REGRESSION ANALYSIS OF A COMPLEX DEATH RATE

Jai Won Choi, CDC, National Center for Health Statistics, Hyattsville, MD 20782
Won J. Park, Department of Mathematics and Statistics, Wright State University, Dayton Ohio

Key words: variance model, quasi-likelihood, Iteration, estimation

INTRODUCTION

Estimating regression coefficients requires var(y). One often assumes normal, binomial, or Poisson distribution to obtain var(y). Such a simple assumption on distribution is not correct for complex data such as the death rates obtained by Vital Statistics Division at National Center for Health Statistics. Therefore var(y) should include not only random errors of deaths, but also those caused by sampling, classification, and weighing. This paper presents var(y) including these four sources of errors. The regression coefficients β based on such var(y) are estimated, and the limiting var(β) are presented.

Consider outcome variables, $\mathbf{y}_i = (y_{i1}, \ldots, y_{ij}, \ldots, y_{iJ})'$ are weighted death rates 10,000 persons, and vector of covariate, $\mathbf{x}_i = (x_{i1}, \ldots, x_{ij}, \ldots, x_{iJ})'$ is p x J matrix, observed in the ith month, i = 1, ..., M, for the J groups, j=1,..., J. one may be interested in change of rate y in the dependence of outcome on the covariate **x**. For instance, death rates of outcome variables y's may depend on such covariates x's as age, sex and race for certain causes of deaths. We assume that months are independent. Often such rates are obtained through multistage process of occurrence of death, sampling from all deaths, classification of sample deaths, and finally estimation of population rates by weigh. The major purpose of this paper is to present a tractable parametric form of the variance of estimated rates, which accounts for four impacts: occurrence of deaths, sampling, classification, and weighing.

In setting a generalized linear model with independent covariates x and dependent variable y, we utilized the actual variance and the quasi-likelihood for death rate analysis. Liang and Zeger (1986), and Thall and Vail (1990) used quasi-likelihood when they could not assume one of the well known distributions.

1.1. POISSON DEATH

The occurrence of death is a rather rare happening, and may be assumed to be distributed as Poisson. It can be approximated by normal distribution for a large number of occurrences. Although other distribution may be possible, deaths frequently follow the Poisson process. Thus we assume that D_i deaths occurred in the month i (i=1,...,M) among N_i population are distributed as Poisson.

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. 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. For this data, we assume that the sample is taken by simple random sample without replacement.

1.3. POST-CLASSIFICATION

After a sample of d_i is taken, d_i sample deaths are post-classified into J groups, giving d_{ij} deaths belong to the jth age-sex-race-cause group. We assume the post-stratification follows multinomial distribution for this data. However it could be other distribution, for example, we can consider the dependence among the deaths in the same group as the persons of the same group may likely have common cause of death for certain diseases. We may assume a model for such correlation for members in the same group (Choi and McHugh, 1989), and adjust the multinomial distribution to correct such correlation problem.

1.4. WEIGHT

Sample data are often weighted to estimate population parameters from which the sample is taken. Weighing takes place in various forms, and we denote the weight by w. The weighted death rate, y_{ij}, is the post-stratified yearly death d_{ij}

multiplied by appropriate weight w_i, that is $y_{ij} = w_i d_{ij}$, where $w_i = f_i (365/a_i)(D_i/d_i)(1/N_i)$ 10,000. The weight w_i is assumed to be known at this time and $a_i = 28$ for February, 30 or 31 for the rest of the months. $365/a_i$ is about 12 by which annual rate is obtained from monthly rate. D_i/d_i is the sampling weight or inverse of the sampling probability, N_i is the population in the month i, and f_i is the processed counts divided by estimated counts which is almost one and ignorable. f_i is the factor that makes the rate equal to that of the population. The rate y presented in the NCHS publications is the annualized age-race-cause specific rate of deaths per 100,000 persons. Only change made in this paper is the rate per 10,000 persons instead of 100,000 persons.

In Section 2 we obtain the var(y) when a multiplicative model is assumed. In Section 3 we present the estimation of the parameter β with var(y) obtained in Section 2. A numerical example of U.S. homicide rate is presented in Section 4.

