NBER WORKING PAPER SERIES

ECONOMETRIC RISK ADJUSTMENT, ENDOGENEITY
AND EXTRAPOLATION BIAS

John Mullahy

Working Paper 12236
http://www.nber.org/papers/w12236

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 2006

Econometric Risk Adjustment, Endogeneity, and Extrapolation Bias
John Mullahy
NBER Working Paper No. 12236
May 2006
JEL No. I1

ABSTRACT

In econometric risk-adjustment exercises, models estimated with one or more included endogenous explanatory variables ("risk adjusters") will generally result in biased predictions of outcomes of interest, e.g. unconditional mean healthcare expenditures. This paper shows that a first-order contributor to this prediction bias is the difference between the distribution of explanatory variables in the estimation sample and the prediction sample -- a form of "extrapolation bias." In the linear model context, a difference in the means of the respective joint marginal distributions of observed covariates suffices to produce bias when endogenous explanatory variables are used in estimation. If these means do not differ, then the "endogeneity-related" extrapolation bias disappears although a form of "standard" extrapolation bias may persist. These results are extended to some of the nonlinear models in common use in this literature with some provisionally-similar conclusions. In general the bias problem will be most acute where risk adjustment is most useful, i.e. when estimated risk-adjustment models are applied in populations whose characteristics differ from those from which the estimation data are drawn.

John Mullahy
University of Wisconsin-Madison
787 WARF, 610 Walnut Street
Madison, WI 53726
and NBER
jmullahy@wisc.edu

I. Introduction

In econometric risk-adjustment exercises, models estimated with one or more included endogenous explanatory variables ("adjusters") will generally result in biased predictions of outcomes of interest, e.g. unconditional mean expenditures. Taken at face value, this result hardly seems surprising.

Interestingly, a first-order contributor to this prediction bias turns out to be the difference between the distribution of explanatory variables in the estimation sample and that in the prediction sample -- a form of "extrapolation bias." In the linear model context, a difference in the means of the respective joint marginal distributions of observed covariates suffices to produce bias when endogenous explanatory variables are used in estimation. If these respective means do not differ, then the "endogeneity-related" extrapolation bias ("EEB") characterized below disappears even though a form of "standard" extrapolation bias ("SEB") may persist. These results suggest that the bias problem will tend to be most acute where risk adjustment exercises are likely to be most useful, i.e. when estimated risk-adjustment models are applied in populations whose characteristics -- observed and unobserved -- differ from those from which the estimation data are drawn.

The plan of the paper is as follows. Section II provides the basic econometric analysis and results. Using these results, Section III demonstrates how these extrapolation biases arise in the linear model. Section IV sketches briefly how related extrapolation biases might be manifested in some nonlinear models. Section V concludes.

II. Analysis

To the extent that there is a typical econometric risk-adjustment exercise, it might entail:

(a) regressing some measure of period-$(t+1)$ expenditures (y) on a set of period-t "adjusters" or covariates (x,z) using data characterizing a reference population ("A") described by the distribution $\phi_A(x_A, z_A, u_A)$, where u are unobserved determinants of y and $y = y(x, z, u; \boldsymbol{\beta})$;

(b) deploying these results to formulate a generic prediction model for an outcome like mean expenditures as $\widehat{m}(x, z) = m(x, z; \widehat{\boldsymbol{\beta}}_A)$ (note that $m(x, z; \boldsymbol{\beta})$ does not necessarily represent a correct specification of $E[y|x,z]$; more on this below); and

(c) applying this estimated model in a different population ("B") to predict unconditional mean expenditures in that population by utilizing covariate data

drawn from the joint marginal $\phi_B(x_B, z_B)$ to drive the prediction.[1]

See Ettner et al., 2001, Glazer and McGuire, 2002, Schokkaert and Van de Voorde, 2004, Veazie et al., 2003, and Zhao et al., 2005 for some important recent contributions to the conceptual and empirical risk adjustment literature.

