QSEP RESEARCH INSTITUTE FOR QUANTITATIVE STUDIES IN ECONOMICS AND POPULATION

# ORDINARY LEAST SQUARES BIAS AND BIAS CORRECTIONS FOR *iid* SAMPLES

LONNIE MAGEE

QSEP Research Report No. 419

McMaster University
SOCIAL SCIENCES

# ORDINARY LEAST SQUARES BIAS AND BIAS CORRECTIONS FOR *iid* SAMPLES

**LONNIE MAGEE**

**QSEP Research Report No. 419**

**June 2007**

**Lonnie Magee is a QSEP Research Associate and a faculty member in the McMaster University Department of Economics.**

This report is cross-listed as No. 207 in the McMaster University SEDAP Research Paper Series.

# Ordinary Least Squares Bias and Bias Corrections for *iid* Samples

Lonnie Magee
Department of Economics
McMaster University
magee@mcmaster.ca [*]

**Abstract:**
The $O(n^{-1})$ bias and $O(n^{-2})$ MSE of OLS are derived for *iid* samples. An approach is suggested for handling nonexistent finite sample moments. Bias corrections based on plug-in, weighting, jackknife and pairs bootstrap methods are equal to $O_p(n^{-3/2})$. Sometimes they are effective at lowering bias and MSE, but not always. In simulations, the bootstrap correction removes more bias than the others, but has a higher MSE. A hypothesis test is given for the presence of this bias. The techniques are applied to survey data on food expenditure, and the estimated bias is small and statistically insignificant.

**Key words:** OLS bias; finite sample moments; Nagar approximation; bias correction; pairs bootstrap

**JEL Classification:** C13, C29, C49

[*] I would like to thank Jeff Wooldridge and Jeff Racine for helpful advice, and Tom Crossley and Yuqian Lu for supplying the food expenditure data. This research was supported by the Social and Economic Dimensions of an Aging Population II (SEDAP) Research Program — a SSHRC MCRI grant with Byron Spencer the Principal Investigator.

# 1 Introduction

This paper examines the $O(n^{-1})$ bias and $O(n^{-2})$ mean squared error (MSE) of the most common estimator, OLS, in one of the most common statistical frameworks, *iid* sampling. This bias is different from the more familiar $O(1)$ omitted-variables or endogenous-regressor bias. It is caused by correlations between the linear regression disturbances and nonlinear functions of the regressor variables $x$ that arise from misspecification of the conditional mean, $E(y|x)$. Section 5.1 contains a simple illustration.

The parameter of interest is $\beta = E(xx')^{-1}E(xy)$. $\beta$ minimizes the mean square specification error $E(x'\beta^* - E(y|x))^2$ (Goldberger (1991), Poirier (1995, p.69)). It is a weighted average of the gradients $\partial E(y|x)/\partial x$ (Peracchi (2001, p.46)). Angrist *et al.* (2006, p.540) state "OLS provides a meaningful and well-understood summary statistic for conditional expectations under almost all circumstances."

OLS may not have finite sample moments even with finite population moment assumptions (e.g. Schmidt (1976, pp.93-96)). *iid* sampling often leaves a non-zero probability of a singular $X'X$ matrix. In this paper, restricted OLS is used when $X'X$ is singular or nearly so. The resulting modified OLS estimator has finite-sample moments while retaining the same $O_p(n^{-3/2})$ expansion as the un-modified OLS. As a result, Nagar (1959) approximations are valid for the modified estimator.

Section 2 presents the model, the $O(n^{-1})$ bias, and the finite-moment modification. In Section 3, bias corrections are suggested, and MSE approximations are given. The effect of bias correction on the MSE is studied for special cases. A procedure is suggested for testing the null hypothesis that the $O(n^{-1})$ bias is zero in Section 4. Section 5 contains examples and numerical evaluations. Section 6 illustrates these bias corrections and tests with an application to data on food consumption used by Crossley and Lu (2004). Remarks are collected in Section 7, and Section 8 summarizes. Proofs and other details are in the Appendix.

# 2 OLS Bias

## 2.1 Notation and Assumptions

$y$ is a scalar random variable and $x$ is a $K$-element random vector. $Exx' = A$ is nonsingular. Let

$$\beta = A^{-1}E(xy) \tag{1}$$

$y$, $x$ and $\beta$ are related as

$$y = x'\beta + v \tag{2}$$

The disturbance $v$ satisfies the unconditional moment restriction $Exv = 0$ due to (1), but the conditional mean of $v$ is nonzero,

$$E(v|x) = E(y|x) - x'\beta \;\neq 0 \tag{3}$$

The data, $(y_1, x_1), (y_2, x_2), \ldots, (y_n, x_n)$ are *iid* draws of $(x, y)$, generated according to an unknown population distribution. The OLS estimator is

$$b = (\sum_i x_i x_i')^{-1} \sum_i x_i y_i = \hat{A}^{-1}(n^{-1} \sum_i x_i y_i) \tag{4}$$

where $\hat{A} = n^{-1} \sum_i x_i x_i'$. $b$ is a consistent estimator of $\beta$ (White (1984, p.17)). In finite samples, $\hat{A}$ may be singular and $b$ may be undefined.

## 2.2 Existence of Finite-Sample Mean and Variance: Modified OLS

Let $x = [\ 1 \quad d\ ]'$ where $\text{Prob}(d = 1) = p$ and $\text{Prob}(d = 0) = 1 - p$. $\hat{A}$ is singular when the sampled $d_i$'s all equal 0 or all equal 1. This occurs with probability $p^n + (1 - p)^n$, which is strictly positive for finite $n$ and any $p \in [0, 1]$. In practice (the `reg` command in Stata for example) a singular $\hat{A}$ often is managed by deleting regressors. Similarly, define a modified OLS estimator as

$$b_M = \begin{cases} b_R & \text{if } \lambda_1(\hat{A}) < \tau \\ b & \text{otherwise} \end{cases} \tag{5}$$

where $0 < \tau < \lambda_1(A)$, $\lambda_1(A)$ is the smallest characteristic root of $A$, and $b_R$ is any $K \times 1$ vector having a finite mean and variance given $\lambda_1(\hat{A}) < \tau$. In the dummy variable regressor example given above, $b_R$ might be $b_R = [\ \bar{y} \quad 0\ ]'$ (if $\bar{y}$ has a finite mean and variance) or even $b_R = 0$. $\tau$ must be small enough to assure $\tau < \lambda_1(A)$. For example, set $\tau$ just large enough to account for numerical computing precision. This truncation rule is similar to that suggested for bootstrap MSE estimation by Liu and Singh (1992, p.382). Their rule uses the determinant instead of $\lambda_1$.

Theorem 1(a) describes assumptions sufficient for the existence of $b_M$'s finite-sample mean and variance. Sufficient conditions for $b_M - b = 0 + o_p(n^{-3/2})$ are given in Theorem 1(b).

**Theorem 1**

(a) *If the moments $E(x^2y^2)$ are finite, then $E(b_M)$ and $Var(b_M)$ are finite.*

(b) *If the moments $E(x^5)$ are finite, and $A$ has a distinct minimum characteristic root $\lambda_1(A)$, then $b_M = b + o_p(n^{-3/2})$.*

Theorem 1(a) implies that the $O(n^{-1})$ bias and $O(n^{-2})$ MSE approximations for $b_M$ are valid in Sargan's (1974) sense. Theorem 1(b) implies that these approximations are the same as for $b_M$ as for $b$.

## 2.3 OLS Expansion

To reduce notation, let $H$ be a matrix satisfying $HH' = A^{-1}$ and $H'AH = I$. Then

$$\begin{aligned} y &= x'\beta + v \\ &= z'(H^{-1}\beta) + v \end{aligned}$$

where $z = H'x$ and $Ezz' = I$. The OLS estimator of $H^{-1}\beta$ is

$$H^{-1}b = (\sum_i z_i z_i')^{-1} \sum_i z_i y_i = H^{-1}\beta + (n^{-1}\sum_i z_i z_i')^{-1}(n^{-1}\sum_i z_i v_i) \tag{6}$$

Let

$$\Delta_{zz} = n^{-1}\sum_i z_i z_i' - I \quad \text{and} \quad \Delta_{xv} = n^{-1}\sum_i z_i v_i$$

$\Delta_{zz}$ and $\Delta_{zv}$ are the $O_p(n^{-1/2})$ differences between the sample moments appearing in (6) and their population counterparts. Substituting in (6) and expanding gives

$$\begin{aligned} H^{-1}b - H^{-1}\beta &= (I + \Delta_{zz})^{-1}\Delta_{zv} \\ &= \Delta_{zv} - \Delta_{zz}\Delta_{zv} + \Delta_{zz}^2\Delta_{zv} + O_p(n^{-2}) \end{aligned} \tag{7}$$

When appearing in an expectation, a term such as $z_i$ represents the $i^{th}$ random variable $z$ rather than its realized value. The expectations of the terms in (7) are

$$\begin{aligned} E\Delta_{zv} &= En^{-1}\sum_i z_i v_i = Ezv = 0 \\ E\Delta_{zz}\Delta_{zv} &= En^{-2}\sum_{ij}(z_i z_i' - I)(z_j v_j) \\ &= n^{-2}E\sum_i z_i(z_i' z_i)v_i - n^{-1}E\sum_i z_i v_i \\ &= n^{-1}E(z'z)zv - 0 \\ &= n^{-1}\gamma \\ E\Delta_{zz}^2\Delta_{zv} &= En^{-3}\sum_{ijk}(z_i z_i' - I)(z_j z_j' - I)(z_k v_k) \\ &= n^{-2}E(z'z)^2 zv = 0 + O(n^{-2}) \end{aligned}$$

where $\gamma = E(z'z)zv$.

## 2.4 OLS Bias to $O(n^{-1})$

The $O(n^{-1})$ bias of $H^{-1}b$ is then

$$E(H^{-1}b) - H^{-1}\beta = -n^{-1}\gamma + O(n^{-2}) \tag{8}$$

In the original parametrization (1) to (4), the result is

$$E(b) - \beta = -n^{-1}A^{-1}\gamma^* + O(n^{-2}) \tag{9}$$

where $\gamma^* = E(x'A^{-1}x)xv$.

# 3 Bias Corrections and $O(n^{-2})$ Mean Squared Errors

## 3.1 Bias Corrections

A consistent estimator of $n$ times the $O(n^{-1})$ bias term from (9) is

$$n\widehat{\text{Bias}(b)} = -(n\hat{A})^{-1}\sum_i (np_{ii})x_i e_i \tag{10}$$

which uses

$$\begin{aligned}\hat{A}^{-1} &= A^{-1} + o_p(1)\\ np_{ii} &= x_i'A^{-1}x_i + o_p(1)\\ v_i &= e_i + o_p(1)\end{aligned}$$

where $p_{ii} = x_i'(\sum_i x_i x_i')^{-1}x_i = n^{-1}x_i'\hat{A}^{-1}x_i$ and $e_i = y_i - x_i'b$. A plug-in bias-corrected estimator (BC) is then

$$b_{c1,M} = \begin{cases} b_R & \text{if } \lambda_1(\hat{A}) < \tau \\ b_{c1} & \text{otherwise, where}\end{cases} \tag{11}$$

$$b_{c1} = b - \widehat{\text{Bias}(b)} = b + (\sum_i x_i x_i')^{-1}\sum_i x_i p_{ii} e_i$$

Another BC is the pairs bootstrap correction

$$b_{boot,M} = \begin{cases} b_R & \text{if } \lambda_1(\hat{A}) < \tau \\ b_{boot} & \text{otherwise, where}\end{cases} \tag{12}$$

$$\begin{aligned}& b_{boot} = 2b - \bar{b}_B\\ & \bar{b}_B = B^{-1}\sum_{\ell=1}^{B} b_{B\ell,M} \quad \text{and}\\ & b_{B\ell,M} = \begin{cases} b & \text{if } \lambda_1(\hat{A}_\ell) < \tau \\ b_{B\ell} & \text{otherwise, and}\end{cases}\\ & b_{B\ell} = \hat{A}_\ell^{-1}n^{-1}\sum_i x_{\ell i}y_{\ell i}\\ & \hat{A}_\ell = n^{-1}\sum_i x_{\ell i}x_{\ell i}'\end{aligned}$$

$b_{c1,M}$ and $b_{boot,M}$ have been modified in the same way as was $b_M$. Each $b_{B\ell,M}$ is constructed from a pairs bootstrap sample, indexed by $\ell$. The $i^{th}$ observation of the $\ell^{th}$ pairs bootstrap sample, $(x_{\ell i}, y_{\ell i})$, is sampled with replacement from the original data. The $b_{B\ell,M}$'s are modified using the same truncation rule as in (5), applied to $\hat{A}_\ell$.

Theorem 2 is the counterpart to Theorem 1 for $b_{c1,M}$ and $b_{boot,M}$.

**Theorem 2**

(a) *If the moments $E(x^8y^2)$ are finite, then $E(b_{c1,M})$ and $Var(b_{c1,M})$ are finite. If the moments $E(x^2y^2)$ are finite, then $E(b_{boot,M})$ and $Var(b_{boot,M})$ are finite.*

(b) *If the moments $E(x^5)$ are finite, and $\lambda_{min}(A)$ is a distinct minimum characteristic root, then $b_{c1,M} = b_{c1} + o_p(n^{-3/2})$ and $b_{boot,M} = b_{boot} + o_p(n^{-3/2})$.*

Three other BCs that are the same to $O_p(n^{-3/2})$ are

$$\begin{aligned} b_{c2} &= (\sum_i x_i x_i'(1+p_{ii}))^{-1} \sum_i x_i y_i (1+p_{ii}) &(13)\\ b_{c3} &= (\sum_i x_i x_i'(1-p_{ii})^{-1})^{-1} \sum_i x_i y_i (1-p_{ii})^{-1} &(14)\\ b_{jack} &= b + (n-1)n^{-1}(\sum_i x_i x_i')^{-1} \sum_i (1-p_{ii})^{-1} x_i e_i &(15)\end{aligned}$$

$b_{c2}$ and $b_{c3}$ are weighted least squares, and $b_{jack}$ is the standard or balanced jackknife estimator (Hinkley (1977, equation (2.4)).

## 3.2 $O_p(n^{-3/2})$ expansions

The expansion shared by the five BCs defined above is given in Theorem 3.

**Theorem 3**

$$\begin{aligned} H^{-1}b_c - H^{-1}\beta &= \Delta_{zv} - (\Delta_{zz}\Delta_{zv} - n^{-1}\gamma) \\ &\quad + n^{-1}(\Delta_\gamma - \Delta_{zz}\gamma - n^{-1}\sum_i z_i(z_i'\Delta_{zz}z_i)v_i - \Omega_{z'z}\Delta_{zv}) + o_p(n^{-3/2}) \qquad (16)\end{aligned}$$

*where $\Omega_{z'z} = E(zz')(z'z)$. $b_c$ refers to any of $b_{c1}, b_{c1,M}, b_{c2}, b_{c3}$, or $b_{jack}$. $b_c$ includes $b_{boot}$ and $b_{boot,M}$ if $B^{-1}$ in (12) is $o(n^{-2})$.*

The condition on $B$ reduces the bootstrap sampling error to $o_p(n^{-3/2})$.

## 3.3 $O(n^{-2})$ Mean Squared Errors

Let $\Omega_{g(v,z)} \equiv E(zz')g(v,z)$, where $g(v,z)$ is some scalar random variable. $z_a$ and $\gamma_a$ denote the $a^{th}$ elements of $z$ and $\gamma$.

**Theorem 4**

*Given the assumptions in Theorems 1 and 2,*

(a)

$$\begin{aligned} \mathrm{MSE}(H^{-1}b) &= n^{-1}\Omega_{v^2} + n^{-2}(\gamma\gamma' + \Omega_{z'\Omega_{v^2}z} - \Omega_{v^2} + 3\sum_{a=1}^{K}\Omega^2_{z_a v} + \Omega_{v^2}\Omega_{z'z} + \Omega_{z'z}\Omega_{v^2} \\ &\quad - 2\Omega_{z'zv^2} + 2\sum_{a=1}^{K}\Omega_{z_a v}\gamma_a + o(n^{-2}) \end{aligned} \qquad (17)$$

(b)

$$\mathrm{MSE}(H^{-1}b_{c1}) = \mathrm{MSE}(H^{-1}b_{boot}) = n^{-1}\Omega_{v^2} + n^{-2}(\Omega_{z'\Omega_{v^2}z} + \sum_{a=1}^{K}\Omega^2_{z_a v} - \Omega_{v^2}) + o(n^{-2}) \qquad (18)$$

Letting $b_c$ refer to $b_{c1}$ or $b_{boot}$, the difference between (18) and (17) is

$$\begin{aligned} &\mathrm{MSE}(H^{-1}b_c) - \mathrm{MSE}(H^{-1}b) \\ = &\ -n^{-2}(\gamma\gamma' + 2\sum_{a=1}^{K}\Omega^2_{z_a v} + 2\sum_{a=1}^{K}\Omega_{z_a v}\gamma_a + \Omega_{v^2}\Omega_{z'z} + \Omega_{z'z}\Omega_{v^2} - 2\Omega_{z'zv^2}) + o(n^{-2}) \end{aligned} \qquad (19)$$

## 3.4 Special Cases

The $\Omega$ matrices and $\gamma\gamma'$ in (19) all are nonnegative. The minus sign on the very last term rules out a general result about the BCs reducing the $O(n^{-2})$ MSE. That last term can dominate the others, and the BCs can increase MSE. To gain some intuition on when the bias is large and when the BCs do not reduce MSE, two special cases are considered.

