NBER TECHNICAL PAPER SERIES

# CONSISTENT ESTIMATION USING DATA FROM MORE THAN ONE SAMPLE

William T. Dickens

Brian A. Ross

Technical Working Paper No. 33

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge MA 02138
March 1984

We would like to thank Kevin Lang, Paul Ruud, and James Stock for helpful discussions and the Institute of Industrial Relations at Berkeley for generous research support. The research reported here is part of the NBER's research program in Labor Studies. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

NBER Technical Working Paper #33
March 1984

# Consistent Estimation Using Data From More Than One Sample

## ABSTRACT

This paper considers the estimation of linear models when group average data from more than one sample is used. Conditions under which OLS coefficient estimates are consistent are indentified. The standard OLS covariance estimate is shown to be inconsistent and a consistent estimator is proposed. Finally, since the conditions under which OLS is consistent are quite restrictive, several estimators which are consistent in many cases where OLS is not are developed. The large sample distribution properties and an estimator for the asymptotic covariance matrix for the most general of these alternative estimators is also presented.

One important application of these findings is to estimating compensating wage differences. Past authors, beginning with Thaler and Rosen (1976) have argued that finer classification schemes would reduce errors-in-variable bias. The analysis presented here suggests that the opposite is true if finer classification results in fewer observations per classification. This could explain why authors using the broader (industry) classification schemes have found larger compensating differences and suggests that these estimates may be closer to the true values.

William T. Dickens
Department of Economics
University of California, Berkeley
Berkeley, CA 94720

Brian A. Ross
Department of Economics
Miami University
Oxford, Ohio 45056

## I. Introduction

The econometric literature has thoroughly discussed the estimation of the linear model when data from one sample are grouped and where group means are used as both dependent and independent variables.[1,2] However the literature has yet to discuss extensively the estimation of the linear model when the data are grouped and drawn from two different samples.

For example, researchers frequently estimate linear models in which industry, occupation or state are the unit of observation; group means are used as data, but the means are calculated in different samples.[3] Another common estimation procedure augments micro data from one data set with group averages from another data set. Often the researcher finds no micro data set complete. Each excludes at least one regression variable. Consequently, for each variable not contained within the primary data set, the researcher attributes to each individual, firm, or other micro observational unit that unit's occupation, industry or other identifiable group average, which the researcher computes in an auxillary data set.

The latter procedure has been used frequently to estimate the value of job safety implicit in workers' job choice.[4] Despite its common use, the literature appears confused about the OLS estimator's statistical properties. Several authors have suggested that such estimates are subject to conventional errors in variables bias while at least one other has claimed that they are unbiased but that the error variance will be heteroskedastic.[5]

The following analysis presents necessary and sufficient conditions for the OLS estimator to be consistent, whenever one employs either of these aforementioned estimation techniques. The analysis suggests that for either technique, OLS should provide reasonably accurate parameter estimates whenever: 1) For the mean-valued independent variables, whose values are derived in the sample from which the dependent variable values are not drawn, their sampling error is inconsequential compared to these variables' between group variation. 2) Very little within group correlation exists between the independent variables which are drawn from distinct samples.[6] 3) The number of groups is large or the number of observations in each group is large. Also since either of these first two conditions may not be expected to hold, our analysis presents alternative estimators, which are consistent in many cases where OLS is not.

Finally, even if the OLS estimator is consistent the standard regression program's estimate of the estimator's covariance matrix will generally understate the asymptotic covariance of the coefficients.[7] Consequently, our analysis includes the asymptotic covariances of both the OLS and the most generally applicable alternative estimator, as well as consistent estimates for each.

The balance of the paper proceeds as follows: Section II sets up the formal model used in the paper. Section III presents necessary and sufficient conditions for consistent OLS coefficient estimates, their large sample distributional properties, and a consistent estimator for the asymptotic convariance. Section IV develops the alternative estimators, describes large sample properties of the most generally applicable alternative estimator, and presents a consistent estimate of its asymptotic convariance.

## II. Model

Assume that one believes there exists a linear relationship between y and V:

$$y = V\beta + \varepsilon$$

where $V$ is a matrix of N observations of K variables,

$y$ is an N by one vector of observations,

$\beta$ is a K by one vector of unobserved constants, and

$\varepsilon$ is an unobserved N by one random variable with

the properties

$$E(\varepsilon\varepsilon') = \sigma^2 I$$

$$E(\varepsilon|V) = 0.$$

Further, assume that only two data sets (labelled 1 and 2) can be used to estimate the unknown $\beta$. Each constitute a random sample of independent multivariate observations. Each sample is drawn from the same population, and each is independent of the other. However, neither sample is capable of providing all the data for the regression, since each sample has no observed values for at least one variable of V (assumed different for each data set so that each variable has recorded values in at least one of the two data sets).

To account for this, we partition V into two subsets: Z, which includes the variables which are present in data set 1, and W which includes the variables absent in 1 but present in data set 2.[8] An equivalent representation of the model is then:

$$y = Z\beta_Z + W\beta_W + \varepsilon$$

In addition to observing y and the variables of Z in the first data set and those of W in the second data set, we also observe a group membership for each observation in each data set. To distinguish the sample and group we adopt the following notation. We denote group membership by subscripts, and data set or sample by superscript. Thus we write:

$$y^{(j)} = \begin{bmatrix} y_1^{(j)} \\ y_2^{(j)} \\ \cdot \\ \cdot \\ \cdot \\ y_G^{(j)} \end{bmatrix} \quad Z^{(j)} = \begin{bmatrix} Z_1^{(j)} \\ Z_2^{(j)} \\ \cdot \\ \cdot \\ \cdot \\ Z_G^{(j)} \end{bmatrix} \quad W^{(j)} = \begin{bmatrix} W_1^{(j)} \\ W_2^{(j)} \\ \cdot \\ \cdot \\ \cdot \\ W_G^{(j)} \end{bmatrix} \quad \varepsilon^{(j)} = \begin{bmatrix} \varepsilon_1^{(j)} \\ \varepsilon_2^{(j)} \\ \cdot \\ \cdot \\ \cdot \\ \varepsilon_G^{(j)} \end{bmatrix} \quad j = 1,2$$

Represented in this fashion, $y_g^{(j)}$ and $\varepsilon_g^{(j)}$ are respectively $n_g^{(j)}$ dimensional vectors of the jth sample observations of $y_g$ and the unobserved error of $y_g^{(j)}$ for the gth group. $Z_g^{(j)}$ is an $n_g^{(j)}$ x L matrix composed of the jth sample observations of the L variables of Z for the gth group, and $W_g^{(j)}$ is an $n_g^{(j)}$ x M matrix defined similarly.[9] We assume these matricies (and all others defined in the paper) are of full rank, and $z_{g\ell h}^{(j)}$ and $w_{imh}^{(j)}$ are their respective h$\ell$ th and hm th elements. Also, let $N^{(j)}$ be the jth sample size,

$$N^{(j)} = \sum_{g=1}^{G} n_g^{(j)} \quad j = 1,2$$

where G is the number of groups. For notational convenience, henceforth all omitted superscripts will imply sample 1, and the expectation operator is conditional upon the group membership index.

Finally, we also make the following distributional assumptions. First, we assume that for each group there exists a multivariate distribution which generates each observation ($y_g^{(j)}$, $z_{g1k}^{(j)}$, ... $z_{gLk}^{(j)}$, $w_{g1k}^{(j)}$ ... $w_{gMk}^{(j)}$) and that the sample moments of each group distribution, up to and including all fourth moments, converge in probability to their respective distributional moments. Also we assume that all absolute sixth moments are $O(\sqrt{n_g^{(j)}})$, that all absolute fourth moments are O(1) and that these limits are uniform in g.

Secondly, we assume that for all m = 1,2,...M there exists no real valued constant, $c_m$, such that $E(W_{gm}) = c_m$ for all g = 1,2,..G, and that $E(\varepsilon_g | Z_g, W_g) = 0$ for all g = 1,2,..G. Lastly, we assume that each $n_g^{(j)}$ is known for each value of $N^{(j)}$, being determined by a group sampling rule.

## III. Consistent OLS Estimation

We begin by considering the case where group averages from sample 2 are used to augment micro data from sample 1. Since the data available for analysis are restricted to $y^{(1)}$, $Z^{(1)}$, and $W^{(2)}$, we use $\overline{W}^{(2)}$ as a proxy for the unobserved $W^{(1)}$, where we define

$$\overline{W}^{(2)} = D_1 \overline{w}^{(2)}$$

$$D_j = \begin{bmatrix} e_{n_1^{(j)}} & & & & 0 \\ & e_{n_2^{(j)}} & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \cdot \\ 0 & & & & e_{n_G^{(j)}} \end{bmatrix} \qquad j = 1,2$$

$$\overline{w}^{(2)} = (D_2' D_2)^{-1} D_2 W^{(2)}$$

$$= (\overline{w}_1^{(2)'}, \overline{w}_2^{(2)'}, \ldots \overline{w}_G^{(2)'})'$$

and where $e_{n_g^{(j)}}$ is an $n_g^{(j)}$ x 1 vector of ones. Thus $\overline{W}^{(2)}$ is a N by M matrix of group means computed in sample 2.

We may again rewrite the model:

$$y = Z\beta_Z + \overline{W}^{(2)}\beta_W + \xi$$

$$= X\beta + \xi$$

where $\xi = (W - \overline{W}^{(2)})\beta_W + \varepsilon$

$$X = \left[ Z \;\vdots\; \overline{W}^{(2)} \right]$$

and define the OLS estimator:

$$\hat{\beta} = (X'X)^{-1} X'y.$$

$\hat{\beta}$ is consistent if and only if

$$\text{(III.1)} \quad \operatorname*{plim}_{N\to\infty} \frac{X'\xi}{N} = \begin{bmatrix} \Omega_{ZW} \\ \hline -\Omega_{\overline{W}} \end{bmatrix} \beta_W + \operatorname*{plim}_{N\to\infty} \frac{X'\varepsilon}{N} = 0$$

where

$$\text{(III.2)} \quad \Omega_{ZW} = \operatorname*{plim}_{N\to\infty} \frac{1}{N} \sum_g Z'_g (W_g - \overline{W}^{(2)}_g)$$

$$\text{(III.3)} \quad \Omega_{\overline{W}} = \operatorname*{plim}_{N\to\infty} \frac{1}{N} \sum_g \overline{W}^{(2)\prime}_g (\overline{W}^{(2)}_g - W_g)$$

If the within group correlation between Z and W is zero, $\Omega_{ZW}$ will equal the zero matrix, and if $\overline{W}^{(2)}_g$ is known to equal $E(\overline{W}^{(2)}_g)$ for all g, $\Omega_{\overline{W}}$ will also equal the zero matrix. Thus if Z is empty or if W is relatively uncorrelated with Z and the group size in sample 2 is sufficiently large to insure accurate estimation of the group means, the OLS estimates of $\beta$ should be accurate for large G or where all $n_g$ are large. A formal statement of these sufficient conditions is given in proposition 1.[10]

<u>Proposition 1</u>

$$\operatorname*{plim}_{N\to\infty} \hat{\beta} = \beta \text{ if}$$

i) $\mathrm{cov}(Z_g, W_g) = 0 \;\; \forall g$

ii) as $N\to\infty$, $n_g^{(2)} \to\infty$ with $\dfrac{n_g}{n_g^{(2)}} \to a_g \in \mathbb{R}^+ \;\; \forall g$

and either

iiia) as $N\to\infty$, G is finite and $n_g \to\infty \; \forall g$

or

iiib) as $N\to\infty$, $G\to\infty$

proof: given in appendix.

These conditions are also sufficient to insure consistent estimation of the linear model when all data are grouped. To see this note that the standard WLS estimator for grouped data is computationally equivalent to estimating $\hat{\beta}$ replacing Z with $\bar{Z}$ -- the matrix of group means defined as

$$\bar{Z} = D_1 \, (D_1{}'D_1)^{-1} \, D_1{}'Z.$$ [11]

The condition necessary and sufficient for consistency is similar to III.1. The difference is that $\bar{Z}$ replaces Z. Thus, the matrix $\Omega_{\bar{W}}$ would be unaffected by this change, while the matrix associated with $\Omega_{ZW}$ would then become:

$$\Omega_{\bar{Z}W} = \operatorname*{plim}_{N\to\infty} \frac{1}{N} \sum_g \bar{Z}_g{}' \, (W_g - \bar{W}_g^{(2)}) \; .$$

Whenever the within group correlation between Z and W is zero, $\Omega_{\bar{Z}W} = 0$. In addition, even if the within group correlation is not zero, $\Omega_{\bar{Z}W} = 0$ as well, if as $N\to\infty$, $n_g\to\infty \; \forall g$, since each element of $\bar{Z}$ would converge in probability to its respective within group moment.

