Catalogue no. 11-522-XIE

Statistics Canada International Symposium Series - Proceedings

Symposium 2006 : Methodological Issues in Measuring Population Health

2006

Statistics Canada Statistique Canada

Canada

Proceedings of Statistics Canada Symposium 2006
Methodological Issues in Measuring Population Health

Estimation of Regression Parameters with Survey Data

Wayne A. Fuller and Yu Y. Wu[1]

Abstract

The coefficients of regression equations are often parameters of interest for health surveys and such surveys are usually of complex design with differential sampling rates. We give estimators for the regression coefficients for complex surveys that are superior to ordinary expansion estimators under the subject matter model, but also retain desirable design properties. Theoretical and Monte Carlo properties are presented.

KEY WORDS: Instrumental variables; probability weighting; complex surveys

1. Introduction

We consider estimation of regression coefficients using data collected under a complex survey design. It is assumed that the regression equation is part of a subject matter model that specifies the finite population to be generated by some stochastic mechanism, where the stochastic mechanism is called the superpopulation model. We use script $\mathcal{F}$ to denote the finite population, *U* to denote the set of indices of the finite population, and *A* to denote the set of indices of the sample. We assume that there is a function *p*(.) such that *p*(A) gives the probability of selecting sample *A* from *U*.

A superpopulation model for regression is

$$y_i = \mathbf{x}_i \boldsymbol{\beta} + e_i ,$$

where $(y_i, \mathbf{x}_i)$ are independent and identically distributed (*iid*) vectors, and e_i is independent of $\mathbf{x}_i$. Let a set of *N* vectors define a finite population. The model for the finite population can then be written as

$$\begin{aligned} \mathbf{y}_N &= \mathbf{X}_N \boldsymbol{\beta} + \mathbf{e}_N , \\ e_N &\sim (\mathbf{0}, \mathbf{I}_N \sigma^2), \end{aligned} \tag{1}$$

where $\mathbf{y}_N = (y_1, y_2, ..., y_N)'$ is the N dimensional vector of values for the dependent variable, $\mathbf{X}_N = (x_1', x_2', ..., x_N')'$ is the $N \times k$ matrix of values of the explanatory variables, and the error vector $\mathbf{e}_N = (e_1, e_2, ..., e_N)'$ is an N dimensional vector which is independent of $\mathbf{X}_N$. Assume a probability sample is selected from the finite population with selection probabilities π_i.

2. Estimators

The ordinary least squares estimator of $\boldsymbol{\beta}$ is

[1] Wayne A. Fuller is Distinguished Professor Emeritus, , Department of Statistics, Iowa State University, Ames, Iowa, U.S.A. 50011 (email: waf@iastate.edu); Yu Y. Wu, Department of Statistics, Iowa State University, Ames, Iowa, U.S.A. 50011 (email: yuwu@iastate.edu)

$$\hat{\boldsymbol{\beta}}_{ols} = \left(\sum_{i\in A}\mathbf{x}_i'\mathbf{x}_i\right)^{-1}\sum_{i\in A}\mathbf{x}_i' y_i = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}, \tag{2}$$

where $\mathbf{y} = (y_1, y_2, \ldots, y_n)'$ is the *n* dimensional column vector of observations, and $\mathbf{X} = (\mathbf{x}_1', \mathbf{x}_2', \ldots, \mathbf{x}_n')'$ is the $n \times k$ matrix of observations on the explanatory variables. The error in the ordinary least squares estimator is

$$\hat{\boldsymbol{\beta}}_{ols} - \boldsymbol{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{e}\,, \tag{3}$$

where $\mathbf{e}$ is the *n* dimensional vector of errors, and under mild assumptions

$$\hat{\boldsymbol{\beta}}_{ols} - \boldsymbol{\beta} = \left(\sum_{i\in U}\mathbf{x}_i'\pi_i\mathbf{x}_i\right)^{-1}\sum_{i\in U}\mathbf{x}_i'\pi_i e_i + O_p(n^{-1})\,. \tag{4}$$

Therefore, if $\mathbf{x}_i\pi_i$ and e_i are correlated, the OLS estimator is biased.

