JOINT EFFECTS OF ERRORS OF MEASUREMENT AND GROUPING ON THE ESTIMATED PRODUCT-MOMENT CORRELATION

John J. Chai, and Chung Chen, Syracuse University
School of Management, Syracuse University, Syracuse, NY 13244-2130

Key words: Ecological Data, Measurement Error Models, Sample Survey.

1. Introduction

Despite the problems associated with grouped ("ecological") data, they are still used by many economists, sociologists, and others. The reasons for this practice vary from case to case. In general, grouped data may be employed to study the behavior of grouped properties, to examine the relationship between group characteristics and individual behaviors, and/or to save the cost of enumerating individual values.

Suppose that there is a population S of size N and there are M subpopulations $S_1, S_2, \ldots, S_M$, of sizes $N_1, N_2, \ldots,$ and N_M respectively, $\sum_{i=1}^{M} N_i = N$. Simple random samples of size n_i, $i = 1, \ldots, M$, were drwan from each subpopulation. Let (x_{ijk}, y_{ijk}) be the measurement taken on the kth subject by the jth interviewer from the ith subpopulation. Let (μ_X, μ_Y) be the population means of X and Y and $(\mu_X^{(i)}, \mu_Y^{(i)})$ be the population means of (X, Y) in the subpopulation i. Define ρ as the usual correlation coeifficient between X and Y,

$$\rho = \frac{E(X - \mu_X)(Y - \mu_Y)}{\sqrt{\sigma_X^2 \sigma_Y^2}} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} \tag{1}$$

where σ_X, σ_Y are the population standard deviations of (X, Y) respectively. Let ρ_G be the ecological (grouped) correlation between X and Y with respect to subpopulations S_1, S_2, ..., S_M. Then

$$\rho_G = \frac{\sum_{i=1}^{M} N_i(\mu_X^{(i)} - \mu_X)(\mu_Y^{(i)} - \mu_Y)/N}{\bar{\sigma}_X \bar{\sigma}_Y} \tag{2}$$

where $\mu_X^{(i)}$ and $\mu_Y^{(i)}$ are the means of variables X and Y respectively in subpopulation i, $\bar{\sigma}_X^2 = \sum_{i=1}^{M} N_i(\mu_X^{(i)} - \mu_X)^2/N$ and $\bar{\sigma}_Y^2 = \sum_{i=1}^{M} N_i(\mu_Y^{(i)} - \mu_Y)^2/N$ are between groups variations of X and Y respectively. It is well known that ρ_G and ρ are related by

$$\begin{aligned}\rho_G =& (1 - \ddot{\sigma}_X^2/\sigma_X^2)^{-1/2}(1 - \ddot{\sigma}_Y^2/\sigma_Y^2)^{-1/2} \\ & \times (1 - \ddot{\sigma}_{XY}/\sigma_{XY})\rho,\end{aligned} \tag{3}$$

where $\ddot{\sigma}_X^2$, $\ddot{\sigma}_Y^2$ and $\ddot{\sigma}_{XY}$ are within group variations and covariance respectively such that

$$\ddot{\sigma}_X^2 = \sum_{i=1}^{M} N_i \sigma_X^2(i)/N$$

$$\ddot{\sigma}_Y^2 = \sum_{i=1}^{M} N_i \sigma_Y^2(i)/N$$

$$\ddot{\sigma}_{XY} = \sum_{i=1}^{M} N_i \sigma_{XY}(i)/N$$

and $\sigma_X^2(i) = E(X - \mu_X^{(i)})^2$, $\sigma_Y^2(i) = E(Y - \mu_Y^{(i)})^2$, and $\sigma_{XY}(i) = E(X - \mu_X^{(i)})(Y - \mu_Y^{(i)})$. Hence the effect of grouping is mainly associated with the ratios of within group characteristics to the individual characteristics.

In practice errors of measurement in a survey are very common. There has been a considerable amount of work published in this area. A good source of references may be obtained in Fuller (1987). The purpose of this paper is to study the combined effects of errors of measurement and grouped ("ecological") data on the estimation of the ungrouped correlation in a sample survey design. A model incorporating a special type of measurement error, namely the interviewer effect, is considered in section 2. Section 3 derives the joint effects of measurement errors and grouping on the ecological correlation.

