REGRESSION COEFFICIENT OBTAINED WITH INFORMATION ON VARIABLE TYPE, SAMPLING, WEIGHTING, AND CLASSIFICATION

Jai W. Choi, National Center for Health Statistics, 6525 Belcrest Rd. Hyattsville, MD 20782

Key words: Estimating Equation; sampling design; weighted and classified data; covariance, quasi-likelihood; Fishers Scoring.

SUMMARY

Estimating regression coefficients requires var(y). One often assumes normal, binomial, or Poisson distribution for y to estimate its moments. Such a simple assumption on distribution is not correct for a complex data y such as the sample deaths obtained by Vital Statistics Division at National Center for Health Statistics, and therefore the variance should be adjusted, including the error arising not only from random death, but also from other sources such as sampling, classification, and weighting. This paper presents the variance of y obtained according to four sources: death, sampling, classification, and weighting, and obtain the regression coefficients using this result. The estimator of regression coefficients is approximately unbiased and has a special form of limiting variance.

1. INTRODUCTION

Dichotomous data set, consisted of an outcome variable, y_{ijh}, and a p x 1 vector of covariate, x_{ijh}, observed in the ith month, $i = 1, \ldots, M$, for the jth age-sex group, $j = 1, \ldots, J_i$, and for the hth cause or category, $h = 1, \ldots, r_{ij}$, arise often in the death rate analysis. To simplify the notation, we set $J_i = J$ and $r_{ij} = r$ without loss of generality.

Typically, one may be interested in the pattern of rate changes in the dependence of outcome on the covariates. For instance, death rates of outcome variables y's might depend on such covariates x's as age, sex and race for certain causes of deaths.

In a generalized linear model, the independent observations on each independent subject can be linked to a variety of continuous or discrete outcome variables through a typical assumption of distribution. However, when dependent observation y's are correlated, Liang and Zeger (1986) used a model corrected for time dependence in a longitudinal data. Thall and Vail (1990) presented a covariance model for overdispersed and correlated y's. This research further extends these results to a covariance model for complex data that are sampled, classified, and weighted. We apply resulting covariance to regression model.

1.1 POISSON DEATHS

The occurrence of death may be distributed as Poisson or Binomial, and both distributions are approximated by normal distribution for a large number of occurrences. We assume that D_i deaths occur among P_i people according to Poisson distribution.

1.2 SAMPLE DESIGN

A sample of d_i deaths is selected from all D_i deaths in the i-th month according to a certain sample design. Denote design variable by z_{ik}. It may be a simple random sample, or stratified simple random sample, or cluster sample, with or without replacement, or any other design used to take a sample from a well defined population D_i.

1.3 POSTCLASSIFICATION

After a sample is taken, d_i sample deaths are postclassified into J age-sex groups and r causes, giving d_{ijh} deaths for the jth age-sex group, and died by cause h. Denote poststratification variable by z_{ijhk}.

1.4 WEIGHTING

Sample data are often weighted to estimate population parameters. Weighting takes place in various forms, and we denote the weight by w. The weighted death rate, y_{ijh}, is the poststratified yearly death d_{ijh} multiplied by appropriate weight w_{ij}, that is

$$y_{ijh} = (365/a_i)(D_i/d_i)\,(d_{ijh}/P_i) = w_{ij}\, d_{ijh}$$

where $w_{ij} = (365\, D_i)/(a_i\, d_i\, P_i)$, which is assumed to be known at this time. $a_i = 28$ for February, 30 or 31 for the rest of the months, D_i/d_i is the sampling weight, and P_i is the population in the month i.

1.5 OBJECTS

We assume that occurrences of deaths in each month are independent. Let $y_i = (y_{i11}, \ldots, y_{ijh}, \ldots, y_{iJr})'$ be 1 x T vector (T = Jr) of the weighted death rates. The primary purpose of this paper is to present a tractable parametric form of the $\text{cov}(y_i)$ or V_i, that accounts for four impacts: occurrence of deaths, sampling, classification, and weight.

1.6 CORRELATION

We can consider dependence among the ementary units in the same age-sex group for the persons of same age-sex group may likely have common cause of death for some typical diseases.