However, even if $\hat{\beta}$ consistently estimates $\beta$ the estimator's covariance estimate based solely upon the sum of squared errors

$$y'(I-X(X'X)^{-1}X')y(X'X)^{-1}/(N-L-M)$$

may be both a poor large and small sample approximation to the asymptotic covariance.

Intuitively, this follows because we construct $\hat{\beta}$ from a two-stage procedure. First, we estimate group means; then we include these means in the regression as independent variables. Consistent estimation of the asymptotic covariance requires that we account for any error introduced to the second stage (the regression) by the first stage (mean estimation.) The following proposition develops this argument formally,[12]

Proposition 2

Under assumptions 1(i,ii,iii) and

i) $E(z_{g\ell}^2|w_{gm}) = Ez_{g\ell}^2 \quad \ell = 1,2,..L, \; m = 1,2,..M$

ii) $\lim_{N\to\infty} \sup_{g,h} \frac{n_g}{n_h} \in \mathbb{R}_{++}$

iii) $G = o(n_g^{(2)}) \quad \forall g$

then

$$\sqrt{N}(\hat{\beta}-\beta) \xrightarrow{d} N(0,Q^{-1}\Delta Q^{-1})$$

where

$$Q = \operatorname*{plim}_{N\to\infty} \frac{1}{N} X'X$$

$$\Delta = \operatorname*{plim}_{N\to\infty} \frac{1}{N} X'\xi\xi'X$$

proof: given in appendix

Proposition 3

Under the assumptions 1(i,ii,iii) and 2(i,ii,iii)

$$\hat{Q}^{-1} \hat{\Delta} \hat{Q}^{-1} \xrightarrow{P} Q^{-1} \Delta Q.$$

where $\hat{Q} = \frac{1}{N} X'X$

$$\hat{\Delta} = \frac{1}{N} \sum_{g=1}^{G} \hat{C}_g \left( X_g' X_g + \frac{n_g}{n_g^{(2)}} X_g' \tilde{A}_g X_g \right) + \hat{\sigma}^2 \frac{1}{N} X' X$$

$$\hat{C}_g = \hat{\beta}_w' \hat{\text{cov}}(W_g) \hat{\beta}_w$$

$$\hat{\text{cov}}(W_g) = (n_g^{(2)} - 1)^{-1} W_g^{(2)'} \left( I - \frac{e_{n_g}^{(2)} e_{n_g}^{(2)'}}{n_g^{(2)}} \right) W_g^{(2)}$$

$$X_g = \left[ Z_g \,\vdots\, \overline{W}_g^{(2)} \right]$$

$$\tilde{A}_g = (n_g - 1)^{-1} (e_{n_g} e_{n_g}' - I)$$

$$\hat{\sigma}^2 = \frac{1}{N} \hat{\xi}' \hat{\xi} - \frac{1}{N} \Sigma_g n_g \hat{C}_g$$

$$\hat{\xi} = y - X \hat{\beta}$$

proof: given in appendix

The difference between the consistent estimator proposed above and the standard OLS covariance estimate can be written[13]

$$\left[\frac{1}{N} X' X\right]^{-1} \left[\frac{1}{N} \Sigma_g X_g' \left[ \Sigma_g \frac{N - n_g}{N} \hat{C}_g \; I + \hat{C}_g \frac{n_g}{n_g^{(2)}} \tilde{A}_g \right] X_g \right] \left[\frac{1}{N} X' X\right]^{-1}$$

Since $\forall \hat{\beta}$ and $\forall g$, $\hat{C}_g \geq 0$, $\tilde{A}_g$ is positive semi-definite, and the sum of positive semi-definite matrices is positive semi-definite, therefore, the difference between the covarience estimates is positive semi-definite. This is also true in the limit, where $N \to \infty$.[14] Thus one would expect the standard regression package estimate of the OLS estimator's covariance to both underestimate the OLS asymptotic covariance and understate its consistent estimator for all N.[15]

IV. Consistent esimation when Cov $(Z_g,W_g) \neq 0$ and/or group means are measured with error in the limit.

The preceding section shows that OLS may be an appropriate estimator for the coefficients $\beta$ when the within group covariance of W and Z is negligible, when the number of observations in the sample used to estimate the group means is large and either the number of groups is large or for each group the number of observations is large. As noted previously, since the first two conditions are unlikely to hold in many situations, we consider estimators which are consistent when OLS is not.[16]

We begin by considering the violation of each of these conditions separately for the case where group averages computed in one sample are used to augment micro data from another sample. If for all $g, n_g^{(2)}$ is large but $\text{cov}(Z_g,W_g) \neq 0$ for some g, each individual's unobserved deviation from the group mean for at least one W will be correlated with at least one Z. Thus the error $\xi$ will be correlated with at least one Z, $\Omega_{ZW}$ can not be expected to be equal to zero, and OLS will be biased and inconsistent.

One solution to this problem is to find instruments which are correlated with Z but not with the individual deviations from the group means of W. Fortunately, one can derive instruments from the original data with only one additional assumption--that the means of Z differ between groups and between the variables of Z.

The instruments we propose are modified group means. If for all g, $n_g$ were large, the group sample means

$$\overline{z} = (D_1{'}D_1)^{-1} D_1{'}Z$$

would be nearly uncorrelated with $(W_g - \overline{W}_g^{(2)})$ and the correlation would go to zero as $n_g \to \infty$. Yet, for small samples this correlation may not be inconsequential. Therefore, we propose a slight modification to the mean, which eliminates its correlation with $(W_g - \overline{W}_g^{(2)})$, and consequently $\xi$, but retains its correlation with $Z_g$. Note that under independent sampling the kth row of $(W_g - \overline{W}_g^{(2)})$ is correlated only with the kth row of $Z_g$. Eliminating the kth row of $Z_g$ in constructing the kth row of the instrument matrix will thus eliminate the correlation between the means and $(W_g - \overline{W}_g^{(2)})$. Therefore we define the instruments as follows:

$$\tilde{X} = \tilde{A}X$$

$$\tilde{A} = \begin{bmatrix} \tilde{A}_1 & & & 0 \\ & \tilde{A}_2 & & \\ & & \ddots & \\ 0 & & & \tilde{A}_G \end{bmatrix}$$

$$\tilde{A}_g = (e_{n_g} e'_{n_g} - I_{n_g})/(n_g - 1)$$

with

$$\tilde{X}_g = \left[ \tilde{Z}_g \,\vdots\, \overline{W}_g^{(2)} \right]$$

Now consider the case where cov $(Z_g, W_g) = 0 \;\forall g$ but each $n_g^{(2)}$ is small and bounded as $N \to \infty$. Under these assumptions $\overline{\Omega_W} \neq 0$ and $\hat{\beta} \overset{P}{\nrightarrow} \beta$. Inspection of equation III.3 suggests that the bias this causes will be proportional to the sampling variation of $\overline{W}_g$. Since this sampling variation is estimatable, our approach is to use it to construct a correction to $\hat{\beta}$.

This problem corresponds to the classical, errors-in-variables situation. Although the correspondence is not exact the similarity motivates the approach. For small $n_g^{(2)}$ one could consider $\overline{W}_g^{(2)}$ to be an imprecise measure of $EW_g$. Thus th inclusion of $\overline{W}_g^{(2)}$ in place of $EW_g$ in the regression creates an "errors-in-variables problem[17] giving rise to the familiar probability limit:

$$\operatorname*{plim}_{N\to\infty} \hat{\beta} = (I - Q^{-1}\Omega)\beta$$

where

$$\Omega = \begin{bmatrix} 0 & 0 \\ 0 & \Omega_{\overline{w}} \end{bmatrix} .$$

Since one can construct $\frac{1}{N}$ S, a consistent estimate of $\Omega$, we define

$$R = I - (X'X)^{-1} S$$

where

$$S = \begin{bmatrix} 0 & 0 \\ 0 & S_{22} \end{bmatrix}$$

$$S_{22} = \sum_g \frac{n_g}{n_g^{(2)}} \hat{\text{cov}} (W_g)$$

$$\hat{\text{cov}} (W_g) = (n_g^{(2)} - 1)^{-1} W_g^{(2)'} [I - n_g^{(2)^{-1}} e_{n_g}^{(2)} e_{n_g}^{(2)'}] W_g^{(2)} ,$$

we then compute $R^{-1}\hat{\beta}$ which is a consistant estimate of $\beta$.[18]

To confront both problems simultaneously the two estimators may be combined.[19] For that case we define [20,21]

$$\tilde{\beta} = \tilde{R}^{-1} (\tilde{X}'X)^{-1} \tilde{X}' y$$

where

$$\tilde{R} = I - (\tilde{X}'X)^{-1} S$$

Proposition 4 states formally the conditions under which $\tilde{\beta}$ is a consistent estimator for $\beta$, and for its asymptotic normality, while Proposition 5 considers the consistent estimation of its asymptotic covariance matrix.

Proposition 4

Assume:

i) as $N \to \infty$, $G \to \infty$

ii) $\lim_{N \to \infty} n_g(j) = d_g(j) \in \mathbb{R}_+ \quad \forall g$

Then

$$\operatorname*{plim}_{N \to \infty} \tilde{\beta} = \beta$$

and

$$\sqrt{N}(\tilde{\beta} - \beta) \xrightarrow{d} N(0,\ P^{-1} \Gamma\, P^{-1})$$

where

$$\Gamma = (\lim_{N \to \infty} \sum_{g=1} \frac{1}{N} E\ \gamma_g \gamma'_g)$$

$$P = \operatorname*{plim}_{N \to \infty} \frac{1}{N} [\tilde{X}'X - S]$$

and

$$\gamma_g = \begin{bmatrix} Z'_g(W_g - \bar{W}_g^{(2)})\ \beta_w + Z'_g\ \varepsilon_g \\ (\bar{W}_g^{(2)\prime}\ (W_g - \bar{W}_g^{(2)}) + S_{22})\ \beta_w + \bar{W}_g^{(2)\prime} \varepsilon_g \end{bmatrix}$$

Proof: given in Appendix

Proposition 5

Under assumptions 4(i), 4(ii), and

(i) $\text{cov}(z_{g\ell}\ w_{gm}) = \text{cov}(z_{h\ell}\ w_{hm}) \qquad \forall\ g,h = 1, 2, \ldots G$

$\ell = 1, 2, \ldots L$

$m = 1, 2, \ldots M$

then

$$\tilde{P}^{-1}\ \hat{\Gamma}\ \tilde{P}^{-1} \xrightarrow{P} P^{-1}\ \Gamma\ P^{-1}$$

where

$$\hat{P} = \frac{1}{N} [X'X-S]$$

$$\hat{\Gamma} = \begin{bmatrix} \hat{\Gamma}_{11} & \hat{\Gamma}_{12} \\ \hat{\Gamma}_{21} & \hat{\Gamma}_{22} \end{bmatrix}$$

and where $\hat{\Gamma}_{11}$, $\hat{\Gamma}_{12}$, and $\hat{\Gamma}_{22}$ are respectively $L \times L$, $L \times M$, and $M \times M$ matrices with respective $\ell p$, $\ell m$, and $mq$ elements given by[22]

$$(\hat{\Gamma}_{11})_{\ell p} = \sum_i \sum_j \tilde{\beta}_{w_i} \tilde{\beta}_{w_j} \sum_g \frac{n_g}{N} \left[ \left( \frac{n_g-2}{n_g-1} + \frac{n_g-1}{n_g(2)} \right) \widehat{Ez_{g\ell} Ez_{gp}} + \right.$$

$$\left. \left( [n_g-1]^{-1} + n_g(2)^{-1} \right) \widehat{Ez_{g\ell} z_{gp}} \right] \widehat{\mathrm{cov}} (w_{gi}, w_{gj})$$

$$+ \tilde{\sigma}^2 \frac{1}{N} \sum_g \sum_k \tilde{z}_{g\ell k} \tilde{z}_{gpk} + [\Omega_{zw} \hat{\beta}_w]_\ell [\Omega_{zw} \hat{\beta}_w]_p \frac{1}{N} \sum_g n_g/(n_g-1)$$

$$(\hat{\Gamma}_{12})_{\ell m} = \sum_i \sum_j \tilde{\beta}_{w_i} \tilde{\beta}_{w_j} \sum_g \frac{n_g}{N} (1 + n_g/n_g^{(2)}) \bar{z}_{g\ell} \, \mathrm{cov} \, \widehat{(w_{gi} w_{gj}) E \, w_{gm}}$$