2. A Model of Measurement Errors

Following the same framework introduced in section 1, let (x_{ijk}, y_{ijk}) be the observed variables.

$$x_{ijk} = X_{ijk} + \epsilon_{ijk} \tag{4.a}$$

$$y_{ijk} = Y_{ijk} + \eta_{ijk} \tag{4.b}$$

$$i = 1, \ldots, M, j = 1, \ldots, m_i, \quad \text{and} \quad k = 1, \ldots l_{ij}$$

where X_{ijk} and Y_{ijk} are the true values of the variables to be measured, ϵ_{ijk} and η_{ijk} are measurement errors produced by the jth interviewer in the ith subpopulation pertaining to variables X_{ijk} and Y_{ijk} respectively, m_i is the number of interviewers employed in the ith subpopulation, and l_{ij} is the number of individuals interviewed by the jth interviewer in the ith subpopulation. We further assume that $E(\epsilon_{ijk}) = E(\eta_{ijk}) = 0$, $\sigma_\epsilon^2(ij) = E\epsilon_{ijk}^2$, $\sigma_\eta^2(ij) = E\eta_{ijk}^2$, and the errors are mutually uncorrelated except when they are measured by the same interviewer. (i.e. ϵ_{rst} and ϵ_{ijk}; η_{rst} and η_{ijk}; and ϵ_{rst} and η_{ijk} are mutually uncorrelated when $r \neq i$ or $s \neq j$.) For the errors made by the same interviewer, we have

$$\begin{aligned}\rho_\epsilon(ij) &= \sigma_\epsilon^2(ij)E(\epsilon_{ijr}\epsilon_{ijs}), \qquad r \neq s \\ \rho_\eta(ij) &= \sigma_\eta^2(ij)E(\eta_{ijr}\eta_{ijs}), \qquad r \neq s \\ \rho_{\epsilon\eta}(ij) &= \sigma_\eta(ij)\sigma_\epsilon(ij)E(\epsilon_{ijr}\eta_{ijs}).\end{aligned}$$

The probabilistic structure of measurement errors is formulated in a way to reflect a common situation that the errors made by the same interviewer tend to be correlated to each other.

3. Joint Effects of Measurement Errors and Grouping

Define the (individual) correlation between x and y and the ecological correlation between x and y with errors of measurement as

$$\tilde{\rho} = \frac{E\sum_{ijk}(x_{ijk}-\bar{x})(y_{ijk}-\bar{y})}{[E\sum_{ijk}(x_{ijk}-\bar{x})^2]^{1/2}[E\sum_{ijk}(y_{ijk}-\bar{y})^2]^{1/2}} \tag{5.a}$$

$$\tilde{\rho}_G = \frac{E\sum_{i=1}^{M} n_i(\bar{x}_i-\bar{x})(\bar{y}_i-\bar{y})}{[E\sum_{i=1}^{M} n_i(\bar{x}_i-\bar{x})^2]^{1/2}[E\sum_{i=1}^{M} n_i(\bar{y}_i-\bar{y})^2]^{1/2}}, \tag{5.b)}$$

where $n_i = \sum_{j=1}^{m_i} l_{ij}$. Using the standard technique of sum of squares decomposition, terms on the right hand side of equation (5.b) before taking the expectation may be expressed as the difference between total sum of squares and within subpopulation sum of squares.

$$\sum_{i=1}^{M} n_i(\bar{x}_i-\bar{x})^2 = [\sum_{ijk}(x_{ijk}-\bar{x})^2] - [\sum_{ijk}(x_{ijk}-\bar{x}_i)^2] \tag{6.a}$$

$$\sum_{i=1}^{M} n_i(\bar{y}_i-\bar{y})^2 = [\sum_{ijk}(y_{ijk}-\bar{y})^2] - [\sum_{ijk}(y_{ijk}-\bar{y}_i)^2] \tag{6.b}$$

$$\sum_{i=1}^{M} n_i(\bar{x}_i-\bar{x})(\bar{y}_i-\bar{y}) = [\sum_{ijk}(x_{ijk}-\bar{x})(y_{ijk}-\bar{y})] - [\sum_{ijk}(x_{ijk}-\bar{x}_i)(y_{ijk}-\bar{y}_i)] \tag{6.c}$$