The probability weighted (PW) estimator, constructed with the inverses of the selection probabilities, is

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{PW} &= \left(\sum_{i\in A}\mathbf{x}_i'\pi_i^{-1}\mathbf{x}_i\right)^{-1}\sum_{i\in A}\mathbf{x}_i'\pi_i^{-1}y_i \\ &= (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\,\mathbf{X}'\mathbf{W}\mathbf{y},\end{aligned} \tag{5}$$

where $\mathbf{W} = diag(\pi_1^{-1}, \pi_2^{-1}, \ldots, \pi_n^{-1}) =: diag(w_1, w_2, \ldots, w_n)$. Under mild assumptions on the population and for many designs,

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{PW} - \boldsymbol{\beta} &= \left(\sum_{i\in U}\mathbf{x}_i'\mathbf{x}_i\right)^{-1}\sum_{i\in U}\mathbf{x}_i'e_i + O_p(n^{-1}) \\ &= O_p(n^{-1/2}).\end{aligned} \tag{6}$$

Thus $\hat{\boldsymbol{\beta}}_{PW}$ is a consistent estimator of the parameter of interest. Also see Fuller (2002).

Let the subject matter model specify vectors $\mathbf{q}_i$ such that

$$E\left\{\sum_{i\in A}\mathbf{q}_i'e_i\right\} = E\left\{\sum_{i\in U}\pi_i\mathbf{q}_i'e_i\right\} = \mathbf{0}. \tag{7}$$

We define an instrumental variable estimator by

$$\hat{\boldsymbol{\beta}}_{IV} = [(\mathbf{Q}'\mathbf{X})'\hat{\mathbf{V}}_{bb}^{-1}\mathbf{Q}'\mathbf{X}]^{-1}(\mathbf{Q}'\mathbf{X})'\hat{\mathbf{V}}_{bb}^{-1}\mathbf{Q}'\mathbf{y} \tag{8}$$

where $\hat{\mathbf{V}}_{bb}$ is a symmetric positive definite matrix. The preferred choice for $\hat{\mathbf{V}}_{bb}$ is an estimator of the variance of $\mathbf{Q}'\mathbf{e}$. As an example, let $\mathbf{x}_i = (1, \mathbf{x}_{1,i})$, let $\mathbf{q}_i = (w_i, w_i\mathbf{x}_{1i})$, where $w_i = \pi_i^{-1}$, and note that $E\{\mathbf{Q}'\mathbf{e}\} = \mathbf{0}$. If there is modest correlation between e_i and π_i, and if $e_i \sim ind\ (0, \sigma^2)$, then $\mathbf{V}_{bb} = V\{\mathbf{Q}'\mathbf{e}\} \doteq \mathbf{Q}'\mathbf{Q}\sigma^2$. The estimator (8) with $\mathbf{V}_{bb} = \mathbf{Q}'\mathbf{Q}\sigma^2$ is the two stage least squares estimator,

$$\hat{\boldsymbol{\beta}}_{IV} = (\hat{\mathbf{X}}'\hat{\mathbf{X}})^{-1}\hat{\mathbf{X}}'\mathbf{y}, \tag{9}$$

where $\hat{\mathbf{X}} = \mathbf{Q}(\mathbf{Q}'\mathbf{Q})^{-1}\mathbf{Q}'\mathbf{X}$. For example, see Wooldridge (2000). Using (9) and $\mathbf{q}_i = (w_i, w_i\mathbf{x}_{1,i})$, the instrumental variable estimator is

$$\hat{\boldsymbol{\beta}}_{IV} = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{y} \tag{10}$$

Thus the PW estimator is an instrumental variable estimator with $\mathbf{q}_i = (w_i, w_i\mathbf{x}_{1,i})$. The instrumental variable framework permits one to add instruments and to perform tests on potential instruments. For example, the Pfeffermann – Sverchkov (1999) estimator for the model with $\mathbf{x}_i = (1, \mathbf{x}_{1,i})$ is an instrumental variable estimator with $\mathbf{q}_i = (w_i / \hat{w}_i, w_i / \hat{w}_i \mathbf{x}_{1,i})$, where $\hat{w}_i$ is the least squares predictor of w_i based on $\mathbf{x}_i$.