2. MULTIPLICATIVE MODEL

We adopt the symbols E, V, and C for expectation, variance, and covariance operator respectively throughout this paper. The first stage is the occurrence of death. Define indicator

$\delta_{ik} = 1$ if the k-th person died, and $= 0$ otherwise $(1 \leq k \leq N_i)$).

We assume that δ_{ik}'s are independent and distributed as Poisson with mean $m_i = D_i/N_i$, which does not depend on the other subscript j.

The stage 2 is a sampling of d_i deaths out of D_i. Define $z_{ik} = 1$ if the k-th file is sampled, and $= 0$ otherwise.

The top stage 3 is the classification of d_i sample deaths into different age-sex-race-cause groups, giving d_{ij} cell counts.

Define $z_{ijk} = 1$ if k-th file classified into j-th group, and $= 0$ otherwise.

Let the variable $y_{ij} = \sum_{k=1}^{N_i} y_{ijk}$, where

$y_{ijk} = w_i d_{ijk}$ and

$d_{ijk} = \delta_{ik} z_{ik} z_{ijk}$.

We consider the weights w_i's are fixed numbers, and d_{ijk}'s are the only variables $(1 \leq k \leq N_i)$.

2.1. ASSUMPTIONS

Conditioning on the z_{ik} and z_{ijk}, we assume that responses d_{ijk}'s are conditionally not correlated, and that the z_{ik}'s and z_{ijk}'s are mutually independent, and at the stage 1,

$E(d_{ijk} / z_{ik} z_{ijk}) = m_i z_{ik} z_{ijk}$ and the expectation over all stages is $E(m_i z_{ik} z_{ijk}) = m_i p_i \pi_{ij}$, where $m_i = D_i/N_i$, $p_i = d_i/D_i$, and $\pi_{ij} = D_{ij}/D_i$ are the expectations of δ_{ik}, z_{ik}, and z_{ijk}, respectively, and

$E(y_{ij}) = \mu_{ij}(\beta) = w_i N_i m_i p_i \pi_{ij} = w_i d_i D_{ij}/D_i$

Thall and Vail (1990) used a multiplicative Poisson model $E(y_{ij}) = \mu_{ij} E(z_i) E(z_j)$ for two independent random variables z_i and z_j. Morton (1987) and Firth and Harris (1991) used a multiplicative error model, $y_{ijk} = \mu_{ijk} \varepsilon_i \varepsilon_{ij} \varepsilon_{ijk}$, where ε_i, ε_{ij}, and ε_{ijk} are the errors arising from three stage nested data, and the $E(y_{ijk}) = \mu_{ijk} = \mu(\beta)_{ijk}$. Note that our model is somewhat different from these two models. Chiang (1967) also obtained the var(**y**) without using such a multiplicative model.

If a sample was taken by simple random sample without replacement, then we may define the variance of z_{ik} at the 2nd stage

$$(1) V(z_{ik} z_{ik'}) = \begin{cases} e_i = p_i(1-p_i)D_i/(D_i-1) & \text{for } k=k', \\ e_{ii} = -p_i(1-p_i)/(D_i-1) & \text{for } k \neq k'. \end{cases}$$

This variance depends on the actual sampling design and type of variable. Some common variances are presented in most text book (for example, Cochran, 1979) for simple random sample, stratified simple random sample, cluster sample, and systematic sample with or without replacement.

We may assume multinomial distribution for classified cells in the 3rd stage, and write its variance

$$(2) V(z_{ijk} z_{ij'k'}) = \begin{cases} e_{ij} = \pi_{ij}(1-\pi_{ij}) & \text{for } j=j' \text{ and } k=k' \\ e_{ijj'} = -\pi_{ij} \pi_{ij'} & \text{for } j \neq j' \text{ and } k=k', \end{cases}$$

and all other covariances are zero for $k \neq k'$. This covariance would not be zero if they are dependent on each other, and one may use an adjusted multinomial distribution discussed in Section 1.3.