We assume a model for the correlation between the members within a age-sex category. For the k-th and k'-th units in the j-th age-sex group as

$$(3) \quad E(z_{ijhk} z_{ijh'k'}) = \begin{cases} \theta\pi_{ijh} + (1-\theta)\pi^2_{ijh} \\ (1-\theta)\pi_{ijh}\pi_{ijh'} \end{cases}$$

where θ ($0 \le \theta \le 1$) is the correlation between these two members, and this model has been used in other research (see Choi and McHugh, 1989). The first part is the probability that both members fall in the same cell (h=h'), and the second part is the probability that two members falls into two different cells ($h \neq h'$).

In Section 2, we obtain the covariance $V(y_i)$ when multiplicative stages of errors are assumed. In Section 3, we present the estimation of the parameter β with covariance derived in Section 2.

2. MULTIPLICATIVE ERROR MODEL

2.1. ASSUMPTIONS

Consider three stages labelled 2, 1, and 0. The bottom stage 0 is the occurrence of death (δ_{ik} = 1 if the k-th person died among the P_i people in the month i, and = 0 otherwise ($1 \le k \le P_i$)). We assume that δ_{ik}'s are independent with Poisson mean $\mu_i = D_i/P_i$.

The stage 1 is sampling (z_{ik} = 1 if the k-th person who died is sampled and = 0 otherwise) and has error s_{ik}. A sample of d_i is obtained from D_i.

The top stage 2 is the classification (z_{ijhk} = 1 if the k-th person who died and sampled, is classified into the j-th age-sex and cause h, and = 0 otherwise) and has error c_{ijhk}. A sampled d_i is classified into d_{ijh}'s.

Let the variable $y_{ijh} = \sum_{k=1}^{P_i} y_{ijhk}$, where $y_{ijhk} = w_{ij}\, d_{ijhk}$ and $d_{ijhk} = \delta_{ik}\, z_{ik}\, z_{ijhk}$. $d_{ijh} = \sum_{k=1}^{P_i} d_{ijhk}$. We consider the weights w_{ij}'s are fixed numbers, and d_{ijhk}'s are the variables ($1 \le k \le P_i$).

We adopt the symbols E, V, and C for expectation, variance, and covariance operator respectively.

Conditioning on the s_{ik} and c_{ijhk}, we assume an error model at the stage 0 for the data d:

$$(4) \quad E(d_{ijhk} / s_{ik}\, c_{ijhk}) = \mu_i\, s_{ik}\, c_{ijhk}$$
$$V(d_{ijhk} / s_{ik}\, c_{ijhk}) = \mu_i\, s_{ik}\, c_{ijhk}.$$

We also assume that d_{ijhk}'s are conditionally uncorrelated, and that the s_{ik}'s and c_{ijhk}'s are mutually independent with

$$(5) \quad E(c_{ijhk}) = \pi_{ijh} = D_{ijh}/D_i, \quad E(s_{ik}) = d_i/D_i,$$
$$V(c_{ijhk}) = \sigma_h^2, \qquad V(s_{ik}) = S_i^2.$$

and the marginal $E(d_{ijh}) = d_i\, D_{ijh}/D_i$.

Thall and Vail (1990) and Morton (1987) used somewhat similar Poisson model for overdispersed data. Chiang (1967) obtained the variance based on non-multiplicative model.

If the classification errors at the stage 2 behave according to adjusted multinomial with model (3), we may write its variance and covariance:

$$(6) \quad V(c_{ijhk}) = \sigma_h^2 = \pi_{ijh}(1-\pi_{ijh}),$$
$$C(c_{ijhk}\, c_{ijhk'}) = \sigma_{hh} = \theta\pi_{ijh}(1-\pi_{ijh}) \text{ for } h=h' \text{ and } k \neq k',$$
$$C(c_{ijhk}\, c_{ijh'k}) = \sigma_{hh'} = -\pi_{ijh}\, \pi_{ijh'} \text{ for } h \neq h' \text{ and } k=k',$$
$$C(c_{ijhk}\, c_{ijh'k'}) = \sigma_{hh''} = -\theta\pi_{ijh}\, \pi_{ijh'} \text{ for } h \neq h' \text{ and } k \neq k',$$