In a number of situations it may be reasonable to believe that the selection probability is correlated with the error e_i, but that

$$E\{(\mathbf{x}_{1,i} - \bar{\mathbf{x}}_{1,N})'e_i \mid i \in A\} = \mathbf{0}. \tag{11}$$

Consider the hypothesis that the selection probabilities have the representation

$$\pi_i = g_1(\mathbf{x}_i) + g_2(e_i) + u_i, \tag{12}$$

where $g_1(.)$ and $g_2(.)$ are continuous differentiable functions and u_i is independent of $(\mathbf{x}_i, e_i)$. An example where model (12) is reasonable is one in which selection probabilities are related to a previous *y*-value. Given (12),

$$E\left\{\sum_{i\in A}(\mathbf{x}_{1,i} - \bar{\mathbf{x}}_{1,N})' e_i\right\} = E\left\{\sum_{i\in U}\pi_i(\mathbf{x}_{1,i} - \bar{\mathbf{x}}_{1,N})' e_i\right\} = \mathbf{0}, \tag{13}$$

because under model (1), e_i is independent of $\mathbf{x}_i$. It follows from (13) that (11) holds and that the estimator defined by

$$\left[\sum_{i\in A}\left[w_i, (\mathbf{x}_{1,i} - \bar{\mathbf{x}}_{1,N})\right]'(1, x_{1i})\right]\hat{\boldsymbol{\beta}}_{IV} = \sum_{i\in A}\left[w_i, (\mathbf{x}_{1,i} - \bar{\mathbf{x}}_{1,N})\right]' y_i \tag{14}$$

is consistent for $\boldsymbol{\beta}$. If $\bar{\mathbf{x}}_{1,N}$ is unknown, $\bar{\mathbf{x}}_{1,N}$ can be replaced with a consistent estimator.

To study tests for instruments, we partition the vector $\mathbf{q}_i$ as $(\mathbf{q}_{1i}, \mathbf{q}_{2i})$, where it is assumed that $E\{\Sigma_{i\in A}\mathbf{q}'_{1i}e_i\} = \mathbf{0}$ and we wish to test

$$E\left\{\sum_{i\in A}\mathbf{q}'_{2i}e_i\right\} = \mathbf{0}. \tag{15}$$

To test that $E\{\mathbf{Q}'_2\mathbf{e}\} = \mathbf{0}$, using the two-stage least squares estimator (9) as our basic estimator, we compute

$$\hat{\boldsymbol{\gamma}} = [(\hat{\mathbf{X}}, \mathbf{R}_2)'(\hat{\mathbf{X}}, \mathbf{R}_2)]^{-1}(\hat{\mathbf{X}}, \mathbf{R}_2)'\mathbf{y}, \tag{16}$$

where $\mathbf{R}_2 = \mathbf{Q}_2 - \mathbf{Q}_1(\mathbf{Q}'_1\mathbf{Q}_1)^{-1}\mathbf{Q}_2$ and $\mathbf{Q} = (\mathbf{Q}_1, \mathbf{Q}_2)$. If the finite population correction can be ignored, an estimated covariance matrix for $\hat{\boldsymbol{\gamma}}$ is

$$\hat{V}\{\hat{\boldsymbol{\gamma}}\} = [(\hat{\mathbf{X}}, \mathbf{R}_2)'(\hat{\mathbf{X}}, \mathbf{R}_2)]^{-1}\hat{V}\{(\hat{\mathbf{X}}, \mathbf{R}_2)'\mathbf{e}\}[(\hat{\mathbf{X}}, \mathbf{R}_2)'(\hat{\mathbf{X}}, \mathbf{R}_2)]^{-1}. \tag{17}$$

An estimator of $V\{(\hat{\mathbf{X}}, \mathbf{R}_2)'\mathbf{e}\}$ is the Horvitz-Thompson estimator calculated with $(\hat{\mathbf{X}}, \mathbf{R}_2)'\hat{\mathbf{e}}$, where $\hat{e}_i = y_i - (\mathbf{x}_i, \mathbf{r}_{2i})\hat{\boldsymbol{\gamma}}$ and $\mathbf{r}_{2i}$ is the *i*-th row of $\mathbf{R}_2$. Under the null hypothesis that $E\{\mathbf{Q}'_2\mathbf{e}\} = \mathbf{0}$, $\hat{\boldsymbol{\gamma}}_1$, the coefficient for $\hat{\mathbf{X}}$, is estimating $\boldsymbol{\beta}$ and $\hat{\boldsymbol{\gamma}}_2$, the coefficient for $\mathbf{r}_{2i}$, is estimating **0.** Therefore a test statistic is

$$F(k_2, n-k) = k_2^{-1}\hat{\boldsymbol{\gamma}}'_2\hat{\mathbf{V}}^{-1}_{\gamma\gamma 22}\hat{\boldsymbol{\gamma}}_2, \tag{18}$$

where $\hat{\mathbf{V}}_{\gamma\gamma 22}$ is the lower right $k_2 \times k_2$ block of $\hat{V}\{\hat{\boldsymbol{\gamma}}\}$, k is the dimension of $\mathbf{q}_i$ and k_2 is the dimension of $\mathbf{r}_{2i}$. Under the null hypothesis, the test statistic is approximately distributed as the tabulated *F*-distribution with k_2 and $n-k$ degrees of freedom.