It is assumed in this section that the model relating (x,z,u) to y in a population described by $\phi_j(u_j, x_j, z_j)$ is a linear model with parameters $\boldsymbol{\beta}_j = \left[\beta_{0_j}, \beta_x, \beta_z\right]'$:

$$y_j = y(x_j, z_j, u_j; \boldsymbol{\beta}_j) = \beta_{0_j} + \beta_x x_j + \beta_z z_j + u_j, \tag{1}$$

where for simplicity of exposition x and z are specified for now as scalar "endogenous" and "exogenous" covariates, respectively; the general k-covariate case is presented below.[2] While the parameters $\left[\beta_x, \beta_z\right]$ are assumed invariant across populations, the constant term is permitted to vary across populations to accommodate distributions of unobservables that may have different unconditional means in different populations. Then

$$\begin{aligned} E\left[y_j\right] &= E[\beta_{0_j} + \beta_x x_j + \beta_z z_j + u_j] \\ &= \beta_{0_j} + \beta_x E\left[x_j\right] + \beta_z E\left[z_j\right] + E\left[u_j\right] \\ &= \beta_{0_j} + \beta_x \mu_x^j + \beta_z \mu_z^j, \end{aligned} \tag{2}$$

that is $E\left[u_j\right]=0$ *unconditionally* even though $E[u_j|x_j,z_j]=E\left[u_j|x_j\right] = \xi_j(x_j) \neq 0$ for particular $\left(x_j, z_j\right)$, as β_{0_j} absorbs any nonzero $E\left[u_j\right]$, i.e. $E_{x_j}\left[\xi_j(x_j)\right] = 0$. Exogeneity of z and endogeneity of x in population j are characterized here as, respectively,

---

[1] For present purposes, a "population" is a collection of individuals defined by some geographic, demographic, temporal, administrative, membership, or other relevant exogenous characteristics. Thus, for instance, a set of individuals defined by a common geographic criterion but having time-varying characteristics that is observed at two distinct points in time might usefully be described as two distinct populations.

[2] In general, the "j" notation will be used here to index populations (e.g. $j \in \{A, B\}$); except where useful for clarity, subscripts indexing individual observations are suppressed. The symbols $\{y_j, x_j, z_j, u_j\}$ represent either scalars or $N_j \times 1$ vectors; the particular meaning should be obvious from context.

$$\lim_{N_j \to \infty} \left(z_j{}' u_j / N_j \right) = \sigma_{zu}^j = 0 \tag{3}$$

and

$$\lim_{N_j \to \infty} \left(x_j{}' u_j / N_j \right) = \sigma_{xu}^j \neq 0. \tag{4}$$

For particular characterizations of x and u, a suggestion that $\sigma_{xu}^j \neq 0$ would be compelling under any number of reasonable assumptions about the nature of observed and unobserved determinants of healthcare expenditures.

Define $\boldsymbol{\mu}_{xz}^A = \left[\mu_x^A, \mu_z^A \right]$ and $\boldsymbol{\Sigma}_{xz}^A = \begin{bmatrix} \sigma_{xx}^A & \sigma_{xz}^A \\ \sigma_{zx}^A & \sigma_{zz}^A \end{bmatrix}$ as the finite mean vector and covariance

matrix of the population joint marginal distribution $\phi_A(x_A, z_A)$. With $\sigma_{zu}^A = 0$, the OLS estimator of (1) based on a suitable sample drawn from population A has the following familiar large-sample properties:

$$\text{plim}(\hat{\beta}_x) = \beta_x + \Delta_A^{-1} \left\{ \sigma_{xu}^A \sigma_{zz}^A - \sigma_{zu}^A \sigma_{xz}^A \right\}$$

$$= \beta_x + \Delta_A^{-1} \sigma_{xu}^A \sigma_{zz}^A \tag{5}$$

$$= \beta_x + \kappa_x$$

$$\text{plim}(\hat{\beta}_z) = \beta_z + \Delta_A^{-1} \left\{ \sigma_{zu}^A \sigma_{xx}^A - \sigma_{xu}^A \sigma_{xz}^A \right\}$$