### 3.4.1 No misspecification or heteroskedasticity

Let $E(v|x) = 0$, then $Eb = \beta$. Let $E(v^2|x) = \sigma^2$ for all $x$, The terms in (19) simplify to $\gamma = 0$, $\Omega_{z_{(a)}v} = 0$, $\Omega_{v^2} = \sigma^2 I$, and $\Omega_{z'zv^2} = \sigma^2\Omega_{z'z}$. (19) becomes

$$\begin{aligned} \mathrm{MSE}(H^{-1}b_c) - \mathrm{MSE}(H^{-1}b) &= -n^{-2}(0 + 0 + 0 + \sigma^2\Omega_{z'z} + \sigma^2\Omega_{z'z} - 2\sigma^2\Omega_{z'z}) + o(n^{-2}) \\ &= 0 + o(n^{-2}) \end{aligned} \qquad (20)$$

Thus the BCs have no effect on the $O(n^{-2})$ MSE when there is no misspecification or heteroskedasticity.

### 3.4.2 Single regressor with misspecification and heteroskedasticity

Let $x$ be scalar, with moments $Ex^\omega \equiv \mu_\omega$, standardized so that $\mu_1 = 0$ and $\mu_2 = 1$. Then $z = x$ and $Exx' = H = 1$. Let the conditional mean of the disturbance be $E(v|z) = \alpha(z^\omega - \mu_{\omega+1}z)$ where $\omega$ is a nonnegative integer. This model satisfies $Evz = 0$ although $E(v|z) \neq 0$ unless $\alpha = 0$ or $\omega = 1$. Before specifying heteroskedasticity, note that $E(v^2|z)$ is the sum of two components, $E(v|z)^2$ and $E((v - E(v|z))^2|z)$. $E(v|z)^2$ is the square of the error in specification of the mean that already has been specified. The second component is the conditional variance of the disturbance $\epsilon$ in the well-specified regression $y = E(y|z) + \epsilon$. Let it be

$E((v-E(v|z))^2|z) = \sigma^2(1+\theta z^{2\phi})$ where $\theta \geq 0$ and $\phi$ is a non-negative integer. The terms in (19) become

$$\begin{aligned}
\gamma &= Ez^3v = \alpha(\mu_{\omega+3} - \mu_{\omega+1}\mu_4) \\
\sum_{a=1}^{K}\Omega^2_{z_a v} &= E(z^3v)^2 = \gamma^2 \\
\Omega_{v^2} &= E(z^2v^2) = \alpha^2(\mu_{2\omega+2} - 2\mu_{\omega+1}\mu_{\omega+3} + \mu^2_{\omega+1}\mu_4) + \sigma^2(1+\theta\mu_{2\phi+2}) \\
\Omega_{z'z} &= E(z^4) = \mu_4 \\
\sum_{a=1}^{K}\Omega_{z_a v}\gamma_a &= E(z^3v)\gamma = \gamma^2 \\
\Omega_{z'zv^2} &= E(z^4v^2) = \alpha^2(\mu_{2\omega+4} - 2\mu_{\omega+1}\mu_{\omega+5} + \mu^2_{\omega+1}\mu_6) + \sigma^2(\mu_4 + \theta\mu_{2\phi+4})
\end{aligned}$$

Applying (8) or (9), the OLS bias is

$$E(b) - \beta = -n^{-1}\alpha(\mu_{\omega+3} - \mu_{\omega+1}\mu_4) + O(n^{-2})$$

Not surprisingly, the bias is larger when $n$ is small and when the scale of the misspecification, $\alpha$, is large. The third factor, $\mu_{\omega+3} - \mu_{\omega+1}\mu_4$, depends on higher moments of $z$ through the function

$$q_\rho = \mu_{\rho+2} - \mu_\rho\mu_4, \quad \rho = 1, 2, \ldots$$

evaluated at $\rho = \omega + 1$. At one extreme, when $z$ is symmetric and $\omega$ is even, then $\mu_{\omega+3} = \mu_{\omega+1} = 0$. Even though $E(v|z)$ may differ from zero, it is not correlated with $z^3$, so the $O(n^{-1})$ bias is zero. At the other extreme, when $\omega$ is large and odd, and the probability function of $z$ has thick enough tails, then $\mu_{\omega+3} \gg \mu_{\omega+1}$, leading to a large bias. For example, if $z \sim N[0,1]$, then $\mu_{2r} = \frac{(2r)!}{2^r r!}$ and

$$q_{\omega+1} = \mu_{\omega+3} - \mu_{\omega+1}\mu_4 = \frac{(\omega+1)!}{2^{(\omega+1)/2}(\frac{\omega+1}{2})!}(\omega-1) = \begin{cases} q_4 = 6 & \text{when } \omega = 3 \\ q_6 = 60 & \text{when } \omega = 5 \\ q_8 = 630 & \text{when } \omega = 7 \end{cases} \tag{21}$$

Turning to the MSE comparison, (19) becomes

$$\begin{aligned}
n^2(\mathrm{MSE}(b_c) - \mathrm{MSE}(b)) &= -(\gamma^2 + 2\gamma^2 + 2\gamma^2 + \Omega_{v^2}\mu_4 + \mu_4\Omega_{v^2} - 2\Omega_{z'zv^2}) + o(1) \\
&= -(5\gamma^2 + 2(\mu_4\Omega_{v^2} - 2\Omega_{z'zv^2})) + o(1) \\
&= \alpha^2(-5(\mu_{\omega+3} - \mu_{\omega+1}\mu_4)^2 + 2(\mu_{2\omega+4} - \mu_{2\omega+2}\mu_4) \\
&\quad -4\mu_{\omega+1}(\mu_{\omega+5} - \mu_{\omega+3}\mu_4) + 2\mu^2_{\omega+1}(\mu_6 - \mu_4^2)) + \sigma^2\theta(\mu_{2\phi+4} - \mu_{2\phi+2}\mu_4) + o(1) \\
&= \alpha^2\left(-5q^2_{\omega+1} + 2q_{2\omega+2} - 4\mu_{\omega+1}q_{\omega+3} + 2\mu^2_{\omega+1}q_4\right) + \sigma^2\theta q_{2\phi+2} + o(1)
\end{aligned} \tag{22}$$

Setting $\alpha = \theta = 0$ recalls result (20).

Two cases can be identified from (22) in which BC will increase MSE. One is the large-bias situation mentioned above. Let $\theta = 0$. When $\omega$ is odd, the arguments in the $q$ functions are even. $q_{2\rho}$ increases dramatically with $\rho$ when $z \sim N[0,1]$, as shown in (21). With thicker-than-normal-tailed $z$ distributions, this increase would be even more dramatic. The $q_{2\omega+2}$ term has the largest argument, hence it may dominate the other terms in (22), and it has a positive sign.

A second situation in which BC increases MSE involves heteroskedasticity. Let $\alpha = 0$. When $\theta > 0$ and $\phi > 0$, the last term of (22) shows that $\mathrm{MSE}(b_c) > \mathrm{MSE}(b)$. Again, (21) suggests that this MSE increase is large when $\phi$ is large and $z$ is sufficiently thick-tailed. The weighted least squares corrections defined in (13) and (14) show that the BCs place a higher weight on the larger-$p_{ii}$ observations, whereas a GLS estimator would lower the weight on those observations with this form of heteroskedasticity. Hence it is not surprising that the BCs increases MSE in this case, particularly when there is no bias to correct.

The RMSE ratios shown in Figures 1, 3 and 6 of Section 5 show that the BCs can reduce the MSE in other situations.

## 4 A Test for OLS Bias

Although many commonly-used significance tests have power against the misspecification that causes this bias, a direct test should have highest power in large samples. The following theorem enables Wald tests to be constructed from the bias estimator defined in (10).

**Theorem 5**

(a) *If the moments $Ex^6y^2$ are finite, then*

$$n^{1/2}(n\widehat{\mathrm{Bias}(b)} - n\mathrm{Bias}(b)) \to N[0, V_{Bias}]$$

*where $n\widehat{\mathrm{Bias}(b)}$ is defined in (10).*

(b) *A consistent estimator of $V_{Bias}$ is*

$$\hat{V}_{Bias} = n^{-1}\sum_i (\widehat{H\psi})_i(\widehat{H\psi})_i' \tag{23}$$

*where*

$$(\widehat{H\psi})_i = \hat{A}^{-1}(np_{ii})x_ie_i - \hat{A}^{-1}(\hat{A}_{np})\hat{A}^{-1}x_ie_i - c_i - \hat{A}^{-1}x_ix_i'\hat{A}^{-1}\hat{\gamma}^* \tag{24}$$

*and*

$$\hat{A} = n^{-1}\sum_i x_ix_i'$$

$$\hat{A}_{np} = n^{-1}\sum_i x_ix_i'(np_{ii})$$

$$np_{ii} = x_i'\hat{A}^{-1}x_i$$

$$c_{ia} = x_i'\hat{A}^{-1}(\sum_b (\hat{A}^{-1})_{ab}\hat{A}_{x_bv})\hat{A}^{-1}x_i - (\hat{A}^{-1})_{a\cdot}\hat{\gamma}^*$$

$$\hat{A}_{x_b v} = n^{-1} \sum_i x_i x_i' (x_{ib} e_i)$$

$$\hat{\gamma}^* = n^{-1} \sum_i x_i (n p_{ii}) e_i$$

*$x_{ia}$ and $c_{ia}$ are the $a^{th}$ elements of $x_i$ and $c_i$. $(\hat{A}^{-1})_{a\cdot}$ and $(\hat{A}^{-1})_{ab}$ are the $a^{th}$ row and $(a,b)^{th}$ element of $\hat{A}^{-1}$.*

A joint test for the overall absence of OLS bias can be based on the Wald statistic

$$W = (n\widehat{\text{Bias}(b)})'(n^{-1}\hat{V}_{Bias})^{-1}(n\widehat{\text{Bias}(b)})$$

$W$ is asymptotically distributed as $\chi^2$ with $K$ *d.f.* under the no-bias null hypothesis. $t$ statistics can be obtained in the usual way to test for biases in individual elements of $b$.

# 5 Examples and Simulations

In this Section, the OLS bias and the performance of the BCs and specification tests are examined for some simple models. The $X'X$ matrices cannot be singular in any of the models, so modifications like (5) are unnecessary.

## 5.1 $x$ takes only two values

Let $x$ be a scalar, distributed as

$$\Pr(x = 1) = p \text{ and } \Pr(x = a) = 1 - p$$

where $a > 1$ and the unknown probability $p$ satisfies $0 < p < 1$. The no-intercept regression model is $y = x\beta + v$. Let $\beta = 0$. Departing from the conditional mean zero assumption allows $E(y|x = 1) \equiv \mu_1 \neq 0$ and $E(y|x = a) \equiv \mu_a \neq 0$. Since $\beta = 0$, then $E(xy) = p\mu_1 + (1 - p)a\mu_a = 0$, so that $\mu_a = -p\mu_1/((1 - p)a)$.

### 5.1.1 $x = 1$ or $x = 2$, each with probability .5, and n = 1 or 2

First, let $n = 1$ and $\mu_1 = 2$. Then

| $x$ | $E(y\|x)$ | Prob$(x)$ | $x'x$ | $E(x'y\|x)$ | $E(b\|x)$ |
|---|---|---|---|---|---|
| 1 | 2 | .5 | 1 | 2 | 2 |
| 2 | -1 | .5 | 4 | -2 | -.5 |
| unconditional expectations: | | | 2.5 | 0 | .75 |

The OLS bias arises because $E(b)$ places equal weight on the $x = 1$ and the $x = 2$ values of $y$, whereas $\beta$ weights each $y$ by $E(xx')^{-1}x$. Thus $\beta$ places a higher weight than $E(b)$ does on the higher-leverage $(x = 2)$ values of $y$.

For an example where a BC reduces the OLS bias but increases the MSE, consider $n = 2$ and $\mu_1 = 2$.

| $x'$ | Prob$(x)$ | $x'x$ | $E(x'y\|x)$ | $E(b\|x)$ | $E(b_{c2}\|x)$ |
|---|---|---|---|---|---|
| (1,1) | .25 | 2 | 4 | 2 | 2 |
| (1,2) | .25 | 5 | 0 | 0 | -.143 |
| (2,1) | .25 | 5 | 0 | 0 | -.143 |
| (2,2) | .25 | 8 | -4 | -.5 | -.5 |
| unconditional expectations: | | 5 | 0 | .375 | .3035 |

$b_{c2}$, defined in (13), equals OLS when the two observed $x$ values are equal. Although $b_{c2}$ has a smaller unconditional bias, it has a larger bias than $b$ conditional on $x$ when $x_1 \neq x_2$. This bias correction increases the MSE if, for example, $y = E(y|x)$, since then the MSE is the average of the four squared conditional means.

#### 5.1.2 $x = 1$ or $x = a$

For general values of $n$, $p$ and $a$, exact means and MSEs can be computed for OLS and all of the BC estimators, including the bootstrap-based one when the bootstrap sampling error is ignored. Computational details are in Appendix A.7.

The left panel of Figure 1 plots the biases for various sample sizes. The BCs are effective unless $n$ is very small. The right panel plots the ratio of the square root of the MSEs (RMSEs) of the BCs to the RMSE of OLS. They all are less than unity, showing that the BCs reduce the MSE. When the sample size is 25 or greater, the BCs have nearly identical means and RMSEs.

#### 5.1.3 Controlling misspecification detectability

A fixed amount of mean misspecification is more likely to be detected by specification tests when $n$ is larger. An OLS user might respond to evidence of misspecification by changing the regressors in order to reduce the gap between $x'\beta$ and $E(y|x)$. The new estimation problem probably would have a smaller OLS bias.

Suppose this detectability is not controlled for when varying coefficients in these simulations, and that the bias is higher in situation A than B, because the mean in A is "more" misspecified than it is in B. This finding may not be relevant. A researcher may be less likely to end up in situation A. She may be more likely to change model A after estimating it, if its greater misspecification is easier to detect. To address this, the non-centrality parameter ($ncp$) of a commonly-used specification test is used as an indicator of bias detectability in many of the following simulations. Parameter values are calibrated to hold this $ncp$ fixed.

In the two-$x$-value model of this Section, one such test is a Wald test of the restriction $\beta_0 = 0$ in the augmented model $y = \beta_0 + x\beta + v^*$, allowing $\text{Var}(v^*|x = x_{(j)})$ to depend on $x$. Details are in Appendix A.8. In Figures 2 to 7, the conditional means $\mu_1$ and $\mu_a$ are adjusted to keep the $ncp$ fixed at $ncp = 5$.

#### 5.1.4 Extreme x values

In Figure 2 the larger value taken by $x$ (called "$a$") is varied, while fixing the smaller $x$ value at unity and adjusting $p$ such that $E(x) = 1.5$ for all $a$. $ncp = 5$, and $n = 15$. A large value of $a$ gives a high probability

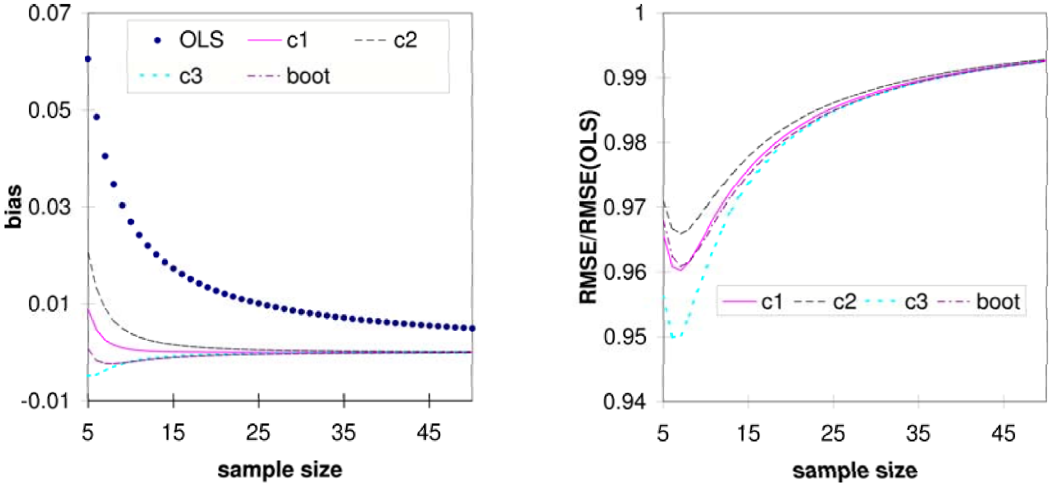

Figure 1: Biases and RMSE ratios under fixed alternative. $\Pr(x = 1) = .5$, $\Pr(x = 2) = .5$, $E(y|x = 1) = 2$, $E(y|x = 2) = -1$, $\mathrm{Var}(y|x = 1) = \mathrm{Var}(y|x = 2) = 1$.

that $x = 1$ and a low probability of a large $(x = a)$ value of $x$.

At the largest plotted $a$ value, $a = 34.3$, the distribution of $x$ is extremely skewed. Then $x = 1$ with probability $p = .985$, and $x = 34.3$ with probability $1 - p = .015$. Since $n = 15$, $p^n = 80\%$ of the samples do not contain a single large-$x$ observation, leaving no possibility of bias correction in those samples. This explains the poor bias-removal performance of the BCs. While $b_{boot}$ removes more bias than the other BCs, it also has the largest RMSE.