Hence it is sufficient to evaluate the expectations of total sum of squares and the within subpopulation sum of squares. Based on the error structure defined in section 2, we summarize the results of taking expectation of sum of squares and cross products below. Detailed derivation of them is given in the appendices.

$$\begin{aligned} E\sum_{ijk}(x_{ijk}-\bar{x}_i)^2 = \sum_i\{[\sum_j(l_{ij}-1)(\sigma_X^2(ij) &+ \sigma_\epsilon^2(ij)(1-\rho_\epsilon(ij))] \\ &+ (m_i-1)\sigma_X^2(i) \\ &+ \sum_j(1-\frac{l_{ij}}{n_i})\sigma_\epsilon^2(ij) \\ &\times (1+(l_{ij}-1)\rho_\epsilon(ij))\} \end{aligned} \tag{7}$$

$$E\sum_{ijk}(x_{ijk}-\bar{x})^2 = (n-1)\{\sigma_X^2 + \sum_i\sum_j \frac{l_{ij}}{n}\sigma_\epsilon^2(ij) \times [1-\frac{(l_{ij}-1)}{(n-1)}\rho_\epsilon(ij)]\} \tag{8}$$

$$\begin{aligned} E\sum_{ijk}(x_{ijk}-\bar{x}_i)(y_{ijk}-\bar{y}_i) = \sum_i\{\sum_j[(l_{ij}-1)\sigma_{XY}(ij) &+ ((m_i-1)l_{ij}+n_i)\sigma_{\epsilon\eta}(ij))/n_i] \\ &+ (m_i-1)\sigma_{XY}(i)\} \end{aligned} \tag{9}$$

$$\begin{aligned} E\sum_{ijk}(x_{ijk}-\bar{x})(y_{ijk}-\bar{y}) = &(n-1)\sigma_{XY} \\ &+ \sum_i\sum_j l_{ij}\sigma_{\epsilon\eta}(ij) \\ &\times [1+(1-2n)l_{ij}/n^2] \end{aligned} \tag{10}$$

Consider two sets of conditions that simplify the above results and hence provide a clear picture on the impact of measurement errors and grouping.

Condition (I):
$\sigma_X^2(ij) = \sigma_X^2(i), \sigma_\epsilon^2(ij) = \sigma_\epsilon^2(i), \rho_\epsilon(ij) = \rho_\epsilon(i)$,
$\sigma_Y^2(ij) = \sigma_Y^2(i), \sigma_\eta^2(ij) = \sigma_\eta^2(i), \rho_\eta(ij) = \rho_\eta(i)$, and $l_{ij} = l_i$.

Condition (II):
$\sigma_X^2(ij) = \sigma_X^2(i), \sigma_\epsilon^2(ij) = \sigma_\epsilon^2, \rho_\epsilon(ij) = \rho_\epsilon$,
$\sigma_Y^2(ij) = \sigma_Y^2(i), \sigma_\eta^2(ij) = \sigma_\eta^2, \rho_\eta(ij) = \rho_\eta$
$m_i = m$ and $l_{ij} = l$.

Under condition (I), equations (7) to (10) may be expressed as:

$$\begin{aligned} E\sum_{ijk}(x_{ijk}-\bar{x}_i)^2 = \sum_i(n_i-1)\{\sigma_X^2(i) \\ + \sigma_\epsilon^2(i)[1-\frac{l_i-1}{n_i-1}\rho_\epsilon(i)]\}, \end{aligned} \tag{7*}$$

$$\begin{aligned} E\sum_{ijk}(x_{ijk}-\bar{x})^2 = (n-1)\{\sigma_X^2 + \sum_i\frac{n_i}{n}\sigma_\epsilon^2(i) \\ \times [1-\frac{(l_i-1)}{n-1}\rho_\epsilon(i)]\} \end{aligned} \tag{8*}$$

$$\begin{aligned} E\sum_{ijk}(x_{ijk}-\bar{x}_i)(y_{ijk}-\bar{y}_i) = \sum_{i=1}^{M}(n_i-1)\sigma_{XY}(i) \\ + 2(m_i-1)\sigma_{\epsilon\eta}(i), \end{aligned} \tag{9*}$$

$$\begin{aligned} E\sum_{ijk}(x_{ijk}-\bar{x})(y_{ijk}-\bar{y}) = (n-1)\sigma_{XY} + \sum_{i=1}^{M} n_i\sigma_{\epsilon\eta}(i) \\ \times (1+l_i(1-2n)/n^2), \end{aligned} \tag{10*}$$

where $n_i = m_i l_i$, $n = \sum_{i=1}^{M} m_i$. Under condition (II), the above results may be further simplified.