$$+ \tilde{\sigma}^2 \frac{1}{N} \sum_g \bar{w}_{gm}^{(2)} \sum \tilde{z}_{g\ell k} - [\Omega_{zw} \hat{\beta}_w]_\ell \frac{1}{N} \sum_i \tilde{\beta}_{w_i} \sum_g [n_g/n_g^{(2)}] \widehat{\mathrm{cov}} (w_{gm} w_{gi})$$

$$(\hat{\Gamma}_{22})_{mq} = \sum_i \sum_j \tilde{\beta}_{w_i} \tilde{\beta}_{w_j} \sum_g \frac{n_g}{N} 1/n_g^{(2)} \{ [1 + n_g/(n_g^{(2)}-1)] \, E \, \widehat{w_{gm} w_{gq} \, \mathrm{cov} \, (w_{gi}, w_{gj})}$$

$$+ [n_g^{(2)}-1+n_g(n_g^{(2)}-2)/(n_g^{(2)}-1)] \, E \, \widehat{w_{gm} Ew_{gq} \, \mathrm{cov} \, (w_{gi}, w_{gj})}$$

$$+ [n_g/(n_g^{(2)}-1)] \, \widehat{\mathrm{cov} \, (w_{gm}, w_{gj}) \, \mathrm{cov} \, (w_{gq} w_{gi})} \} + \tilde{\sigma}^2 \frac{1}{N} \sum_g n_g \bar{w}_{gm}^{(2)} \bar{w}_{gq}^{(2)}$$

where

$$\widehat{Ez_{g\ell} Ez_{gp}} = [n_g(n_g-1)]^{-1} \sum\sum_{h \neq k} z_{g\ell h} z_{gpk}$$

$$E \, \widehat{z_{g\ell} z_{gp}} = n_g^{-1} \sum_k z_{g\ell k} z_{gpk}$$

$$\widehat{\mathrm{cov}} (w_{gi}, w_{gj}) = [\widehat{\mathrm{cov}} (W_g)]_{ij}$$

$$\hat{\Omega}_{zw}\ \beta_w = J(\frac{1}{N}X'X)\ (\hat{\beta}-\tilde{\beta})$$

$$J = [I_L \quad 0]$$

$$\bar{z}_\ell = \frac{1}{N}\sum_g n_g\ \bar{z}_{g\ell}$$

$$\bar{w}_m^{(2)} = \frac{1}{N^{(2)}}\sum_g n_g^{(2)}\ \bar{w}_{gm}^{(2)}$$

$$\widehat{\mathrm{cov}\ (w_{gi},\ w_{gj})\ E\ w_{gm}} = [n_g^{(2)}(n_g^{(2)}-1)]^{-1}\ [\sum_{k\neq h}\sum w_{gik}^{(2)}\ w_{gjk}^{(2)}\ w_{gmh}^{(2)} - (n_g^{(2)}-2)^{-1}$$

$$\sum_{\substack{k\neq r\neq h\\ k\neq h}}\sum\sum w_{gik}^{(2)}\ w_{gjr}^{(2)}\ w_{gmh}^{(2)}]$$

$$\widehat{E\ w_{gm}\ w_{gq}\ \mathrm{cov}\ (w_{gi},\ w_{gj})} = a - b - d$$

$$\widehat{E\ w_{gm}\ E\ w_{gq}\ \mathrm{cov}\ (w_{gi},\ w_{gj})} = c$$

$$\widehat{\mathrm{cov}\ (w_{gm},\ w_{gq})\ \mathrm{cov}\ (w_{gi},\ w_{gj})} = a-b-c-d$$

(where the subscripts on a, b, c, and d are dropped)

$$a = [n_g^{(2)}\ (n_g^{(2)}-1)]^{-1}\sum_{k\neq h}\sum w_{gmk}^{(2)}\ w_{gqk}^{(2)}\ w_{gih}^{(2)}\ w_{gjh}^{(2)}$$

$$b = [n_g^{(2)}\ (n_g^{(2)}-1)(n_g^{(2)}-2)(n_g^{(2)}-3)]^{-1}\sum_{\substack{k\neq h\neq r\neq s\\ k\neq r,s\\ h\neq s}}\sum\sum\sum w_{gmk}^{(2)}\ w_{gqh}^{(2)}\ w_{gir}^{(2)}\ w_{gjs}^{(2)}$$

$$c = [n_g^{(2)}\ (n_g^{(2)}-1)(n_g^{(2)}-2)]^{-1}\sum_{\substack{k\neq h\neq r\\ k\neq r}}\sum\sum w_{gmk}^{(2)}\ w_{gqh}^{(2)}\ w_{gir}^{(2)}\ w_{gjr}^{(2)} - b$$

$$d = [n_g^{(2)}\ (n_g^{(2)}-1)(n_g^{(2)}-2)]^{-1}\sum_{\substack{k\neq h\neq r\\ k\neq r}}\sum\sum w_{gmk}^{(2)}\ w_{gqk}^{(2)}\ w_{gih}^{(2)}\ w_{gjr}^{(2)} - b$$

$$\tilde{\sigma}^2 = \frac{1}{N}\tilde{\xi}'\ \tilde{\xi} - \frac{1}{N}\sum_g n_g\ (1+n_g^{(2)})^{-1}\ \tilde{\beta}_w'\ \widehat{\mathrm{cov}}\ (W_g)\ \tilde{\beta}_w$$

$$\tilde{\xi} = y - X\tilde{\beta}$$

and where $\widehat{\mathrm{cov}}\ (W_g)$ is defined in proposition 2.

Proof: In appendix

The procedure for correcting for sampling error in the group means which one computes in the auxiliary data (assuming cov $(Z_g, W_g) = 0$) extends rather simply to where the regression uses group means from both samples. The correction factor, $R^{-1}$, is identical. Also even if cov $(Z_g, W_g) \neq 0$ and $n_g \to \infty$, the corrected WLS estimator is consistent. In fact, one can show its difference from $\tilde{\beta}$ converges in probability to zero as all $n_g \to \infty$ .

However, if for most groups there exists a non-zero correlation between $Z_g$ and $W_g$ and if both $n_g$ and $n_g^{(2)}$ remain small $\forall g$ as N becomes large, then the estimators' difference does not possess a zero probability limit. $\tilde{\beta}$ is consistent and the corrected WLS estimator is not. Thus $\tilde{\beta}$ would be a preferred estimator, even when the regression data are group. Of course, this presumes that for each group both the individual data of $Z_g$ as well as the sample variances and covariances of $W_g^{(2)}$ are available. If either are not, a standard instrumental variable approach would be the only alternative.

Two final caveats remain. First, even if $n_g \to \infty$ for each group so that the corrected WLS estimator and $\tilde{\beta}$ are both consistent as well as asymptotically equivalent, the use of either estimator would be ill-advised. This is due to a somewhat novel result. If group sizes in sample 1 increase much more rapidly than their respective sample 2 counterparts, then these estimators' asymptotic covariance, which depends upon the ratio of $n_g$ to $n_g^{(2)}$, contains unbounded elements.[23]

Finally, under a reasonable set of assumptions, ignoring the covariance problem and correcting for the sampling error in the estimated group averages from the second sample could produce lower MSE estimates of $\beta$ than would obtain if one did not correct for the error. Specific exploration of small sample properties is left for future work.

## Conclusion

Although we can not say for certain how much of a difference these corrections will make, there is some evidence that it may be substantial. Thaler and Rosen (1976) have argued that finer classification schemes would reduce errors-in-variables bias in estimating compensating wage differences. The analysis above suggests the opposite may be true in most situations. While making the classification finer will reduce any bias due to the correlation between the errors and other variables in the equation, if it also results in fewer observations per classification, the accuracy of the estimated group means will decline. This will probably lead to a greater downward bias of the estimated compensating wage differential. This could account, in part, for the differences between the Thaler and Rosen study and later studies which use broader classification schemes. These studies have generally found compensating wage differences about 3 to 20 times higher than those found by Thaler and Rosen. Correcting for any remaining error in the later studies might increase estimates even further.

## FOOTNOTES

1. Prais and Aitchinson is the seminal article. See also Johnston for a discussion of further issues and additional references.

2. We employ the terminology of independent and dependent variable because of its familiarity. We do not mean to imply causality by its use.

3. For example see Cogan (1982).

4. For example, see R.E.B. Lucas, Thaler and Rosen, Viscusi (1978ab, 1979, and 1980), and Olson. Also, both Brown and Smith review the literature.

5. Thaler and Rosen were first to suggest that there may be an errors-in-variables problem.

> However, using death and injury statistics in that manner implies inducing a huge component of measurement error for individuals, because job risks in each industry are not uniform across occupations. Hence any estimates of the risk premium obtained in this way will probably be biased. (p. 286)

Viscusi also believes that using group averages produces a "...conventional errors-in-variable situation in which the empirical estimates are biased downward." (1979, p. 251) See also his discussion on p. 372, 1978a. Olson mentions the measurement error problem on p. 177. Bartel also comments on this problem.

Lucas argued that OLS estimates will be unbiased but that errors will be heteroskedastic.

6. This condition is of more importance for the technique which augments micro data with means. If, for where all data are means, the number of observations per group in the sample, from which the dependent variable is drawn, is large, the within group correlation would produce little error in the estimates.

7. By asymptotic covariance, we mean the covariance matrix of the distribution of the random vector to which the OLS estimator converges. Also by understate, we mean that any quadratic form of the matrices' difference is non-negative.

8. Of course $y^{(2)}$ $Z^{(2)}$ and $W^{(1)}$ are assumed to be unobserved.

9. We assume that Z contains a column of ones or that all variables are measured as deviations from their grand means.

10. These conditions are only sufficient because they exclude the following. For $\Omega_{zw}$ they exclude the unlikely possibility that non-zero group moment values might exactly cancel in the sum. For $\Omega_{\bar{w}}$ there would naturally exist no error in the estimate of any group mean if the econometrician knew the mean's true value a priori.

11. The WLS model is

$$\sqrt{n_g}\,\bar{y}_g = \sqrt{n_g}\,\bar{z}_g\,\beta_z + \sqrt{n_g}\,\bar{w}_g^{(2)}\,\beta_w + \sqrt{n_g}\,\bar{\xi}_g \qquad g = 1,2,..G$$

$$\bar{y}_g = n_g^{-1}\,e_{n_g}'\,y_g$$

$$\bar{z}_g = n_g^{-1}\,e_{n_g}'\,Z_g$$

$$\bar{w}_g^{(2)} = n_g^{(2)^{-1}}\,e_{n_g^{(2)}}'\,w_g^{(2)}$$

12. Note that consistent estimation of the covariance matrix for the coefficients requires an estimate of the within group covariance matrix W. Since the individual observations of W are often unavailable when one is using grouped data it may sometimes be necessary to use another procedure to estimate those moments. For instance, one could use a third data set. Or, if the within group variance of W is known but the covariance is not, it will be possible to estimate those covariances if one can assume that the process generating the between group covariation is the same as that generating the within group covariance. In that case, the covariance can be estimated using the grouped data.

If it is impossible to estimate the moments it may still be possible to put an upper and/or lower bound or the covariance matrix for β. If the variances are known and the correlations can be bounded on the basis of a priori information the values of the correlations which minimize and maximize each quadratic form $C_g\ (\hat{\beta})$ subject to the a priori bounds can be choosen and upper and lower limits for $\hat{\Delta}$ constructed. If the variances are unknown but can be bounded on the basis of a priori information a similar method can be used.

13. More precisely, these are covariance estimates of $\sqrt{N}(\hat{\beta}-\beta)$.

14. $n_g$ and $n_g^{(2)}$ of equivalent order is sufficient for the difference to be positive definite.

15. We have not considered the possibility of a group specific error component. If such components are present they will completely dominate the individual specific components as group size goes to infinity. Thus the unweighed OLS covariance matrix would be asymtoticly equivalent to the correct asymptotic covariance matrix for this problem and would be unlikely to seriously

understate the true covariance for large samples. Dickens (1984) gives evidence suggesting that group specific error components may dominate individual components in many applications using grouped data.

16. For instance, in the case of using labor market evidence to measure the value of life, economic theory suggests that the danger of an individual's job should be correlated with many individual characteristics and the other characteristics of the person's job. In addition, industry job risk sometimes is measured with a great deal of error.

17. Unlike the errors-in-variables case where the variable measured with error is an independent variable, here $EW_g$ is a proxy for a left out independent variable. Thus the problems are not equivalent even though the same analysis may be applied.