### 5.1.5 Heteroskedasticity

Figure 3 shows the effect of heteroskedasticity by changing $\sigma_2^2$ while fixing $\sigma_1^2 = 1$. The larger $x$ value is 3.5, the same as in the left ends of both panels of Figure 2. The *ncp* is fixed at 5 and the sample size at $n = 15$. When $\sigma_2^2$ is larger, more misspecification is required to result in a given *ncp*, causing a larger OLS bias. The biases of the BCs also increase with $\sigma_2^2$, but remain much smaller than the OLS bias. As anticipated in Section 3.4.2, however, the RMSEs of the BCs exceed the OLS RMSE when $\sigma_2^2$ is much larger than $\sigma_1^2$, although this excess RMSE reaches a maximum of roughly 2% to 4%.

## 5.2 $x$ contains an intercept and a continuous variable

Let the population consist of two groups, $j = 1, 2$. The distribution of the regressor $x^*$ and the conditional expectation function $E(y|x^*, j)$ differs across groups. Group identities are not observed, and the researcher estimates a single pooled regression. The regressor vector is $x = (1\ x^*)'$, where $x^*$ is a mixture of two random variables $x^*_{(1)}$ and $x^*_{(2)}$. $x^* = x^*_{(j)}$ with probability $p_{(j)}$, $j = 1, 2$. $x^*_{(1)}$ is distributed as exponential with a

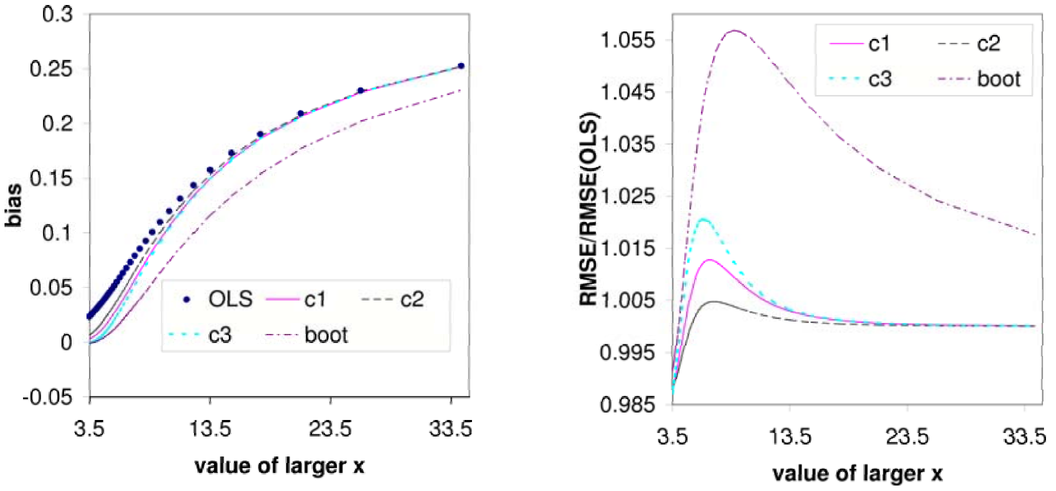

Figure 2: Biases and RMSE ratios vs. value of larger $x$. $\Pr(x = 1) = p$, $\Pr(x = \text{larger } x) = 1 - p$. $p$ is set such that $E(x) = 1.5$. $ncp = 5$, $n = 15$, $\text{Var}(y|x = 1) = \text{Var}(y|x = \text{larger } x) = 1$.

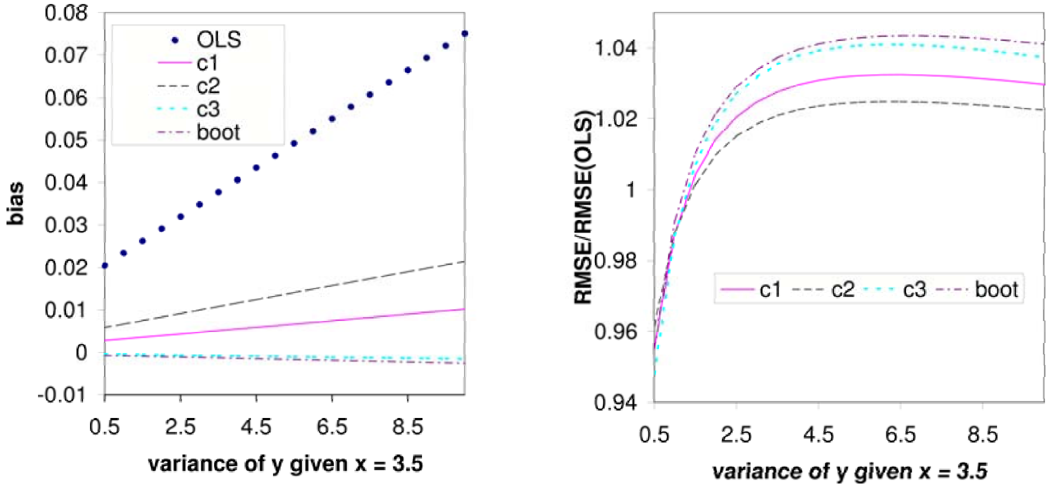

Figure 3: Biases and RMSE ratios vs. $\sigma_2^2$. $\Pr(x = 1) = .8$, $\Pr(x = 3.5) = .2$, $ncp = 5$, $n = 15$, $\text{Var}(y|x = 1) = 1$.

mean of one, and $x^*_{(2)}$ is exponential with mean $a$, $a \geq 1$. The probability that an observation belongs to group 1 given $x$, is

$$\Pr(j = 1|x) = \frac{p_{(1)}e^{-x^*}}{p_{(1)}e^{-x^*} + \frac{p_{(2)}}{a}e^{-x^*/a}}$$

The expectation of $y$ conditional on $x$ and group $j$ is $E(y|x, j) = \alpha + x^*\beta_{(j)}$. Then the conditional mean of $y$ is $E(y|x) = E(y|x, j = 1) \Pr(j = 1|x) + E(y|x, j = 2) \Pr(j = 2|x)$. Let $\text{Var}(y|x, j) = 1$ until Section 5.2.2.

The model is

$$y_i = [1 \ x^*_i]\beta + v_i$$

where $\beta = (Exx')^{-1}Exy$.

Figure 4 shows the difference between $E(y|x)$ and the linear projection $x'\beta$ when $p_{(1)} = .8, p_{(2)} = .2$, and $a = 3.5$. The (scaled) density of $x^*$ also is shown.

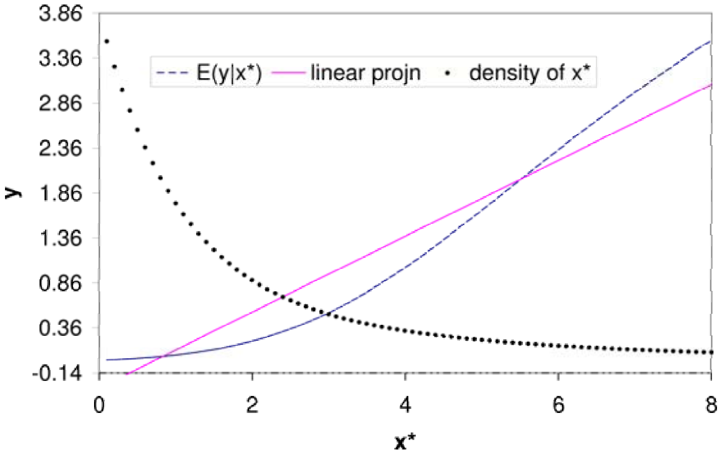

Figure 4: $E(y|x)$, $x'\beta$, and the scaled density of $x^*$. $p_{(1)} = .8, p_{(2)} = .2$, and $a = 3.5$.

Unless $a = 1$, the population slope coefficient, $\beta_2$, does not equal the population average "effect" or mean population response, which is $E_{(x^*,j)}(\frac{\partial E(y|x^*,j)}{\partial x^*}) = p_{(1)}\beta_{(1)} + p_{(2)}\beta_{(2)}$. If a researcher interprets $\beta_2$ as the mean population response, this is an example of what Freedman (2006, p.302) succinctly refers to as "estimating the wrong parameter". I will proceed nevertheless to examine the OLS bias in estimating $\beta_2$.

To calibrate the detectability of the misspecification, an *ncp* (details in Appendix A.9) was taken from the Wald test of $\theta = 0$ in the augmented linear regression model

$$y_i = \alpha + x^*_i\beta_{(1)} + (x^*_i)^2\theta + \text{error} \qquad (25)$$

Figure 5 plots the biases and RMSE ratios against the sample size. The BCs reduce the bias, but not nearly as effectively as they did in the discrete-$x$ example of Figure 1. $b_{boot}$ reduces the OLS bias the most, but it also has the highest RMSE, as it did in Figure 2. The BCs increase the MSE.

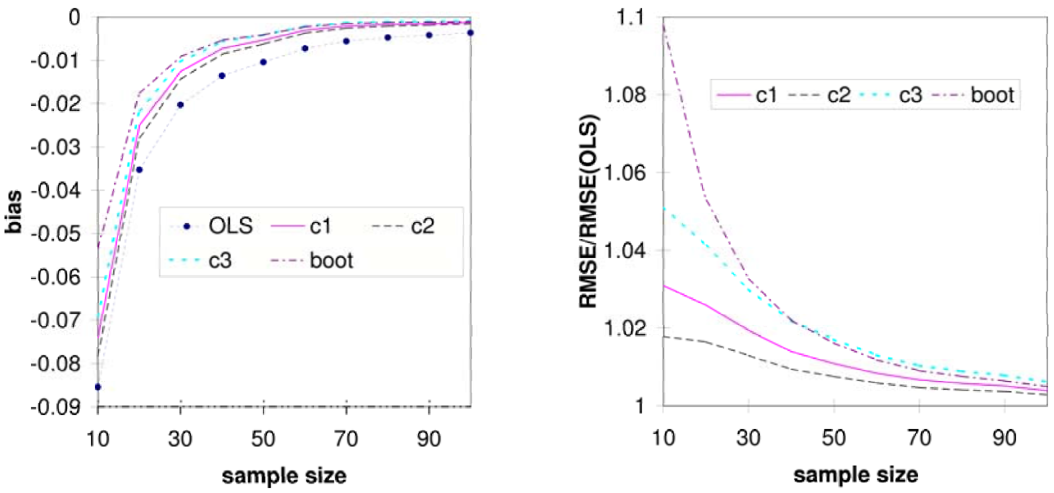

Figure 5: Biases and RMSE ratios with *ncp* fixed equal to 5. With probability .5, regressor $x$ is exponential with mean = 1 and $E(y|x) = 0 \times x$; With probability .5, $x$ is exponential with mean = 2 and $E(y|x) = \beta_2 \times x$. $\beta_2$ is set such that $ncp = 5$. $\text{Var}(y|x) = 1$ for all $x$.

### 5.2.1 Skewness in the Regressor

Figure 6 shows the effects of skewness in the $x$ distribution by increasing the mean of the larger-mean ($j = 2$) component of the distribution of $x$ while adjusting the $p_{(j)}$'s to hold the mean of $x$ constant at 1.5 and adjusting the $\beta_{(j)}$'s to fix the $ncp$ at 5. With higher skewness, the source of bias is more concentrated in the right tail of $x$, and becomes harder to detect. Higher skewness then leads to a larger OLS bias because we are holding the $ncp$ of test (25) constant. The BCs do not remove much of the bias. Once again, $b_{boot}$ removes the most, yet has the highest RMSE. The RMSE ratios show that the non-bootstrap BCs can provide some RMSE reduction over OLS, but this improvement fades as the skewness becomes more extreme.

### 5.2.2 Heteroskedasticity

Figure 7 shows the effect of increasing heteroskedasticity by increasing the error variance of the $j = 2$ disturbances. As in Figure 3, the relative RMSEs of the BCs grow with the heteroskedasticity, but again it appears to have an upper limit. The BCs provide some RMSE improvement unlike Figure 5, because in Figure 7 there is more skewness in $x$. The results (not reported) are similar when the error variance depends on $x$ as $\text{Var}(y|x^*, j) = \sigma^2 e^{\omega x^*}$. This requires $\omega < 1$ for finite moments.

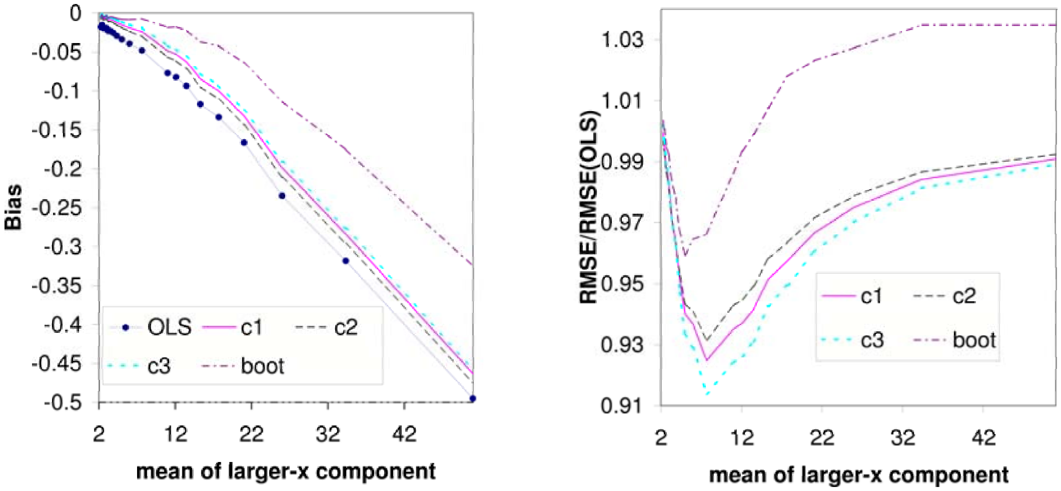

Figure 6: Biases and RMSE ratios vs. $a$, the mean of the larger-$x$ component. With probability $p$, regressor $x$ is exponential with mean 1 and $E(y|x) = 0 \times x$; With probability $1-p$, $x$ is exponential with mean $a = \frac{1.5-p}{1-p}$ and $E(y|x) = \beta_2 \times x$. $\beta_2$ is set such that $ncp = 5$. $\text{Var}(y|x) = 1$. $n = 100$.

Because the $E_x$ operation is simulated, the plots in this Section are not as smooth as in previous Sections. I used 5,000 replications and 5,000 bootstrap samples for each replication for $b_{boot}$.

## 5.3 Specification Tests

Two tests for the presence of OLS bias are examined here by simulating the model of Section 5.2. One ("$x^2$") is a $t$ test for the significance of $\theta$ in $y_i = \alpha + x_i^* \beta_{(1)} + (x_i^*)^2 \theta + u_i$ using the basic unweighted HCCME. The other ("direct") is the Wald significance test for bias in the OLS estimator of the $x_i^*$ coefficient in $y_i = [1 \; x_i^*]\beta + v_i$ using the bias estimator and its variance given in Section 4. The $x^2$ test is easier to compute with standard software, while the direct test should have higher power in large samples. For this comparison, $p = .9$, $a = 6$, and $\sigma_1^2 = \sigma_2^2 = 1$.

Figure 8 shows the sizes of these two tests at the 5% nominal significance level as a function of the square root of the sample size. Although both tests over-reject, the direct test is reasonably close to its nominal size.

In the left panel of Figure 9, size-corrected powers of these tests are shown against the square root of the sample size, with the $ncp$ fixed at 40. To facilitate power comparisons, size-corrected 5% critical values based on null-hypothesis simulations were used. At small sample sizes, the $x^2$ test has higher power, but the direct test has higher power for larger $n$.

The right panel of Figure 9 shows size-corrected powers as a function of the square root of the $ncp$, with the sample size fixed at $n = 1000$. With a small $ncp$, the $x^2$ test has higher power, while the direct test has

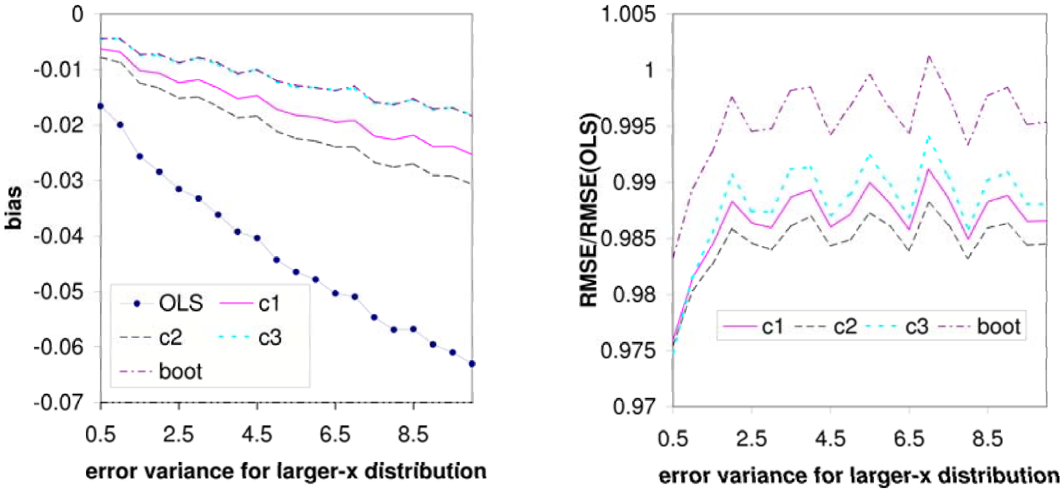

Figure 7: Biases and RMSE ratios vs. $\sigma_2^2 \equiv \mathsf{Var}(y|x, j=2)$. $\text{Var}(y|x, j=1) = 1$. With probability .75, regressor $x$ is exponential with mean 1 and $E(y|x) = 0 \times x$; With probability .25, $x$ is exponential with mean $a = \frac{1.5-p}{1-p} = 3$ and $E(y|x) = \beta_2 \times x$. $\beta_2$ is set such that $ncp = 5$. $n = 100$.