$$\begin{aligned} E\sum_{ijk}(x_{ijk}-\bar{x}_i)^2 = (ml-1)\sum_i\sigma_X^2(i) \\ + M\sigma_\epsilon^2(1-\frac{l-1}{ml-1}\rho_\epsilon) \end{aligned} \tag{7**}$$

$$\begin{aligned} E\sum_{ijk}(x_{ijk}-\bar{x})^2 = (n-1)[\sigma_X^2 \\ + \sigma_\epsilon^2(1-\frac{l-1}{ml-1}\rho_\epsilon)] \end{aligned} \tag{8**}$$

$$E\sum_{ijk}(x_{ijk}-\bar{x}_i)(y_{ijk}-\bar{y}_i) = (ml-1)[\sum_i \sigma_{XY}(i)] + 2M(m-1)\sigma_{\epsilon\eta} \quad (9^{**})$$

$$E\sum_{ijk}(x_{ijk}-\bar{x})(y_{ijk}-\bar{y}) = (n-1)\sigma_{XY} + n\sigma_{\epsilon\eta}(1+l(1-2n)/n^2) \quad (10^{**})$$

Effect of Measurement Errors on ρ

From equations (7) and (9),

$$\begin{aligned}\tilde{\rho} &= \frac{\sigma_{XY}+A}{(\sigma_X^2+B)^{1/2}(\sigma_Y^2+C)^{1/2}} \\ &= \frac{\sigma_{XY}}{\sigma_X\sigma_Y}\frac{(1+A/\sigma_{XY})}{(1+B/\sigma_X^2)^{1/2}(1+C/\sigma_Y^2)^{1/2}} \\ &= \rho\frac{(1+A/\sigma_{XY})}{(1+B/\sigma_X^2)^{1/2}(1+C/\sigma_Y^2)^{1/2}}\end{aligned} \quad (11)$$

where $A/\sigma_{XY} = [(n^2-2nl+l)/(n^2-n)](\sigma_{\epsilon\eta}/\sigma_{XY})$, $B/\sigma_X^2 = [1-(l-1)/(n-1)\rho_\epsilon](\sigma_\epsilon^2/\sigma_X^2)$, $C/\sigma_Y^2 = [1-(l-1)/(n-1)\rho_\eta](\sigma_\eta^2/\sigma_Y^2)$ under condition (II). The terms A, B and C may be interpreted as the effects of measurement errors. The measurement errors make $\tilde{\rho}$ bias toward zero. One interesting implication is that when the errors associated with the same interviewer are positively correlated, ($\sigma_{\epsilon\eta} > 0$, $\rho_\epsilon > 0$, and $\rho_\eta > 0$), the extent of bias toward zero may be reduced. However the interviewer effect is multiply by a factor of l/n where l is the number of subjects interviewed by a surveyor and n is the total sample size of the whole survey.

Effect of Measurement Errors and Grouping on ρ_G

Similarly $\tilde{\rho}_G$ may be expressed as:

$$\begin{aligned}\tilde{\rho}_G &= \frac{\sigma_{XY}+A'+D}{(\sigma_X^2+B'+E)^{1/2}(\sigma_Y^2+C'+F)^{1/2}} \\ &= \frac{\sigma_{XY}}{\sigma_X\sigma_Y}\frac{1+\frac{A'+D}{\sigma_{XY}}}{(1+\frac{B'+E}{\sigma_X^2})^{1/2}(1+\frac{C'+F}{\sigma_Y^2})^{1/2}} \\ &= \rho\varepsilon_1\varepsilon_2\varepsilon_3\end{aligned} \quad (12)$$