If the sample is taken by simple random sample without replacement, then we can write the variance and covariance of error at stage 1:

$$(7) \quad V(s_{ik}) = S_i^2 = (D_i/(D_i-1))(d_i/D_i)(1-(d_i/D_i))$$
$$C(s_{ik}\, s_{ik'}) = S_{ii} = -\, d_i(D_i - d_i)/\, (D_i^2\, (D_i-1)).$$

However, the form of these variance and covariance depends on the sampling design actually used and type of variable.

2.2 A VARIANCE-COVARIANCE MATRIX V_i

From the above assumptions, we present a variance-covariance matrix V_i or $C(y_i)$. The V_i is a block diagonal matrix with J (r x r) submatrices V_{i1}, ... , V_{ij}, ..., V_{iJ} on the main diagonal, and zero for all elements for $j \neq j'$. The inverse of V_i is obtained by inverting each submatrix V_{ij}. Following the conditional variance and covariance (Appendix 1):

$$(8) \quad V(y_{ijhk}) = w_{ij}^2[\mu_i(d_i/D_i)\pi_{ijh} + \mu_i^2\{S_i^2(\sigma_h^2+\pi_{ijh}^2)+(d_i/D_i)^2\sigma_h^2\}],$$

$$(9) \quad C(y_{ijhk,}\, y_{ijh'k}) = w_{ij}^2(d_i/D_i)^2\mu_i^2\sigma_{hh'}$$
$$= -\, w_{ij}^2\, (\mu_i\, d_i/D_i)^2\, \pi_{ijh}\, \pi_{ijh'}.$$

$$(10) \quad C(y_{ijhk}\, y_{ijh'k'}) = w_{ij}^2\, \mu_i^2\, [C(c_{ijhk}\, c_{ijh'k'})$$
$$\{C(s_{ik}\, s_{ik'})+(d_i/D_i)^2\} + C(s_{ik}\, s_{ik'})\, \pi_{ijh}\, \pi_{ijh'}],$$
$$= w_{ij}^2\, \mu_i^2\, [\sigma_{hh''}S_{ii} + \sigma_{hh''}(d_i/D_i)^2 + S_{ii}\, \pi_{ijh}\, \pi_{ijh'}],$$

$$(11) \quad C(y_{ijhk}\, y_{ijhk'}) = w_{ij}^2\, \mu_i^2\, \{\, C(s_{ik}\, s_{ik'})E(c_{ijhk}\, c_{ijhk'}) + (d_i/D_i)^2\, C(c_{ijhk}\, c_{ijhk'})\},$$
$$= w_{ij}^2\mu_i^2\{S_{ii}(\sigma_{hh} + \pi_{ijh}^2) + (d_i/D_i)^2\, \sigma_{hh}\, \}.$$

Summing over the subscript k ($1 \le k \le P_i$) gives variance-covariance matrix V_{ij} with $V(y_{ijh})$ on the diagonal and $C(y_{ijh}\, y_{ijh'})$ on the off-diagonal:

$$(12) \quad V(y_{ijh}) = P_iV(y_{ijhk}) + P_i(P_i-1)C(y_{ijhk}\, y_{ijhk'})$$
$$= P_iw_{ij}^2[\mu_i(d_i/D_i)\pi_{ijh} + \mu_i^2\{s_i^2(\sigma_h^2+\pi_{ijh}^2)+(d_i/D_i)^2\sigma_h^2\}]$$
$$+ P_i(P_i-1)w_{ij}^2[\mu_i^2\{S_{ii}(\sigma_{hh}+\pi_{ijh}^2)+ (d_i/D_i)^2\, \sigma_{hh}\, \}]$$

(13) $C(y_{ijh}\ y_{ijh'}) = P_i C(y_{ijhk}\ y_{ijh'k}) + P_i(P_i-1)C(y_{ijhk}\ y_{ijh'k'})$

$= P_i\ w_{ij}^2 \mu_i^2 (d_i/D_i)^2 \sigma_{hh'}$

$+ P_i(P_i-1) w_{ij}^2 [\mu_i^2 \{S_{ii}(\sigma hh'' + \pi_{ijh}\pi_{ijh'}) + (d_i/D_i)^2\ \sigma_{hh''}\}]$.