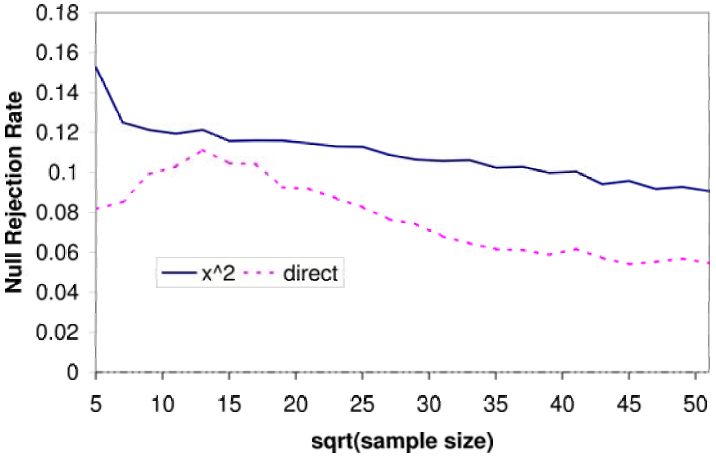

Figure 8: Rejection Rates vs. square root of sample size under null hypothesis of no OLS bias. Nominal size = 5%. $\text{Var}(y|x, j=1) = 1$. With probability .9 (.1), the regressor $x$ is exponential with mean 1 (6).

higher power when $ncp$ is larger.

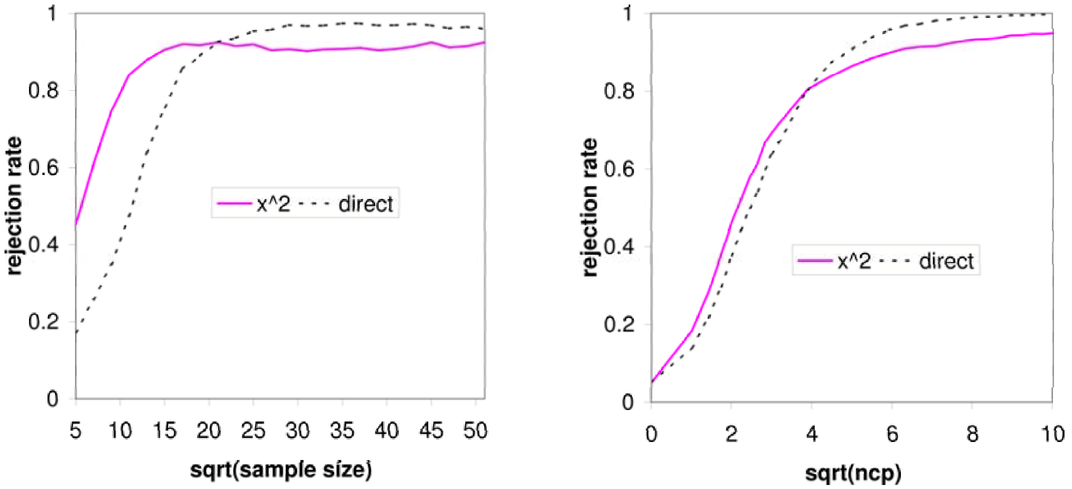

Figure 9: Rejection Rates vs. square root of sample size (left panel) and vs. square root of *ncp* (right panel). Tests are size-corrected to size $= 5\%$. $\text{Var}(y|x, j = 1) = 1$. With probability .9 (.1), the regressor $x$ is exponential with mean 1 (6). Left panel: $ncp = 40$. Right panel: $n = 1000$.

## 6 Empirical Example

Linear regressions similar to those reported in Crossley and Lu (2004) (CL) are reported here. Their sample contains 1158 single-person households and 915 childless-couple households. All individuals are working full time and between the ages of 25 and 55. It is taken from the 1992 and 1996 Canadian Food Expenditure Surveys and was kindly supplied by the authors. CL use linear regressions to examine differences between single and couple household spending on food and food ingredients. When combined with predictions from microeconomic models of household expenditure and other empirical results, these regressions shed light on the presence of economies of scale in home production, and on the relation between household size, food expenditure, and food consumption.

Four linear regressions are estimated. The four dependent variables are (1) Food (purchased from store) budget share (2) Ratio of expenditure on prepared food to expenditure on ingredients (3) Ratio of expenditure on take-out fast-food to expenditure on ingredients, and (4) Ingredients budget share. The same regressors appear in each equation. The two key regressors are the couples dummy variable and the log of household income per person ($LY$ below). The other regressors are the age of the household head, and dummies for sex of household head, education (four dummies) of the household head, as well as season (three dummies) and region (four dummies). A constant term is included.

CL check their specification by testing the significance of the square of the income variable and inspect nonlinear regressions of these dependent variables on household income, done separately for the single and couple households. Their theoretical model predicts that the coefficient of the couple dummy will be negative in the first three regressions and positive in the fourth.

There are several reasons to expect the OLS bias to be small. First, the sample size is fairly large at $n = 2073$. Second, CL carried out the model specification tests already mentioned. Third, the results of Section 3.4.2 suggest a larger bias may be expected when there are extreme-valued regressors or high-leverage observations. In this regressor set, the highest value of the commonly-used leverage measure $p_{ii}/(\frac{k}{n})$, is only 3.87. The two regressors that are not dummy variables are $LY$ and age, and age ranges from 25 to 55, while $LY$ has less kurtosis than does a Gaussian random variable (1.81 < 3).

Table 1: Bias-Corrected Estimates

| dependent variable | OLS (st.err.) | BC1 | BC2 | BC3 | B-boot |
|---|---|---|---|---|---|
| | coefficient estimate on couple dummy × 100 | | | | |
| Food (purchased from store) budget share | -1.236 (0.295) | -1.246 | -1.246 | -1.247 | -1.246 |
| Ratio of prepared food to ingredients | -9.078 (2.610) | -9.080 | -9.080 | -9.080 | -9.083 |
| Ratio of take-out fast-food to ingredients | -11.549 (3.259) | -11.541 | -11.541 | -11.540 | -11.542 |
| Ingredients budget share | -0.832 (0.248) | -0.840 | -0.840 | -0.840 | -0.840 |
| | coefficient estimate on $LY$ × 100 | | | | |
| Food (purchased from store) budget share | -10.129 (2.056) | -10.180 | -10.179 | -10.180 | -10.183 |
| Ratio of prepared food to ingredients | 0.710 (3.262) | 0.746 | 0.746 | 0.748 | 0.753 |
| Ratio of take-out fast-food to ingredients | 11.988 (3.765) | 11.907 | 11.909 | 11.902 | 11.911 |
| Ingredients budget share | -8.206 (1.621) | -8.250 | -8.250 | -8.251 | -8.250 |

Table 1 reports the estimated coefficients on the couple dummy and the $LY$ variable. The OLS estimates differ a bit from CL because they used sampling weights and a slightly different sample. The difference between OLS and the BCs is always much smaller than the OLS standard error. The OLS variance was estimated using Stata's "robust" option, which is Davidson and MacKinnon's (1993, p.554) $HC_1$. The OLS biases indeed appear to be very small. When only $B = 10,000$ bootstrap replications were used, the bootstrap sampling errors were large compared to the differences between the BCs, so $B = 500,000$ bootstrap replications are used for $b_{boot}$.

Table 2 reports tests for the existence of OLS bias using the test procedure described in Section 4. The estimated biases are scaled up by multiplying by the sample size. They are not statistically significant at the 5% level. Neither are any of the four Wald statistics, which test the joint null hypothesis that none of the OLS estimates in an equation are biased. OLS bias is not a concern in this application.

Table 2: Direct Bias Tests

| dependent variable | $n\times$ estimated bias | standard error | asymptotic $t$-ratio |
|---|---|---|---|
| | couple dummy | | |
| Food (purchased from store) budget share | 0.200 | 0.142 | 1.410 |
| Ratio of prepared food to ingredients | 0.039 | 0.191 | 0.203 |
| Ratio of take-out fast-food to ingredients | -0.174 | 0.120 | -1.451 |
| Ingredients budget share | 0.151 | 0.109 | 1.387 |
| | log of household income per person ($LY$) | | |
| Food (purchased from store) budget share | 1.045 | 1.368 | 0.764 |
| Ratio of prepared food to ingredients | -0.756 | 0.548 | -1.379 |
| Ratio of take-out fast-food to ingredients | 1.673 | 1.031 | 1.622 |
| Ingredients budget share | 0.913 | 1.030 | 0.887 |
| | Wald statistic (joint test) (P-value) | | |
| Food (purchased from store) budget share | 10.64 (.8314) | | |
| Ratio of prepared food to ingredients | 11.36 (.7867) | | |
| Ratio of take-out fast-food to ingredients | 13.13 (.6632) | | |
| Ingredients budget share | 10.56 (.8360) | | |

# 7 Remarks

1. The bias approximation (9) is captured by the first right-hand side term of Proposition 3.2 of Rilstone et al. (1996) and the $B_I$ term of Newey and Smith's (2004) equation (4.3). Notation is compared in the following table.

| Newey and Smith | Rilstone et al. | Section 2 |
|---|---|---|
| $g_i$ | $q_i$ | $x_i(y_i - x_i'\beta) = x_i v_i$ |
| $G_i$ | $\nabla q_i$ | $-x_i x_i'$ |
| $G$ | $\overline{\nabla q_i}$ | $-A$ |
| $a$ | $\overline{H_2}$ | $0$ |
| $\Sigma$ | | $A^{-1}\Omega_{v^2}A^{-1}$ |
| $H$ | $Q$ | $-A^{-1}$ |
| | $V_i$ | $-(x_i x_i' - A)$ |
| | $d_i$ | $-A^{-1}x_i v_i$ |
| $HE\left[G_i H g_i\right]$ | $Q(\overline{V_1 d_1})$ | $-A^{-1}Ex(x'A^{-1}x)v$ |

Rilstone et al. are concerned with nonlinear estimators, while Newey and Smith compare GMM and GEL estimation techniques in a more general setting. Newey and Smith (2004, p.233) present a bias correction for GMM estimators. The non-zero part of their result for OLS bias correction purposes is, in their notation, $-\hat{H}\sum_{i=1}^{n}\hat{G}_i\hat{\psi}_i^{\beta}/n$, which equals $(n\widehat{\text{Bias}(b)})/n$ from equation (10) of the present paper.

2. If it is assumed that the $N$ finite-population members themselves have been independently drawn from a superpopulation satisfying the above assumptions, then $\beta$ is the superpopulation regression coefficient, and

the expectations in (1) are taken with respect to the distribution of $(x, y)$ in the superpopulation. Then the *iid* assumption is appropriate for sampling without replacement (SWOR) from this finite population. $\beta$ differs from the census regression coefficient $\beta_{CEN}$, which is the OLS coefficient applied to the $N$ population members.

Now suppose the parameter of interest is $\beta_{CEN}$ instead of $\beta$. For the approximations in this paper to apply to OLS estimation of $\beta_{CEN}$ with SWOR, $(n^2/N) \to 0$ is required. To see this, consider $\Delta_{zv}$ in (7). The sample is drawn from a finite population $(x_r, y_r)$, $r = 1, \ldots, N$, where $y_r = x_r'\beta_{CEN} + v_r$, and $E_P xv = N^{-1}\sum_{r=1}^{N} x_r v_r = 0$, where $E_P$ denotes expectation under SWOR with finite $N$. To simplify, let $z_i = x_i$ and $z_i$ be scalar. The leading $O(n^{-1})$ term in the MSE expansions involves the term

$$E_P\Delta_{zv}^2 = n^{-2}\sum_i (E_P z_i^2 v_i^2) + n^{-2}\sum_{i\neq j} E_P(z_i v_i \times E(z_j v_j | z_i v_i))$$

SWOR implies $E_P(z_j v_j | z_i v_i) = -z_i v_i/(N-1)$ when $i \neq j$, therefore

$$\begin{aligned} E_P\Delta_{zv}^2 &= n^{-1}E_P z^2 v^2 + n^{-2}\sum_{i\neq j} E_P(z_i v_i \times (\frac{-z_i v_i}{N-1})) \\ &= n^{-1}E_P z^2 v^2 - n^{-2}n(n-1)\frac{E_P z^2 v^2}{N-1} \\ &= n^{-1}E_P z^2 v^2 - \left(\frac{n(n-1)}{n^2(N-1)}\right) E_P z^2 v^2 \qquad (26) \end{aligned}$$

Under SWOR, the last term must be $o(n^{-2})$ for this paper's $O(n^{-2})$ MSE approximations to remain valid. This requires $n^2((\frac{n(n-1)}{n^2(N-1)}) \to 0$, or $(n^2/N) \to 0$. (A first-order approximation only would require $(n/N) \to 0$.) Therefore in large samples, an extremely large $N$ may be needed for these *iid*-based $O(n^{-2})$ MSE results to be applicable to SWOR-based estimation of a census regression coefficient. With *iid* sampling, $E(z_j v_j | z_i v_i) = 0$ when $i \neq j$, and the last term of (26) disappears.

3. MacKinnon and Smith (1998) consider bias corrections and their effects on the MSE for a general class of models and estimators. They assume that the distribution generating the data is known up to an unknown parameter. Their bias corrections may reduce or increase the MSE, as do the corrections in this paper. It follows that bias corrections that tackle both sources of bias, such as Newey and Smith's (2004), share this property.

4. The weighted OLS versions of bias correction, $b_{c2}$ and $b_{c3}$ defined in (13) and (14), show that the corrections make the estimator more sensitive to high-leverage observations. As the example in Section 5.1 shows, the bias is caused by a tendency for OLS to place a lower weight on this class of observation than it receives in the corresponding population regression that defines $\beta$. Many robust regression estimators weight in the opposite direction, and are estimating a parameter that also places a lower weight on high-leverage population members, than does $\beta$.

   While on the topic of weighted least squares, note that GLS versions of WLS are not consistent estimators of $\beta$ with conditional mean misspecification, because of correlations between $v$ and functions of $x$ like those that cause the OLS bias studied in this paper. Boothe and MacKinnon (1988) provide a misspecification test based on this fact.

5. When conditional mean misspecification is concentrated at the tail values of skewed regressor variables that have thick-tailed distributions, the bias is large, and the corrections do not work well. Unfortunately this is a familiar characteristic of corrections based on asymptotic expansions. Phillips and Park (1988, p.1066) write that Edgeworth corrections "tend to work well when the error on the crude asymptotic is small (when they are least needed) and are poor when the error is large (when they are most needed)". Some consolation may be found in their next remark, that these corrections "still provide a valuable source of information about the adequacy of asymptotic theory".

   A simple but unambitious remedy to this bias problem is to redefine the parameter of interest as $\beta_X = (\sum_i x_i x_i')^{-1} \sum_i x_i E(y|x_i)$, assuming that $\sum_i x_i x_i'$ is invertible. $x'\beta_X$ is the minimum-MSE linear approximation to $E(y|x)$ when the expectation $E_X((E(y|x) - x'\beta^*)^2)$ is taken with respect to the empirical distribution of the $x_i$'s instead of the population distribution of $x$. $b$ is an unbiased estimator of $\beta_X$.

6. The finite-sample behaviour of tests and confidence intervals concerning $\beta$ have not been studied here. The shape of the distribution of $y$ given $x$, $f(y|x)$, becomes relevant, not just the first two moments. Although the methods of this paper can be extended to obtain bias-corrected estimators of $\text{Var}(b)$ or $\text{Var}(b_c)$, the results of Davidson and Flachaire (2001) and Flachaire (2005) suggest that accounting for features of $f(y|x)$ by bootstrapping is more useful for improving inference than is correcting biases in the variance estimators.

   Since the bias and bias corrections of this paper take place at the $O_p(n^{-1})$ level, the asymptotic validity of the standard HCCMEs designed for the $O_p(n^{-1/2})$ part of OLS still are applicable.

7. The bias results for equal-probability random sampling assumption made throughout this paper could be extended in a straightforward fashion to variable probability sampling and standard stratified sampling, based on the asymptotic theory contained in Wooldridge (1999, 2001).

# 8 Summary

The $O(n^{-1})$ bias and $O(n^{-2})$ MSE of OLS is derived under *iid* sampling. A modification to OLS is suggested to handle the problem of nonexistent finite sample moments. Bias corrections based on plug-in, weighting, jackknife and pairs bootstrap methods, are equal to $O_p(n^{-3/2})$. Sometimes they are effective at lowering bias and MSE, but not always.

For the results to be applicable to census regression coefficient estimation in a finite population of size $N$ using sampling without replacement, the condition $n^2/N \to 0$ is required. Similarly, in order for the sampling error of a bootstrap bias correction to belong in the remainder term of the expansion, $n^2/B \to 0$ is required, where $B$ is the number of bootstrap replications. When $n$ is large, these requirements on $N$ and $B$ may be considerable.

In the simulations, the bootstrap correction removes more bias than the others, but has a higher MSE. Exact means and MSEs are given for the single binary regressor case. Parameters are calibrated in an effort to hold constant the detectability of the misspecification across specifications.