where

$$\begin{aligned}\varepsilon_1 &= \frac{1+\frac{A'}{\sigma_{XY}}}{(1+\frac{B'}{\sigma_X^2})^{1/2}(1+\frac{C'}{\sigma_Y^2})^{1/2}}, \\ \varepsilon_2 &= \frac{1+\frac{D}{\sigma_{XY}}}{(1+\frac{E}{\sigma_X^2})^{1/2}(1+\frac{F}{\sigma_Y^2})^{1/2}}, \\ \varepsilon_3 &= \{1-\frac{A'D}{\sigma_{XY}^2}(1+\frac{A'}{\sigma_{XY}})^{-1}(1+\frac{D}{\sigma_{XY}})^{-1}\} \\ &\times\{1-\frac{B'E}{\sigma_X^4}(1+\frac{B'}{\sigma_X^2})^{-1}(1+\frac{E}{\sigma_X^2})^{-1}\}^{-1/2} \\ &\times\{1-\frac{C'F}{\sigma_Y^4}(1+\frac{C'}{\sigma_Y^2})^{-1/2}(1+\frac{F}{\sigma_Y^2})^{-1/2}\}^{-1/2}.\end{aligned}$$

Under Condition (II),

$$\begin{aligned}A' &= \frac{n^2-2n(l+M(m-1))+l}{n(n-1)}\sigma_{\epsilon\eta} \\ B' &= \sigma_\epsilon^2[\frac{M-(n-1)}{n-1}+(M-1)(l-1)/(n-1)\rho_\epsilon] \\ C' &= \sigma_\eta^2[\frac{M-(n-1)}{n-1}+(M-1)(l-1)/(n-1)\rho_\eta] \\ D &= (1-ml)/(n-1)\sum_i\sigma_{XY}(i) \\ E &= (1-ml)/(n-1)\sum_i\sigma_X^2(i) \\ F &= (1-ml)/(n-1)\sum_i\sigma_Y^2(i)\end{aligned}$$

Hence ε_1, ε_2, and ε_3 represent the effects due to measurement errors, grouping and their interactions respectively.

An empirical issue is to estimate these three bias factors and then take them into account when the ecological correlation is computed based on the variables with measurement errors. On the other hand, it is also desirable to investigate the impact of measurement errors on the variance of the estimated $\tilde{\rho}$ and $\tilde{\rho}_G$.

References

Chai, John J., (1971), "Correlated Measurement Errors and the Least Squares Estimator of the Regression Coefficient," *Journal of the American Statistical Association*, 66

Chai, John J., (1971), " Errors of Measurement and 'Ecological' Correlation," American Statistical Association, Proceedings of the Social Statistics Section.

Cochran, William G., (1977) Sampling Techniques, third edition, New York: John Wiley & Sons, Inc.

Fuller, Wayne A. (1987) Measurement Errors Models, New York: John Wiley & Sons, Inc.

Appendix A

Derivation of equation (3)

The variance of each variable and the covariance between X and Y may be expressed as the sum of two components associated with the within subpopulation and the between subpopulations variations respectively.

$$\begin{aligned}\sigma_X^2 &= \sum_{i=1}^{M}\sum_{j=1}^{N_i}(X_{ij}-\mu_X)^2/N \\ &= \sum_{i=1}^{M}\sum_{j=1}^{N_i}(X_{ij}-\mu_X^{(i)})^2/N \\ &+ \sum_{i=1}^{M}N_i(\mu_X^{(i)}-\mu_X)^2 \\ &= \ddot{\sigma}_X^2+\bar{\sigma}_X^2\end{aligned} \quad (A.1)$$

$$\begin{aligned}Cov(X,Y) &= \sum_{i=1}^{M}\sum_{j=1}^{N_i}(X_{ij}-\mu_X)(Y_{ij}-\mu_Y)/N \\ &= \sum_{i=1}^{M}\sum_{j=1}^{N_i}(X_{ij}-\mu_X^{(i)})(Y_{ij}-\mu_Y^{(i)})/N\end{aligned}$$

$$+\sum_{i=1}^{M} N_i(\mu_X^{(i)}-\mu_X)(\mu_Y^{(i)}-\mu_Y)/N$$
$$=\ddot{\sigma}_{XY}+\bar{\sigma}_{XY} \qquad (A.2)$$

Hence from (1) and (2), we have

$$\rho=\frac{\ddot{\sigma}_{XY}+\bar{\sigma}_{XY}}{(\ddot{\sigma}_X^2+\bar{\sigma}_X^2)^{1/2}(\ddot{\sigma}_Y^2+\bar{\sigma}_Y^2)^{1/2}}$$
$$=\rho_G\frac{(1+\ddot{\sigma}_{XY}/\bar{\sigma}_{XY})}{(1+\ddot{\sigma}_X^2/\bar{\sigma}_X^2)^{1/2}(1+\ddot{\sigma}_Y^2/\bar{\sigma}_Y^2)^{1/2}} \qquad (A.3)$$

