero as the sample size increases. Under these conditions, predictor (12) is design-consistent because $\omega_{it} \rightarrow 1$ as $\psi_{it} \rightarrow 0$. Of course, (12) no longer has minimum MSE and the formula in (15) for the MSE of (12) is no longer appropriate. Predictor (20) has minimum MSE and is also design-consistent because $\omega_{it}^* \rightarrow 1$ as $\psi_{it} \rightarrow 0$. Of the two design-consistent predictors, (12) and (20), only (12) is affected by the misspecifications in the GLM model in (11). In small samples, i.e., when n is small, these effects can only be limited by replacing the GLM model by the improved area level model in (19), but not by using the design-consistent predictor (12). Thus, not all models that yield design-consistent predictors limit the effects of model misspecifications in small samples.

The MSE of (20) about y_{it}^* is

$$\omega_{it}^* \psi_{it} + (1 - \omega_{it}^*)^2 (z'_{it} \otimes x'_{it}) \left[\sum_{i=1}^n \frac{(z_{it} \otimes x_{it})(z'_{it} \otimes x'_{it})}{(x'_{it} \{ \Delta + L(\phi_{it} \Gamma_{it} \phi'_{it} + \sigma_i^2 \Delta_{it}) L' \} x_{it} + \psi_{it})} \right]^{-1} (z_{it} \otimes x_{it}) \quad (23)$$

where the second term arises as a direct consequence of using an estimator of π^{Long} in predictor (20).

This MSE is less than or equal to the MSE of (12) derived under the assumption that the improved area level model in (19) is appropriate. The inappropriate formula in (15) leads to the following conclusions: Under some regularity conditions stated in J.N.K. Rao (2003, p. 117), the second term in (15) goes to zero as n goes to ∞ . Therefore, this term can be ignored when n is large. Comparison of the first term, $\omega_{it} \psi_{it}$, in (15) with the design variance, ψ_{it} , of the direct estimator, \hat{y}_{it} , shows that predictor (12) leads to large gains in efficiency when ω_{it} is small, that is, when the variability of the GLM model's error, $b_i v_i$, is small relative to the total variability of $b_i v_i + e_{it}$. This result due to J.N.K. Rao (2003, p. 117)) arises as a direct consequence of using the GLM model in (11) and ignoring the bias components contained in the coefficients of the TVC model. The MSE (15) of predictor (12) derived under the assumption that the GLM model in (11) holds without the u_{it} 's exaggerates the gains in efficiency resulting from the "strength" the GLM model "borrows" from the area-specific auxiliary variables, x_{jit} , if (23) is larger than (15).

We can conclude that in small samples, the linking model in (9) permits us to "borrow" strength from the area-specific auxiliary variables, x_{it} , if (23) is smaller than ψ_{it} . This condition is unlikely to hold in small samples. Therefore, the indirect model-dependent estimator (20) can have larger MSE than the direct estimator in (1) in small samples. Even the inappropriate MSE in (15) can be larger than ψ_{it} . However, for large n , the second term in (23) can be close to zero and MSE (23) can be smaller than ψ_{it} , since $\omega_{it}^* < 1$. Therefore, in large samples, the linking model in (9) permits us to "borrow" strength from the area-specific auxiliary variables.

2.5 Empirical BLUP

Predictor (20) depends on the distinct nonzero elements of Δ , ϕ_{it} , Δ_{it} , and σ_i^2 which are unknown in practical applications. Hence it is not operational. Let δ be a vector consisting of these elements. Swamy et al.'s (2003) IRSGLS method may be used to estimate simultaneously the parameter vectors π^{Long} and δ of the improved area level model in (19). The predictor of y_{it}^* obtained by using this estimate of δ in place of the "true" value of δ used in (20) is called "the empirical BLUP (EBLUP)" and is denoted by $\hat{y}_{it}^*(\hat{\delta})$. The error in the EBLUP may be decomposed into $\hat{y}_{it}^*(\hat{\delta}) - y_{it}^* = (\hat{y}_{it}^* - y_{it}^*) + (\hat{y}_{it}^*(\hat{\delta}) - \hat{y}_{it}^*)$ where \hat{y}_{it}^* is given in (20). Therefore, the MSE of EBLUP is

$$MSE[\hat{y}_{it}^*(\hat{\delta})] = E[\hat{y}_{it}^*(\hat{\delta}) - y_{it}^*]^2 = E(\hat{y}_{it}^* - y_{it}^*)^2 + E[\hat{y}_{it}^*(\hat{\delta}) - \hat{y}_{it}^*]^2 \quad (24)$$

where use is made of the conditions under which $E[\hat{y}_{it}^* - y_{it}^*][\hat{y}_{it}^*(\hat{\delta}) - \hat{y}_{it}^*] = 0$. These conditions are given in J.N.K. Rao