Using the notations (6) for σ_h^2, σ_{hh}, $\sigma_{hh'}$, and $\sigma_{hh''}$ and (7) for s_i^2 and S_{ii}, and defining

$\alpha_{ij} = P_i w_{ij}^2 [\mu_i (d_i/D_i)\{1 + \mu_i d_i/D_i\} + \mu_i^2 S_i^2]$,

$+ P_i(P_i-1) w_{ij}^2 \mu_{i2} \theta (S_{ii} + (d_i/D_i)^2)$,

$\beta_{ij} = P_i\ w_{ij}^2\ \mu_i^2 (d_i/D_i)^2 + P_i(P_i-1) w_{ij}^2 \mu_{i2} [(d_i/D_i)^2 \theta - S_{ii}(1-\theta)]$,

we can express

(14) $$V_{ij} = D + \beta_{ij}\ U\ U'$$

where D is $diag(\alpha_{ij}\pi_{ijh})$ and $U' = (\pi_{ij1}, ..., \pi_{ijr})$.

Denoting $u_h = \pi_{ijh}^2/(\alpha_{ij}\ \pi_{ijh})$ and $u_+ = u_1 + ... + u_r$, V_{ij} being nonsingular, Householder's formula yields

(15) $$V_{ij}^{-1} = D^{-1} - (\beta_{ij}/(1 + \beta_{ij}\ u_+)\ D^{-1}\ UU'\ D^{-1}.$$

Note that V_{ij} becomes the variance-covariance matrix of multinomial variables if $\alpha_{ij} = 1$ and $\beta_{ij} = -1$. α_{ij} and β_{ij} reflect the impacts of death process, sampling and weighting.

Considering only sampling and classification variable for given death records, we can simplify the multiplier $\alpha_{ij} = -\beta_{ij} = w_{ij}^2\ d_{ij}\ [1 + \theta(d_{ij} - 1)]$ which reflects impacts of weighting, sampling, and correlation of members on the variance. Further if $\theta = 0$, or no correlation among members, $\alpha_{ij} = -\beta_{ij} = d_{ij}\ w_{ij}^2$. The V_{ij} can also be obtained by assuming that the data are not weighted by setting $w_{ij} = 1$. Other sample design can be easily reflected on the variance, using the conditional argument.

Therefore it is essential to use correct variance in the estimation of regression coefficients. When we consider each y_{ijh} as function of ratio, d_{ijh}/d_i, it is somewhat more involved as both d_i and d_{ijh} are random variables.

3. ESTIMATION EQUATIONS

3.1 NOTATIONS

Let $y_i = (y_{i11}, ..., y_{ijh}, ..., y_{iT})'$ (T = Jr) of the weighted death rates with mean $m_i = (m_{i11}, ..., m_{ijh}, ..., m_{iT})'$, and $X_i = (x_{i11}, ..., x_{ijh}, ..., x_{iT})'$ be the p x T matrix of covariates, each $x_{ijh} = (x_{ijh1}, ..., x_{ijht}, ..., x_{ijhp})$.

We require that some link function g, linking each mean $m_{ijh} = E(y_{ijh})$ to p x 1 parameter vector $\beta = (\beta_0, ..., \beta_t, ..., \beta_{p-1})'$: $g(m_{ijh}) = \eta_{ijh} = x_{ijh}\ \beta$, where η_{ijh} is a correct linear function. Denote predictor $\eta_i = (\eta_{i11}, ..., \eta_{ijh}, ..., \eta_{iT})'$.

3.2 ESTIMATION

Often there is no probability distribution available on the variables, especially complex survey data; however, under mild assumptions, quasi-likelihood function has similar properties as those of ordinary log likelihood. We assume that $y_1, ..., y_I$, are independent and (14) is correct variance.