18. The use of this estimator requires that a researcher know the within group covariance of W or have access to the original data to compute it. Often when one is dealing with grouped data this information is unavailable. Footnote 12 proposes some alternative approaches to this problem.

19. If one believes that either of these conditions fail while the other still holds, an alternative estimator can be constructed for just the single problem. For example, if $cov(W_g Z_g) \neq 0$ and $\overline{\Omega_W} = 0$ define

$$\hat{\beta} = (\tilde{X}'X)^{-1}\tilde{X}'y$$

and if $\Omega_{ZW} = 0$ but $n_g^{(2)}$ bounded for some g as $N \to \infty$, define

$$\hat{\beta} = R^{-1}\hat{\beta}.$$

The large sample distributions of these estimators and consistent estimates of their asymptotic covariances are available from the authors upon request.

20. $\tilde{\beta}$ can be equivalently defined as an instrumental variable estimator of the following regression model:

$$y = X\tilde{R}\beta + \xi$$

where the instrument is $\tilde{X}\tilde{R}$. Therefore

$$\tilde{\beta} = (\tilde{R}'\tilde{X}'X\tilde{R})^{-1}\tilde{R}'\tilde{X}'y$$

$$= \tilde{R}^{-1}(\tilde{X}'X)^{-1}\tilde{X}'y \quad \text{for } \tilde{R} \text{ of full rank.}$$

21. Once again, the individual observations of W must be available to construct this estimator. See footnote 12 for a discussion of alternatives.

22. In the actual estimation of $\Gamma$, one would not need to estimate the terms involving $\sigma^2$ on an element by element basis. Rather one can write $\Gamma = \Phi + \sigma^2 Q$, and estimate $\sigma^2 Q$ with $\tilde{\sigma}^2 \frac{1}{N} X'X$ separately from $\Phi$.

23. The intuition for this is that if $n_g \to \infty$ as $n_g^{(2)}$ remains bounded the increased group size in sample 1 amplifies the noise inherent in the mean estimates of the second sample such that it obscures the increasing informational content of the first sample. The asymptotic covariance is a function of terms $E\gamma_g\gamma_g'$ where

$$\gamma_g = \tilde{X}g'\xi g + \begin{bmatrix} 0 \\ \text{-----} \\ S_{22}\beta_w \end{bmatrix}.$$ Since one term of $\xi_g$ is $(W_g - \overline{W}_g^{(2)})\beta_w$, letting $n_g \to \infty$

without also decreasing the sampling error in $\overline{W}_g^{(2)}$, makes $(W_g - \overline{W}_g^{(2)})\beta_w$ an infinite dimensional vector in which each element is correlated with every other element of the vector. Thus in the limit any quadratic form of $E\tilde{X}_g'\xi_g \xi_g'\tilde{X}_g$ is $O(n_g^2)$, which implies $E\gamma_g\gamma_g' = O(n_g^2)$.

# BIBLIOGRAPHY

Bartel, Ann P. "Wages, Nonwage Job Characteristics and Labor Mobility," Industrial and Labor Relations Review, Vol. 35, No. 4 (July 1982), pp. 578-589.

Billingsley, Patrick. Probability and Measure (Wiley, 1979).

Brown, Charles. "Equalizing Differences in the Labor Market," Quarterly Journal of Economics, Vol. 94, No. 1 (February 1980), pp. 113-134.

Cogan, John F., "The Decline in Black Teenage Employment: 1950-1970," American Economic Review, Vol. 72, No. 4 (September 1982).

Dickens, William T. "Random Group Effects and Grouped Data," mimeo, Berkeley 1984.

Heckman, James J. "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and A Sample Estimator for Such Models," Annals of Economic and Social Measurement (May 1976).

Johnston, J. Econometric Methods (McGraw-Hill, 1972), pp. 228-238.

Lucas, Robert E. B. "Hedonic Wage Equations and Psychic Wages in the Return to Schooling," American Economic Review, Vol. 67, No. 4 (September 1977). pp. 549-558.

Olson, Craig A. "An Analysis of Wage Differentials Received by Workers on Dangerous Jobs," Journal of Human Resources, Vol. 16, No. 2 (Spring 1981), pp. 167-185.

Prais, S. J. and Aitchinson, J. "The Grouping of Observations in Regression Analysis," Revue International Statistical Institute, Vol. 1 (1954), pp. 1-28.

Rothenberg, Thomas J., "Limiting Distributions," University of California, mimeo., 1980.

Smith, Robert S. "Compensating Wage Differentials and Public Policy: A Review," Industrial and Labor Relations Review, Vol. 32 (April 1979), pp. 339-352.

Thaler, Richard and Rosen, Sherwin. "The Value of Saving a Life: Evidence from the Labor Market," Household Production and Consumption, Nestor Terleckyj (ed.), NBER STudies in Income and Wealth No. 90 (New York: Columbia Press, 1976), pp. 265-298.

Viscusi, W. Kip. "Labor Market Valuation of Life and Limb: Empirical Evidence and Policy Implications," Public Policy, Vol. 26, No. 3 (Summer 1978), pp. 359-386.

____________. "Wealth Effects and Earnings Premiums for Job Hazards," Review of Economics and Statistics, Vol. 60, No. 3 (August 1978), pp. 408-416.

____________. Employment Hazards: An Investigation of Market Performance (Harvard University Press, 1979).

____________. "Union, Labor Market Structure, and the Welfare Implications of the Quality of Work," Journal of Labor Research, Vol. No. 1 (Spring 1980), pp. 175-192.

## Appendix

Proof of the following propositions are made easier by use of two Lemmas.

Lemma 1

If the sequence of random variables

$\alpha_n \quad n = 1,2,...$

converges in quadratic mean

$$\lim_{n\to\infty} E(\alpha_n - \alpha)^2 = 0$$

then

$$\operatorname{plim}_{n\to\infty} \alpha_n = \alpha \quad .$$

Proof: Rothenberg gives a proof for random variables with continuous distributions. The Lemma may also be proven by straightforward application of the Chebyshev's Inequality.

Lemma 2

Let ($\alpha_n$, n = 1,2,...m) be independent but not an identically distributed sequence of random variables. Then if there exists an $\alpha$ such that

$$(1) \quad \lim_{m\to\infty} \frac{1}{m} \sum_{n=1}^{m} E\alpha_n = \alpha$$

and

$$(2) \quad \lim_{m\to\infty} \frac{1}{m} \sum_{n=1}^{m} E(\alpha_n - E\alpha_n)^2 < \infty$$

then $\frac{1}{m} \sum_{n=1}^{m} \alpha_n$ converges in quadratic mean to $\alpha$ as $m \to \infty$.

Proof:

$$\lim_{m\to\infty} E\left[\frac{1}{m}\sum_{n=1}^{m}\alpha_n - \alpha\right]^2 =$$

$$\lim_{m\to\infty} E \quad \frac{1}{m^2}\sum_{n=1}^{m}(\alpha_n - E\alpha_n)^2 + \frac{1}{m^2}\sum_{n\neq \ell}\sum(\alpha_n - E\alpha_n)(\alpha_\ell - E\alpha_\ell) + \frac{2}{m}(\sum_n(\alpha_n - E\alpha_n))(\frac{1}{m}\sum_\ell E\alpha_\ell - \alpha)$$

$$+ (\frac{1}{m}\sum_n E\alpha_n - \alpha)^2$$

Since $(\alpha_n - E\alpha_n)^2 \geq 0 \quad _n$ by Billingsley 16.6

$$E\sum_{n=1}^{m}(\alpha_n - E\alpha_n)^2 = \sum_{n=1}^{m} E(\alpha_n - E\alpha_n)^2$$

and by (2)

$$\lim_{m\to\infty}\frac{1}{m^2}\sum_{n=1}^{m} E(\alpha_n - E\alpha_n)^2 = 0$$

Also by (2)

$$\max\{\sup_n(\alpha_n - E\alpha_n)^2, 1\}$$

is an integrable function with respect to the probability measure for $\alpha_n$.

This function bounds

$(\alpha_n - E\alpha_n)(\alpha_\ell - E\alpha_\ell)$.

Therefore by Billingsley 16.7, and independence

$$\lim_{m\to\infty}\frac{1}{m^2} E\sum_{n\neq \ell}\sum(\alpha_n - E\alpha_n)(\alpha_\ell - E\alpha_\ell) = \lim_{n\to\infty}\frac{1}{m^2}\sum_{n\neq \ell}\sum E(\alpha_n - E\alpha_n)E(\alpha_\ell - E\alpha_\ell) = 0$$

Finally by (1)

$$\lim_{m\to\infty}(\frac{1}{m}\sum_n E\alpha_n - \alpha)2 = 0$$

and the Lemma is proven.

Proposition 1

We first prove the proposition for (i), (ii), and (iiia). Since G is fixed, we need only show for each g that

(1) $\frac{1}{N} Z_g' (W_g - \overline{W}_g^{(2)}) \overset{P}{\rightarrow} 0$

(2) $\frac{1}{N} \overline{W}_g^{(2)'} (W_g - \overline{W}_g^{(2)}) \overset{P}{\rightarrow} 0.$

Consider the ℓm th element of (1) and the mq th element of (2), which we denote respectively:

(3) $\frac{1}{N} \left[ Z_g' (W_g - \overline{W}_g^{(2)}) \right]_{\ell m}$

(4) $\frac{1}{N} \left[ \overline{W}_g^{(2)'} (W_g - \overline{W}_g^{(2)}) \right]_{mq}$

For (3) we have by (i)

$$\underset{N\to\infty}{\text{plim}} \frac{1}{N} \left[ Z_g' W_g \right]_{\ell m} = \underset{N\to\infty}{\text{plim}} \sum_{k=1}^{n_g} \frac{n_g}{N} \left(\frac{1}{n_g} z_{g\ell k} w_{gmk}\right) = \frac{n_g}{N} Ez_{g\ell} Ew'_{gm}$$

and by sampling independence

$$\underset{N\to\infty}{\text{plim}} \frac{1}{N} \left[ Z_g' \overline{W}_g^{(2)} \right]_{\ell m} = \underset{N\to\infty}{\text{plim}} \sum_{k=1}^{n_g} \frac{1}{N} z_{g\ell k} \overline{w}_{gm}^{(2)} = \frac{n_g}{N} Ez_{g\ell} Ew_{gm}.$$

Therefore (1) is proved.

For (4) by sampling independence

$$\underset{N\to\infty}{\text{plim}} \frac{1}{N} \left[ \overline{W}_g^{(2)'} W_g \right]_{mq} = \frac{n_g}{N} Ew_{gm} Ew_{gq}$$

and

$$\underset{N\to\infty}{\text{plim}} \frac{1}{N} \left[ \overline{W}_g^{(2)'} \overline{W}_g^{(2)} \right]_{mq} = \underset{N\to\infty}{\text{plim}} \frac{n_g}{N} \overline{w}_{gm}^{(2)} \overline{w}_{gq}^{(2)} = \lim_{N\to\infty} \frac{n_g}{N} \left( \frac{1}{n_g^{(2)}} \text{cov}(w_{gm}, w_{gq}) + Ew_{gm} Ew_{gq} \right)$$

$$= \frac{n_g}{N} Ew_{gm} Ew_{gq}.$$

Therefore (2) is proved.

To prove the proposition with (i), (ii), and (iiib) first note that by sampling independence the sequence of random matrices

$$\begin{bmatrix} Z_g'(W_g-\overline{W}_g^{(2)}) \\ \\ \overline{W}_g^{(2)\prime}(W_g-\overline{W}_g^{(2)}) \end{bmatrix} \quad g = 1,2\ldots G$$

are independent but not identically distributed. Thus to show

$$\begin{bmatrix} \Omega_{zw} \\ \text{-----} \\ \Omega_{\bar{w}} \end{bmatrix} = \operatorname*{plim}_{N\to\infty} \frac{1}{N} \sum_{g=1}^{G} \begin{bmatrix} Z_g'(W_g-\overline{W}_g^{(2)}) \\ \text{-------------} \\ \overline{W}_g^{(2)\prime}(W_g-\overline{W}_g^{(2)}) \end{bmatrix} = 0$$

we show

(1) $\operatorname*{plim}_{N\to\infty} \frac{1}{N} \sum_{g=1}^{G} (Z_g'(W_g-\overline{W}_g^{(2)}))_{\ell m} = 0$

(2) $\operatorname*{plim}_{N\to\infty} \frac{1}{N} \sum_{g=1}^{G} (\overline{W}_g^{(2)\prime}(W_g-\overline{W}_g^{(2)}))_{mq} = 0$

for all $\ell = 1,2\ldots L$, $m,q = 1,2\ldots M$ where the subscripts $\ell m$ and $mq$ denote the $\ell m$ and $mq$ elements of the matrix.