3. Monte Carlo Study

A simulation study was used to assess the performance of two IV estimators and a two-step pretest estimator. The model is

$$y_i = \beta_0 + x_{1i}\beta_1 + e_i$$
$$= \mathbf{x}_i\boldsymbol{\beta} + e_i,$$

where $\mathbf{x}_i = (1, x_{1i})$.

We create each sample by generating the vector (x_{1i}, e_i, a_i, u_i), where x_{1i} is a normal (0, 0.5) random variable, e_i is a normal (0, 0.5) random variable, a_i is a normal (0, 0.5) random variable, u_i is a uniform (0, 1) random variable, and the variables x_{1i}, e_i, a_i, and u_i are mutually independent. The selection probability p_i is a function of x_{1i}, e_i and a_i,

$$p_i = p(x_{1i}, e_i, a_i) = 0.25r(x_{1i}) + 1.75r(\psi^{0.5}e_i + [1-\psi]^{0.5}a_i), \tag{19}$$

where

$$r(x) = \begin{cases} 0.025 & \text{if } x < 0.2 \\ 0.475(x-0.20)+0.025 & \text{if } 0.2 \le x \le 1.2 \\ 0.5 & \text{if } x > 1.2 \end{cases}$$

and ψ is a parameter that is varied in the experiment. The parameter ψ determines the correlation between p_i and e_i. If $u_i \le p_i$ the vector is retained for the sample; otherwise it is discarded.

The first IV estimator uses a vector of four instrumental variables, $\mathbf{z}_{1,i} = (w_i, w_i x_{1i}, w_i\hat{p}_i, w_i\hat{p}_i x_{1i})$, where $\hat{p}_i$ is the predicted value from the OLS regression of p_i on $(1, r(x_{1i}))$. The second IV estimator is based on the vector $\mathbf{z}_{2,i} = (w_i, w_i x_{1i}, w_i\hat{p}_i, w_i\hat{p}_i x_{1i}, x_{1i})$. The IV estimators of $\boldsymbol{\beta}$ are

$$\hat{\boldsymbol{\beta}}_{IVj} = [\mathbf{X}'\mathbf{Z}_j(\mathbf{Z}'_j\mathbf{Z}_j)^{-1}\mathbf{Z}'_j\mathbf{X}]^{-1}\mathbf{X}'\mathbf{Z}_j(\mathbf{Z}'_j\mathbf{Z}_j)^{-1}\mathbf{Z}'_j\mathbf{y}, \tag{20}$$

where $\mathbf{Z}_j = (\mathbf{z}'_{j,1}, \mathbf{z}'_{j,2}, \ldots, \mathbf{z}'_{j,n})'$, for $j = 1, 2$. The estimated covariance matrix of $\hat{\boldsymbol{\beta}}_{IVj}$ is

$$\hat{V}(\hat{\boldsymbol{\beta}}_{IVj}) = (\hat{\mathbf{X}}'_j\hat{\mathbf{X}}_j)^{-1}\hat{\mathbf{X}}'_j\hat{\mathbf{D}}_{ee,IVj}\hat{\mathbf{X}}_j(\hat{\mathbf{X}}'_j\hat{\mathbf{X}}_j)^{-1}, \tag{21}$$

where $\hat{\mathbf{X}}_j = \mathbf{Z}_j(\mathbf{Z}'_j\mathbf{Z}_j)^{-1}\mathbf{Z}'_j\mathbf{X}$, $\hat{\mathbf{D}}_{ee,IVj} = diag(\hat{e}^2_{1,IVj}, \hat{e}^2_{2,IVj}, \ldots, \hat{e}^2_{n,IVj})$, and $\hat{e}_{i,IVj} = y_i - \mathbf{x}_i\hat{\boldsymbol{\beta}}_{IVj}$.