Let $S_i = y_i - m_i$ with $E(S_i) = 0$, and V_i be the T x T matrix of $V(y_i)$. $X_i = \partial\eta_{ij}/\partial\beta = (Xi_{11}, ..., X_{iT})$ is p x T matrix, and the $\Delta_i = diag(\partial m_{ij}/\partial\eta_{ij}) = diag(\Delta_{ij})$ is T x T matrix. $P_i = \partial m_{ij}/\partial\beta = \Delta_i\ X_i$.

The entire set of data is expressed as $y = (y_1, ..., y_i, ..., y_M)'$ is a 1 x A vector (A = MJr), and $X = (X_1, ..., X_i, ..., X_M)'$ is p x A matrix of covariates, S is A x 1 vector of (y - m), Δ is A x A matrix of $diag(\Delta_i)$, $P = \Delta X$. V is A x A matrix of $diag(V_i)$.

Following the quasi likelihood (McCullagh and Nelder, 1983), the p estimating score equations for regression parameters β are given as

(16) $$U(\beta) = \sum_{i=1}^{M} P_i^T\ V_i^{-1}\ S_i = 0$$

β is defined to be the solution of equation (16). We obtain the estimator $\hat{\beta}$ of β with sample data. Approximate unbiasedness and asymptotic Normality of $\hat{\beta}$ may be stated as follow:

Define $H = M^{1/2}(\hat{\beta} - \beta)$. H can be approximated by $\Gamma^{-1}\ \Lambda$ where $\Lambda = M^{-1/2}\ U(\hat{\beta})$ and $\Gamma = -\{\partial U(\beta)/\partial\beta\}/M$ from Taylor expansion of $U(\hat{\beta})$ around true β. Under certain regularity conditions (Thall and Vail, 1990), H is asymptotically Normal mean 0 and variance

(17) $$C(H) = \Gamma^{-1}\ cov(\Lambda)\ \Gamma^{-1},$$

where $cov(\Lambda) = M^{-1}\ \Sigma_i(P_i^T\ V_i^{-1} cov(y_i) V_i^{-1}\ P_i)$ and $\Gamma = M^{-1}\ \Sigma_i(P_i^T\ V_i^{-1}\ P_i)$.

Thall and Vail (1990) use overdispersed Poisson model, while current V uses the error model for complex data. If we assume $cov(y_i) = V_i$, (17) takes a simple form $C(H) = \Gamma^{-1}$. When a link function is specified, we can obtain the explicit form of V_i.

Variance of $\hat{\beta}$ may be correctly estimated by replacing $cov(y_i)$ with $S^T_i S_i$. $S^T_i S_i$ may be more efficient then $\hat{V}_i$ when the model used for the derivation of V_i is not correct.

3.3 ITERATIVE METHOD

We begin with $\hat{\beta}^0$ substantially close to $\hat{\beta}$ in (18). The sequence of parameter estimates generated by Newton-Raphson method with Fishers

Scoring is

$$(18)\quad \beta^1=\beta^0+(\hat{P}^T\hat{V}^{-1}\hat{P})^{-1}(\hat{P}^T\ \hat{V}^{-1}\hat{S})$$

The estimate β may be obtained by iterating until it converges. The convergence criterion is to stop the iteration at (r + 1) step when $MAX|(\beta^{r+1} - \beta^r)/\beta^r| \leq 10^{-5}$, .

Provided that the eigenvalues of $\hat{P}^T\hat{V}^{-1}\hat{p}$ are sufficiently large, the second term of (18) is negligible. Then, we may take the first round approximation $\beta^1=\beta$.

When V_i is set equal to usual multinomial form and $w_{ij} = 1$, existing GLIM software provides the estimates of the parameter β.

4. COMMENTS

More study has to be done to complete this research. First the equation (17) has to be proved in a form of theorem. Secondly we need to illustrate this result with NCHS data as we have originally planned.

The variance models used for the stage 1 and 2 can be replaced if necessary. Especially stage 1 of sampling has to be changed when other sampling design is used. The adjusted multinomial model for stage 2 also needs to be changed if the classification rule is different.