2.2. A VARIANCE-COVARIANCE

From the above assumptions, we present a variance-covariance matrix $\mathbf{V}_i$ or var($\mathbf{y}_i$)on the main diagonal of $\mathbf{V}$. The $\mathbf{V}_i$ is a block diagonal matrix with var(y_{ij}) on the main diagonal, and

$cov(y_{ij}\ y_{ij'})$ on the off-diagonal for $j \neq j'$. Following the conditional variance (Appendix 1), and defining

$\mu_{ij}(\beta) = w_i N_i m_i p_i \pi_{ij}$,

$\alpha_i = w_i [1 + m_i]$,

and $\zeta_i = -[1/N_i - (1-1/N_i)(1-p_i)D_i/(d_i(D_i-1)]$, it is elementary to show that the variance is

$\mathbf{V}_i = \alpha_i \operatorname{diag}(\mu_{ij}) + \zeta_i \mu_i \mu_i'$

where $\mu_i' = (\mu_{11}, \ldots, \mu_{ij}, \ldots, \mu_{iJ})$.

Here we assumed that months were independent, and the cells of age-sex-race-cause were multinomial with restriction of $\Sigma_j \pi_{ij} = 1$ and $\pi_{ij} > 0$ for $i = 1, \ldots, M$. Deleting one redundant column and row, and define $\mathbf{D}_i = \alpha_i \operatorname{diag}(\mu_{ij})$, $V_i^{-1} = D_i^{-1} - \{\zeta_i/(1 + \zeta_i \mu_{i+}/\alpha_i)\} \alpha_i^{-2} \mathbf{1}_i \mathbf{1}_i'$. where $\mu_{i+} = \Sigma \mu_{ij}$ and $\mathbf{1}$ is the column vector of ones. The impacts of death process, sampling, and weighing are reflected on the variance $\mathbf{V}_i$.

The $\mathbf{V}_i$ can be simplified when the data are not weighted by setting $w_i = 1$. Other sample design can be easily reflected on the variance $\mathbf{V}_i$ by adjusting sampling error (1) accordingly. Similarly we can also modify V_i for other cell distribution. For independent months, var(y) is a $\operatorname{diag}\{\mathbf{V}_1, \ldots, \mathbf{V}_M\}$, and zero on the off-diagonal.

3. ESTIMATION

Let $\mathbf{y}_i = (y_{i1}, \ldots, y_{ij}, \ldots, y_{iJ})'$ of the weighted death rates with mean $\mu_i = (\mu_{i1}, \ldots, \mu_{ij}, \ldots, \mu_{iJ})'$. $\mathbf{X}_i = (\mathbf{x}_{i1}, \ldots, \mathbf{x}_{ij}, \ldots, \mathbf{x}_{iJ})'$ be the p x J matrix of covariates as each $\mathbf{x}_{ij} = (x_{ij1}, \ldots, x_{ijp})$ is 1 x p vector. Denote linear predictor $\eta_i = (\eta_{i1}, \ldots, \eta_{ij}, \ldots, \eta_{iJ})'$. The link function g relates the predictor η_i to the expected value $\mu_i = E(y_i)$. i.e. $g(\mu_i) = \eta_i = \mathbf{X}_i \beta$. If the link function is wrong or $\eta_i = \mathbf{X}_i \beta$ is not a correct predictor, the variance will be distorted as the variance depends on μ_i.

Let $\mathbf{S}_i = \mathbf{y}_i - \mu_i$ with $E(\mathbf{S}_i) = \mathbf{0}$, and $\mathbf{V}_i(\mu_{ij})$ be the J x J matrix of var(y_i). $\mathbf{D}_i = \partial\mu_i/\partial\beta = \partial\mu_i/\partial\eta_i\ \partial\eta_i/\partial\beta$. 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. Following the quasi likelihood (McCullagh and Nelder, 1989), the p estimating score equations for p regression parameters β are given as

(3) $\quad U(\beta) = \Sigma_i \mathbf{D}_i^T \mathbf{V}_i^{-1} \mathbf{S}_i$

β is defined to be the solution of this equation. Here if the variance of sampling and classification, and weighing were set to one, the variance $\mathbf{V}_i$ in (3) reduces to the usual case of Poisson var(y).