A test is given for the $O(n^{-1})$ OLS bias. In a simulation it has a power advantage over a mean misspecification test at higher sample sizes and larger biases.

When these techniques are applied to survey data on food expenditure, the estimated bias is very small and statistically insignificant.

# 9 References

Angrist, J., Chernozhukov, V., and Fernàndez-Val, I. (2006), "Quantile Regression under Misspecification, with an Application to the U.S. Wage Structure," *Econometrica*, 74, 539-563.

Baum, L.E., and Katz, M. (1965), "Convergence Rates in the Law of Large Numbers," *Transactions of the American Mathematical Society*, 120, 108-123.

Boothe, P., and MacKinnon, J.G. (1988), "A Specification Test for Models Estimated by GLS," *Review of Economics and Statistics*, 68, 711-714.

Crossley, T.F. and Lu, Y. (2004), "Exploring the Returns to Scale in Food Preparation (Baking Penny Buns at Home)," Social and Economic Dimensions of an Aging Population (SEDAP) Research Paper No. 121, McMaster University.

Davidson, R. and Flachaire, E. (2001), "The Wild Bootstrap, Tamed at Last," working paper IER#1000, Queen's University.

Davidson, R. and MacKinnon, J.G. (1993), *Estimation and Inference in Econometrics*, New York: Oxford University Press.

Flachaire, E. (2005) "Bootstrapping Heteroskedastic Regression Models: Wild Bootstrap versus Pairs Bootstrap," *Computational Statistics and Data Analysis*, 49, 361-376.

Freedman, D.A. (2006), "On the So-Called "Huber Sandwich Estimator" and "Robust Standard Errors"", *The American Statistician*, 60, 299-302.

Goldberger, A.S. (1991), *A Course in Econometrics*, Cambridge: Harvard University Press.

Hinkley, D.V. (1977), "Jackknifing in Unbalanced Situations," *Technometrics*, 19, 285-292.

Liu, R.Y. and Singh, K. (1992), "Efficiency and Robustness in Resampling," *The Annals of Statistics*, 20, 370-384.

MacKinnon, J.G., and Smith, A.A. (1998), "Approximate Bias Correction in Econometrics," *Journal of Econometrics*, 85, 205-230.

Magnus, J.R., and Neudecker, H. (1988), *Matrix Differential Calculus with Applications in Statistics and Econometrics*, Chichester: Wiley.

Nagar, A.L. (1959), "The Bias and Moment Matrix of the General $k$-class Estimators of the Parameters in Simultaneous Equations," *Econometrica*, 27, 573-595.

Newey, W.K. and Smith, R.J. (2004), "Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators," *Econometrica*, 72, 219-255.

Peracchi, F. (2001), *Econometrics*, Chichester: John Wiley and Sons Ltd.

Phillips, P.C.B., and Park, J.Y. (1988), "On the Formulation of Wald Tests of Nonlinear Restrictions," *Econometrica*, 56, 1065-1083.

Poirier, D.J. (1995), *Intermediate Statistics and Econometrics, A Comparative Approach*, Cambridge: MIT Press.

Rilstone, P., V.K. Srivastava, and A. Ullah (1996), "The Second-Order Bias and Mean Squared Error of Nonlinear Estimators," *Journal of Econometrics*, 75, 369-395.

Sargan, J.D. (1974), "The Validity of Nagar's Expansion for the Moments of Econometric Estimators," *Econometrica*, 42, 169-176.

Schmidt, P. (1976), *Econometrics*, New York: Marcel Dekker.

White, H. (1984), *Asymptotic Theory for Econometricians*, Orlando: Academic Press.

Wooldridge, J.M. (1999), "Asymptotic Properties of Weighted $M$-Estimators for Variable Probability Samples," *Econometrica*, 67, 1385-1406.

Wooldridge, J.M. (2001), "Asymptotic Properties of Weighted $M$-Estimators for Standard Stratified Samples," *Econometric Theory*, 17, 451-470.

# A Appendix

## A.1 Proof of Theorem 1(a)

Let the random variable $g = 1$ if $\lambda_1(\hat{A}) < \tau$, and $g = 0$ otherwise. An $a$ subscript denotes the $a^{th}$ element of a vector. Then

$$\begin{aligned} E(b_{Ma}^2) &= E(g)E(b_{Ra}^2|g=1) + E(1-g)E(b_a^2|g=0) \\ &\leq \max(E(b_{Ra}^2|g=1), E(b_a^2|g=0)) \text{ since } 0 \leq E(g) \leq 1 \end{aligned}$$

$E(b_{Ra}^2|g=1)$ is finite by construction. It is left to show that $E(b_a^2|g=0)$ is finite. Let

$$\hat{A} = \sum_{k=1}^{K} \lambda_k c_k c_k' \tag{27}$$

where $\lambda_k$ and $c_k$, $k = 1, \ldots, K$, are the eigenvalues and eigenvectors of $\hat{A}$. If $g = 0$, then $\lambda_k > \tau$ for all $k$. Defining $w = n^{-1} \sum_i x_i y_i$, then

$$
\begin{aligned}
E(b_a^2|g=0) &= E\left(\sum_{j=1}^{K} \lambda_j^{-1} c_{ja} c_j' w\right)\left(\sum_{k=1}^{K} \lambda_k^{-1} c_{ka} c_k' w\right) \\
&= E \sum_{jk} \lambda_j^{-1} \lambda_k^{-1} c_{ja} c_{ka} (c_j' w)(c_k' w) \\
&\leq E \sum_{jk} |\lambda_j^{-1}| |\lambda_k^{-1}| |c_{ja}| |c_{ka}| |c_j' w| |c_k' w| \\
&\leq K^2 \max_{jk} E |\lambda_j^{-1}| |\lambda_k^{-1}| |c_{ja}| |c_{ka}| |c_j' w| |c_k' w|
\end{aligned}
$$

$g = 0$ implies $|\lambda_j^{-1}| < \tau^{-1}$, and $c_j' c_j = 1$ implies $|c_{ja}| \leq 1$. Therefore

$$
\begin{aligned}
E(b_a^2|g=0) &\leq K^2 \tau^{-2} \max_{c} E(c'w)^2 \\
&\leq K^2 \tau^{-2} \max_{j} E w_j^2
\end{aligned}
$$

$\max_j E w_j^2 = \max_j E(n^{-1} \sum_i x_{i,j} y_i)^2 = \max_j (n^{-1} E(x_j^2 y^2) + (1 - n^{-1}) E(x_j y)^2)$, where $x_j$ refers to the $j^{th}$ element of the random vector $x$. $E(x_j^2 y^2)$ (hence $E(x_j y)^2)$) is assumed finite, therefore $E(b_a^2|g = 0)$ and $E(b_{Ma}^2)$ are finite.

## A.2 Proof of Theorem 1(b)

$b_M = g b_R + (1-g) b = b + g(b_R - b) = b + g \times O(1)$, where $g$ is defined at the beginning of the proof of Theorem 1(a). A sufficient condition for $b_M = b + o_p(n^{-3/2})$ is $g = o_p(n^{-3/2})$, or $\text{Prob}(\lambda_1(\hat{A}) < \tau) = 0 + o_p(n^{-3/2})$. The proof below is similar to the proof of Lemma 2 of Liu and Singh (1992, p.383).

Since $\lambda_1(A) > \tau$, the event $\lambda_1(\hat{A}) < \tau$ can be written as $\lambda_1(\hat{A}) - \lambda_1(A) < -d$ for a positive $d$ bounded away from zero. Let $\text{vec}(A)$ and $\text{vec}(\hat{A})$ be the $K(K-1)/2$ element vectors of distinct elements of the symmetric matrices $A$ and $\hat{A}$. Let $a_j$ and $\hat{a}_j$ be the $j^{th}$ elements of $\text{vec}(A)$ and $\text{vec}(\hat{A})$. Then $\hat{a}_j - a_j = n^{-1} \sum_i (x_{ik} x_{i\ell} - E x_{ik} x_{i\ell})$ where $j = 1, \ldots, K(K+1)/2$ indexes $(k, \ell)$ combinations. Let $G = x_{ik} x_{i\ell} - E x_{ik} x_{i\ell}$ and $\bar{G} = \hat{a}_j - a_j = n^{-1} \sum_i (x_{ik} x_{i\ell} - E x_{ik} x_{i\ell})$. From Baum and Katz (1965, p.113),

$$
E(G) = 0 \text{ and } E|G|^t < \infty \implies n^{t-1} \text{Prob}(|\bar{G}| > \epsilon) \to 0 \text{ for all } \epsilon > 0
$$

The assumption of finite moments of order 5 in $x$ implies $E|G|^t < \infty$ for $t = \frac{5}{2}$. Therefore

$$
n^{\frac{5}{2}-1} \text{Prob}\left(|\hat{a}_j - a_j| > \epsilon\right) \to 0 \text{ for all } \epsilon > 0 \qquad (28)
$$

When $\lambda_1(A)$ is a single root, it is a smooth real-valued function of $A$ with all derivatives (Magnus and Neudecker (1988, Theorem 7, p.158)). Choose $\epsilon$ in (28) to be small enough that these derivatives exist over $(a_j - \epsilon, a_j + \epsilon)$ for each $j$, and small enough that $\epsilon < d^*$, where $d^* = d \times (\max_j |\frac{\partial \lambda_1(A)}{\partial a_j}|)$. Let $a_j^*$ be some scalar that lies between $a_j$ and $\hat{a}_j$. Then the vector mean value theorem gives

$$\begin{aligned}
\text{Prob}\left(\lambda_1(\hat{A}) < \tau\right) &= \text{Prob}\left(\lambda_1(\hat{A}) - \lambda_1(A) < -d\right) \\
&= \text{Prob}\left(\sum_{j=1}^{K(K-1)/2} (\frac{\partial \lambda_1(A)}{\partial a_j})(a_j^* - a_j) < -d^*\right) \\
&< \text{Prob}\left(K^2 \max_j \left|\frac{\partial \lambda_1(A)}{\partial a_j}\right| \max_j |a_j^* - a_j| > d^*\right) \\
&< \text{Prob}\left(\max_j |a_j^* - a_j| > d^{**}\right) \\
&< \sum_j \text{Prob}\left(|a_j^* - a_j| > d^{**}\right) \\
&< K^2 \max_j \text{Prob}\left(|a_j^* - a_j| > d^{**}\right) \qquad (29)
\end{aligned}$$

where $d^{**} = d^*/(K^2 \max_j |\frac{\partial \lambda_1(A)}{\partial a_j}|)$ is bounded away from zero. Letting $\epsilon$ in (28) equal $d^{**}$, the probabilities in (29) are $o(n^{-3/2})$. Hence

$$\text{Prob}\left(\lambda_1(\hat{A}) < \tau\right) \leq 0 + o(n^{-3/2})$$

## A.3 Proof of Theorem 2

Following the proof of Theorem 1(a) and its notation, it is sufficient for Theorem 2(a) to show that $E(b_{c1,a}^2 | g = 0)$ and $E(b_{boot,a}^2 | g = 0)$ are finite. For $b_{c1,a}$, since $E(b_a^2 | g = 0)$ is finite from Theorem 1(a), it is enough to show that $E((b_{c1,a} - b_a)^2 | g = 0)$ is finite. Using (11) and (27),

$$\begin{aligned}
b_{c1,a} - b_a &= -(\sum_{j=1}^{K} \lambda_j^{-1} c_{ja} c_j')(n^{-1} \sum_i x_i p_{ii} e_i) \\
n p_{ii} &= x_i'(n^{-1} \sum_i x_i x_i')^{-1} x_i = x_i'(\sum_{k=1}^{K} \lambda_k^{-1} c_k c_k') x_i = \sum_{k=1}^{K} \lambda_k^{-1} (c_k' x_i)^2 \quad \text{and} \\
e_i &= y_i - x_i' b = y_i - x_i'(\sum_{\ell=1}^{K} \lambda_\ell^{-1} c_\ell c_\ell') w = y_i - \sum_{\ell=1}^{K} \lambda_\ell^{-1} (c_\ell' x_i)(c_\ell' w)
\end{aligned}$$

where $w = n^{-1} \sum_i x_i y_i$. Therefore

$$\begin{aligned}
b_{c1,a} - b_a &= -\sum_{jk\ell=1}^{K} \lambda_j^{-1} c_{ja} c_j' \left(n^{-1} \sum_{i=1}^{n} x_i (n^{-1} \lambda_k^{-1} (c_k' x_i)^2)(y_i - \lambda_\ell^{-1} (c_\ell' x_i)(c_\ell' w))\right) \\
&= -n^{-2} \sum_i \left(\sum_{jkl} \lambda_j^{-1} \lambda_k^{-1} c_{ja} (c_j' x_i)(c_k' x_i)^2 y_i - \lambda_j^{-1} \lambda_k^{-1} \lambda_\ell^{-1} c_{ja} (c_j' x_i)(c_k' x_i)^2 (c_\ell' x_i)(c_\ell' w)\right)
\end{aligned}$$

$E(b_{c1,a} - b_a)^2$ can be shown to be finite when $g = 0$ following steps similar to the last part of the proof of Theorem 1(a). The $\lambda^{-1}$'s and elements of the $c$'s are bounded. The above expression has $(x^3 y)$ and $(x^3 w)$ or $(x^4 y)$ terms, so the expectation of its square is finite if the $(x^8 y^2)$ moments are finite.

Turning to the finiteness of $E(b_{boot,a}^2|g=0)$, it is enough to show that $E((\bar{b}_{B,a})^2|g=0)$ is finite. Writing $\bar{b}_{B,a}$ as

$$\bar{b}_{B,a} = B^{-1}\sum_{\ell=1}^{B}(b_a g_\ell + b_{B\ell,a}(1-g_\ell))$$

where $g_\ell = 1$ if $\lambda_1(n^{-1}\sum_i x_{\ell i}x'_{\ell i}) < \tau$ and $g_\ell = 0$ otherwise, and noting that $E(b_a^2|g=0, g_\ell=1)$ is finite, then it is enough to show that $E((b_{B\ell,a})^2|g=0, g_\ell=0)$ is finite. This follows from Theorem 1(a), with $b_{B\ell,a}$ replacing $b_a$, $g_\ell = 0$ replacing $g=0$, and the pairs bootstrap sampling being a special case of *iid*-population sampling which satisfies the finite-$E(x^2y^2)$ moment condition. This completes the proof of Theorem 2(a).

Using the same argument as the first paragraph of the proof of Theorem 1(b), then Theorem 2(b) holds if $\text{Prob}(\lambda_1(\hat{A}) < \tau) = 0 + o_p(n^{-3/2})$. That result also is contained in Theorem 1(b).

## A.4 Proof of Theorem 3

First, pairwise comparisons are used to show the $O_p(n^{-3/2})$ equivalence of all of the BCs. Then an $O(n^{-3/2})$ expansion for $H^{-1}b_{c1}$ is derived. Given their $O(n^{-3/2})$ equivalence, it also applies to the other BCs.