$$= \beta_z - \Delta_A^{-1} \sigma_{xu}^A \sigma_{xz}^A \tag{6}$$

$$= \beta_z + \kappa_z$$

$$\text{plim}(\hat{\beta}_{0_A}) = \beta_{0_A} + \Delta_A^{-1} \left\{ \sigma_{xu}^A \left(\sigma_{xz}^A \mu_z^A - \sigma_{zz}^A \mu_x^A \right) + \sigma_{zu}^A \left(\sigma_{xz}^A \mu_x^A - \sigma_{xx}^A \mu_z^A \right) \right\}$$

$$= \beta_{0_A} + \Delta_A^{-1} \sigma_{xu}^A \left(\sigma_{xz}^A \mu_z^A - \sigma_{zz}^A \mu_x^A \right) \tag{7}$$

$$= \beta_{0_A} - \left(\kappa_x \mu_x^A + \kappa_z \mu_z^A \right),$$

where $\Delta_A = \det\left(\boldsymbol{\Sigma}_{xz}^A \right) > 0$ and the $\kappa_{(.)}$ terms describe the large-sample biases. Defining the generic prediction (or "risk adjustment") model

$$\widehat{m}(x, z) = m(x, z; \hat{\boldsymbol{\beta}}_A) = \hat{\beta}_{0_A} + \hat{\beta}_x x + \hat{\beta}_z z, \tag{8}$$

and using (5)-(7), it is instructive to note that

$$p\lim\left(m(\mu_x^A, \mu_z^A; \hat{\boldsymbol{\beta}}_A)\right) = \beta_{0_A} + \beta_x \mu_x^A + \beta_z \mu_z^A ,\tag{9}$$

i.e. that the inconsistent OLS estimator $\hat{\boldsymbol{\beta}}_A$ does *not* translate into an inconsistent estimator of $E[y_A]$ in (2). $\widehat{m}(x,z)$ evaluated at $\left(\mu_x^A, \mu_z^A\right)$ consistently predicts $E[y_A]$ even though $\widehat{m}(x,z)$ is in general inconsistent for $E\left[y_j | x, z\right] = \beta_{0_j} + \beta_x x + \beta_z z + \xi_j(x)$ for arbitrary values (x,z). The next section shows, however, that this consistency property is fragile, holding only at the particular point of evaluation $\left\{\left(x = \mu_x^A\right), \left(z = \mu_z^A\right)\right\}$; it is this fragility that ultimately has potentially serious implications for empirical risk adjustment exercises.

III. Extrapolation Bias in the Linear Model

The policymaker or risk adjuster is presumed to be interested primarily in the quantity $E[y_B]$ -- i.e. mean expenditure in population B in period $t+1$ -- but has access only to period-t data from $\phi_B(x_B, z_B)$. (The quantity of interest might be total expenditures, $N_B \times E[y_B]$, rather than mean expenditures, but this distinction is not important here.) The forecasting stage of the risk-adjustment process would then proceed by taking the generic prediction model estimated using population-A data, $\widehat{m}(x,z)$, and predicting $E[y_B]$ as

$$\hat{E}\left[y_B\right] = \sum_{i=1}^{N_B} w_i \widehat{m}(x_{i_B}, z_{i_B}) \tag{10}$$
$$= \hat{\beta}_{0_A} + \hat{\beta}_x \hat{\mu}_x^B + \hat{\beta}_z \hat{\mu}_z^B ,$$

where $\left(\hat{\mu}_x^B, \hat{\mu}_z^B\right)$ are the sample-based estimators of $\left(\mu_x^B, \mu_z^B\right)$ that are presumed to be consistent owing to suitable specification of the weights w_i (e.g. $w_i = 1/N_B$ in random sampling). From Slutsky's Theorem,

$$p\lim\left(\hat{E}\left[y_B\right]\right) = p\lim\left(\hat{\beta}_{0_A}\right) + p\lim\left(\hat{\beta}_x\right)\mu_x^B + p\lim\left(\hat{\beta}_z\right)\mu_z^B ,\tag{11}$$