The regression coefficient is properly estimated only through the correct form of covariance matrix. Although data analysis at NCHS assumed a simple distribution, this result can be used to reflect the distribution of NCHS data better for statistical inference.

REFERENCES

JAI WON CHOI AND RICHARD B. MCHUGH (1989). A Reduction Factor in Goodness of Fit and Independence Tests for Clustered and Weighted Observations. Biometrics 45, 979-996.

C. L. CHIANG (1967). Variance and Covariance of Life table functions estimated from a sample of deaths. National Center for Health Statistics, Series 2, Number 20.

KUNG-YEE LIANG AND SCOTT L. ZEGER (1986). Longitudinal data analysis using generalized linear models, Biometrika 73. 1. 13-22.

RICHARD MORTON (1987). A generalized linear model with nested strata of extra-Poisson variation. Biometrika, 74, 2, 247-57.

PETER F. THALL AND STEPHEN C. VAIL (1990). Some Covariance Models for Longitudinal Count Data with Overdispersion. Biometrics 46. 657-671.

PETER MCCULLAGH AND A.J. NELDER (1983). Generalized Linear Models. London: Chapman and Hall.

J.N.K RAO AND A.J.SCOTT (1981). The Analysis of Categorical Data from Complex Sample Surveys: Chi-squared tests for Goodness of fit and Independence in Two Way Tables. JASA 76, 221-230.

Author is thankful to Dr. Randy Curtin for his help in this research.

APPENDIX 1

The subscripts 2, 1, and 0 under E, V, and C symbolize the respective operation conditioning on stage 2, 1, or 0.

(A1.1) $V(y_{ijhk})= w_{ij}^2\ V(d_{ijhk})$

$= w_{ij}^2\ [E_2E_1V_0(d_{ijhk}) + E_2V_1E_0(d_{ijhk}) + V_2E_1E_0(d_{ijhk})]$

$= w_{ij}^2\ [E_2E_1(\mu_i\ s_{ik}\ c_{ijhk}) + E_2V_1(\mu_i\ s_{ik}\ c_{ijhk})$

$+ V_2E_1(\mu_i\ s_{ik}\ c_{ijhk})]$,

$= w_{ij}^2\ [E_2(\mu_i(d_i/D_i)c_{ijhk})+ E_2(\mu_i^2S_i^2c_{ijhk}2)$

$+ V_2(\mu_i(d_i/D_i)c_{ijhk})]$,

$= w_{ij}^2\ [\mu_i(d_i/D_i)\pi_{ijh}+\mu_i^2S_i^2E(c_{ijhk}^2)+\mu_i^2(d_i/D_i)^2V(c_{ijhk})]$,

$= w_{ij}^2\ [\mu_i(d_i/D_i)\pi_{ijh}+\mu_{i2}\{S_i2(\sigma_h^2+\pi_{ijh}^2)+(d_i/D_i)^2\ \sigma_h^2\ \}]$.

Letting $D_i/(D_i-1) \approx 1$ for large D_i in the definitions shown in (7), we can write (A1.1) as

$= w_{ij}^2[\mu_i(d_i/D_i)\pi_{ijh} + \mu_i^2\{S_i^2(\sigma_h^2+\pi_{ijh}^2)+(d_i/D_i)^2\sigma_h^2\}]$,

(A1.2) $C(y_{ijhk}\ y_{ijh'k}) = w_{ij}^2\ C(d_{ijhk}\ d_{ijh'k})$,

$= w_{ij}^2\ [E_2E_1\ C_0(d_{ijhk}\ d_{ijh'k}) + E_2\ C_1\{E_0(d_{ijhk})\ E_0(d_{ijh'k})\}$