Let β be the solution of $U(\beta)$ in (3), in which the variance $\mathbf{V}_i$ is not only a function of β and σ, but also that of m_i, p_i, and π_{ij} as well. Assuming that σ are known at this time, the equation (3) may be expressed as $\Sigma_i\ U_i(\beta\ m^*(\beta)\ \pi^*(\beta)) = \mathbf{0}$ where $m^*(\beta) = \hat{m}(\beta, \hat{p}(\beta))$, $\pi^*(\beta) = \hat{\pi}(\beta, \hat{p}(\beta), \hat{m}*(\beta))$. $\hat{\beta}$ is now defined to be the solution of this equation.

Theorem 1. Under some basic conditions for Taylor expansion, and

(i) $\hat{m}_i$, $\hat{\pi}_{ij}$, $\hat{p}_i$ are the $M^{1/2}$-consistent estimates of m_i, π_{ij}, and p_i respectively given β, that is $o_p(1)$;

(ii) $\frac{\partial}{\partial p}\hat{m}(\beta, \hat{p}(\beta))$, $\frac{\partial}{\partial p}\hat{\pi}(\beta, \hat{p}(\beta), \hat{m}(\beta))$, and $\frac{\partial}{\partial m}\hat{\pi}(\beta, \hat{p}(\beta), \hat{m}(\beta))$ are bounded in probability or $O_p(1)$;

then $M^{1/2}(\hat{\beta} - \beta)$ has asymptotically normal with mean 0 and variance $\mathbf{\Gamma}^{-1} \Lambda \mathbf{\Gamma}^{-1}$, where

$\Lambda = \Sigma_{i\epsilon M} \mathbf{D}_i^T \mathbf{V}_i^{-1} \operatorname{cov}(\mathbf{y}_i) \mathbf{V}_i^{-1} \mathbf{D}_i$ and

$\mathbf{\Gamma} = \Sigma_{i\epsilon M}(\mathbf{D}_i^T \mathbf{V}_i^{-1} \mathbf{D}_i)$.

The proof is outlined in Appendix 2. The matrix $\mathbf{V}_i$ is given in Section 2.2. If we assume $\operatorname{cov}(\mathbf{y}_i) = \mathbf{V}_i$, then the variance becomes $[\Sigma_{i\epsilon M}(\mathbf{D}_i^T \mathbf{V}_i^{-1} \mathbf{D}_i)]^{-1}$. When a link function is specified, we can obtain the explicit form of $\mathbf{V}_i$. Variance of $\hat{\beta}$ may be correctly estimated by replacing $\operatorname{cov}(y_i)$ with $\mathbf{S}^T_i\mathbf{S}_i$ which may be more efficient then $\mathbf{V}_i$ when the model used for the derivation of $\mathbf{V}_i$ is not correct.

3.2. ITERATIVE METHOD

We may begin iteration with $\hat{\beta}^0$ substantially close to $\hat{\beta}$. The sequence of parameter estimates are generated by Newton-Raphson method, using the sum,

$$(5) \quad \hat{\beta}^{r+1} = \hat{\beta}^r + (\hat{D}^T\hat{V}^{-1}\hat{D})^{-1}(\hat{D}^T\ \hat{V}^{-1}\hat{S})$$

The estimate $\hat{\beta}$ may be obtained by iterating until it converges. We may start the iteration with ordinary least square estimate of β. Convergence criterion is to stop the iteration at (r+1) step when

$$MAX|(\hat{\beta}^{r+1} - \hat{\beta}^{r})/\hat{\beta}^{r}| \leq 10^{-5},.$$

Provided that the eigenvalues of $\hat{D}^{T}\hat{V}^{-1}\hat{D}$ are sufficiently large, the second term of (5) is negligible. Then, we may take the first round approximation $\hat{\beta}^{1}=\hat{\beta}$, even when $\hat{\beta}^{1}$ is not a computable statistics. When V_i is set equal to usual Poisson form, existing GLIM software (A.V. Swan et al, release 3.77, 1987) provides the estimates of the parameters.

4. U.S. HOMICIDE DATA

The monthly data of mortality of U.S. population are reported from 50 states and Washington D.C. to the National Center for Health Statistics. Ten percent sample is taken each month and used to estimate the U.S. death rate according to age-sex-race-causes. Only homicide records are extracted from the ten percent sample. Since the highest rate was reported from the age group of 15-24 years, we used only this age group. The homicides are also more frequently reported from the blacks and males, therefore our study concerns only the black and male in this age group. The year is included in the model to see any change of rate during these five years from 1985 to 1989. The preliminary data analysis shows that the race and sex are the good candidate, and years are in a lesser degree.