By Lemmas 1 and 2 sufficient conditions for (1) and (2) are:

a.1) $\lim_{N\to\infty} \frac{1}{N} \sum_{g=1}^{G} E(Z_g'(W_g-\overline{W}_g^{(2)}))_{\ell m} = 0$

a.2) $\lim_{N\to\infty} \frac{1}{N} \sum_{g=1}^{G} E(\overline{W}_g^{(2)\prime}(W_g-\overline{W}_g^{(2)}))_{mq} = 0$

b.1) $\lim_{N\to\infty} \frac{1}{N} \sum_{g=1}^{G} E\left[(Z_g'(W_g-\overline{W}_g^{(2)}))_{\ell m}\right]^2 = O(1)$

b.2) $\lim_{N\to\infty} \frac{1}{N} \sum_{g=1}^{G} E\left[(\overline{W}_g^{(2)\prime}(W_g-\overline{W}_g^{(2)}))_{mq}\right]^2 = O(1)$

for all $\ell = 1,2,\ldots L$; $m,q = 1,2,\ldots M$.

By (i) and sampling independence,

(3) $E(Z'_g(W_g-\overline{W}_g^{(2)}))_{\ell m} = 0 \;\forall g.$

Thus (a.1) necessarily holds.

For the absolute value of the LHS of (a.2) we have

(4) $$\lim_{N\to\infty} \frac{1}{N} \left| \sum_{g=1}^{G} (E\overline{w}_{gm}^{(2)} \sum_{k=1}^{n_g} w_{gqk} - n_g E\overline{w}_{gm}^{(2)} \overline{w}_{gq}^{(2)})\right| = \lim_{N\to\infty} \left| \sum_{g=1}^{G} \frac{n_g}{N} (\mathrm{cov}(w_{gm},w_{gq})/n_g^{(2)})\right|$$

$$\leq \lim_{N\to\infty} \sup_g |\mathrm{cov}(w_{gm}w_{gq})| / \inf_g n_g^{(2)} .$$

Since by (ii)

$$\lim_{N\to\infty} \inf_g n_g^{(2)} = \infty$$

and since by the assumption of uniformly bounded moments

$\lim_{N\to\infty} \sup_g |\mathrm{cov}(w_{gm}w_{gq})| < K$ for some finite K (a.2) must hold.

Finally, to show (b.1) and (b.2), we add and subtract $Ew_{gm}$ within $(w_{gmk} - \overline{w}_{gm}^{(2)})$ which yields by the intragroup and intersample sampling independence:

$$\lim_{N\to\infty} \frac{1}{N} \sum_g E\left[\sum_{k=1}^{n_g} z_{g\ell k}^2 (w_{gmk} - Ew_{gm})^2 + (\overline{w}_{gm}^{(2)} - Ew_{gm})^2 (\sum_{k=1}^{n_g} z_{g\ell k})^2\right]$$

$$\lim_{N\to\infty} \frac{1}{N} \sum_g E\left[\overline{w}_{gm}^{(2)\,2} \sum_{k=1}^{n_g} (w_{gqk} - Ew_{gq})^2 + n_g^2 (\overline{w}_{gm}^{(2)} (\overline{w}_{gq}^{(2)} - Ew_{gq}))^2\right]$$

for the LHS of (b.1) and (b.2) respectively. By (ii) and the uniform bound that exists for all fourth moments, (b.1) and (b.2) hold and consequently does the proposition.

Proposition 2

We prove the proposition first with the use of 1(iiia), then with 1(iiib) and finally show the equivalence of $\Delta$ under either of these assumptions.

We begin by noting:

$$\sqrt{N}\,(\hat{\beta}-\beta) = \left(\frac{1}{N}X'X\right)^{-1} \frac{1}{\sqrt{N}} X'\xi$$

and that by assumption both $\frac{1}{N}X'X$ and Q are of full rank. Therefore, $Q^{-1}$ exists and thus

$$(1) \quad \left(\frac{1}{N}X'X\right) \overset{p}{\rightarrow} Q^{-1}.$$

Using (iiia) one can rewrite $X_g'\xi_g$ as follows

$$X_g'\xi_g = \delta_{g1} + \delta_{g2} + \delta_{g3}$$

where

$$\delta_{g1} = \begin{bmatrix} Z_g' \\ \bar{W}_g^{(2)'} \end{bmatrix} (W_g - EW_g)\beta_W$$

$$\delta_{g2} = \begin{bmatrix} Z_g' \\ \bar{W}_g^{(2)'} \end{bmatrix} (\bar{W}_g^{(2)} - EW_g)\beta_W$$

$$\delta_{g3} = \begin{bmatrix} Z_g' \\ \bar{W}_g^{(2)'} \end{bmatrix} \varepsilon_g$$

Each element of $\frac{1}{n_g}\delta_{gj}$ $j = 1, 2, 3$ is either a weighted sum of independent mean zero random variables or a weighted sum of independent mean zero random variables multiplied by a sample mean. The $\ell$ th and $(L + M)$ th elements of each respectively are:

$$\left.\begin{array}{l} \sum_i \beta_{w_i} \frac{1}{n_g} \sum_{k=1}^{n_g} z_{g\ell k}(w_{gik} - Ew_{gi}) \\ \\ \bar{w}_{gm}^{(2)} \sum_i \beta_{w_i} \frac{1}{n_g} \sum_{k=1}^{n_g} (w_{gik} - Ew_{gi}) \end{array}\right\} \quad j = 1$$

$$\left.\begin{array}{l} \bar{z}_{g\ell} \sum_i \beta_{w_i} \frac{1}{n_g^{(2)}} \sum_{k=1}^{n_g^{(2)}} (w_{gik}^{(2)} - Ew_{gi}) \\ \\ \bar{w}_{gm}^{(2)} \sum_i \beta_{w_i} \frac{1}{n_g^{(2)}} \sum_{k=1}^{n_g^{(2)}} (w_{gik}^{(2)} - Ew_{gi}) \end{array}\right\} \quad j = 2$$

$$\left.\begin{array}{l} \frac{1}{n_g} \sum_k z_{g\ell k} \, \varepsilon_{gk} \\ \\ \bar{w}_{gm}^{(2)} \frac{1}{n_g} \sum_k \varepsilon_{gk} \end{array}\right\} \quad j = 3$$

Thus by 1(i, ii, iiia), and the convergence of sample to population moments

$$\frac{1}{n_g} \delta_{gj} \overset{p}{\to} 0 \quad \forall g \quad j = 1, 2, 3$$

and thus along with (i)

$$\frac{1}{\sqrt{n_g}} \delta_{g1} \overset{d}{\to} N(0 \;\; \Delta_{g1})$$

where

$$\Delta_{g1} = \beta_w{}' \text{cov}(W_g) \beta_w \, Q_g$$

$$\frac{1}{\sqrt{n_g}} \delta_{g2} \overset{d}{\to} N(0, \; \Delta_{g2})$$

where

$$\Delta_{g2} = a_g \; \beta_w{}' \text{cov}(W_g) \beta_w \, \tilde{Q}_g$$

and

$$\frac{1}{\sqrt{n_g}} \delta_{g3} \xrightarrow{d} N(0, \Delta_{g3})$$

$$\Delta_{g3} = \sigma^2 Q_g$$

where

(2) $Q_g = \frac{1}{n_g} E X_g' X_g$

(3) $\tilde{Q}_g = \frac{1}{n_g} E X_g' \tilde{A}_g X_g.$

Therefore since

$$\frac{1}{n_g} E \delta_{gi} \delta_{gj}' = 0 \quad i \neq j$$

we have by the multivariate extension to the Lindeberg Central Limit Theorem

$$\frac{1}{\sqrt{n_g}} X_g' \xi_g \xrightarrow{d} N(0, \Delta_g)$$

$$\Delta_g = \sum_j \Delta_{gj}$$

and thus by sampling independence

(4) $$\frac{1}{\sqrt{N}} \sum_g X_g' \xi_g \xrightarrow{d} N(0, \Delta)$$

(5) $$\Delta = \sum_g \frac{n_g}{N} \Delta_g = \lim_{N\to\infty} \sum_g \frac{n_g}{N} [\beta_w' \operatorname{cov}(W_g) \beta_w [Q_g + a_g \tilde{Q}_g]] + \sigma^2 Q$$

Using (iiib)

$$\frac{1}{\sqrt{N}} X' \xi = \frac{1}{\sqrt{N}} \sum_g X_g' \xi_g$$

is the sum of a series of independent but not identically distributed random vectors. Define

$$\delta_g = X_g' \xi_g + \begin{bmatrix} 0 \\ \hline \frac{n_g^{(2)}}{n_g} \operatorname{cov}(W_g) \beta_w \end{bmatrix}.$$

By the multivariate extension (Rao p. 128) to the Liapunov central limit theorem

(6) $\frac{1}{\sqrt{N}} X'\xi \xrightarrow{d} N(0,\Delta)$

where

$$\Delta = \lim_{N\to\infty} \frac{1}{N} \sum_g E\ \delta_g\ \delta_g'$$

if

(a.1) $E\ \delta_g = 0\ \forall g$

(a.2) $$\lim_{G\to\infty} \frac{(\sum_g E|\ell'\delta_g|^3)^{1/3}}{(\sum_g E(\ell'\delta_g)^2)^{1/2}} = 0 \quad \forall\ \ell \in \mathbb{R}^{L+M},$$

(a.3) $\left(\frac{1}{\sqrt{N}} X'\xi - \frac{1}{\sqrt{N}} \sum_g \delta_g\right) \xrightarrow{P} 0$

Conditions (a.1) and (a.2) are sufficient for $\frac{1}{\sqrt{N}} \sum_g \delta_g$ to converge in distribution, while (a.3) insures that $\frac{1}{\sqrt{N}} X'\xi$ converges in distribution to exactly the same limiting distribution as does $\frac{1}{\sqrt{N}} \sum_g \delta_g$.

By inspection of equations (3) and (4) of the preceding proof (a.1) holds. One can rewrite (a.3) as

(a.3)' $\lim_{N\to\infty} \frac{1}{\sqrt{N}} \sum_g \frac{n_g}{n_g^{(2)}} \operatorname{cov}(W_g)\ \beta_w = 0.$

By the assumptions 1(ii), 2(ii), 2(iii), and of uniformly bounded moments, one can bound (a.3)' by

(7) $$\lim_{N\to\infty} \left[\sup_{g,h} \frac{\sqrt{n_g}}{\sqrt{n_h}}\right]\left[\sup_g \frac{\sqrt{n_g}}{\sqrt{n_g^{(2)}}}\right]\left[\sup_g \operatorname{cov}(W_g)\ \beta_w\right] \lim_{N\to\infty}\left[\frac{\sqrt{G}}{\inf_g \sqrt{n_g^{(2)}}}\right] = 0$$

and therefore (a.3) holds. Note that (7) implies

(8) $$\Delta = \lim_{N\to\infty} \frac{1}{N} \sum_g E\, X_g' \xi_g \xi_g' X_g = \lim_{N\to\infty} \frac{1}{N} \sum_g E\, \delta_g \delta_g'$$

and that (7) along with 2(ii) imply the following sufficient conditions for (a.2):

(a.2.1) $$\sup_g E \left| \frac{1}{\sqrt{n_g}} \ell' X_g' \xi_g \right|^3 = O(1)$$

(a.2.2) $[\frac{1}{n_g} E\, X_g' \xi_g \xi_g' X_g]$ is positive definite $\forall g$.

To motivate these conditions, we note that given (8) the LHS of (a.2) can be bounded by the following expression:

$$\lim_{N\to\infty} G^{-1/6} \sup_{g,h} \left(\frac{n_g}{n_h}\right)^{1/2} \frac{\left(\sup_g E\left|\frac{1}{\sqrt{n_g}} \ell' X_g' \xi_g\right|^3\right)^{1/3}}{\left(\inf_g E\left(\frac{1}{\sqrt{n_g}} \ell' X_g' \xi_g\right)^2\right)^{1/2}}, \quad \forall \ell \in \mathbb{R}^{L+M}$$

and that $\inf_g E(\frac{1}{\sqrt{n_g}} \ell' X_g' \xi_g)^2 > 0 \ \forall \ell$ is implied by (a.2.2).