$+ C_2\{E_1E_0(d_{ijhk})\ E_1E_0(d_{ijhk})\}]$;

where the first term $E_2E_1\ C_0(d_{ijhk}\ d_{ijh'k})$

$= E_2E_1(0)$

$= 0$;

the second term $E_2\{C_1[E_0(d_{ijhk})\ E_0(d_{ijh'k})]\}$

$= E_2\ C_1[\mu_i\ c_{ijhk}\ s_{ik}, \mu_i\ c_{ijh'k}\ s_{ik}]$

$= E_2\ [\mu_i^2\ c_{ijhk}\ c_{ijh'k}\ \{E_1(s_{ik}^2)-(d_i/D_i)^2\ \}]$

$= E_2\ [\mu_i^2\ c_{ijhk}\ c_{ijh'k}\ \sigma_h^2] = 0$;

the third term $C_2\ [E_1E_0(d_{ijhk}),E_1E_0(d_{ijh'k'})]$

$= C_2\ [(d_i/D_i)\mu_i\ c_{ijhk},(d_i/D_i)\mu_i\ c_{ijh'k'}],$

$= (d_i/D_i)^2\ \mu_i^2\ C(c_{ijhk}\ c_{ijh'k'}),$

$= -\ (\mu_i\ d_i/D_i)^2\ \pi_{ijh}\ \pi_{ijh'}.$

Combining these three terms, we can write (A1.2) as

$$C(y_{ijhk}\ y_{ijh'k'}) = w_{ij}^2\ (d_i/D_i)^2\ \mu_i^2\ \sigma_{hh'},$$

$$= -\ w_{ij}^2\ (\mu_i\ d_i/D_i)^2\ \pi_{ijh}\ \pi_{ijh'}.$$

(A1.3) $C(y_{ijhk}\ y_{ijh'k'}) = w_{ij}^2\ C(d_{ijhk}\ d_{ijh'k'}),$

$= w_{ij}^2\ [E_2E_1\ C_0(d_{ijhk}\ d_{ijh'k'}) + E_2C_1[E_0(d_{ijhk})E_0(d_{ijh'k'})]$

$+ C_2\{E_1E_0(d_{ijhk})E_1E_0(d_{ijh'k'})\}];$

the first term $E_2E_1\ C_0(d_{ijhk}\ d_{ijh'k'})$

$= E_2E_1\ [E_0(d_{ijhk}\ d_{ijh'k'}) - \mu_i^2\ s_{ik}\ c_{ijhk}\ s_{ik'}c_{ijh'k'}]$

$= E_2E_1\ [\{E(\delta_{ik}\ \delta_{ik'}\}s_{ik}s_{ik'}\ c_{ijhk}\ c_{ijh'k'} - \mu_i\ \mu_i\ s_{ik}\ s_{ik'}\ c_{ijhk}\ c_{ijh'k'}]$

$= E_2E_1\ [C(\delta_{ik}\delta_{ik'})\ s_{ik}\ s_{ik'} -\ c_{ijhk}\ c_{ijh'k'}],$

$= 0\ ;$

the second term $E_2C_1\ [E_0(d_{ijhk})\ E_0(d_{ijh'k'})]$

$= E_2[E_1\{E_0(d_{ijhk}),E_0(d_{ijh'k'})\} - \mu_{ik}\ \mu_{ik'}E(s_{ik})E(s_{ik'})\ c_{ijhk}\ c_{ijh'k'}]$

$= E_2[\{(\mu_i^2\ E(s_{ik}\ s_{ik'})\ c_{ijhk}\ c_{ijh'k'}\} - \mu_i^2\ (d_i/D_i)^2\ c_{ijhk}\ c_{ijh'k'}]$

$= \mu_i^2\ C(s_{ik}\ s_{ik'})\ E(c_{ijhk}\ c_{ijh'k'}),$

$= \mu_i^2\ C(s_{ik}\ s_{ik'})\ \{C(c_{ijhk}\ c_{ijh'k'}) +\ \pi_{ijh}\ \pi_{ijh'}\},$