so it follows that

$$\text{plim}\left(\hat{E}[y_B]\right) = \beta_{0_A} + \beta_x\mu_x^B + \beta_z\mu_z^B + \left[\left(\mu_x^B - \mu_x^A\right)\kappa_x + \left(\mu_z^B - \mu_z^A\right)\kappa_z\right]$$

$$= E[y_B] + \left\{\left(\beta_{0_A} - \beta_{0_B}\right) + \left[\left(\mu_x^B - \mu_x^A\right)\kappa_x + \left(\mu_z^B - \mu_z^A\right)\kappa_z\right]\right\} \qquad (12)$$

$$= E[y_B] + \left\{\text{ SEB } + \text{ EEB }\right\}.$$

This is the central result of the paper: In cases where endogeneity is in play in the estimation population ($\sigma_{xu}^A \neq 0$), the departure of the estimator of unconditional mean expenditures in population B from its true value will be governed in part by how much the marginal means of the explanatory variables differ between the A and B populations, i.e. $\left(\mu_x^B - \mu_x^A\right)$ and $\left(\mu_z^B - \mu_z^A\right)$. Note that this large-sample "endogeneity-related" extrapolation bias ("EEB") may or may not have the same sign as the "standard" extrapolation bias ("SEB"), $\left(\beta_{0_A} - \beta_{0_B}\right)$, that arises even when $\sigma_{xu}^A = 0$ owing simply to the distributions of unobservables in populations A and B having different unconditional means. Finally, note from (5)-(7) that the parameters underlying the bracketed term in (12) except for $(\beta_{0_A} - \beta_{0_B})$ and σ_{xu}^A can be identified from the population-A and population-B samples, and in some instances theory or prior information might shed light on at least the signs of $(\beta_{0_A} - \beta_{0_B})$ and σ_{xu}^A if not their magnitudes.

These results generalize to the more empirically-relevant cases that specify $(k+1)$-vectors of explanatory variables, subsets of which may be endogenous. The model of interest is now

$$y_j = \beta_{0_j} + \mathbf{Q}_j\boldsymbol{\beta_Q} + u_j, \qquad (13)$$

where $\mathbf{Q}_j$ is a k-vector or -matrix of explanatory variables and $\boldsymbol{\beta_Q}$ is a conformable k-vector of slope parameters. $\mathbf{Q}_j$ may contain an arbitrary mix of endogenous x's and exogenous z's. Let $\boldsymbol{\mu_Q^j}$ and $\boldsymbol{\Sigma_Q^j}$ represent the finite mean vector and covariance matrix of the marginal distribution $\phi_j\left(\mathbf{Q}_j\right)$.

In this case, the OLS estimator $\left[\hat{\beta}_{0_A}, \hat{\boldsymbol{\beta}}_{\mathbf{Q}}'\right]$ based on a sample drawn from population A can be shown[3] to have the following large-sample properties:

$$\text{plim}\left(\hat{\boldsymbol{\beta}}_{\mathbf{Q}}\right) = \boldsymbol{\beta}_{\mathbf{Q}} + \left(\boldsymbol{\Sigma}_{\mathbf{Q}}^A\right)^{-1} \boldsymbol{\Sigma}_{\mathbf{Q}u}^A \tag{14}$$

$$= \boldsymbol{\beta}_{\mathbf{Q}} + \boldsymbol{\kappa}_{\mathbf{Q}}$$

and

$$\text{plim}\left(\hat{\beta}_{0_A}\right) = \beta_{0_A} - \left(\frac{\boldsymbol{\mu}_{\mathbf{Q}}^A{}'\left(\boldsymbol{\Sigma}_{\mathbf{Q}}^A + \boldsymbol{\mu}_{\mathbf{Q}}^A\boldsymbol{\mu}_{\mathbf{Q}}^A{}'\right)^{-1}\boldsymbol{\Sigma}_{\mathbf{Q}u}^A}{1 - \boldsymbol{\mu}_{\mathbf{Q}}^A{}'\left(\boldsymbol{\Sigma}_{\mathbf{Q}}^A + \boldsymbol{\mu}_{\mathbf{Q}}^A\boldsymbol{\mu}_{\mathbf{Q}}^A{}'\right)^{-1}\boldsymbol{\mu}_{\mathbf{Q}}^A}\right) \tag{15}$$