The pretest estimator is a two-step estimator based on the OLS estimator and the two IV estimators. The first test is a test for the importance of weights and is obtained from two regressions: the regression of y_i on $(1, x_{1i}, w_i, w_i x_{1i})$ (full model) and the regression of y_i on $(1, x_{1i})$ (reduced model). If the *F* statistic

$$F^2_{n-4} = \frac{(SSE_{red} - SSE_{full})/2}{MSE_{full}} \tag{22}$$

is not statistically significant, $\hat{\boldsymbol{\beta}}_{ols}$ is the estimator, otherwise a second test is performed. The OLS regression of y_i on $(\tilde{\mathbf{x}}_i, x_{1i} - \hat{x}_{1i})$ is computed, where $\tilde{\mathbf{x}}_i$ is the predicted value from the OLS regression of $\mathbf{x}_i$ on $\mathbf{z}_{2,i}$, and $\hat{x}_{1i}$ is the predicted value from the OLS regression of x_{1i} on $\mathbf{z}_{1,i}$. The test statistic for $\gamma_2 = 0$ is defined in (18) where γ_2 is the OLS coefficient for $x_{1i} - \hat{x}_{1i}$. Because γ_2 is a scalar, let t^2 denote the statistic of (18). Then the two-step pretest estimator is

$$\hat{\boldsymbol\beta}_{pre} = \begin{cases} \hat{\boldsymbol\beta}_{ols} & \text{if } F < F_{2,n-4}(\alpha) \\ \begin{cases} \hat{\boldsymbol\beta}_{IV2} & \text{if } |t| < Z(\alpha/2) \\ \hat{\boldsymbol\beta}_{IV1} & \text{if } |t| \geq Z(\alpha/2) \end{cases} & \text{and } F \geq F_{2,n-4}(\alpha), \end{cases} \tag{23}$$

where α is the size of the test.

The standard error for $\hat{\boldsymbol\beta}_{pre}$ uses the variance estimation procedure appropriate for the estimator chosen. An estimated variance is

$$\hat{V}\{\hat{\boldsymbol\beta}_{pre}\} = \begin{cases} \hat{V}\{\hat{\boldsymbol\beta}_{ols}\} & \text{if } F < F_{2,n-4}(\alpha) \\ \begin{cases} \hat{V}\{\hat{\boldsymbol\beta}_{IV2}\} & \text{if } |t| < Z(\alpha/2) \\ \hat{V}\{\hat{\boldsymbol\beta}_{IV1}\} & \text{if } |t| \geq Z(\alpha/2) \end{cases} & \text{and } F \geq F_{2,n-4}(\alpha), \end{cases} \tag{24}$$

where $\hat{V}\{\hat{\boldsymbol\beta}_{IVj}\}$ is defined in (21). The estimated variance $\hat{V}\{\hat{\boldsymbol\beta}_{pre}\}$ is not an unbiased variance estimator. We call the statistic

$$t_{\beta_{pre,m}} = [\hat{V}\{\hat{\beta}_{pre,m}\}]^{-1/2}(\hat{\beta}_{pre,m} - \beta_m) \tag{25}$$

for β_m, $m = 0, 1,$ the t statistic, although distribution of the statistic is not that of Student's t.

Table 1 and Table 2 contain the mean squared error for the estimators. A sample was created by generating 1000 vectors giving an expected sample size of 221. The pretest estimators used $\alpha = 0.10$. The second column of Table 1 is the correlation between e_i and p_i. For the modest correlation of 0.077 associated with a ψ of 0.01, ordinary least squares is inferior to the PW estimator. The IV1 estimator is more efficient than the PW estimator, because the IV1 estimator contains more instrumental variables than the PW estimator. The IV2 estimator is appropriate for our data generation mechanism and uses the most information. Therefore, the IV2 estimator is always superior to the IV1 estimator. The mean squared errors of the pretest estimator are between the mean squared error of the OLS estimator and the mean squared error of the IV1 estimator. As ψ gets larger, the mean squared errors of the pretest estimator become closer to the mean squared errors of the IV1 estimator, because the pretest procedure rejects the null hypothesis more frequently as the correlation between p_i and e_i increases.