About 23,000 homicides were reported from the 15-24 age group of white and black during the 55 months, 12 months in 1985, 1986, and 1987 respectively, 11 months in 1988, and 8 months in 1989. A ten percent simple random sample is selected without replacement from each of those 55 months, and we have an aggregate of 2,282 sample records which were classified into four categories, two sexes (female and male) and two races (white and black). We assume that deaths in the four categories follow the multinomial distribution, and that the process of death, sampling, and classification is independently contributing to the random errors.

First we used the Poisson variance for the regression analysis of homicide rates per 10,000 persons. Secondly we used the actual variance to study the same data. The deviance and residual sum of square is used to explore the adequacy of model fit in respect to the choice of variance function, link function, and terms in the linear predictor. We decided to use the model:

mean death rate= $1+ \beta_1$ sex+ β_2 race+ β_3 year.

The Poisson assumption gave better fitting than normal assumption for error. There was no evidence of significant interaction between sex, race, and year when the ordering of the variable are rotated, thus dropping interaction terms. All the 220 residuals of above model were less than 1 except 12 when the Poisson error and log link are used for the model. Even those 12 were less than 1.7.

It took nine iterations to obtain the estimates of the coefficients under the assumption of Poisson variances, giving

log(rate)=-5.07+1.59sex+1.85race+0.09 year.

Although all terms are significant, the influence of sex and race to the log-rate is about 1.6 and 1.9 units respectively while the year is only 0.09 unit. Scaled deviance is 0.3036, and means quare error is 1.3328.

It took seven iterations to obtain the estimates under the assumption of actual variance, giving

log(rate)=-3.98+1.25sex+1.52race+0.05year.

The contributions of sex and race to the log-rate is 1.25 units and 1.52 units respectively. Both are significant factors at alpha = 0.01 level. But the year contributes only 0.05 unit to the log-rate, and is not significant at this level. The scaled deviance is 3.5045, and mean squared error is 3.159.

The variance of y estimated only by Poisson assumption is significantly underestimating the real variance. As a result of the underestimation of variance, it provides the coefficient estimates (- 5.0719, 1.5872, 1.8504, 0.0927) with the standard errors of (0.183, 0.067, 0.053, 0.018), while the actual variance gives more realistic estimates (- 3.9766, 1.2458, 1.5177, 0.0507) with the standard errors of (0.269, 0.080, 0.097, 0.038). In both situations the goodness of fit of the model is acceptable, and the sex and race are the most influential terms of all reducing the deviance greatly.

From the model under the actual variance, we have the mean death rate for male=2, white=1 and first year=1 is 1.08, while the mean rate of male=2, black=2, and first year=1 is 4.95. Thus, the difference between the male-white-1985 and male-black-1985 is about 3.87 = 4.95-1.08 annualized deaths per 10,000 persons. Similarly the difference between the male-black-1989 and male-white-1989 is 4.71 = 6.05-1.34 annualized deaths per 10,000 persons.

The rate difference between black and white is 3.87 =4.95-1.08 in 1985 and 4.71= 6.05-1.34 in 1989. Both black and white rate are increased during these five years, 22 percent for black and 24 percent for white. Although the percent increase is more for white by 2 percent, the black rate is 4.95, which is 358 percent of white the rate of 1.08 in 1985, and 6.05 or 352 percent of the white rate of 1.34 in 1989, giving much higher rate of black homicide in both years. The homicide rates of both races are increasing during those five years.

It shows that Poisson variance should be adjusted for a meaningful regression analysis of data.

When we considered sex, race, and year as factors of two, two and five levels respectively, it did not improve the model much. When the population is used as offset and actual count of sample deaths as dependent variables, it increased the deviance too much for us to handle it effectively. Although the contribution of the years to the model is not significant, we kept it in the model because the yearly trend is one of the topics of our interest .

It is worthy to mention a possible extension. We obtained the death rate from a population, which remains the same every year except the new born babies and deaths occurred previous year.

The rates obtained from such population may be correlated as a same population may be the basis from which the deaths arise. When the correlation occurs, Liang and Zeger (1986) have further adjusted covariance matrix to correct the time correlation. Following their method, such a correlation problem can be easily corrected.