$$= \beta_{0_A} - \boldsymbol{\mu}_{\mathbf{Q}}^A{}'\boldsymbol{\kappa}_{\mathbf{Q}},$$

where endogeneity is manifested as nonzero elements of $\boldsymbol{\Sigma}_{\mathbf{Q}u}^A = \lim_{N_A \to \infty}\left(\mathbf{Q}_A{}'u_A / N_A\right)$. In this case the generic prediction model is

$$\widehat{m}(\mathbf{Q}) = m(\mathbf{Q}; \hat{\beta}_{0_A}, \hat{\boldsymbol{\beta}}_{\mathbf{Q}}) = \hat{\beta}_{0_A} + \mathbf{Q}\hat{\boldsymbol{\beta}}_{\mathbf{Q}}. \tag{16}$$

Analogous to (9), evaluating $\widehat{m}(\mathbf{Q})$ at $\mathbf{Q} = \boldsymbol{\mu}_{\mathbf{Q}}^A$ gives

---

[3] The derivation of the results that follow is facilitated by reference to two theorems provided by Graybill, 1983. Suppose $\mathbf{J}$ is a nonsingular $k \times k$ matrix and that $\mathbf{v}$ is a $k \times 1$ vector such that $\mathbf{J} + \mathbf{vv}'$ is also nonsingular. Then:

$$\left(\mathbf{J} + \mathbf{vv}'\right)^{-1} = \left(\mathbf{J}^{-1} - \frac{\mathbf{J}^{-1}\mathbf{vv}'\mathbf{J}^{-1}}{1 + \mathbf{v}'\mathbf{J}^{-1}\mathbf{v}}\right) \qquad \text{(Graybill, Theorem 8.9.3)}$$

and

$$\mathbf{v}'\mathbf{Jv} = \text{tr}\left(\mathbf{Jvv}'\right). \qquad \text{(Graybill, Theorem 9.1.20)}$$

Equations (14) and (15) are derived starting with familiar expositions of OLS estimator properties when the latter are presented in terms of matrix partitions (e.g. Amemiya, 1985, equations (1.2.12) and (1.2.13)).

$$plim\left(\widehat{m}(\boldsymbol{\mu_Q^A})\right) = \beta_{0_A} + \boldsymbol{\mu_Q^A}{}'\boldsymbol{\beta_Q}, \tag{17}$$

i.e. the inconsistent OLS estimator $\left[\hat{\beta}_{0_A}, \boldsymbol{\hat{\beta}_Q}{}'\right]$ still yields a consistent estimator of $E\left[y_A\right]$.

As in (12), however, there will arise in general large-sample extrapolation bias when $\widehat{m}(\boldsymbol{Q})$ is evaluated at the covariate means from population B. Specifically,

$$plim\left(\widehat{m}\left(\boldsymbol{\mu_Q^B}\right)\right) = plim\left(\hat{E}\left[y_B\right]\right) = \beta_{0_A} + \boldsymbol{\mu_Q^B}{}'\boldsymbol{\beta_Q} + \left(\boldsymbol{\mu_Q^B} - \boldsymbol{\mu_Q^A}\right)'\boldsymbol{\kappa_Q}$$

$$= E\left[y_B\right] + \left\{\left(\beta_{0_A} - \beta_{0_B}\right) + \left(\boldsymbol{\mu_Q^B} - \boldsymbol{\mu_Q^A}\right)'\boldsymbol{\kappa_Q}\right\} \tag{18}$$