To show (a.2.1) holds note that its LHS can be bounded by

$$\sup_g E \Big( |\ell_1 \beta_{L+1}| \Big| \frac{1}{\sqrt{n_g}} \sum_{k=1}^{n_g} z_{g1k}(w_{g1k} - Ew_{g1}) \Big| + |\ell_1 \beta_{L+1}| \Big| \frac{1}{\sqrt{n_g}} (\overline{w}_{g1}^{(2)} - Ew_{g1}) \sum_k z_{g1k} \Big|$$

$$+ |\ell_1| \Big| \frac{1}{\sqrt{n_g}} \sum_k z_{g1k}\, Ew_{gk} \Big| + |\ell_1 \beta_{L+2}| \Big| \frac{1}{\sqrt{n_g}} \sum_k z_{g1k} (w_{g2k} - Ew_{g2}) \Big| + \ldots$$

$$+ |\ell_L \beta_{L+M}| \Big| \frac{1}{\sqrt{n_g}} (\overline{w}_{gM}^{(2)} - Ew_{gM}) \sum_k z_{gLk} \Big| + |\ell_L| \Big| \frac{1}{\sqrt{n_g}} \sum_k z_{gLk} \varepsilon_{gk} \Big| + \ldots$$

$$+ |\ell_{L+M} \beta_{L+M}| \Big| \sqrt{n_g}\, \overline{w}_{gM}^{(2)} (\overline{w}_{gM}^{(2)} - Ew_{gM}) \Big| + |\ell_{L+M}| \Big| \frac{1}{\sqrt{n_g}} \overline{w}_{gM}^{(2)} \sum_k \varepsilon_{gk} \Big| \Big)^3$$

Upon expansion there are $[(L+M)(2M+1)]^3$ terms (without collecting similar terms) of the following form:

(9) $$\text{constant} \cdot \sup_g E \Big|\sum_{i=1}^{n_g} a_{1i}\Big| \Big|\sum_{i=1}^{n_g} a_{2i}\Big| \Big|\sum_{i=1}^{n_g} a_{3i}\Big| / n_g^{3/2}$$

If $a_{ji} \in \{z_{g\ell i}(w_{gmi} - Ew_{gm}), z_{g\ell i}\,\varepsilon_{gi}:\ \ell=1,2,..L,\ m=1,2,...M\}$ for all $j=1,2,3$ then by intragroup sampling independence and the Holder Inequality, (9) is bounded by

(10) $$\text{constant} \cdot \sup_g \Big[\frac{1}{n_g}\sum_i \frac{1}{\sqrt{n_g}} E|a_{1i}\,a_{2i}\,a_{3i}| + \Big(\frac{1}{n_g}\sum_{i\neq j} E|a_{1i}\,a_{2i}|\Big) \times$$

$$\Big(\frac{1}{n_g}\sum_{j\neq i} Ea_{3j}^2\Big)^{1/2} + \ldots + \Big(\frac{1}{n_g}\sum_{i\neq j,k} Ea_{1i}^2\Big)^{1/2}\Big(\frac{1}{n_g}\sum_{j\neq i,k} Ea_{2j}^2\Big)^{1/2}\Big(\frac{1}{n_g}\sum_{k\neq i,j} Ea_{3k}^2\Big)^{1/2}\Big]$$

which by the moment assumptions is $O(1)$.

If $a_{ji} \in \{z_{g\ell i}(\bar{w}^{(2)}_{gmi} - E\bar{w}_{gm}),\ \bar{w}^{(2)}_{gq}(\bar{w}^{(2)}_{gm} - Ew_{gm}),\ \bar{w}^{(2)}_{gm}\,\varepsilon_{gi}:\ \ell=1,2,..L,\ m,q=1,2,..M\}$ for any $j=1,2,3$

then by intersample sampling independence (9) is bounded by

(11) $$\text{constant} \cdot \sup_g \Big[\sum_{j=1}^{5} E\,h_j(W_g^{(2)})\,Ef_j(Z_g, \varepsilon_g)\Big]/n_g^{3/2}$$

By (5) at most

$$\sup_g Ef_j(Z_g, \varepsilon_g) = O(n_g^{3/2})\ \forall_j$$

and one can show that at most

$$\sup_g Eh_j(W_g^{(2)}) = O(1)\ \forall_j$$

Consequently (a.2.1) holds.

Finally to prove (a.2.2) we note that by assumption $X_g$ is of full rank and $\xi$ is a mean zero random variable which takes values other than zero with probability greater than zero.

Thus, by either (1) and (4) or (1) and (6) the proposition holds. All that remains is to show the equivalence of the $\Delta$ definitions given by (5) and (8). Since $\Delta$ is a function of fourth moments we consider $\Delta$ on an element by element basis. Using the definition of (8), we first partition $\Delta$ as follows:

$$\Delta = \begin{bmatrix} \Delta_{11} & \Delta_{12} \\ \Delta_{21} & \Delta_{22} \end{bmatrix}$$

where

$$\Delta_{11} = \lim_{N\to\infty} \frac{1}{N} [\sum_g E\, Z_g' \,(W_g - \overline{W}_g^{(2)})\, \beta_W \beta_W' \,(W_g - \overline{W}_g^{(2)})' Z_g + \sum_g E\, Z_g' \,\varepsilon_g \varepsilon_g' \,Z_g]$$

$$\Delta_{12} = \lim_{N\to\infty} \frac{1}{N} [\sum_g E\, Z_g' \,(W_g - \overline{W}_g^{(2)})\, \beta_W \beta_W' \,(W_g - \overline{W}_g^{(2)})' \,\overline{W}_g^{(2)} + \sum_g E\, Z_g' \,\varepsilon_g \varepsilon_g' \,\overline{W}_g^{(2)}]$$

$$\Delta_{22} = \lim_{N\to\infty} \frac{1}{N} [\sum_g E\, \overline{W}_g^{(2)'} (W_g - \overline{W}_g^{(2)})\, \beta_W \beta_W' \,(W_g - \overline{W}_g^{(2)})' \,\overline{W}_g^{(2)} + \sum_g E\, \overline{W}_g^{(2)'} \,\varepsilon_g \varepsilon_g' \,\overline{W}_g^{(2)}]$$

and consider their respective $\ell p$, $\ell m$, and mq elements

$$(\Delta_{11})_{\ell p} = \lim_{N\to\infty} \sum_{i=1}^{M} \sum_{j=1}^{M} \beta_{w_i} \beta_{w_j} \frac{1}{N} \sum_g E \sum_{k=1}^{n_g} \sum_{h=1}^{n_g} z_{g\ell k}\, (w_{gik} - \overline{w}_{gi}^{(2)})\, z_{gph}\, (w_{gjh} - \overline{w}_{gj}^{(2)})$$

$$+ \lim_{N\to\infty} \frac{1}{N} \sum_g E \sum_k \sum_h z_{g\ell k}\, z_{gph}\, \varepsilon_{gk}\, \varepsilon_{gh}$$

which by sampling independence, the conditional independence of $\varepsilon_g$ given $Z_g$, and 2(i)

$$= \lim_{N\to\infty} \sum_i \sum_j \beta_{w_i} \beta_{w_j} \frac{1}{N} \sum_g [n_g\, E\, z_{g\ell}\, z_{gp}\, \mathrm{cov}(w_{gi}\, w_{gj}) + (\frac{n_g}{n_g^{(2)}}\, \mathrm{cov}(z_{g\ell}\, z_{gp})$$

$$+ \frac{n_g^2}{n_g^{(2)}}\, Ez_{g\ell}\, Ez_{gp})\, \mathrm{cov}(w_{gi}\, w_{gj})] + \lim_{N\to\infty} \frac{1}{N} \sum_g n_g\, \sigma^2\, Ez_{g\ell}\, z_{gp}$$

$$(12) \quad = \lim_{N\to\infty} \sum_g \frac{n_g}{N} [\beta_W'\, \mathrm{cov}(w_g)\, \beta_W\, (E\, z_{g\ell}\, z_{gp} + a_g\, Ez_{g\ell}\, Ez_{gp}) + \sigma^2\, Ez_{g\ell}\, z_{gp}]$$

$$(\Delta_{12})_{\ell m} = \lim_{N\to\infty} \sum_i \sum_j \beta_{w_i} \beta_{w_j} \frac{1}{N} \sum_g E\, \overline{w}_{gm}^{(2)} \sum_k \sum_h z_{g\ell k}\, (w_{gik} - \overline{w}_{gi}^{(2)})(w_{gjh} - \overline{w}_{gj}^{(2)})$$

$$+ \lim_{N\to\infty} \frac{1}{N} \sum_g E \sum_k \sum_h z_{g\ell k}\, \overline{w}_{gm}^{(2)}\, \varepsilon_{gk}\, \varepsilon_{gh}$$

which by sampling independence, conditional independence of $\varepsilon$ given Z, and 1(i)

$$= \lim_{N\to\infty} \sum_i \sum_j \beta_{w_i} \beta_{w_j} \frac{1}{N} \sum_g [n_g\, Ez_{g\ell}\, Ew_{gm}\, \mathrm{cov}(w_{gi}\, w_{gj}) +$$

$$n_g^2\, E\, z_{g\ell}\, [n_g^{(2)^{-2}}\, E\, w_{gm}\, (w_{gi} - E\, w_{gi})\, (w_{gj} - E\, w_{gj}) +$$

$$\frac{n_g^{(2)}-1}{n_g^{(2)^2}} E\, w_{gm}\, cov(w_{gi}\, w_{gj})]] + \lim_{N\to\infty} \frac{1}{N} \sum_g n_g\, E\, z_{g\ell}\, E\, w_{gm}\, \sigma^2$$

and by 2(iii)

$$(13) = \lim_{N\to\infty} \sum_g \frac{n_g}{N} [\beta_w'\, cov(W_g)\, \beta_w\, (1+a_g) + \sigma^2]\, E\, z_{g\ell}\, E\, w_{gm}$$

$$(\Delta_{22})_{mq} = \lim_{N\to\infty} \sum_i \sum_j \beta_{w_i}\, \beta_{w_j} \frac{1}{N} \sum_g E\, \bar{w}_{gm}^{(2)}\, \bar{w}_{gq}^{(2)} \sum_k \sum_h (w_{gik}-\bar{w}_{gi}^{(2)})(w_{gjh}-\bar{w}_{gj}^{(2)})$$

$$+ \lim_{N\to\infty} \frac{1}{N} \sum_g E\, \bar{w}_{gm}^{(2)}\, \bar{w}_{gq}^{(2)} \sum_k \sum_h \varepsilon_{gk}\, \varepsilon_{gh}$$

which by sampling independence

$$= \lim_{N\to\infty} \sum_i \sum_j \beta_{w_i}\, \beta_{w_j} \frac{1}{N} \sum_g [n_g\, cov(w_{gi}\, w_{gj})(n_g^{(2)^{-1}}\, cov(w_{gm}\, w_{gq}) +$$

$$E\, w_{gm}\, E\, w_{gq}) + n_g^2\, [n_g^{(2)^{-3}}\, E\, w_{gm}\, w_{gq}\, (w_{gi}-E\, w_{gi})(w_{gi}-E\, w_{gj}) +$$

$$\frac{n_g^{(2)}-1}{n_g^{(2)^3}} [E\, w_{gm}\, (w_{gi}-E\, w_{gi})\, (w_{gj}-E\, w_{gj})\, E\, w_{gq} + E\, w_{gq}\, (w_{gi}-E\, w_{gi})\, (w_{gj}-E\, w_{gj})\, E\, w_{gm}$$

$$+ cov(w_{gm}\, w_{gi})\, cov(w_{gq}\, w_{gj}) + cov(w_{gm}\, w_{gj})\, cov(w_{gq}\, w_{gi})$$

$$+ cov(w_{gi}\, w_{gj})\, E\, w_{gm}\, w_{gq}] + \frac{(n_g^{(2)}-1)(n_g^{(2)}-2)}{n_g^{(2)^3}}\, cov\, (w_{gi}\, w_{gj})\, E\, w_{gm}\, E\, w_{gq}]]$$

$$+ \lim_{N\to\infty} \sum_g n_g\, \sigma^2\, (n_g^{(2)^{-1}}\, cov(w_{gm}\, w_{gq}) + E\, w_{gm}\, E\, w_{gq})$$

and by 2(iii)

$$(14) \quad = \lim_{N\to\infty} \sum_g \frac{n_g}{N} [\beta_w'\, cov(W_g)\, \beta_w\, (1+a_g) + \sigma^2]\, E\, w_{gm}\, E\, w_{gq}$$

Combining (12), (13) and (14) to reconstruct $\Delta$, shows the definitions' equivalence.