Equivalently

$$\rho_G=(1+\ddot{\sigma}_{XY}/\bar{\sigma}_{XY})^{-1}(1+\ddot{\sigma}_X^2/\bar{\sigma}_X^2)^{1/2}(1+\ddot{\sigma}_Y^2/\bar{\sigma}_Y^2)^{1/2}\rho$$
$$=(\frac{\bar{\sigma}_X^2+\ddot{\sigma}_X^2}{\bar{\sigma}_X^2})^{1/2}(\frac{\bar{\sigma}_Y^2+\ddot{\sigma}_Y^2}{\bar{\sigma}_Y^2})^{1/2}(\frac{\bar{\sigma}_{XY}+\ddot{\sigma}_{XY}}{\bar{\sigma}_{XY}})^{-1}\rho$$
$$=(1-\ddot{\sigma}_X^2/\sigma_X^2)^{-1/2}(1-\ddot{\sigma}_Y^2/\sigma_Y^2)^{-1/2}$$
$$(1-\ddot{\sigma}_{XY}/\sigma_{XY})\rho \qquad (A.4)$$

Appendix B

Derivation of equation (7)

First we substitute x with $X+\epsilon$. For fixed i,

$$\sum_{j=1}^{m_i}\sum_{k=1}^{l_{ij}}(x_{ijk}-\bar{x}_i)^2=\sum_{j=1}^{m_i}\sum_{k=1}^{l_{ij}}(X_{ijk}-\bar{X}_i)^2$$
$$+(\epsilon_{ijk}-\bar{\epsilon}_i)^2+2(X_{ijk}-\bar{X}_i)$$
$$\times(\epsilon_{ijk}-\bar{\epsilon}_i) \qquad (B.1)$$

The first two terms on the right hand side of (B.1) may be further decomposed into within interviewer sum of squares and between interviewer sum of squares.

$$\sum_{j=1}^{m_i}\sum_{k=1}^{l_{ij}}(X_{ijk}-\bar{X}_i)^2=\underbrace{\sum_{j=1}^{m_i}\sum_{k=1}^{l_{ij}}(X_{ijk}-\bar{X}_{ij})^2}_{(B.2.1)}+\underbrace{\sum_{j=1}^{m_i}l_{ij}(\bar{X}_{ij}-\bar{X}_i)^2}_{(B.2.2)} \qquad (B.2)$$

$$\sum_{j=1}^{m_i}\sum_{k=1}^{l_{ij}}(\epsilon_{ijk}-\bar{\epsilon}_i)^2=\underbrace{\sum_{j=1}^{m_i}\sum_{k=1}^{l_{ij}}(\epsilon_{ijk}-\bar{\epsilon}_{ij})^2}_{(B.3.1)}+\underbrace{\sum_{j=1}^{m_i}l_{ij}(\bar{\epsilon}_{ij}-\bar{\epsilon}_i)^2}_{(B.3.2)} \qquad (B.3)$$

It is not hard to obtain the expectations of (B.2.1) and (B.2.2).

$$E((B.2.1))=\sum_{j=1}^{m_i}(l_{ij}-1)\sigma_X^2(ij) \qquad (B.4)$$
$$E((B.2.2))=(m_i-1)\sigma_X^2(i) \qquad (B.5)$$

The derivation of (B.5) employs the following identity.

$$(\bar{X}_{ij}-\bar{X}_i)^2=[\sum_k(X_{ijk}-\mu_X^{(i)})/l_{ij}$$
$$-\sum_j\sum_k(X_{ijk}-\mu_X^{(i)})/n_i]^2$$
$$=[(\frac{1}{l_{ij}}-\frac{1}{n_i})\sum_k(X_{ijk}-\mu_X^{(i)})$$
$$-\frac{1}{n_i}\sum_{r\neq j}\sum_k(X_{irk}-\mu_X^{(i)})]^2$$