$$= E\left[y_B\right] + \left\{\ SEB\ +\ EEB\ \right\}.$$

IV. Extrapolation Bias in Nonlinear Models

The use of linear risk-adjustment models like (1) or (13) may be more exception than rule in practice. A reasonable conjecture is that the conclusions reached here for the linear model will hold *a forteriori* in the variety of nonlinear approaches ($y = \varphi(\boldsymbol{Q}, u)$) to econometric risk-adjustment that have been proposed and applied in the literature: transformation-retransformation models; two-part models with or without part-two transformation-retransformation; generalized linear models under log-link assumptions; and others. In these contexts it would be expected that the aforedescribed biases will depend on properties of the joint marginals $\phi_j\left(x_j, z_j\right)$ beyond simply their marginal means. This section sketches briefly such an approach for one prominent nonlinear model structure used in the risk-adjustment literature; however, further research into this issue would probably be valuable. Suppose that the standard log-transformation-retransformation approach provides the basis for estimation, i.e.

$$y_j = \exp\left(\beta_{0_j} + \beta_x x_j + \beta_z z_j + u_j\right), \qquad y_j > 0 \tag{19}$$

with

$$\ln\left(y_j\right) = \beta_{0_j} + \beta_x x_j + \beta_z z_j + u_j, \tag{20}$$

where $\sigma_{xu}^j \neq 0$ and $\sigma_{zu}^j = 0$ as before. Such a model may be appropriate when positive expenditures are observed for all subjects, or alternatively as part two of a two-part model. With covariate endogeneity, an explicit functional-form characterization of $E[y_j]$ is typically more difficult to derive than in the linear case.[4]

On the basis of OLS estimation of (20) using population-A data, a researcher might posit the generic prediction model

$$\widehat{m}(x,z) = m(x,z;\hat{\boldsymbol{\beta}}_A) = \exp\left(\hat{\beta}_{0_A} + \hat{\beta}_x x + \hat{\beta}_z z\right)\hat{\rho}_A , \tag{21}$$

and then (in light of Jensen's inequality) use a sample average of this quantity rather than evaluating it at the covariates' means to estimate mean expenditures in population j as

$$\hat{E}[y_j] = \sum_{i=1}^{N_j} w_i \exp\left(\hat{\beta}_{0_A} + \hat{\beta}_x x_{i_j} + \hat{\beta}_z z_{i_j}\right)\hat{\rho}_A , \tag{22}$$

with $w_i = 1/N_j$, for example. In (21) and (22) $\hat{\rho}_A$ is an estimated retransformation factor that will be discussed below. Specifying (22) as a generic prediction model under the (incorrect) assumption that all covariates are exogenous would be sensible under an assumption of mean-independence of u conditional on (x,z), i.e.

---

[4] For illustrative purposes, the structure of $E[y_j]$ can be shown in the case where the observables and unobservables (x,z,u) are jointly normally-distributed. That is, suppose $(x_j, z_j, u_j) \sim N\left(\boldsymbol{\mu}_{xzu}^j, \boldsymbol{\Sigma}_{xzu}^j\right)$ with $\boldsymbol{\mu}_{xzu}^j = \left[\mu_x^j, \mu_z^j, 0\right]'$ and $\boldsymbol{\Sigma}_{xzu}^j = \begin{bmatrix} \sigma_{xx}^j & \sigma_{xz}^j & \sigma_{xu}^j \\ \sigma_{xz}^j & \sigma_{zz}^j & 0 \\ \sigma_{xu}^j & 0 & \sigma_{uu}^j \end{bmatrix}$. Define $\boldsymbol{\theta} = \left[\beta_x, \beta_z, 1\right]'$. Then (Aitchison and Brown, 1969), y_j follows a lognormal distribution $y_j \sim \Lambda\left(\left(\beta_{0_j} + \boldsymbol{\theta}'\boldsymbol{\mu}_{xzu}^j\right), \boldsymbol{\theta}'\boldsymbol{\Sigma}_{xzu}^j\boldsymbol{\theta}\right)$ with

$$E[y_j] = \exp\left(\left\{\beta_{0_j} + \beta_x\mu_x^j + \beta_z\mu_z^j\right\} + .5\left\{\beta_x^2\sigma_{xx}^j + \beta_z^2\sigma_{zz}^j + \sigma_{uu}^j + 2\beta_x\beta_z\sigma_{xz}^j\right\} + \left\{\beta_x\sigma_{xu}^j\right\}\right).$$