Proposition 3

Given assumptions 1(ii), and 2(iii), Proposition 1, equations (1), (2), and (3) of the preceding proof, and the unbiasedness of $\widehat{\text{cov}}\,(W_g)$, one need only show $\text{plim}_{N\to\infty}\ \hat{\sigma}^2 = \sigma^2$

By definition

$$\hat{\sigma}^2 = \frac{1}{N}\hat{\xi}'\hat{\xi} - \sum_g \frac{n_g}{N}\hat{\beta}_w' \widehat{\text{cov}}\,(W_g)\ \hat{\beta}_w$$

$$\frac{1}{N}\hat{\xi}'\hat{\xi} = (y-X\hat{\beta})'(y-X\hat{\beta}) = \frac{1}{N}[(V-X)\beta - X(\hat{\beta}-\beta) + \varepsilon)]'[(V-X)\beta - X(\hat{\beta}-\beta) + \varepsilon]$$

and by the consistency of $\hat{\beta}$

$$= \frac{1}{N}\beta'\ (V-X)'(V-X)\beta + \frac{1}{N}\varepsilon'\varepsilon + o_p(N).$$

By assumption

$$\frac{1}{N}\varepsilon'\varepsilon \overset{P}{\to} \sigma^2$$

while

$$\frac{1}{N}\beta'\ (V-X)'(V-X)\beta = \beta_w'\ [\frac{1}{N}\sum_g (W_g-\overline{W}_g^{(2)})'(W_g-\overline{W}_g^{(2)})]\ \beta_w$$

and by sampling independence and addition and subtraction of $E\,W_g$ within $(W_g - \overline{W}_g^{(2)})$

$$\overset{P}{\to} \lim_{N\to\infty}\frac{1}{N}\sum_g (n_g + \frac{n_g}{n_g^{(2)}})\ \beta_w'\ \text{cov}(W_g)\ \beta_w$$

which by 2(iii)

$$= \lim_{N\to\infty}\sum_g \frac{n_g}{N}\beta_w'\ \text{cov}(W_g)\ \beta_w.$$

Finally, by Lemma 2, Proposition 1, and since

$$E\ \widehat{\text{cov}}(W_g) = \text{cov}(W_g)$$

$$\text{plim}_{N\to\infty}\sum_g \frac{n_g}{N}\hat{\beta}_w'\ \widehat{\text{cov}}\,(W_g)\ \hat{\beta}_w = \lim_{N\to\infty}\sum_g \frac{n_g}{N}\beta_w'\ \text{cov}(W_g)\ \beta_w$$

and thus

$$(1)\quad \hat{\sigma}^2 \overset{P}{\to} \sigma^2$$

and the proposition is proved.

Proposition 4

To prove consistency, consider

$$(\tilde{\beta} - \beta) = \tilde{R}^{-1} (\tilde{X}'X)^{-1} \tilde{X}' (y - X\tilde{R}\beta)$$

$$= (\tilde{X}'X - S)^{-1} [(\tilde{X}' (V-X) + S)\beta + \tilde{X}'\varepsilon].$$

By $E(\varepsilon|V) = 0$,

$$(\tilde{\beta} - \beta) \overset{P}{\to} 0 \text{ if and only if}$$

$$\frac{1}{N} [\tilde{X}' (V-X) + S] = \frac{1}{N} \begin{bmatrix} \tilde{Z}' (W-\overline{W}^{(2)}) \\ \overline{W}^{(2)\prime} (W-\overline{W}^{(2)}) + S_{22} \end{bmatrix} \overset{P}{\to} 0$$

Since $\tilde{z}_{g\ell k}$ is uncorrelated with $w_{gmk}$ by the independence of observations

$$\frac{1}{N} \tilde{Z}'(W-\overline{W}^{(2)}) \overset{P}{\to} 0$$

Similar to the proof of Proposition 1, by the assumption of uniformly bounded moments and sampling independence:

$$\lim_{N\to\infty} \frac{1}{N} \sum_g \frac{n_g}{n_g^{(2)}} E\, [\widehat{\text{cov}}\, (W_g) - \text{cov}(W_g)]_{mq} = 0$$

$$\lim_{N\to\infty} \frac{1}{N} \sum_g \frac{n_g}{n_g^{(2)}} E\, [\widehat{\text{cov}}\, (W_g) - \text{cov}(W_g)]_{mq}^2 = O(1) \text{ at most}$$

where $[\cdot]_{mq}$ represents the mq matrix element.

Thus by (i), Lemmas 1 and 2,

$$\frac{1}{N} S_{22} \overset{P}{\to} \Omega_{\overline{W}} \, .$$

Since $\frac{1}{N} \overline{W}^{(2)} (W-\overline{W}^{(2)}) \overset{P}{\to} -\Omega_{\overline{w}}$

$$\frac{1}{N} [\overline{W}^{(2)\prime} (W-\overline{W}^{(2)}) + S_{22}] \overset{P}{\to} 0$$

and consistency is proved.

For asymptotic normality

$$\sqrt{N}\,(\tilde{\beta} - \beta) = (\tfrac{1}{N}\,[\tilde{X}'X-S])^{-1}\,\frac{1}{\sqrt{N}}\,([\tilde{X}'\,(V-X) + S]\beta + \tilde{X}'\varepsilon).$$

We define $P^{-1}$ by the following:

$$(1)\quad \frac{1}{N}\,[\tilde{X}'X - S]^{-1} \overset{P}{\to} P^{-1}$$

and we write

$$\frac{1}{\sqrt{N}}\,([\tilde{X}'\,(V-X) + S]\beta + \tilde{X}'\varepsilon) = \frac{1}{\sqrt{N}}\sum_g \gamma_g$$

where

$$\gamma_g = \begin{bmatrix} \tilde{Z}_g'(W_g - \overline{W}_g^{(2)})\beta_W + \tilde{Z}_g'\varepsilon_g \\ (\overline{W}_g^{(2)\prime} W_g - n_g S_g^*)\beta_W + \overline{W}_g^{(2)\prime}\varepsilon_g \end{bmatrix}$$

$$S_g^* = [n_g^{(2)}(n_g^{(2)}-1)]^{-1}\, W_g^{(2)\prime}\,[e_{n_g}^{(2)}\, e_{n_g}^{(2)\prime} - I]\, W_g^{(2)}$$

To verify this, note that

$$\frac{n_g}{n_g^{(2)}}\,\widehat{\text{cov}}\,(W_g) - \overline{W}_g^{(2)\prime}\overline{W}_g^{(2)} = n_g[\frac{1}{n_g^{(2)}}\,\widehat{\text{cov}}\,(W_g) - \overline{w}_g^{(2)\prime}\overline{w}_g^{(2)}]$$

$$= n_g W_g^{(2)\prime}\,[[n_g^{(2)}(n_g^{(2)}-1)]^{-1}(I-\frac{ee'}{n_g^{(2)}}) - n_g^{(2)^{-2}}\,ee']\,W_g^{(2)}$$

$$= n_g[n_g^{(2)^2}(n_g^{(2)}-1)]^{-1}\,W_g^{(2)\prime}[n_g I - ee' - (n_g^{(2)}-1)\,ee']\,W_g^{(2)}$$

$$= -n_g[n_g^{(2)}(n_g^{(2)}-1)]^{-1}\,W_g^{(2)\prime}[ee' - I]W_g^{(2)}$$

where the subscripts of e have been suppressed.

From proposition 3 and inspection of $S_g^*$,

(a.1) $E\ \gamma_g = 0\ \forall g$

and the sequence of vectors $\gamma_g$, g=1,2...G are independent.

Following the proof of proposition 2, by (a.1) and if

$$\text{(a.2)} \quad \lim_{N\to\infty} \frac{(\sum_g E|\ell'\gamma_g|^3)^{1/3}}{(\sum_g E(\ell'\gamma_g)^2)^{1/2}} = 0 \qquad \ell \in \mathbb{R}^{(L+M)}$$

then

$$\text{(2)} \quad \frac{1}{\sqrt{N}} \sum_g \gamma_g \xrightarrow{d} N(0,\Gamma)$$

where

$$\Gamma = \lim_{N\to\infty} \frac{1}{N} \sum_g E\, \gamma_g\, \gamma_g' = \operatorname*{plim}_{N\to\infty} \frac{1}{N} \sum_g \gamma_g\, \gamma_g'.$$

Thus, (1) and (2) will prove the proposition.

To show (a.2) and therefore (2), we first note that the LHS can be bounded by

$$\lim_{N\to\infty} G^{-1/6} \frac{(\sup_g E|\ell'\gamma_g|^3)^{1/3}}{(\inf_g E(\ell'\gamma_g)^2)^{1/2}}$$

By 4(ii) and the assumption of bounded moments

$$\lim_{N\to\infty} \sup_g E\,|\ell'\gamma_g|^3$$

is some positive constant. Similar to the proof of proposition 2, $\frac{1}{n_g} E\, \gamma_g'\gamma_g$ is a positive definite matrix so $\lim_{N\to\infty} \inf_g E(\ell'\gamma_g)^2$ is a non-zero constant. Thus by 4(i) (a.2) holds and the proposition is proved.

Proposition 5

We prove the proposition by offering an element by element consistency argument.

$$\Gamma = \begin{bmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{bmatrix}$$

where

$$\Gamma_{11} = \lim_{N\to\infty} \frac{1}{N} \sum_g [E\ \tilde{Z}_g' (W_g - \overline{W}_g^{(2)})\ \beta_W \beta_W' (W_g - \overline{W}_g^{(2)})' \tilde{Z}_g + \tilde{Z}_g' \varepsilon_g \varepsilon_g' \tilde{Z}_g]$$

$$\Gamma_{12} = \lim_{N\to\infty} \frac{1}{N} \sum_g [\tilde{Z}_g' (W_g - \overline{W}_g^{(2)})\ \beta_W \beta_W' (\overline{W}_g^{(2)\prime} W_g - n_g S_g^*)' + \tilde{Z}_g' \varepsilon_g \varepsilon_g' \overline{W}_g^{(2)}]$$

$$\Gamma_{22} = \lim_{N\to\infty} \frac{1}{N} \sum_g E\ [(\overline{W}_g^{(2)\prime} W_g - n_g S_g^*) \beta_W \beta_W (\overline{W}_g^{(2)\prime} W_g - n_g S_g^*)' + \overline{W}_g^{(2)\prime} \varepsilon_g \varepsilon_g' \overline{W}_g^{(2)}]$$

Consider their respective $\ell p$, $\ell m$ and mq the elements:

$$(\Gamma_{11})_{\ell p} = \lim_{N\to\infty} \sum_i \sum_j \beta_{w_i} \beta_{w_j} \frac{1}{N} \sum_g (n_g - 1)^{-2}\ E \sum_{k\neq h} \sum \sum_{s\neq t} \sum z_{g\ell k}\ z_{gps}\ (w_{gih} - \overline{w}_{gi}^{(2)})$$

$$(w_{gjt} - \overline{w}_{gj}^{(2)}) + \lim_{N\to\infty} \frac{1}{N} \sum_g E \sum_h \sum_t \tilde{z}_{g\ell h}\ \varepsilon_{gh}\ \varepsilon_{gt}\ \tilde{z}_{gpt}$$

$$= \lim_{N\to\infty} \sum_i \sum_j \beta_{w_i} \beta_{w_j} \frac{1}{N} \sum_g [\frac{n_g}{n_g - 1}\ E\ z_{g\ell}\ z_{gp}\ \mathrm{cov}(w_{gi}\ w_{gj})$$

$$+ \mathrm{cov}(z_{g\ell}\ w_{gj})\ \mathrm{cov}(z_{gp}\ w_{gi})] + \frac{n_g(n_g-2)}{n_g-1}\ E\ z_{g\ell}\ E\ z_{gp}\ \mathrm{cov}(w_{gi}\ w_{gj})$$

$$+ n_g^{(2)^{-1}} (n_g\ E\ z_{g\ell}\ z_{gp} + n_g(n_g-1)\ E\ z_{g\ell}\ E\ z_{gp})\ \mathrm{cov}(w_{gi}\ w_{gj})]$$

$$+ \lim_{N\to\infty} \frac{1}{N} \sum_g n_g \; \sigma^2 \; [(n_g-1)^{-1} E\; z_{g\ell}\, z_{gp} + \frac{n_g-2}{n_g-1} E\; z_{g\ell}\; E\; z_{gp}]$$