Acknowledgement: Authors thank Dr. Randy Curtin for introducing this problem and for his kind comments.

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

W. G. Cochran (1979). Sampling Techniques. 2nd Edition, John Wiley and Sons, Inc. New York.

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.

D. Firth and I.R. Harris (1991) Quasi-likelihood for multiplicative random effects. Biometrika 78, 3, pp. 545-55.

Kung-Yee Lang and Scott L. Zeger (1986). Longitudinal data analysis using generalized linear models, Biometrika 73. 1. pp 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 (1989). 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.

Scott L. Zeger and Kung-Yee Liang (1986). Longitudinal data analysis for discrete and continuous outcomes. Biometrics 42, 121-130.

Scott L. Zeger (1988). A regression model for time series of counts. Biometrika 75, 621-629.

APPENDIX 1

The subscript 1, 2, and 3 under E, V, and C symbolize the respective stage. $d_{ijk} = \delta_{ik} z_{ik} z_{ijk}$, $i = 1$ to M months, $j = 1$ to J, $k = 1$ to N_i

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

$$= w_i^2 [E_3E_2V_1(d_{ijk}) + E_3V_2E_1(d_{ijk}) + V_3E_2E_1(d_{ijk})]$$

$$w_i^2 [E_3E_2(m_i z_{ik} z_{ijk}) + E_3V_2(m_i z_{ik} z_{ijk}) + V_3E_2(m_i z_{ik} z_{ijk})],$$

$$= w_i^2 [m_i p_i \pi_{ij} + m_i^2 e_i E_3(z_{ijk}^2) + m_i^2 p_i^2 V_3(z_{ijk})],$$

$= w_i^2 [m_i p_i \pi_{ijh} + m_{i2} \{e_i(e_{ij} + \pi_{ij}^2) + p_i^2 e_{ij}\}].$
$= w_i^2 [m_i p_i \pi_{ij} + m_i^2 \{e_i(e_{ij}+\pi_{ij}^2)+ p_i^2 e_{ij}\}],$
$= w_i^2 [m_i p_i \pi_{ij} + m_i^2 \{e_i \pi_{ij} + p_i^2(\pi_{ij}-\pi_{ij}^2)\}],$

(A1.2) $C(y_{ijk}\, y_{ij'k}) = w_i^2 C(d_{ijk}\, d_{ij'k})$, where

$$C(d_{ijk}\, d_{ij'k})=[E_3E_2\, C_1(d_{ijk},\, d_{ij'k}) + E_3\, C_2\{E_1(d_{ijk}),\, E_1(d_{ij'k})\} + C_3\, \{E_2E_1(d_{ijk}),\, E_2E_1(d_{ijk})\}];$$
$$=0+E_3\, C_2(m_i\, z_{ijk}\, z_{ik},\, m_i\, z_{ij'k}\, z_{ik}) + C_3\, [p_im_i\, z_{ijhk},\, p_im_i\, z_{ij'k}],$$
$$= 0+E_3\, (m_i^2\, z_{ijk}\, z_{ij'k}\, V_2(z_{ik}) + p_i^2\, m_i^2\, C_3(z_{ijk}\, z_{ij'k}),$$
$$= 0 +0 -(m_i\, p_i)^2\, \pi_{ij}\, \pi_{ij'}.$$

(A1.3) $C(y_{ijk}\, y_{ij'k'}) = w_i^2\, C(d_{ijk}\, d_{ij'k'})$,

$$= w_i^2\, [E_3E_2\, C_1(d_{ijk}\, d_{ij'k'}) + E_3C_2\{E_1(d_{ijk}),\, E_1(d_{ij'k'})\} + C_3\{E_2E_1(d_{ijk}),\, E_2E_1(d_{ij'k'})\}];$$
$$= w_i^2\, [E_3E_2\, [C_1(\delta_{ik}\, \delta_{ik'})\, z_{ijk}\, z_{ij'k'}\, z_{ik}\, z_{ik'}\,] + E_3\, [m_i^2\, C_2(z_{ik}\, z_{ik'})\, z_{ijk}\, z_{ij'k'}] + C_3\, [p_im_i\, z_{ijk},\, p_im_i\, z_{ij'k'}]\,]$$
$$= 0+ m_i^2\, e_{ii}\, \pi_{ij}\, \pi_{ij'}\} + 0= w_i^2\, m_i^2\, e_{ii}\, \pi_{ij}\, \pi_{ij'}.$$