The first and second bracketed terms are familiar, whereas the third bracketed term $\beta_x\sigma_{xu}^j$ arises only in the presence of endogeneity.

$$E\left[y_j\right] = E\left[\exp(\beta_{0_j} + \beta_x x_j + \beta_z z_j + u_j)\right]$$

$$= E_{x_j,z_j}\left[\exp(\beta_{0_j} + \beta_x x_j + \beta_z z_j + u_j)\Big| x_j, z_j\right] \quad (23)$$

$$= E_{x_j,z_j}\left[\exp(\beta_{0_j} + \beta_x x_j + \beta_z z_j)\right]\rho_j,$$

with $E\left[\exp(u_j)\,|\,x_j,z_j\right]$ a constant not depending on (x,z). Note that when the model is estimated by OLS in its log-transformed version, the constant term β_{0_j} is defined by $E\left[u_j\right]$, *not* by $\ln\left(E\left[\exp(u_j)\right]\right)$; this is why the additional retransformation factor ρ_j is required, i.e. $E\left[\exp(u_j)\right] = \exp\left(\beta_{0_j} + \ln(\rho_j)\right)$. Based on the third line of (23), a (perhaps-weighted) sample average of $\exp(\hat{\beta}_{0_j} + \hat{\beta}_x x_{i_j} + \hat{\beta}_z z_{i_j})$ multiplied by an estimated retransformation factor -- i.e. equation (22) -- would be a sensible prediction model to maintain under an assumption of no endogeneity.

To assess how extrapolation bias might influence the estimator (22) in the presence of covariate endogeneity, consider a second-order Taylor expansion of (22) around $\left(x_{i_j} = \hat{\mu}_x^j, z_{i_j} = \hat{\mu}_z^j\right)$:

$$\hat{E}\left[y_j\right] \approx \frac{\hat{\rho}_A}{N_j} \exp\left(\hat{\beta}_{0_A} + \hat{\beta}_x \hat{\mu}_x^j + \hat{\beta}_z \hat{\mu}_z^j\right) \times \quad (24)$$

$$\sum_{i=1}^{N_j}\left\{1 + \left[\delta_{x_i}, \delta_{z_i}\right]\begin{bmatrix}\hat{\beta}_x \\ \hat{\beta}_z\end{bmatrix} + .5\left[\delta_{x_i}, \delta_{z_i}\right]\begin{bmatrix}\hat{\beta}_x^2 & \hat{\beta}_x\hat{\beta}_z \\ \hat{\beta}_x\hat{\beta}_z & \hat{\beta}_z^2\end{bmatrix}\begin{bmatrix}\delta_{x_i} \\ \delta_{z_i}\end{bmatrix}\right\},$$

where $\delta_{x_i} = x_{i_j} - \hat{\mu}_x^j$ and $\delta_{z_i} = z_{i_j} - \hat{\mu}_z^j$. Taking probability limits and assuming that first and second sample moments of $\left(\delta_{x_i}, \delta_{z_i}\right)$ tend to their population counterparts in the limit, it follows that

$$p\lim\left(\hat{E}\left[y_j\right]\right) \approx p\lim\left(\hat{\rho}_A\right) \times T_1 \times T_2 \times T_3, \quad (25)$$

where

$$T_1 = \exp\left(\beta_{0_A} + \beta_x \mu_x^j + \beta_z \mu_z^j\right)$$

$$T_2 = \exp\left(\left(\mu_x^j - \mu_x^A\right)\kappa_x + \left(\mu_z^j - \mu_z^A\right)\kappa_z\right)$$

$$T_3 = 1 + .5 \times \left\{\sigma_{xx}^j \left(\beta_x + \kappa_x\right)^2 + \sigma_{zz}^j \left(\beta_z + \kappa_z\right)^2 + 2\sigma_{xz}^j \left(\beta_x + \kappa_x\right)\left(\beta_z + \kappa_z\right)\right\}.$$