$$(4) \quad = \lim_{N\to\infty} \sum_g \frac{n_g}{N} \beta_w' \operatorname{cov}(W_g)\; \beta_w \; [(\frac{1}{n_g-1} + \frac{1}{n_g^{(2)}}) E\; z_{g\ell}\, z_{gp} + (\frac{n_g-2}{n_g-1} + \frac{n_g-1}{n_g^{(2)}}) E\; z_{g\ell}\; E\; z_{gp}]$$

$$+ \sigma^2 \lim_{N\to\infty} \sum_g \frac{n_g}{N} [(n_g-1)^{-1} (E\; z_{g\ell}\, z_{gp} + (n_g-2)\; E\; z_{g\ell}\; E\; z_{gp})]$$

$$+ \lim_{N\to\infty} \sum_g \frac{n_g}{N} \cdot \frac{1}{n_g-1} \beta_w' \operatorname{cov}(z_{g\ell}\, W_g)' \operatorname{cov}(z_{gp}\, W_g)\; \beta_w$$

$$= \lim_{N\to\infty} \sum_g \frac{n_g}{N} \beta_w' \operatorname{cov}(W_g)\; \beta_w \; [(\frac{n_g}{1+n_g^{(2)}}) E\; z_{g\ell}\; E\; z_{gp} + (\frac{1}{n_g-1} + \frac{1}{n_g^{(2)}}) \operatorname{cov}(z_{g\ell}\, z_{gp})]$$

$$+ \sigma^2 \lim_{N\to\infty} \sum_g \frac{n_g}{N} [(n_g-1)^{-1} (E\; z_{gp}\, z_{g\ell}) + (n_g-2)\; E\; z_{g\ell}\; E\; z_{gp})]$$

$$+ \lim_{N\to\infty} \sum_g \frac{n_g}{N} [(n_g-1)^{-1} \beta_w' \operatorname{cov}(z_{g\ell}, W_g)' \operatorname{cov}(z_{gp}\, W_g)\; \beta_w]$$

$$(\Gamma_{12})_{\ell m} = \lim_{N\to\infty} \sum_i \sum_j \beta_{w_i} \beta_{w_j} \frac{1}{N} \sum_g [(n_g-1) n_g^{(2)}]^{-1} \sum_{k\neq h} \sum \sum_r E z_{g\ell k} w_{gmr} (w_{gih} - \bar{w}_{gi}^{(2)}) (\sum_s w_{gjs} - \frac{n_g}{n_g^{(2)}-1} \sum_{s\neq r} w_{gjs}^{(2)})$$

$$+ \lim_{N\to\infty} \frac{1}{N} \sum_g E\; \bar{w}_{gm}^{(2)} \sum_k \sum_h \tilde{z}_{g\ell k}\; \varepsilon_{gk}\; \varepsilon_{gh}$$

$$= \lim_{N\to\infty} \sum_i \sum_j \beta_{w_i} \beta_{w_j} \sum_g \frac{n_g}{N} [\operatorname{cov}(w_{gi}\, w_{gj})\; E\; z_{g\ell}\; E\; w_{gm}$$

$$- \frac{1}{n_g^{(2)}} \operatorname{cov}(z_{g\ell}\, w_{gj}) \operatorname{cov}(w_{gm}\, w_{gi}) + \frac{n_g}{n_g^{(2)}} \operatorname{cov}(w_{gi}\, w_{gj})\; E\; z_{g\ell}\; E\; w_{gm}]$$

$$+ \sigma^2 \lim_{N\to\infty} \sum_g \frac{n_g}{N} E\; z_{g\ell}\; E\; w_{gm}$$

$$(5) = \lim_{N\to\infty} \sum_g \frac{n_g}{N} \beta_w' \operatorname{cov}(W_g)\ \beta_w\ (1+a_g)\ E\ z_{g\ell}\ E\ w_{gm} + \sigma^2 \sum_g \frac{n_g}{N} E\ z_{g\ell}\ E\ w_{gm}$$

$$- \sum_g \frac{n_g}{N} \frac{1}{n_g^{(2)}} \beta_w' \operatorname{cov}(z_{g\ell}\ W_g)'\ \operatorname{cov}(w_{gm}\ W_g)\ \beta_w]$$

$$(\Gamma_{22})_{mq} = \lim_{N\to\infty} \sum_i \sum_j \beta_{w_i}\ \beta_{w_j} \frac{1}{N} \sum_g n_g^{(2)^{-2}}\ E \sum_{k=1}^{n_g^{(2)}} \sum_{h=1}^{n_g^{(2)}} w_{gmk}^{(2)}\ w_{gqh}^{(2)} \left( \sum_{s=1}^{n_g} w_{gis} - n_g(n_g^{(2)}-1)^{-1} \sum_{s\neq k=1}^{n_g^{(2)}} W_{gis}^{(2)} \right)$$

$$\left( \sum_{t=1}^{n_g} w_{gjt} - n_g(n_g^{(2)}-1)^{-1} \sum_{t\neq h=1}^{n_g^{(2)}} w_{gjt}^{(2)} \right) + \lim_{N\to\infty} \frac{1}{N} \sum_g E\ \overline{w}_{gm}^{(2)}\ \overline{w}_{gq}^{(2)} \sum_k \sum_h \varepsilon_{gk}\ \varepsilon_{gh}$$

$$= \lim_{N\to\infty} \sum_i \sum_j \beta_{w_i}\ \beta_{w_j} \sum_g \frac{n_g}{N} [\operatorname{cov}(w_{gi}\ w_{gj})\ (\frac{1}{n_g^{(2)}} E\ w_{gm}\ w_{gq} + \frac{n_g^{(2)}-1}{n_g^{(2)}} E\ w_{gm}\ E\ w_{gq})$$

$$+ \frac{n_g^2}{n_g^{(2)}(n_g^{(2)}-1)} [\operatorname{cov}(w_{gi}\ w_{gj})\ (E\ w_{gm}\ w_{gq} + (n_g^{(2)}-2)\ E\ w_{gm}\ E\ w_{gq})$$

$$+ \operatorname{cov}(w_{gi}\ w_{gm})\ \operatorname{cov}(w_{gj}\ w_{gq})]] + \sigma^2 \lim_{N\to\infty} \sum_g \frac{n_g}{N} [\frac{1}{n_g^{(2)}} \operatorname{cov}(w_{gm}\ w_{gq}) + E\ w_{gm}\ E\ w_{gq}]$$

$$(6) = \lim_{N\to\infty} \sum_i \sum_j \beta_{w_i}\ \beta_{w_j} \sum_g \frac{n_g}{N} [\operatorname{cov}\ (w_{gi}\ w_{gj})[\frac{1}{n_g^{(2)}} (1 + \frac{n_g}{n_g^{(2)}-1})\ E\ w_{gm}\ w_{gq}$$

$$+ \frac{1}{n_g^{(2)}} (n_g^{(2)} -1 + \frac{n_g(n_g^{(2)}-2)}{n_g^{(2)}-1})\ E\ w_{gm}\ E\ w_{gq}]]$$

$$+ \sigma^2 \lim_{N\to\infty} \sum_g \frac{n_g}{N} [\frac{1}{n_g(2)} \operatorname{cov}(w_{gm}\ w_{gq}) + E\ w_{gm}\ E\ w_{gq}] + \lim_{N\to\infty} \sum_i \sum_j \beta_{w_i} \beta_{w_j} \sum_g \frac{n_g}{N} \frac{n_g}{(n_g(2)(n_g(2)-1)}$$

$$\operatorname{cov}(w_{gi}\ w_{gm})\ \operatorname{cov}(w_{gj}\ w_{gq})$$

Now we show

$$(\hat{\Gamma}_{11})_{\ell p} \xrightarrow{P} (\Gamma_{11})_{\ell p}$$

$$(\hat{\Gamma}_{12})_{\ell m} \xrightarrow{P} (\Gamma_{12})_{\ell m}$$

$$(\hat{\Gamma}_{22})_{mq} \xrightarrow{P} (\Gamma_{22})_{mq}$$

for all $\ell$, p=1, 2,..L and m, q=1, 2,,..M

where

$$\hat{\Gamma} = \begin{bmatrix} \hat{\Gamma}_{11} & \hat{\Gamma}_{12} \\ \hat{\Gamma}_{21} & \hat{\Gamma}_{22} \end{bmatrix}$$

and where the submatrices are of the same dimension of the submatrices of $\Gamma$.

By Proposition 5 and Lemmas 1 and 2,

$$\operatorname*{plim}_{N\to\infty} (\hat{\Gamma}_{11})_{\ell p} = \sum_i \sum_j \beta_{w_i} \beta_{w_j} \lim_{N\to\infty} \frac{1}{N} \sum_g n_g \; E \; \{[(\frac{n_g-2}{n_g-1} + \frac{n_g-1}{n_g^{(2)}}) \; E \; \widehat{z_{g\ell} \; E \; z_{gp}}$$

$$+ ([n_g-1]^{-1} + n_g^{(2)-1}) \; E \; \widehat{z_{g\ell} \; z_{gp}}] \; \widehat{\text{cov}} \; (w_{gi} \; w_{gj})\} +$$

$$+ (\operatorname*{plim}_{N\to\infty} \tilde{\sigma}^2) \lim_{N\to\infty} \frac{1}{N} \sum_g E \sum_k \tilde{z}_{g\ell k} \; \tilde{z}_{gpk}$$

$$+ [\operatorname*{plim}_{N\to\infty} \Omega_{zw} \hat{\beta}_w]_\ell \; [[\operatorname*{plim}_{N\to\infty} \Omega_{zw} \hat{\beta}_w]_p \lim_{N\to\infty} \frac{1}{N} \sum_g \frac{n_g}{n_g-1}$$

By inspection one can note that $E \; \widehat{z_{g\ell} \; E \; z_{gp}}$, $E \; \widehat{z_{g\ell} \; z_{gp}}$, and $\widehat{\text{cov}} \; (w_{gi} \; w_{gj})$ are unbiased, and that

$$\sum_k E \; \tilde{z}_{g\ell k} \; \tilde{z}_{gpk} = (n_g-1)^{-1} \; E \; z_{g\ell} \; z_{gp} + (n_g-2) \; E \; z_{g\ell} \; E \; z_{gp}.$$

Also, by 4(i), 4(ii), and equation III.1,

$$\operatorname*{plim}_{N\to\infty} \Omega_{zw} \hat{\beta}_w = J[\operatorname*{plim}_{N\to\infty} \frac{1}{N} X'X] \; [\operatorname*{plim}_{N\to\infty} (\hat{\beta} - \beta) + \operatorname*{plim}_{N\to\infty} (\tilde{\beta} - \beta)]$$

$$= J \begin{bmatrix} \Omega_{zw} \\ \overline{\Omega_w} \end{bmatrix} \beta_w = \Omega_{zw} \; \beta_w$$

and therefore

$$[\underset{N\to\infty}{\text{plim}}\ \hat{\Omega}_{zw}\ \beta_w]_\ell = \sum_i \beta_{w_i} \sum_g \frac{n_g}{N} \text{cov}(z_{g\ell}\ w_{gi})$$

which by (i)

$$= \sum_i \beta_{w_i}\ \text{cov}(z_{g\ell}\ w_{gi}) \quad \forall g$$

Finally, using an argument similar to that of Proposition 3, one can show

(1) $\tilde{\sigma}^2 \overset{P}{\to} \sigma^2$.

Therefore using equation (4) above, we have

$$\underset{N\to\infty}{\text{plim}}\ (\hat{\Gamma}_{11})_{\ell p} = (\Gamma_{11})_{\ell p}.$$

By similar reasoning $\bar{z}_{g\ell}$, and $\widehat{\text{cov}(w_{gi}\ w_{gj})\ E\ w_{gm}}$, are also unbiased. Using Proposition 4, intersample independence, and equation (5) above one can show

$$\underset{N\to\infty}{\text{plim}}\ (\hat{\Gamma}_{12})_{\ell m} = (\Gamma_{12})_{\ell m}.$$

Finally, by using Proposition 4, (1), the unbiasedness of $E\ \widehat{w_{gm}\ w_{gq}\ \text{cov}(w_{gi}\ w_{gj})}$, $E\ \widehat{w_{gm}\ E\ w_{gq}\ \text{cov}(w_{gi}\ w_{gj})}$, and $\widehat{\text{cov}\ (w_{gm},\ w_{gj})\ \text{cov}\ (w_{gq}\ w_{gi})}$, as well as equation (6) above, we have

$$\underset{N\to\infty}{\text{plim}}\ (\hat{\Gamma}_{22})_{mq} = (\Gamma_{22})_{mq},$$

and the propsition is proved.