Comparing $b_{c1}$ and $b_{c2}$, from (11),

$$\begin{aligned} b_{c1} &= (\sum_i x_i x_i')^{-1}\sum_i x_i(y_i + p_{ii}(y_i - x_i'b)) \\ &= (\sum_i x_i x_i')^{-1}\sum_i x_i y_i(1+p_{ii}) - (\sum_i x_i x_i')^{-1}(\sum_i x_i x_i' p_{ii})b \\ &= (\sum_i x_i x_i')^{-1}\sum_i x_i y_i(1+p_{ii}) - (\sum_i x_i x_i')^{-1}(\sum_i x_i x_i' p_{ii})(b_{c1} + O_p(n^{-1})) \\ &= (\sum_i x_i x_i')^{-1}\sum_i x_i y_i(1+p_{ii}) - (\sum_i x_i x_i')^{-1}(\sum_i x_i x_i' p_{ii})b_{c1} + O_p(n^{-2}) \end{aligned}$$

Grouping the $b_{c1}$ terms, then

$$\begin{aligned} (\sum_i x_i x_i')^{-1}(\sum_i x_i x_i'(1+p_{ii}))b_{c1} &= (\sum_i x_i x_i')^{-1}\sum_i x_i y_i(1+p_{ii}) + O_p(n^{-2}) \\ (\sum_i x_i x_i'(1+p_{ii}))b_{c1} &= \sum_i x_i y_i(1+p_{ii}) + O_p(n^{-1}) \end{aligned}$$

Therefore

$$\begin{aligned} b_{c1} &= (\sum_i x_i x_i'(1+p_{ii}))^{-1}\sum_i x_i y_i(1+p_{ii}) + O_p(n^{-2}) \\ &= b_{c2} + O_p(n^{-2}) \end{aligned}$$

Next, compare $b_{c2}$ and $b_{c3}$. Since $p_{ii}$ is $O_p(n^{-1})$, then $1 + p_{ii} = (1-p_{ii})^{-1} + O_p(n^{-2})$. Therefore

$$\begin{aligned}
b_{c2} &= (\sum_i x_i x_i'(1+p_{ii}))^{-1} \sum_i x_i y_i (1+p_{ii}) \\
&= (\sum_i x_i x_i'((1-p_{ii})^{-1} + O_p(n^{-2})))^{-1} \sum_i x_i y_i ((1-p_{ii})^{-1} + O_p(n^{-2})) \\
&= (\sum_i x_i x_i'(1-p_{ii})^{-1} + O_p(n^{-1}))^{-1} (\sum_i x_i y_i (1-p_{ii})^{-1} + O_p(n^{-1})) \\
&= ((\sum_i x_i x_i'(1-p_{ii})^{-1})^{-1} + O_p(n^{-2}))(\sum_i x_i y_i (1-p_{ii})^{-1} + O_p(n^{-1})) \\
&= (\sum_i x_i x_i'(1-p_{ii})^{-1})^{-1} \sum_i x_i y_i (1-p_{ii})^{-1} + O_p(n^{-2}) \\
&= b_{c3} + O_p(n^{-2})
\end{aligned}$$

Next, consider the standard jackknife estimator defined in (15).

$$\begin{aligned}
b_{jack} &= b + (n-1)n^{-1}(\sum_i x_i x_i')^{-1} \sum_i (1-p_{ii})^{-1} x_i e_i \\
&= b + (1-n^{-1})(\sum_i x_i x_i')^{-1} \sum_i x_i e_i (1 + p_{ii} + O_p(n^{-2})) \\
&= b + (1-n^{-1})(\sum_i x_i x_i')^{-1} \times 0 + (\sum_i x_i x_i')^{-1} \sum_i x_i p_{ii} e_i + O_p(n^{-2}) \\
&= b + (\sum_i x_i x_i')^{-1} \sum_i x_i p_{ii} e_i + O_p(n^{-2}) \\
&= b_{c1} + o_p(n^{-3/2})
\end{aligned}$$

Finally, consider $b_{boot}$. It uses OLS estimates from the bootstrap samples,

$$b_{B\ell} = b + (\sum_{j=1}^{n} x_{\ell j} x_{\ell j}')^{-1} \sum_{j=1}^{n} x_{\ell j} e_{\ell j}, \quad \ell = 1, \ldots, B$$

where invertibility of $\sum_{j=1}^{n} x_{\ell j} x_{\ell j}'$ is ensured by the truncation rule used in defining $b_{B\ell,M}$. Then

$$\begin{aligned}
b_{boot} - b &= (b - (\bar{b}_B - b)) - b \\
&= b - \bar{b}_B \\
&= B^{-1} \sum_{\ell=1}^{B} (b - b_{B\ell}) + o_p(n^{-3/2}) \\
&= -B^{-1} \sum_{\ell=1}^{B} (\sum_{j=1}^{n} x_{\ell j} x_{\ell j}')^{-1} \sum_{j=1}^{n} x_{\ell j} e_{\ell j} + o_p(n^{-3/2}) \\
&= -B^{-1} \sum_{\ell} \left(\hat{A} + (\hat{A}_\ell - \hat{A})\right)^{-1} n^{-1} \sum_j x_{\ell j} e_{\ell j} + o_p(n^{-3/2})
\end{aligned}$$

$$\begin{aligned}
&= -B^{-1}\sum_{\ell}\left(\hat{A}^{-1}-\hat{A}^{-1}(\hat{A}_\ell-\hat{A})\hat{A}^{-1}+o_p(n^{-1/2})\right)n^{-1}\sum_j x_{\ell j}e_{\ell j}+o_p(n^{-3/2})\\
&= -\hat{A}^{-1}(n^{-1}B^{-1}\sum_{\ell}\sum_j x_{\ell j}e_{\ell j})+\hat{A}^{-1}(n^{-1}B^{-1}\sum_{\ell}\sum_j(\hat{A}_\ell-\hat{A})\hat{A}^{-1}x_{\ell j}e_{\ell j})\\
&\quad -B^{-1}n^{-2}\sum_{\ell}o_p(n^{-1/2})(\sum_j x_{\ell j}e_{\ell j})+o_p(n^{-3/2}) \qquad (30)
\end{aligned}$$

The first term of (30) is $O_p(B^{-1/2}n^{-1/2})$, and is non-zero due to bootstrap sampling error. In order to force it into the $o_p(n^{-3/2})$ remainder, $B$ must satisfy

$$B^{-1/2}n^{-1/2} \text{ is } o(n^{-3/2}) \;\Rightarrow\; B^{-1/2} \text{ is } o(n^{-1}) \;\Rightarrow\; B^{-1} \text{ is } o(n^{-2})$$

This assumption is made in the Theorem.

The third term of (30) is $O_p(B^{-1}n^{-2})\times o_p(n^{-1/2})\times O(B^{1/2}n^{1/2}) = o_p(n^{-2}B^{-1/2}) = o(n^{-3/2})$. It also belongs in the remainder.

The second term of (30) is $O_p(n^{-1})$. The large-$B$ condition is sufficient to replace the bootstrap average, $B^{-1}\sum_{\ell=1}^{B}$, by the bootstrap expectation, $E_B$, based on sampling with replacement from the observations $(x_i, y_i),\ i = 1, \ldots, n$.

$$\begin{aligned}
b_{boot}-b &= \hat{A}^{-1}(n^{-1}B^{-1}\sum_{\ell}\sum_j(\hat{A}_\ell-\hat{A})\hat{A}^{-1}x_{\ell j}e_{\ell j})+o_p(n^{-3/2})\\
&= \hat{A}^{-1}(n^{-1}B^{-1}\sum_{\ell}\sum_j\hat{A}_\ell\hat{A}^{-1}x_{\ell j}e_{\ell j}-n^{-1}B^{-1}\sum_{\ell}\sum_j x_{\ell j}e_{\ell j})+o_p(n^{-3/2})\\
&= \hat{A}^{-1}(n^{-2}B^{-1}\sum_{\ell}\sum_j\sum_k x_{\ell k}x'_{\ell k}\hat{A}^{-1}x_{\ell j}e_{\ell j})+o_p(n^{-3/2})\\
&= \hat{A}^{-1}(n^{-2}B^{-1}\sum_{\ell}\sum_j x_{\ell j}x'_{\ell j}\hat{A}^{-1}x_{\ell j}e_{\ell j})+o_p(n^{-3/2})\\
&= \hat{A}^{-1}n^{-1}(E_B x_{\ell j}x'_{\ell j}\hat{A}^{-1}x_{\ell j}e_{\ell j})+o_p(n^{-3/2})\\
&= \hat{A}^{-1}n^{-2}\sum_i x_i x'_i\hat{A}^{-1}x_i e_i+o_p(n^{-3/2})\\
&= \hat{A}^{-1}n^{-1}\sum_i x_i p_{ii}e_i+o_p(n^{-3/2})\\
&= (b_{c1}-b)+o_p(n^{-3/2})
\end{aligned}$$

These equivalences mean that an $O(n^{-3/2})$ expansion for any one of the BCs will apply to the others. An expansion for $H^{-1}b_{c1}$ is obtained next, where

$$\begin{aligned}
H^{-1}b_{c1} &= H^{-1}b+H^{-1}(\sum_i x_i x'_i)^{-1}\sum_i x_i p_{ii}e_i\\
&= H^{-1}b+H^{-1}(\sum_i(H')^{-1}z_i z'_i H^{-1})^{-1}\sum_i(H')^{-1}z_i p_{ii}e_i\\
&= H^{-1}b+(\sum_i z_i z'_i)^{-1}\sum_i z_i p_{ii}e_i \qquad (31)
\end{aligned}$$

Substitute the following two results,

$$\begin{aligned} e_i &= y_i - z_i' H^{-1} b \\ &= y_i - z_i'(H^{-1}\beta + (H^{-1}b - H^{-1}\beta)) \\ &= v_i - z_i'\Delta_{zv} + o_p(n^{-1/2}) \qquad (32) \end{aligned}$$

and

$$\begin{aligned} p_{ii} &= x_i'(\sum_i x_i x_i')^{-1} x_i \\ &= z_i' H^{-1}(\sum_i (H')^{-1} z_i z_i' H^{-1})^{-1}(H')^{-1} z_i \\ &= z_i'(\sum_i z_i z_i')^{-1} z_i \\ &= n^{-1} z_i'(n^{-1}\sum_i z_i z_i')^{-1} z_i \\ &= n^{-1} z_i'(I + \Delta_{zz})^{-1} z_i \qquad (33) \\ &= n^{-1} z_i'(I - \Delta_{zz} + o_p(n^{-1/2})) z_i \\ &= n^{-1} z_i' z_i - n^{-1} z_i'\Delta_{zz} z_i + o_p(n^{-3/2}) \qquad (34) \end{aligned}$$

into (31),

$$\begin{aligned} & H^{-1} b_{c1} - H^{-1} b \\ &= (I - \Delta_{zz} + o_p(n^{-1/2}))n^{-1}\sum_i z_i(n^{-1} z_i' z_i - n^{-1} z_i'\Delta_{zz} z_i + o_p(n^{-3/2}))(v_i - z_i'\Delta_{zv} + o_p(n^{-1/2})) \\ &= n^{-2}\sum_i z_i z_i' z_i v_i + n^{-2}(-\Delta_{zz}\sum_i z_i z_i' z_i v_i - \sum_i z_i z_i'\Delta_{zz} z_i v_i - \sum_i z_i z_i' z_i z_i'\Delta_{zv}) + o_p(n^{-3/2}) \\ &= n^{-1}\gamma + n^{-1}(\Delta_\gamma - \Delta_{zz}\gamma - n^{-1}\sum_i z_i(z_i'\Delta_{zz} z_i) v_i - \Omega_{z'z}\Delta_{zv}) + o_p(n^{-3/2}) \qquad (35) \end{aligned}$$

where $\Delta_\gamma = n^{-1}\sum_i (z_i z_i' z_i v_i - \gamma)$. (35) also applies to the other BCs, given the above $O(n^{-3/2})$ equivalences.

## A.5 Proof of Theorem 4

$$\begin{aligned} \text{MSE}(H^{-1}b) &= E(H^{-1}b - H^{-1}\beta)(H^{-1}b - H^{-1}\beta)' \\ &= E\Delta_{zv}\Delta_{zv}' - E(\Delta_{zv}\Delta_{zv}'\Delta_{zz} + \Delta_{zz}\Delta_{zv}\Delta_{zv}') \\ &\quad + E(\Delta_{zz}\Delta_{zv}\Delta_{zv}'\Delta_{zz} + \Delta_{zv}\Delta_{zv}'\Delta_{zz}^2 + \Delta_{zz}^2\Delta_{zv}\Delta_{zv}') + o(n^{-2}) \qquad (36) \end{aligned}$$

Most of the expectations in (36) simplify from cross-product terms having zero expectation due to independence and from $\Delta_{zz}$ and $\Delta_{zv}$ having mean zero.

(i) The expectation of the first term is

$$E\Delta_{zv}\Delta'_{zv} = n^{-1}Ezz'v^2 = \Omega_{v^2}$$

(ii) For expectations of the next two terms in (36) note that the only nonzero terms in

$$E\Delta_{zz}\Delta_{zv}\Delta'_{zv} = n^{-3}E\sum_{ijk}(z_i z'_i - I)(z_j v_j)(z'_k v_k)$$

occur when $i = k = j$. Therefore

$$\begin{aligned} E\Delta_{zz}\Delta_{zv}\Delta'_{zv} &= n^{-2}E(zz' - I)(zv)(z'v) \\ &= n^{-2}(E(zz'zz'v^2) - E(zz'v^2)) \\ &= n^{-2}(\Omega_{z'zv^2} - \Omega_{v^2}) \end{aligned}$$

The matrices $\Omega_{z'zv^2}$ and $\Omega_{v^2}$ are symmetric, therefore

$$E\Delta_{zv}\Delta'_{zv}\Delta_{zz} = n^{-2}(\Omega_{z'zv^2} - \Omega_{v^2})$$

(iii) The $O(n^{-2})$ terms in

$$E\Delta_{zz}\Delta_{zv}\Delta'_{zv}\Delta_{zz} = n^{-4}E\sum_{ijk\ell}(z_i z'_i - I)(z_j v_j)(z'_k v_k)(z_\ell z'_\ell - I)$$

occur when terms are summed according to $(i = j, k = \ell)$, $(i = k, j = \ell)$, and $(i = \ell, j = k)$.

The $(i = j, k = \ell)$ sum is

$$\begin{aligned} n^{-4}E\sum_{ik}(z_i z'_i - I)(z_i v_i)(z'_k v_k)(z_k z'_k - I) &= n^{-2}(E(zz' - I)(zv))(E(zz' - I)(zv))' + o(n^{-2}) \\ &= n^{-2}\gamma\gamma' + o(n^{-2}) \end{aligned}$$

The $(i = k, j = \ell)$ sum is

$$\begin{aligned} & n^{-4}E\sum_{ij}(z_i z'_i - I)(z_j v_j)(z'_i v_i)(z_j z'_j - I) \\ &= n^{-4}E\sum_{ij}(z_i z'_i)(z_j v_j)(z'_i v_i)(z_j z'_j - I) - n^{-4}E\sum_{ij}(z_j v_j)(z'_i v_i)(z_j z'_j - I) \\ &= n^{-4}E\sum_{ij}(z_i z'_i)(z_j v_j)(z'_i v_i)(z_j z'_j - I) + o(n^{-2}) \\ &= n^{-4}E\sum_{ij}(z_i z'_i)(z_j v_j)(z'_i v_i)(z_j z'_j) - n^{-4}E\sum_{ij}(z_i z'_i)(z_j v_j)(z'_i v_i) + o(n^{-2}) \\ &= n^{-4}E\sum_{ij}(z_i z'_i)(z_j v_j)(z'_i v_i)(z_j z'_j) + o(n^{-2}) \\ &= n^{-4}E\sum_{ij}(z_i z'_i)(z_j v_j)(z'_i v_i z_j)z'_j + o(n^{-2}) \end{aligned}$$

$$
\begin{aligned}
&= n^{-4}E\sum_{ij}(z_iz_i')(z_jv_j)(\sum_{a=1}^{K}z_{ia}v_iz_{ja})z_j' + o(n^{-2}) \\
&= n^{-4}E\sum_{a=1}^{K}\sum_{ij}(z_iz_i'z_{ia}v_i)(z_jz_j'z_{ja}v_j) + o(n^{-2}) \\
&= n^{-2}\sum_{a=1}^{K}\Omega^2_{z_av} + o(n^{-2})
\end{aligned}
$$

The $(i=\ell, j=k)$ sum is

$$
\begin{aligned}
&n^{-4}E\sum_{ij}(z_iz_i'-I)(z_jv_j)(z_j'v_j)(z_iz_i'-I) \\
&\quad= n^{-3}E\sum_{i}(z_iz_i'-I)\Omega_{v^2}(z_iz_i'-I) + o(n^{-2}) \\
&\quad= n^{-3}E\sum_{i}(z_iz_i')\Omega_{v^2}(z_iz_i') - n^{-3}E\sum_{i}\Omega_{v^2}(z_iz_i') - n^{-3}E\sum_{i}(z_iz_i')\Omega_{v^2} + n^{-3}E\sum_{i}\Omega_{v^2} + o(n^{-2}) \\
&\quad= n^{-2}Ezz'(z'\Omega_{v^2}z) - n^{-2}\Omega_{v^2} - n^{-2}\Omega_{v^2} + n^{-2}\Omega_{v^2} + o(n^{-2}) \\
&\quad= n^{-2}\Omega_{z'\Omega_{v^2}z} - n^{-2}\Omega_{v^2} + o(n^{-2})
\end{aligned}
$$

Grouping these sums gives

$$
E\Delta_{zz}\Delta_{zv}\Delta_{zv}'\Delta_{zz} = n^{-2}(\gamma\gamma' + \sum_{a=1}^{K}\Omega^2_{z_av} + \Omega_{z'\Omega_{v^2}z} - \Omega_{v^2}) + o(n^{-2})
$$

(iv) The $O(n^{-2})$ part of

$$
E\Delta_{zv}\Delta_{zv}'\Delta_{zz}^2 = n^{-4}E\sum_{ijk\ell}(z_iv_i)(z_j'v_j)(z_kz_k'-I)(z_\ell z_\ell'-I)
$$

is contained in the $(i=j, k=\ell)$, $(i=k, j=\ell)$, and $(i=\ell, j=k)$ terms.

The $(i=j, k=\ell)$ sum is

$$
\begin{aligned}
n^{-4}E\sum_{ik}(z_iv_i)(z_i'v_i)(z_kz_k'-I)(z_kz_k'-I) &= n^{-2}(Ezz'v^2)(Ezz'(z'z)-I-I+I) + o(n^{-2}) \\
&= n^{-2}\Omega_{v^2}(\Omega_{z'z}-I) + o(n^{-2})
\end{aligned}
$$

The $(i=k, j=\ell)$ sum is

$$
\begin{aligned}
&n^{-4}E\sum_{ij}(z_iv_i)(z_j'v_j)(z_iz_i'-I)(z_jz_j'-I) \\
&\quad= n^{-4}E\sum_{ij}(z_iv_i)(z_j'v_j)(z_iz_i')(z_jz_j') + o(n^{-2}) \\
&\quad= n^{-4}E\sum_{ij}(z_iv_i)(z_j'v_jz_i)(z_i'z_j)z_j' + o(n^{-2})
\end{aligned}
$$

$$\begin{aligned} &= n^{-4}E\sum_{ij}(z_i(z_i'v_iz_j)z_j)z_j'v_j + o(n^{-2}) \\ &= n^{-2}\sum_{a=1}^{K}\Omega_{z_av}^2 + o(n^{-2}) \end{aligned}$$

The $(i = \ell, j = k)$ sum is

$$n^{-4}E\sum_{ij}(z_iv_i)(z_j'v_j)(z_jz_j' - I)(z_iz_i' - I)$$

$$\begin{aligned} &= n^{-4}E\sum_{ij}(z_iv_i)(z_j'v_j)(z_jz_j')(z_iz_i') + o(n^{-2}) \\ &= n^{-4}E\sum_{ij}(z_iv_i)(z_j'v_jz_j)(z_j'z_i)z_i' + o(n^{-2}) \\ &= n^{-4}E\sum_{ij}z_iz_i'v_i(z_i'z_j)z_j'z_jv_j + o(n^{-2}) \\ &= n^{-2}\sum_{a=1}^{K}\Omega_{z_av}\gamma_a + o(n^{-2}) \end{aligned}$$

where $\gamma_a = Ez'zvz_a$ is the $a^{th}$ element of $\gamma$. Grouping these sums gives

$$E\Delta_{zv}\Delta_{zv}'\Delta_{zz}^2 = n^{-2}(\Omega_{v^2}(\Omega_{z'z} - I) + \sum_{a=1}^{K}\Omega_{z_av}^2 + \sum_{a=1}^{K}\Omega_{z_av}\gamma_a) + o(n^{-2})$$

and

$$E\Delta_{zz}^2\Delta_{zv}\Delta_{zv}' = n^{-2}((\Omega_{z'z} - I)\Omega_{v^2} + \sum_{a=1}^{K}\Omega_{z_av}^2 + \sum_{a=1}^{K}\Omega_{z_av}\gamma_a) + o(n^{-2})$$

Result (17) follows from substituting expectations (i) to (iv) into (36) and simplifying.