The (second-order approximate) manifestations of extrapolation bias are apparently more complex here as compared with the linear model. T_2 will equal 1 if the endogeneity-related parameter estimator inconsistency, $\kappa_{(.)}$, equals zero *or* if the covariate means in the prediction population j are equal to those in the estimation population A; otherwise, the magnitude of T_2 will be governed by the magnitudes of the large-sample parameter estimator biases *and* the degree of discrepancy between the respective populations' covariate means. In addition, the form of T_3 indicates that the magnitudes of the $\kappa_{(.)}$ will interact with the second moments of the population-j covariates, so that -- unlike the linear model -- more than just the first moments of the covariate distributions come into play in determining the properties of the estimator (22).

Finally, this discussion has assumed the existence of an estimated retransformation factor $\hat{\rho}_A$. Whether standard methods of estimation like homoskedastic smearing retain their consistency properties in the presence of endogenous covariates seems tenuous; definitive resolution of this issue awaits further research.

V. Conclusions

If population B is similar to population A in the sense that $\phi_B\left(x_B, z_B, u_B\right)$ resembles closely $\phi_A\left(x_A, z_A, u_A\right)$, then the utility of any econometric risk-adjustment exercise is limited: In this case, simply use a statistic like the marginal sample mean of y_A to predict $E[y_B]$. As such and unfortunately, the results presented here suggest that when risk-adjustment exercises are likely to be of greatest utility -- those circumstances where $\phi_B\left(x_B, z_B, u_B\right)$ is dissimilar to $\phi_A\left(x_A, z_A, u_A\right)$ -- they will be most prone to the extrapolation biases described above.

While the paper's algebra is not particularly novel, its application to this important realm of policy research and implementation raises issues that have not -- to the author's

knowledge -- been addressed explicitly in the risk-adjustment literature.[5]   Ultimately the paper's central results underscore the importance of a careful, reasoned approach to risk-adjustment exercises.   The use of endogenous adjusters might in some instances be defensible; for instance, it is imaginable that the inclusion of additional adjuster variables, even if endogenous, could result in a smaller prediction MSE than were they excluded.   Such considerations notwithstanding, the main message to be drawn here is that it can be dangerous to apply a risk-adjustment model estimated using data drawn from one population to a second population whose characteristics -- observed and unobserved -- differ importantly from those of the former population, particularly when maintaining econometric exogeneity of all the adjuster variables is tenuous.

---

[5] While this paper has focused on the specific case of healthcare expenditure risk adjustment, the analytical results presented above are likely to be applicable generally in situations where extrapolations of the sort considered here are of central concern.

References

Aitchison, J. and J.A.C. Brown. 1969. *The Lognormal Distribution with Special Reference to Its Uses in Economics*. Cambridge: Cambridge University Press.

Amemiya, T. 1985. *Advanced Econometrics*. Cambridge, MA: Harvard University Press.

Ettner, S.L. et al. 2001. "Risk Adjustment Alternatives in Paying for Behavioral Health Care under Medicaid." *Health Services Research* 36: 793-811.

Glazer, J. and T.G. McGuire. 2002. "Setting Health Plan Premiums to Ensure Efficient Quality in Health Care: Minimum Variance Optimal Risk Adjustment." *Journal of Public Economics* 84: 153-173.

Graybill, F.A. 1983. *Matrixes with Applications in Statistics, 2nd Edition*. Belmont, CA: Wadsworth.

Schokkaert, E. and C. Van de Voorde. 2004. "Risk Selection and the Specification of the Conventional Risk Adjustment Formula." *Journal of Health Economics* 23: 1237-1259.

Veazie, P.J. et al. 2003. "Improving Risk Adjustment for Medicare Capitated Reimbursement Using Nonlinear Models." *Medical Care* 41: 741-752.

Zhao, Y. et al. 2005. "Predicting Pharmacy Costs and Other Medical Costs Using Diagnoses and Drug Claims." *Medical Care* 31: 34-43.