(2003, p. 103). The first term on the right-hand side of the second equality sign in (24) is equal to (23). The last term in (24) accounts for the variability in the estimator of δ . Since this term is generally intractable except in special cases, J.N.K. Rao (2003, pp. 103-104) finds a second-order approximation to the term. The MSE of EBLUP is larger than (23). An estimator of the MSE of $\hat{y}_i^*(\hat{\delta})$ as a measure of variability in $\hat{y}_i^*(\hat{\delta})$ is given in J.N.K. Rao (2003, pp. 104-105).

3. Conclusions

In making estimates of population characteristics for small domains with adequate level of precision, it is often necessary to use models that relate the true values of population characteristics to auxiliary variables. The major weakness of this approach is that if these models are seriously misspecified, they can yield inferences that are worse than design-based inferences. Misspecifications of models occur when relevant explanatory variables are omitted from the models, when included variables are measured with error, and/or when the unknown functional forms of the models are incorrectly specified. The biasing effects of measurement error, omitted variables, and misspecifications of functional forms are a pervasive problem in applied statistics. It has been shown that wrong inferences can be obtained if these biases are ignored. A method of accounting for these biases is proposed. The strength a statistical model “borrows” from auxiliary variables in making estimates for small areas may be largely offset by the effects of the model’s misspecifications.

References

- Basman, R. L. (1988), “Causality Tests and Observationally Equivalent Representations of Econometric Models,” *Journal of Econometrics, Annals*, 39, 69-104.
- Chang, I., Swamy, P. A. V. B., Hallahan, C. and Tavlas, G. S. (2000), “A Computational Approach to Finding Causal Economic Laws,” *Computational Economics*, 16, 105-136.
- Freedman, D. A. and Navidi, W. C. (1986), “Regression Models for Adjusting the 1980 Census (with discussion),” *Statistical Science*, 1, 3-39.
- Lehmann, E. L. and Casella, G. (1998), *Theory of Point Estimation*, Second edition, New York: Springer.
- Pratt, J. W. and Schlaifer, R. (1984), “On the Nature and Discovery of Structure (with discussion),” *Journal of the American Statistical Association*, 79, 9-21, 29-33.
- Pratt, J. W. and Schlaifer, R. (1988), “On the Interpretation and Observation of Laws,” *Journal of Econometrics*, 39, 23-52.
- Rao, J. N. K. (2003), *Small Area Estimation*, Hoboken, New Jersey: John Wiley & Sons.
- Swamy, P. A. V. B. and Mehta, J. S. (1975), Bayesian and Non-Bayesian Analysis of Switching Regressions and of Random Coefficient Regression Models,” *Journal of the American Statistical Association*, 70, 593-602.
- Swamy, P. A. V. B. and Tinsley, P. A. (1980), “Linear Prediction and Estimation Methods for Regression Models with Stationary Stochastic Coefficients,” *Journal of Econometrics*, 12, 103-142.
- Swamy, P. A. V. B., Mehta, J. S. and Singamsetti, R. N. (1996), “Circumstances in Which Different Criteria of Estimation Can be Applied to Estimate Policy Effects,” *Journal of Statistical Planning and Inferences*, 50, 121-153.
- Swamy, P. A. V. B. and Tavlas, G. S. (2001), “Random Coefficient Models,” in *A Companion to Theoretical Econometrics*, ed. B. H. Baltagi, Malden, Massachusetts: Blackwell Publishers.
- Swamy, P. A. V. B., Chang, I., Mehta, J. S. and Tavlas, G. S. (2003), “Correcting for Omitted-Variables and Measurement-Error Bias in Autoregressive Model Estimation with Panel Data,” *Computational Economics*, 22, 225-253.
- Swamy, P. A. V. B., Tavlas, G. S. and Chang, I. (2005), “How Stable are Monetary Policy Rules: Estimating the Time-Varying Coefficients in Monetary Policy Reaction Function for the U.S.,” *Computational Statistics & Data Analysis*, 49, 575-590.