For the MSE of $H^{-1}b_c$, use

$$\begin{aligned} \text{MSE}(H^{-1}b_c) &= E(H^{-1}b_c - H^{-1}\beta)(H^{-1}b_c - H^{-1}\beta)' \\ &= E((H^{-1}b_c - H^{-1}b) + (H^{-1}b - H^{-1}\beta))((H^{-1}b_c - H^{-1}b) + (H^{-1}b - H^{-1}\beta))' \\ &= E(H^{-1}b_c - H^{-1}b)(H^{-1}b_c - H^{-1}b)' + E(H^{-1}b_c - H^{-1}b)(H^{-1}b - H^{-1}\beta)' \\ &\quad + E(H^{-1}b - H^{-1}\beta)(H^{-1}b_c - H^{-1}b)' + \text{MSE}(H^{-1}b) \end{aligned} \tag{37}$$

MSE$(H^{-1}b)$ is given by (17). From (35), $E(H^{-1}b_c - H^{-1}b)(H^{-1}b_c - H^{-1}b)' = n^{-2}\gamma\gamma' + o(n^{-2})$. The remaining expectations in (37), $E(H^{-1}b_c - H^{-1}b)(H^{-1}b - H^{-1}\beta)'$ and its transpose, require five expectations of products of terms that appear in (35) and (7). These are

$$En^{-1}\gamma(-\Delta_{zz}\Delta_{zv})' = -n^{-2}\gamma\gamma'$$

$$\begin{aligned} En^{-1}\Delta_{\gamma}\Delta'_{zv} &= n^{-3}E\sum_{ij}(z_iz_i'z_iv_i-\gamma)z_j'v_j \\ &= n^{-2}\Omega_{z'zv^2} \end{aligned}$$

$$\begin{aligned} -En^{-1}\Delta_{zz}\gamma\Delta'_{zv} &= n^{-3}E\sum_{ij}(z_iz_i'-I)\gamma)z_j'v_j \\ &= -n^{-2}Ez(z'\gamma)z'v \\ &= -n^{-2}\sum_{a=1}^{K}Ez(z'\gamma)z_a\gamma_a \\ &= -n^{-2}\sum_{a=1}^{K}\Omega_{z_av}\gamma_a \end{aligned}$$

$$\begin{aligned} -En^{-2}\sum_i z_i(z_i'\Delta_{zz}z_i)v_i\Delta'_{zv} &= n^{-2}E\sum_i z_iv_i\Delta'_{zv}(z_i'\Delta_{zz}z_i) \\ &= n^{-4}E\sum_{ijk} z_iv_iz_j'v_j(z_i'z_kz_k'z_i-z_i'z_i) \\ &= n^{-4}E\sum_{ij} z_iv_iz_j'v_j(z_i'z_jz_j'z_i-z_i'z_i) \\ &= n^{-4}E\sum_{ij} z_iv_iz_j'v_j(z_i'z_j)^2 \\ &= n^{-4}E\sum_{ij}\sum_{a=1}^{K} z_iz_i'(v_iz_{ia}z_{ja}v_j)z_jz_j' \\ &= n^{-2}\sum_{a=1}^{K}\Omega^2_{z_av}+o_p(n^{-2}) \end{aligned}$$

$$-En^{-1}\Omega_{z'z}\Delta_{zv}\Delta'_{zv} = -n^{-2}\Omega_{z'z}\Omega_{v^2}$$

(18) follows from substituting these expectations into (37).

### A.6 Proof of Theorem 5

To estimate the variance of $n\widehat{\text{Bias}(b)}$, reparametrize and expand using (32) and (34).

$$\begin{aligned} & n\widehat{\text{Bias}(b)} = -\hat{A}^{-1}n^{-1}\sum_i x_i(np_{ii})e_i \\ = \quad & -((H')(I+\Delta_{zz})H^{-1})^{-1}n^{-1}\sum_i (H')^{-1}z_i(z_i'z_i-z_i'\Delta_{zz}z_i+o_p(n^{-1/2}))(v_i-z_i'\Delta_{zv}+o_p(n^{-1/2})) \\ = \quad & -H(I-\Delta_{zz}+o_p(n^{-1/2}))H'n^{-1}(H')^{-1}(\sum_i z_iz_i'z_iv_i-\sum_i z_iz_i'z_iz_i'\Delta_{zv}-\sum_i z_iz_i'\Delta_{zz}z_iv_i)+o_p(n^{-1/2}) \\ = \quad & -H(I-\Delta_{zz})(n^{-1}\sum_i z_iz_i'z_iv_i-n^{-1}\sum_i z_iz_i'z_iz_i'\Delta_{zv}-n^{-1}\sum_i z_iz_i'\Delta_{zz}z_iv_i)+o_p(n^{-1/2}) \\ = \quad & -H(I-\Delta_{zz})(\gamma+\Delta_\gamma-\Omega_{z'z}\Delta_{zv}-c)+o_p(n^{-1/2}) \end{aligned}$$

$$\begin{aligned} &= -H\gamma - H(\Delta_\gamma - \Omega_{z'z}\Delta_{zv} - c - \Delta_{zz}\gamma) + o_p(n^{-1/2}) \\ &= -H\gamma - H\xi_{-1/2} + o_p(n^{-1/2}) \end{aligned}$$

where

$$\begin{aligned} \xi_{-1/2} &= \Delta_\gamma - \Omega_{z'z}\Delta_{zv} - c - \Delta_{zz}\gamma \\ &= n^{-1}\sum_i \psi_i, \\ \psi_i &= z_i z_i' z_i v_i - \gamma - \Omega_{z'z} z_i v_i - c_i - (z_i z_i' - I)\gamma \\ &= z_i z_i' z_i v_i - \Omega_{z'z} z_i v_i - c_i - z_i z_i'\gamma, \end{aligned}$$

and

$$c_i = n^{-1}\sum_j z_j z_j'(z_i z_i' - I) z_j v_j$$

so the $a^{th}$ element of $c_i$ is

$$\begin{aligned} c_{ia} &= n^{-1}\sum_j z_{ja}\left((z_j' z_i)^2 v_j - (z_j' z_j) v_j\right) \\ &= z_i'(n^{-1}\sum_j z_j z_j' z_{ja} v_j) z_i - \gamma_a + o_p(n^{-1/2}) \\ &= z_i'\Omega_{z_a v} z_i - \gamma_a + o_p(n^{-1/2}) \end{aligned}$$

If the $\psi_i$'s and $H$ were known, a consistent variance estimator could be constructed as

$$\widehat{\mathrm{Var}}(n^{1/2}(n\widehat{\mathrm{Bias}(b)})) = n^{-1}\sum_i (H\psi_i)(H\psi_i)' \tag{38}$$

Asymptotic normality of $n\widehat{\mathrm{Bias}(b)}$ follows from the *iid* assumption, Cramér-Wold device and Lindeberg-Levy central limit theorem (White (1984, p.108)) if the variance of $\psi_i$ is finite. A sufficient condition for this is that the moments $E(z^6 v^2)$ are finite. Since $H$ is fixed and nonsingular, this is the same as the moments $E(x^6 v^2)$ being finite, proving part (a).

$\widehat{H\psi}_i$ in (24) uses the following consistent estimators of the four terms of $H\psi_i$.

$$\begin{aligned} H z_i z_i' z_i v_i &= H H' x_i x_i' H H' x_i v_i \\ &= A^{-1} x_i x_i' A^{-1} x_i v_i \\ &= \hat{A}^{-1} x_i (n p_{ii}) e_i + o_p(1) \\ H\Omega_{z'z} z_i v_i &= H(E z z'(z'z)) H' z_i v_i \\ &= H(E H' x x' H H'(x' H H' x)) x_i v_i \\ &= A^{-1}(E x x'(x' A^{-1} x) A^{-1}) x_i v_i \\ &= \hat{A}^{-1}(\hat{A}_{np})\hat{A}^{-1} x_i e_i + o_p(1) \end{aligned}$$

$$\begin{aligned}
(Hc_i)_a &= \sum_b H_{ab}c_{ib} \\
&= \sum_b H_{ab}(z_i'\Omega_{z_b v}z_i - \gamma_b) \\
&= \sum_b H_{ab}(z_i'Ezz'z_bv)z_i - \sum_b H_{ab}E(z_b(z'z)v) \\
&= \sum_b H_{ab}(x_i'H(EH'xx'H(H_{b\cdot}'x)v)H'x_i) - \sum_b H_{ab}E((H_{b\cdot}'x)(x'HH'x)v) \\
&= x_i'HH'(Exx'(\sum_b H_{ab}H_{b\cdot}'x)v)HH'x_i - E(x'HH'x(\sum_b H_{ab}H_{b\cdot}'x)v) \\
&= x_i'A^{-1}(Exx'((A^{-1})_{a\cdot}x)v)A^{-1}x_i - E(x'A^{-1}x((A^{-1})_{a\cdot}x)v) \\
&= x_i'A^{-1}(Exx'(\sum_b(A^{-1})_{ab}x_b)v)A^{-1}x_i - E(x'A^{-1}x(\sum_b(A^{-1})_{ab}x_b)v) \\
&= x_i'\hat{A}^{-1}(\sum_b(\hat{A}^{-1})_{ab}\hat{A}_{x_bv})\hat{A}^{-1}x_i - (\hat{A}^{-1})_{a\cdot}\hat{\gamma}^* + o_p(1) \\
Hz_iz_i'\gamma &= Hz_iz_i'(E(z'z)zv) \\
&= HH'x_ix_i'HE(H'xv(x'HH'x)) \\
&= A^{-1}x_ix_i'A^{-1}E(x'A^{-1}x)xv \\
&= \hat{A}^{-1}x_ix_i'\hat{A}^{-1}\hat{\gamma}^* + o_p(1)
\end{aligned}$$

The new notation is described below (24).

### A.7 Computing Expectations and MSEs

Let $X$ be the usual $n \times K$ matrix from stacking the $x_i'$ vectors. The estimators each can be expressed in the form $\sum_i f(X, x_i)y_i$, where the $K$-element vector function $f$ is

$$\begin{aligned}
f(X, x_i) &= (X'X)^{-1}x_i && \text{for } b \\
&= (X'X)^{-1}\left(x_i + p_{ii}x_i - (\textstyle\sum_j x_jx_j'p_{jj})(X'X)^{-1}x_i\right) && \text{for } b_{c1} \\
&= \left(\textstyle\sum_j x_jx_j'(1+p_{jj})\right)^{-1} x_i(1+p_{ii}) && \text{for } b_{c2} \\
&= \left(\textstyle\sum_j x_jx_j'(1-p_{jj})^{-1}\right)^{-1} x_i(1-p_{jj})^{-1} && \text{for } b_{c3} \\
&= 2(X'X)^{-1}x_i - B^{-1}\textstyle\sum_{\ell=1}^B \left((\sum_{j=1}^n x_{(\ell j)}x_{(\ell j)}')^{-1}\sum_{j=1}^n x_{(\ell j)}I[(\ell j) = i]\right) && \text{for } b_{boot}
\end{aligned}$$

where $B$ is the number of bootstrap samples, $\ell$ indexes the bootstrap sample, $j$ indexes the observations in the bootstrap sample, and $(\ell j)$ equals $i^*$, say, if the $j^{th}$ observation in the $\ell^{th}$ bootstrap sample is chosen to be the $(i^*)^{th}$ observation in the original sample.

Let the conditional moments be $E(y_i|x_i) \equiv \mu(x_i)$ and $\text{Var}(y_i|x_i) \equiv \sigma^2(x_i)$. Then the conditional moments are

$$E(\hat{\beta}|X) = \sum_i f(X, x_i)\mu(x_i)$$

$$\begin{aligned}
\text{Var}(\hat{\beta}|X) &= E\left((\hat{\beta}-E(\hat{\beta}|X))(\hat{\beta}-E(\hat{\beta}|X))'|X\right) \\
&= E\left(\sum_i f(X,x_i)(y_i-\mu(x_i))\sum_j f(X,x_j)'(y_j-\mu(x_j))\right) \\
&= \sum_i f(X,x_i)f(X,x_i)'\sigma^2(x_i)
\end{aligned}$$

The unconditional expectation is then

$$E(\hat{\beta}) = E_X(E(\hat{\beta}|X))$$

where $E_X$ denotes expectation over $X$. The unconditional variance is

$$\begin{aligned}
\text{Var}(\hat{\beta}) &= E(\hat{\beta}-E(\hat{\beta}))(\hat{\beta}-E(\hat{\beta}))' \\
&= E\left(\hat{\beta}-E(\hat{\beta}|X)+E(\hat{\beta}|X)-E(\hat{\beta})\right)\left(\hat{\beta}-E(\hat{\beta}|X)+E(\hat{\beta}|X)-E(\hat{\beta})\right)' \\
&= E_X\left(E_y(\hat{\beta}-E(\hat{\beta}|X)+E(\hat{\beta}|X)-E(\hat{\beta}))(\hat{\beta}-E(\hat{\beta}|X)+E(\hat{\beta}|X)-E(\hat{\beta}))'|X\right) \\
&= \left(E_X(E_y(\hat{\beta}-E(\hat{\beta}|X))(\hat{\beta}-E(\hat{\beta}|X))'|X)\right)+\left(E_X(E(\hat{\beta}|X)-E(\hat{\beta}))(E(\hat{\beta}|X)-E(\hat{\beta}))'\right) \\
&= E_X(\text{Var}(\hat{\beta})|X)+\left(E_X(E(\hat{\beta}|X)-E(\hat{\beta}))(E(\hat{\beta}|X)-E(\hat{\beta}))'\right)
\end{aligned}$$

Similarly,

$$\text{MSE}(\hat{\beta}) = E_X(\text{Var}(\hat{\beta})|X)+\left(E_X(E(\hat{\beta}|X)-\beta)(E(\hat{\beta}|X)-\beta)'\right)$$

These results can be used to approximate $E(\hat{\beta})$, $\text{Var}(\hat{\beta})$, and $\text{MSE}(\hat{\beta})$ by averaging across simulated $X$'s to estimate the $E_X$ operation. It is not necessary to simulate the $y$'s.