To compute the expectation of (B.3.1), we use the fact that

$$E(\epsilon_{ijk}^2)=\sigma_\epsilon^2(ij)$$

and

$$E(\bar{\epsilon}_{ij}^2)=\sigma_\epsilon^2(ij)[1+(l_{ij}-1)\rho_\epsilon(ij)]/l_{ij}$$

where $\rho_\epsilon(ij)=E_{r\neq s}(\epsilon_{ijr}\epsilon_{ijs})/\sigma_\epsilon^2(ij)$ represents the interviewer effect. Hence

$$E((B.3.1))=E[\sum_{j=1}^{m_i}\sum_{k=1}^{l_{ij}}\epsilon_{ijk}^2-\sum_{j=1}^{m_i}l_{ij}\bar{\epsilon}_{ij}^2]$$
$$=\sum_{j=1}^{m_i}(l_{ij}-1)\sigma_\epsilon^2(ij)(1-\rho_\epsilon(ij)) \qquad (B.6)$$

To compute the expectation of (B.3.2), we need

$$\bar{\epsilon}_{ij}-\bar{\epsilon}_i=\bar{\epsilon}_{ij}-\sum_{r=1}^{m_i}l_{ir}\bar{\epsilon}_{ir}/n_i$$
$$=[(n_i-l_{ij})\bar{\epsilon}_{ij}-\sum_{r\neq j}l_{ir}\bar{\epsilon}_{ir}/n_i]$$

Since the errors associated with different interviewers are assumed uncorrelated, we have

$$E(l_{ij}(\bar{\epsilon}_{ij}-\bar{\epsilon}_i)^2)=[(n_i-l_{ij})^2E(l_{ij}\bar{\epsilon}_{ij}^2)$$
$$+l_{ij}\sum_{r\neq j}l_{ir}E(l_{ir}\bar{\epsilon}_{ir}^2)]/n_i^2$$
$$=n_i^{-2}[(n_i^2-2n_il_{ij})\sigma_\epsilon^2(ij)$$
$$\times(1+(l_{ij}-1)\rho_\epsilon(ij))$$
$$+l_{ij}\sum_{r=1}^{m_i}l_{ir}\sigma_\epsilon^2(ir)(1+(l_{ir}-1)\rho_\epsilon(ir))].$$

Then

$$E((B.3.2))=\sum_{j=1}^{m_i}(1-l_{ij}/n_i)\sigma_\epsilon^2(ij)$$
$$\times(1+(l_{ij}-1)\rho_\epsilon(ij)) \qquad (B.7)$$

Combining the results of (B.4) to (B.7), the equation (7) follows.

Appendix C

Derivation of Equation (8)

$$\sum_{ijk}(x_{ijk}-\bar{x})^2 = \sum_{ijk}(X_{ijk}-\bar{X})^2 + (\epsilon_{ijk}-\bar{\epsilon})^2 + 2(X_{ijk}-\bar{X})(\epsilon_{ijk}-\bar{\epsilon}) \qquad (C.1)$$

The expectation of the first term is readily available.

$$E\sum_{ijk}(X_{ijk}-\bar{X})^2 = (n-1)\sigma_X^2 \qquad (C.2)$$

The sum of squares of the second term may be expressed as

$$\sum_{ijk}(\epsilon_{ijk}-\bar{\epsilon})^2 = \sum_{ijk}\epsilon_{ijk}^2 - n\bar{\epsilon}^2.$$

By definition,

$$\begin{aligned}\bar{\epsilon}^2 =& \{\sum_{ijk}\epsilon_{ijk}\}^2/n^2 \\ =& \sum_{ijk}\epsilon_{ijk}^2/n^2 + n^{-2}\{\sum_i\sum_j\sum_{r\neq s}\epsilon_{ijr}\epsilon_{ijs} \\ &+ \sum_i\sum_{u\neq v}\sum_r\sum_s\epsilon_{iur}\epsilon_{ivs} \\ &+ \sum_{p\neq q}\sum_u\sum_v\sum_r\sum_s\epsilon_{pur}\epsilon_{qvs}\},\end{aligned}$$

$\Rightarrow$

$$E(\bar{\epsilon}^2) = n^{-2}\sum_i\sum_j\sigma_\epsilon^2(ij)l_{ij} \times (1+(l_{ij}-1)\rho_\epsilon(ij)) \qquad (C.3)$$

Then equation (8) folllows.