(A1.4) $C(y_{ijk}\, y_{ijk'}) = w_i^2\, C(d_{ijk}\, d_{ijk'})$, where

$$C(d_{ijk}\, d_{ijk'})= E_3E_2\, C_1(d_{ijk}\, d_{ijk'}) + E_3\, C_2[E_1(d_{ijk})\, E_1(d_{ijk'})] + C_3\, [E_2E_1(d_{ijk})\, E_2E_1(d_{ijk'})],$$
$$=0 + m_i^2\, e_{ii}\, \pi_{ij}^2 + 0.$$

Defining $\mu_{ij} = N_i\, w_i\, m_i\, p_i\, \pi_{ij}$, the variance and covariance matrix V_i includes the diagonal element

$$V(y_{ij}) = \Sigma_{k\in N_i}\, V(y_{ijk}) + \Sigma_{k\neq k'\in N_i}\, V(y_{ijk}\, y_{ijk'})$$
$$= N_i\, w_i^2\, [m_i\, p_i\, \pi_{ij} + m_i^2\, p_i\, \pi_{ij} - m_i^2p_i^2\, \pi_{ij}^2] + N_i(N_i-1)\, w_i^2\, m_i^2\, e_{ii}\, \pi_{ij}^2,$$
$$= \mu_{ij}\, [w_i(1 + m_i)] - \mu_{ij}^2/N_i + \mu_{ij}^2(1-1/N_i)(1-p_i)D_i/(d_i(D_i-1)).$$
$$= \mu_{ij}\, [w_i(1 + m_i)] - \mu_{ij}^2\, [1/N_i - (1-1/N_i)(1-p_i)D_i/(d_i(D_i-1))].$$

and the off-diagonal element $C(y_{ij}\, y_{ij'})$

$$= N_i\, C(y_{ijk}\, y_{ij'k}) + N_i(N_i - 1)\, C(y_{ijk}\, y_{ij'k'}),$$
$$= -N_i\, w_i^2\, m_i^2\, p_i^2\, \pi_{ij}\pi_{ij'} + N_i(N_i - 1)\, w_i^2m_i^2e_{ii}\pi_{ij}\pi_{ij'}$$
$$= - \mu_{ij}\mu_{ij'}\, [1/N_i - (1-1/N_i)(1-p_i)D_i/(d_i(D_i-1))].$$

APPENDIX 2

Denote vectors $m^*(\beta)= \hat{m}(\beta,\hat{p}(\beta))$ and $\pi^*(\beta) = \hat{\pi}(\beta,\, \hat{p}(\beta),\, m^*(\beta))$ and $D^T\, V^{-1}\, S = 0$ may be expressed as

$$U = U(\beta,m^*(\beta),\pi^*(\beta))= 0.$$

Taylor expansion of U, using the sum, can be written as

$$0 = U(\beta,m^*(\beta),\pi^*(\beta)) + (\hat{\beta} - \beta)\, dU/d\beta + o_p(1).$$

$M^{1/2}(\hat{\beta} - \beta)$ can be approximated by

$$[(dU/d\beta)/M]^{-1}\quad [-(U(\beta,m^*(\beta),\pi^*(\beta))/M^{1/2}].$$

where $dU[\beta,m^*(\beta),\pi^*(\beta)]/d\beta =$

$$\partial U[\beta,m^*(\beta),\pi^*(\beta)]/\partial\beta + \partial U[\beta,m^*(\beta),\pi^*(\beta)]/\partial m^*(\beta)\quad \partial m^*(\beta)/\partial\beta + \partial U[\beta,m^*(\beta),\pi^*(\beta)]/\partial\pi^*(\beta)\quad \partial\pi^*(\beta)/\partial\beta$$

$= A + B\, C + D\, E$. It is easy to see that $B = o_p(M^{-1})$, $D = o_p(M^{-1})$, and $C = O_p(1)$ and $E= O_p(1)$ and that the expectation of $-A/M$ converges to $-\mathbf{D}^T\, \mathbf{V}^{-1}\, \mathbf{D}/M$ for large M.