### A.7.1 Results for Section 5.1: $x$ takes only two values

The set of possible X's is small enough that the bias and RMSE results can be computed directly without simulation. Let the possible values of $x$ be $x_{(1)}$ and $x_{(2)}$, which occur in the population and the sample with $\Pr(x = x_{(j)}) = p_{(j)}$, $j = 1, 2$. Let the conditional means and variances of $y$ be $\mu_{(j)}$ and $\sigma^2_{(j)}$. Since the order of observations is irrelevant, then X is fully described by $n$ and $\hat{p}_{(1)}$, which is the proportion of $x_i$'s in the sample that equal $x_{(1)}$. For each of the $n+1$ possible values of $\hat{p}_{(1)}$: 0, $\frac{1}{n}$, $\frac{2}{n}, \ldots, \frac{n-1}{n}$, 1, compute conditional moments

$$\begin{aligned}
E(\hat{\beta}|\hat{p}_{(1)},n) &= \sum_{j=1}^{2} f(\hat{p}_{(1)},n,x_{(j)})\mu_{(j)}(n\hat{p}_{(j)}) \\
\text{Var}(\hat{\beta}|\hat{p}_{(1)},n) &= \sum_{j=1}^{2} f(\hat{p}_{(1)},n,x_{(j)})^2\sigma^2_{(j)}(n\hat{p}_{(j)})
\end{aligned}$$

where, from the definitions in Section 3.1

$$\begin{array}{rcll} f(\hat{p}_{(1)}, n, x_{(j)}) & = & n^{-1}A(\hat{p}_{(1)})^{-1}x_{(j)} & \text{for } b \\ & = & n^{-1}A(\hat{p}_{(1)})^{-1}x_{(j)} + n^{-2}A(\hat{p}_{(1)})^{-2}\left(x_{(j)}^3 - \frac{B(\hat{p}_{(1)})}{A(\hat{p}_{(1)})}\right)x_{(j)} & \text{for } b_{c1} \\ & = & n^{-1}\left(A(\hat{p}_{(1)}) + n^{-1}(\frac{B(\hat{p}_{(1)})}{A(\hat{p}_{(1)})})\right)^{-1}\left(x_{(j)} + n^{-1}A(\hat{p}_{(1)})^{-1}x_{(j)}^3\right) & \text{for } b_{c2} \\ & = & n^{-1}C(\hat{p}_{(1)})^{-1}\left(x_{(j)}/(1 - n^{-1}x_{(j)}^2A(\hat{p}_{(1)})^{-1})\right) & \text{for } b_{c3} \\ & = & 2n^{-1}A(\hat{p}_{(1)})^{-1}x_{(j)} - n^{-1}x_{(j)}\hat{p}_{(j)}^{-1}\sum_{\hat{\hat{p}}_{(1)}}\Pr(\hat{\hat{p}}_{(1)}|\hat{p}_{(1)}, n)A(\hat{\hat{p}}_{(1)})^{-1}\hat{\hat{p}}_{(j)} & \text{for } b_{boot} \end{array}$$

where

$$\begin{array}{rcl} A(\hat{p}_{(1)}) & = & \hat{p}_{(1)}x_{(1)}^2 + \hat{p}_{(2)}x_{(2)}^2 \\ B(\hat{p}_{(1)}) & = & \hat{p}_{(1)}x_{(1)}^4 + \hat{p}_{(2)}x_{(2)}^4 \\ C(\hat{p}_{(1)}) & = & \left(\dfrac{\hat{p}_{(1)}x_{(1)}^2}{1 - n^{-1}x_{(1)}^2A(\hat{p}_{(1)})^{-1}}\right) + \left(\dfrac{\hat{p}_{(2)}x_{(2)}^2}{1 - n^{-1}x_{(2)}^2A(\hat{p}_{(1)})^{-1}}\right) \\ \hat{p}_{(2)} & = & 1 - \hat{p}_{(1)} \end{array}$$

Since $\hat{A}$ cannot be singular in this example, $b_{c1} = b_{c1,M}$ and $b_{boot} = b_{boot,M}$. The above expression for $b_{boot}$ is derived later in this Section. Because of the *iid* sampling, $\hat{p}_{(1)}$ follows a binomial distribution:

$$\Pr(\hat{p}_{(1)}|p_{(1)}, n) = \left(\begin{array}{c} n \\ n\hat{p}_{(1)} \end{array}\right)(n\hat{p}_{(1)})^{p_{(1)}}(n(1 - \hat{p}_{(1)}))^{p_{(2)}} \quad , \ \hat{p}_{(1)} = 0, \frac{1}{n}, \frac{2}{n}, \ldots, \frac{n-1}{n}, 1$$

For large $n$, computational problems were avoided by calculating these probabilities recursively using

$$\Pr(\hat{p}_{(1)} = 0|p_{(1)}, n) = (1 - p_{(1)})^n$$

and

$$\Pr(\hat{p}_{(1)} + n^{-1}|p_{(1)}, n) = \Pr(\hat{p}_{(1)}|p_{(1)}, n)\left(\frac{1 - \hat{p}_{(1)}}{\hat{p}_{(1)} + n^{-1}}\right)\left(\frac{p_{(1)}}{1 - p_{(1)}}\right)$$

The moments are then computed as

$$\begin{array}{rcl} E(\hat{\beta}) & = & \sum_{\hat{p}_{(1)}}\Pr(\hat{p}_{(1)})E(\hat{\beta}|\hat{p}_{(1)}) \\ \text{Var}(\hat{\beta}) & = & \sum_{\hat{p}_{(1)}}\Pr(\hat{p}_{(1)})(\text{Var}(\hat{\beta}|\hat{p}_{(1)}) + (E(\hat{\beta}|\hat{p}_{(1)}) - E(\hat{\beta}))^2) \\ \text{MSE}(\hat{\beta}) & = & \sum_{\hat{p}_{(1)}}\Pr(\hat{p}_{(1)})(\text{Var}(\hat{\beta}|\hat{p}_{(1)}) + (E(\hat{\beta}|\hat{p}_{(1)}) - \beta)^2) \end{array}$$

The $f(\hat{p}_{(1)}, n, x_{(j)})$ expression for $b_{boot}$ given above follows from $\bar{b}_B$ in (12) being the expectation of OLS under *iid* sampling from the empirical distribution. Let $\hat{\hat{p}}_{(1)}$ be the proportion of observations in a bootstrap sample having $x_{ji} = x_{(1)}$, and compute $\Pr(\hat{\hat{p}}_{(1)}|\hat{p}_{(1)}, n)$ in the same way as the $\Pr(\hat{p}_{(1)}|p_{(1)}, n)$'s described earlier. Then applying the procedure just described for $b$ to the empirical distribution gives

$$E(b_{boot}|p_{(1)}=\hat{p}_{(1)},\mu_{(j)}=\hat{\mu}_{(j)},n)=\sum_{\hat{\hat{p}}_{(1)}}\Pr(\hat{\hat{p}}_{(1)}|\hat{p}_{(1)},n)\left(\sum_j(n^{-1}A(\hat{\hat{p}}_{(1)})^{-1}x_{(j)})\hat{\mu}_{(j)}n\hat{\hat{p}}_{(j)}\right)$$

where $\hat{\mu}_{(j)}=\sum_{\{i|x_i=x_{(j)}\}}y_i/(n\hat{p}_{(j)})$. Therefore

$$\begin{aligned}
&E(b_{boot}|p_{(1)}=\hat{p}_{(1)},\mu_{(j)}=\hat{\mu}_{(j)},n)\\
=\;&\sum_{\hat{\hat{p}}_{(1)}}\Pr(\hat{\hat{p}}_{(1)}|\hat{p}_{(1)},n)\left(\sum_j(n^{-1}A(\hat{\hat{p}}_{(1)})^{-1}x_{(j)})n\hat{\hat{p}}_{(j)}\right)\left(\sum_{\{i|x_i=x_{(j)}\}}y_i/(n\hat{p}_{(j)})\right)\\
=\;&\sum_{\hat{\hat{p}}_{(1)}}\Pr(\hat{\hat{p}}_{(1)}|\hat{p}_{(1)},n)\sum_j(n^{-1}A(\hat{\hat{p}}_{(1)})^{-1}x_{(j)})(\hat{\hat{p}}_{(j)}/(\hat{p}_{(j)}))(\sum_{\{i|x_i=x_{(j)}\}}y_i)\\
=\;&\sum_i f_B(\hat{p}_{(1)},n,x_{(j)})y_i
\end{aligned}$$

where $f_B(\hat{p}_{(1)},n,x_{(j)})=\sum_{\hat{\hat{p}}_{(1)}}\Pr(\hat{\hat{p}}_{(1)}|\hat{p}_{(1)},n)n^{-1}A(\hat{\hat{p}}_{(1)})^{-1}x_{(j)}(\hat{\hat{p}}_{(j)}/(\hat{p}_{(j)}))$. Values of zero can occur in the denominators in the computations for $b_{c3}$ and $b_{boot}$, but they always have zero-probability outcomes, so they can be replaced with an arbitrary non-zero number. Finally, assume that the number of bootstrap replications is large enough that the bootstrap sampling error, $\bar{b}_B-E(b_{boot}|p_{(1)}=\hat{p}_{(1)},\mu_{(j)}=\hat{\mu}_{(j)},n)$, can be ignored.

## A.8 The *ncp* used in Sections 5.1.4 and 5.1.5

Test the restriction $\beta_0=0$ in the augmented model $y=\beta_0+x\beta+v^*$ using a Wald test, allowing $\mathrm{Var}(v^*|x=x_{(j)})\equiv\sigma_j^2$ to depend on $x$. Under the local alternative $\beta_0=n^{-1/2}\tau$ for some fixed $\tau$, the asymptotic distribution of this Wald statistic is noncentral $\chi^2$ with an *ncp*

$$ncp=\frac{\tau^2}{\mathrm{AVar}(n^{1/2}\hat{\beta}_0)}$$

where, with $x_{(1)}=1$ and $x_{(2)}=a$,

$$\begin{aligned}
\beta_0=n^{-1/2}\tau\;&=\;\frac{a\mu_1-\mu_2}{x_2-1}\\
\mathrm{AVar}(n^{1/2}\hat{\beta}_0)\;&=\;((p+(1-p)a^2)^2(p\sigma_1^2+(1-p)\sigma_2^2)\\
&\quad-2(p+(1-p)a)(p+(1-p)a^2)(p\sigma_1^2+(1-p)a\sigma_2^2)\\
&\quad+(p\sigma_1^2+(1-p)a^2\sigma_2^2)(p+(1-p)a)^2)/(p+(1-p)a^2-(p+(1-p)a)^2)^2
\end{aligned}$$

and $\hat{\beta}_0$ is the OLS estimator of the intercept.

## A.9 The *ncp* used in Sections 5.2 and 5.3

Under the local alternative $\theta = n^{-1/2}\tau$ for some fixed $\tau$, the asymptotic distribution of a heteroskedasticity-consistent Wald statistic is noncentral $\chi^2$ with non-centrality parameter

$$ncp = \frac{\tau^2}{\mathrm{AVar}(n^{1/2}\hat{\theta})}$$

where $\mathrm{AVar}(n^{1/2}\hat{\theta})$ is element (3,3) of $E(\widetilde{x}\widetilde{x}')^{-1}E(\widetilde{x}\widetilde{x}'u^2)E(\widetilde{x}\widetilde{x}')^{-1}$ and $\widetilde{x} = [1\; x^*\; (x^*)^2]'$ is the regressor vector from an augmented linear regression model described earlier. $\theta$ is the third element of $E(\widetilde{x}\widetilde{x}')^{-1}E(\widetilde{x}y)$. The moment result $Ex^r = r!\phi^r$ for $x$ an exponential random variable with mean $\phi$, lead to the moment matrices given below. Set $\beta_1 = 0$ with no loss of generality.

$$E\widetilde{x}\widetilde{x}' = \begin{bmatrix} 1 & p_{(1)} + p_{(2)}a & 2(p_{(1)} + p_{(2)}a^2) \\ p_{(1)} + p_{(2)}a & 2(p_{(1)} + p_{(2)}a^2) & 6(p_{(1)} + p_{(2)}a^3) \\ 2(p_{(1)} + p_{(2)}a^2) & 6(p_{(1)} + p_{(2)}a^3) & 24(p_{(1)} + p_{(2)}a^4) \end{bmatrix}$$

$$E\widetilde{x}y = p_{(2)}\beta_2 \begin{bmatrix} a \\ 2a^2 \\ 6a^3 \end{bmatrix}$$

Element $(b, c)$ of the $3 \times 3$ matrix $E\widetilde{x}\widetilde{x}'u^2$ is given by

$$E(x^*)^r u^2 = r!(p_{(1)}\sigma^2_{(1)} + p_{(2)}\sigma^2_{(2)}a^r)$$

where $r = b + c - 2$. The size of the *ncp* is controlled by adjusting $\beta_2$.

QSEP RESEARCH REPORTS - Recent Releases

| Number | Title | Author(s) |
|---|---|---|
| (2003) | | |
| No. 380: | Population Aging, Productivity, and Growth in Living Standards | W. Scarth |
| No. 381: | The Transition from Good to Poor Health: An Econometric Study of the Older Population | N.J. Buckley<br>F.T. Denton<br>A.L. Robb<br>B.G. Spencer |
| No. 382: | The Evolution of High Incomes In Canada, 1920-2000 | E. Saez<br>M.R. Veall |
| No. 383: | Population Change and Economic Growth: The Long-Term Outlook | F.T. Denton<br>B.G. Spencer |
| No. 384: | The Economic Legacy of Divorced and Separated Women in Old Age | L. McDonald<br>A.L. Robb |
| No. 385: | National Catastrophic Drug Insurance Revisited: Who Would Benefit from Senator Kirby's Recommendations? | T.F. Crossley<br>P.V. Grootendorst<br>M.R. Veall |
| No. 386: | Wages in Canada: SCF, SLID, LFS and the Skill Premium | A.L. Robb<br>L. Magee<br>J.B. Burbidge |
| No. 387: | Socioeconomic Influence on the Health of Older People: Estimates Based on Two Longitudinal Surveys | N.J. Buckley<br>F.T. Denton<br>A.L. Robb<br>B.G. Spencer |
| No. 388: | An Invitation to Multivariate Analysis: An Example About the Effect of Educational Attainment on Migration Propensities in Japan | A. Otomo<br>K-L. Liaw |
| (2004) | | |
| No. 389: | Financial Planning for Later Life: Subjective Understandings of Catalysts and Constraints | C.L. Kemp<br>C.J. Rosenthal<br>M. Denton |
| No. 390: | Exploring the Use of a Nonparametrically Generated Instrumental Variable in the Estimation of a Linear Parametric Equation | F.T. Denton |
| No. 391: | Borrowing Constraints, the Cost of Precautionary Saving, and Unemployment Insurance | T.F. Crossley<br>H.W. Low |

| Number | Title | Author(s) |
|---|---|---|
| No. 392: | Healthy Aging at Older Ages: Are Income and Education Important? | N.J. Buckley<br>F.T. Denton<br>A.L. Robb<br>B.G. Spencer |
| (2005) | | |
| No. 393: | Where Have All The Home Care Workers Gone? | M. Denton<br>I.S. Zeytinoglu<br>S. Davies<br>D. Hunter |
| No. 394: | Survey Results of the New Health Care Worker Study: Implications of Changing Employment Patterns | I.S. Zeytinoglu<br>M. Denton<br>S. Davies<br>A. Baumann<br>J. Blythe<br>A. Higgins |
| No. 395: | Unexploited Connections Between Intra- and Inter-temporal Allocation | T.F. Crossley<br>H.W. Low |
| No. 396: | Measurement Errors in Recall Food Expenditure Data | N. Ahmed<br>M. Brzozowski<br>T.F. Crossley |
| No. 397: | The Effect of Health Changes and Long-term Health on the Work Activity of Older Canadians | D.W.H. Au<br>T.F. Crossley<br>M. Schellhorn |
| No. 398: | Population Aging and the Macroeconomy: Explorations in the Use of Immigration as an Instrument of Control | F.T. Denton<br>B.G. Spencer |
| No. 399: | Users and Suppliers of Physician Services: A Tale of Two Populations | F.T. Denton<br>A. Gafni<br>B.G. Spencer |
| No. 400: | MEDS-D Users’ Manual | F.T. Denton<br>C.H. Feaver<br>B.G. Spencer |
| No. 401: | MEDS-E Users’ Manual | F.T. Denton<br>C.H. Feaver<br>B.G. Spencer |

| Number | Title | Author(s) |
|---|---|---|
| No. 402: | Socioeconomic Influences on the Health of Older Canadians: Estimates Based on Two Longitudinal Surveys<br>(Revised Version of No. 387) | N.J. Buckley<br>F.T. Denton<br>A.L. Robb<br>B.G. Spencer |
| No. 403: | Population Aging in Canada: Software for Exploring the Implications for the Labour Force and the Productive Capacity of the Economy | F.T. Denton<br>C.H. Feaver<br>B.G. Spencer |
| (2006) | | |
| No. 404: | Joint Taxation and the Labour Supply of Married Women: Evidence from the Canadian Tax Reform of 1988 | T.F. Crossley<br>S.H.Jeon |
| No. 405: | The Long-Run Cost of Job Loss as Measured by Consumption Changes | M. Browning<br>T.F. Crossley |
| No. 406: | Do the Rich Save More in Canada? | S. Alan<br>K. Atalay<br>T.F. Crossley |
| No. 407: | The Social Cost-of-Living: Welfare Foundations and Estimation | T.F. Crossley<br>K. Pendakur |
| No. 408: | The Top Shares of Older Earners in Canada | M.R. Veall |
| No. 409: | Estimating a Collective Household Model with Survey Data on Financial Satisfaction | R. Alessie<br>T.F. Crossley<br>V.A. Hildebrand |
| No. 410: | Physician Labour Supply in Canada: a Cohort Analysis | T.F. Crossley<br>J. Hurley<br>S.H. Jeon |
| No. 411: | Well-Being Throughout the Senior Years: An Issues Paper on Key Events and Transitions in Later Life | M. Denton<br>K. Kusch |
| No. 412: | Satisfied Workers, Retained Workers: Effects of Work and Work Environment on Homecare Workers' Job Satisfaction, Stress, Physical Health, and Retention | I.U. Zeytinoglu<br>M. Denton |
| (2007) | | |
| No. 413: | Gender Inequality in the Wealth of Older Canadians | M. Denton<br>L. Boos |
| No. 414: | Which Canadian Seniors are Below the Low-Income Measure? | M. Veall |

QSEP RESEARCH REPORTS - Recent Releases

| Number | Title | Author(s) |
|---|---|---|
| No. 415: | On the Sensitivity of Aggregate Productivity Growth Rates to Noisy Measurement | F.T. Denton |
| No. 416: | Initial Destination Choices of Skilled-worker Immigrants from South Asia to Canada: Assessment of the Relative Importance of Explanatory Factors | L. Xu<br>K.L. Liaw |
| No. 417: | Problematic Post-Landing Interprovincial Migration of the Immigrants in Canada: From 1980-83 through 1992-95 | K.L. Liaw<br>L. Xu |
| No. 418: | The Adequacy of Retirement Savings: Subjective Survey Reports by Retired Canadians | S. Alan<br>K. Atalay<br>T.F. Crossley |
| No. 419: | Ordinary Least Squares Bias and Bias Corrections for *iid* Samples | L. Magee |