$= \mu_i^2\ S_{ii}\ \{\sigma_{hh''} +\ \pi_{ijh}\ \pi_{ijh'}\},$

the third term $C_2[E_1E_0(d_{ijhk}),E_1E_0(d_{ijh'k'})]$

$= C_2\ [(d_i/D_i)\mu_i\ c_{ijhk},\ (d_i/D_i)\mu_i\ c_{ijh'k'}],$

$= (d_i/D_i)^2\ \mu_i^2\ C(c_{ijhk},\ c_{ijh'k'}),$

$=\ (\mu_i\ d_i/D_i)^2\ \sigma_{hh''}.$

Combining these three terms, we can write (A1.3) as

$$C(y_{ijhk}\ y_{ijh'k'}) = w_{ij}^2\ \mu_i^2\ [\sigma_{hh''}S_{ii} + \sigma_{hh''}(d_i/D_i)^2 + S_{ii}\ \pi_{ijh}\ \pi_{ijh'}].$$

(A1.4) $C(y_{ijhk}\ y_{ijhk'}) = w_{ij}^2\ C(d_{ijhk}\ d_{ijhk'}),$

$= w_{ij}^2\{E_2E_1\ C_0(d_{ijhk}\ d_{ijhk'}) + E_2\ C_1[E_0(d_{ijhk})\ E_0(d_{ijhk'})]$

$+ C_2[E_1E_0(d_{ijhk})\ E_1E_0(d_{ijhk'})]\},$

the first term $E_2E_1\ C_0(d_{ijhk}\ d_{ijhk'})$

$= E_2E_1\ [E_0(d_{ijhk}\ d_{ijhk'}) - \mu_i^2 s_{ik}c_{ijhk}s_{ik'}c_{ijhk'}]$

$= E_2E_1\ [\{E(\delta_{ik}\ \delta_{ik'}\}\ s_{ik}\ s_{ik'}\ c_{ijhk}\ c_{ijhk'} - \mu_i\ \mu_i\ s_{ik}\ s_{ik'}\ c_{ijhk}\ c_{ijhk'}]$

$= E_2E_1\ [C(\delta_{ik}\ \delta_{ik'})\ s_{ik}\ s_{ik'}\ \ c_{ijhk}\ c_{ijhk'}],$

$= 0;$

the second term $E_2\ C_1[E_0(d_{ijhk})\ E_0(d_{ijhk'})]$

$= E_2\ \{\ E_1\ [E_0(d_{ijhk}),E_0(d_{ijhk'})] - \mu_i\ \mu_i\ E(s_{ik})E(s_{ik'})\ c_{ijhk}\ c_{ijhk'}\}$

$= E_2\ \{(\mu_i\ \mu_i\ E(s_{ik}\ s_{ik'})\ c_{ijhk}\ c_{ijhk'}\} - \mu_i^2\ (d_i/D_i)^2\ c_{ijhk}\ c_{ijhk'}$

$= \mu_i^2\ C(s_{ik}\ s_{ik'})\ E(c_{ijhk}\ c_{ijhk'}),$

$= \mu_i^2\ S_{ii}\ \sigma_{hh;}$

the third term $C_2[E_1E_0(d_{ijhk}),E_1E_0(d_{ijhk'})]$

$= C_2[(d_i/D_i)\mu_i\ c_{ijhk},\ (d_i/D_i)\mu_i\ c_{ijhk'}],$

$= (d_i/D_i)^2\ \mu_i^2\ C(c_{ijhk},\ c_{ijhk'}).$

Combining these three terms, we can write (A1.4) as

$$C(y_{ijhk}\ y_{ijhk'}) = w_{ij}^2\ \mu_i^2\ \{\ C(s_{ik}\ s_{ik'})E(c_{ijhk}\ c_{ijhk'}) + (d_i/D_i)^2\ C(c_{ijhk}\ c_{ijhk'})\},$$

$$= w_{ij}^2\ \mu_i^2\ \{\ S_{ii}\ E(c_{ijhk}\ c_{ijhk'}) + (d_i/D_i)^2\ \sigma_{hh}\ \},$$

$$= w_{ij}^2\ \mu_i^2\ \{\ S_{ii}\ (\sigma_{hh}+\pi_{ijh}^2)+(d_i/D_i)^2\ \sigma_{hh}\ \},$$

$$= w_{ij}^2\ \mu_i^2\ [\theta\ \pi_{ijh}(1-\pi_{ijh})\{S_{ii}+(d_i/D_i)^2\} + \pi_{ijh}^2\ S_{ii}].$$