Let β be fixed and again by linear expansion,

$$U(\beta,m^*(\beta),\pi^*(\beta))/M^{1/2} = M^{-1/2}\, U(\beta,m(\beta),\pi(\beta)) + M^{-1}\, \partial U(\beta,m(\beta),\pi(\beta))/\partial m\quad M^{1/2}(m^*-m) + M^{-1}\, \partial U(\beta,m(\beta),\pi(\beta))/\partial\pi\quad M^{1/2}(\pi^*-\pi) + o_p(1)$$
$$= A^* + B^*\, C^* + D^*\, E^* + o_p(1).$$

Note that $B^* = o_p(1)$ and $D^* = o_p(1)$ since $\partial U(\beta,m(\beta),\pi(\beta))/\partial m$ and $\partial U(\beta,m(\beta),\pi(\beta))/\partial\pi$ are linear function of S's whose means are also zero and finite variance. Also

$$C^* = M^{1/2}(m^* - m)$$
$$= M^{1/2}(\hat{m}(\beta,\hat{p}(\beta)) - \hat{m}(\beta,p) + \hat{m}(\beta,p) - m)$$
$$= M^{1/2}[\hat{m}(\beta,\hat{p}(\beta)) - \hat{m}(\beta,p)]+ M^{1/2}[\hat{m}(\beta,p)- m)]$$
$$=\partial\hat{m}(\beta,p)/\partial_p M^{1/2}(\hat{p}-p)+M^{1/2}[\hat{m}(\beta,p)-m)]+o_p(1)$$
$$= A^{**}\, B^{**} + C^{**} + o_p(1) = o_p(1)$$

since $A^{**} = O_p(1)$, $B^{**} = o_p(1)$ and $C^{**} = o_p(1)$ by the assumption (i) and (ii), and

$$E^* = M^{1/2}(\pi^* - \pi)$$
$$= M^{1/2}\, (\hat{\pi}[\beta,\hat{p}(\beta),\hat{m}(\beta,\hat{p}(\beta))] - \pi)$$
$$= M^{1/2}\, \{\hat{\pi}(\beta,\, \hat{p}(\beta),\, \hat{m}(\beta,\, \hat{p}(\beta))) - \hat{\pi}(\beta,\, p(\beta),\, \hat{m}(\beta,\, p(\beta))) + \hat{\pi}(\beta,\, p(\beta),\, \hat{m}(\beta,\, p(\beta))) - \hat{\pi}(\beta,\, p(\beta),\, m(\beta,\, p(\beta))) + \hat{\pi}(\beta,\, p(\beta),\, m(\beta,\, p(\beta))) - \pi(\beta,\, p(\beta),\, m(\beta,\, p(\beta)))\},$$
$$E^* = \partial\hat{\pi}(\beta,\, p(\beta),\, \hat{m}(\beta,\, p(\beta)))/\partial p\quad M^{1/2}(\hat{p} - p) + \partial\hat{\pi}(\beta,\, p(\beta),\, m(\beta,\, p(\beta)))/\partial m\quad M^{1/2}(\hat{m} - m) + M^{1/2}[\hat{\pi} - \pi)] + o_p(1)$$
$$= A^{***}\, B^{***} + C^{***}\, D^{***} + E^{***} + o_p(1) = o_p(1)$$

by the assumption (i) and (ii). Therefore the $A^* + B^*\, C^* + D^*\, E^* + o_p(1)$ is equivalent to A^*, whose covariance matrix is

$$cov(\Lambda) = \Sigma_{i\in M}\, D_i^T\, V_i^{-1}\, cov(y_i)\, V_i^{-1}\, D_i.$$

Therefore

$$M^{1/2}(\hat{\beta} - \beta) \approx (\Sigma_{i\in M}\, A_i/M)^{-1}\, A^* + o_p(1)$$

and has asymptotically normal with mean 0 and covariance matrix $(\Sigma_{i\in M}\, A_i/M)^{-1}cov(\Lambda)(\Sigma_{i\in M}\, A_i/M)^{-1}$.