Table 1: Monte Carlo Mean Squared Error (×1000) for estimators of β_0 (10,000 samples)

ψ	Corr. (p_i, e_i)	$\hat{\beta}_{ols,0}$	$\hat{\beta}_{Pw,0}$	$\hat{\beta}_{IV1,0}$	$\hat{\beta}_{IV2,0}$	$\hat{\beta}_{pre,0}$ $\alpha = 0.10$
0	0.000	2.33	5.92	5.71	5.33	3.39
.01	0.077	6.77	5.71	5.55	5.14	6.97
.02	0.108	10.82	5.75	5.53	5.10	8.94
.05	0.171	23.94	5.60	5.41	4.99	9.35
.07	0.203	32.45	5.65	5.47	5.02	8.01
.10	0.243	45.11	5.58	5.42	5.06	6.55
.20	0.343	88.22	5.67	5.55	5.18	5.41
.30	0.420	131.22	5.44	5.34	4.89	5.11
.50	0.542	217.28	5.26	5.23	4.88	5.07

Table 2: Monte Carlo Mean Squared Error (×1000) for estimators of β_1 (10,000 samples)

ψ	$\hat{\beta}_{ols,1}$	$\hat{\beta}_{PW,1}$	$\hat{\beta}_{IV1,1}$	$\hat{\beta}_{IV2,1}$	$\hat{\beta}_{pre,1}$ $\alpha = 0.10$
0	4.16	9.62	8.53	4.29	5.12
.01	4.30	9.87	8.61	4.32	5.61
.02	4.41	9.71	8.63	4.32	5.93
.05	4.66	9.54	8.49	4.34	6.18
.07	4.94	9.80	8.64	4.46	6.49
.10	5.32	9.69	8.57	4.58	6.52

.20	6.47	9.48	8.39	4.84	6.56
.30	7.91	9.30	8.25	5.20	6.66
.50	10.29	9.10	8.25	5.76	6.97

As the simulation results of Table 3 and Table 4 illustrate, almost all *t*-statistics exceed the tabular $t_{.025}$ value for Student's *t*. One should remember that there is a wide range of selection probabilities so that the variance of the variance estimator is greater than that of a simple random sample. The performance of the test statistic is generally better for the IV1 estimator than for the other estimators. As expected, the *t*-statistic for the pretest estimator is very biased for the true intercept near 1.5 standard deviations of the estimator. As ψ increases, $P(|t_{\beta_{pre,0}}| > t_{.025})$ approaches $P(|t_{\beta_{IV1,0}}| > t_{.025})$.

Table 3: Monte Carlo Probability that $|t_{\beta_0}| > t_{.025}$ (10,000 samples)

ψ	$\hat{\beta}_{ols}$	$\hat{\beta}_{PW}$	$\hat{\beta}_{IV1}$	$\hat{\beta}_{IV2}$	$\hat{\beta}_{pre}$ $\alpha = 0.10$
0	0.049	0.058	0.057	0.053	0.065
.01	0.282	0.065	0.066	0.061	0.237
.02	0.486	0.061	0.061	0.057	0.320
.05	0.870	0.055	0.056	0.051	0.247
.07	0.950	0.065	0.065	0.062	0.167
.10	0.990	0.059	0.059	0.055	0.086
.20	1.000	0.058	0.060	0.055	0.059
.30	1.000	0.059	0.064	0.059	0.063
.50	1.000	0.060	0.064	0.060	0.065

Table 4: Monte Carlo Probability that $|t_{\beta_1}| > t_{.025}$ (10,000 samples)

ψ	$\hat{\beta}_{ols}$	$\hat{\beta}_{PW}$	$\hat{\beta}_{IV1}$	$\hat{\beta}_{IV2}$	$\hat{\beta}_{pre}$ $\alpha = 0.10$
0	0.049	0.069	0.066	0.051	0.063
.01	0.054	0.073	0.070	0.055	0.072
.02	0.057	0.073	0.068	0.053	0.074
.05	0.072	0.070	0.065	0.056	0.080
.07	0.077	0.070	0.068	0.057	0.085
.10	0.083	0.073	0.069	0.054	0.081
.20	0.119	0.071	0.067	0.052	0.074
.30	0.154	0.076	0.072	0.053	0.076
.50	0.233	0.074	0.072	0.054	0.070

Acknowledgements

This work was partially supported by Cooperative Agreement No. 68-3A75-4-122 between the USDA Natural Resources Conservation Service and the Center for Survey Statistics and Methodology at Iowa State University.

References

Fuller, W. A. (2002), "Regression estimation for survey samples", *Survey Methodology,* 28:5-23.

Pfeffermann, D., and M. Sverchkov (1999), "Parametric and semi-parametric estimation of regression models fitted to survey data", *Sankhyā Series B: The Indian Journal of Statistics,* 61:166-186.

Wooldridge, J. M. (2000), *Introductory Econometrics: A Modern Approach,* South-Western Educational Publishing.