Survey Methodology, June 1991
Vol. 17, No. 1, pp. 69–78
Statistics Canada

Marginal and Approximate Conditional Likelihoods for Sampling on Successive Occasions

D.R. BELLHOUSE[1]

ABSTRACT

Marginal and approximate conditional likelihoods are given for the correlation parameters in a normal linear regression model with correlated errors. This general likelihood approach is applied to obtain marginal and approximate conditional likelihoods for the correlation parameters in sampling on successive occasions under both simple random sampling on each occasion and more complex surveys.

KEY WORDS: Likelihood inference; Sampling in time; ARMA models; State space models.

1. INTRODUCTION

Consider a finite population of N units which may be sampled on k occasions. Let y_{tj} denote the measurement on the j^{th} population unit taken on the t^{th} occasion; $j = 1, \ldots, N$ and $t = 1, \ldots, k$. It is assumed that any two units, say j and j', are independent, but that measurements of the same unit across time are correlated. In particular, assume that for any j,

$$(y_{1j}, y_{2j}, \ldots, y_{kj})^T \sim N(\mu, \sigma^2\Omega), \tag{1}$$

where Ω is a $k \times k$ correlation matrix and where μ is the $1 \times k$ vector of fixed means $(\mu_1, \mu_2, \ldots, \mu_k)^T$. In view of the explicit model assumption in (1), a model-based approach to survey estimation is used in this paper. Based on samples taken over the k occasions, it is of interest to estimate $(\mu_1, \mu_2, \ldots, \mu_k)^T$. The form of the model-based estimates $(\hat{\mu}_1, \hat{\mu}_2, \ldots, \hat{\mu}_k)^T$, if obtained by maximum likelihood or generalized least squares, for example, will depend on σ^2 and the parameters in Ω. It is therefore necessary to obtain good estimates of σ^2 and the parameters in Ω.

The notation of Bellhouse (1989) is used to describe the sampling scheme considered here, namely one-level rotation sampling. On any occasion, c rotation groups are sampled. Rotation group $r\,(r = 1, 2, \ldots, k + c - 1)$, denoted by G_r, consists of m_r sample units. On occasion $t\,(t = 1, \ldots, k)$, the sample consists of the units in $G_t, G_{t+1}, \ldots, G_{t+c-1}$, so that the total sample size on occasion t, $n_t = m_t + m_{t+1} + \ldots + m_{t+c-1}$. On occasion $t + 1$, G_t is dropped from the sample and G_{t+c} is added. Each rotation group is chosen without replacement from previously unchosen units in the population. The total sample size over all k occasions is $m = n_1 + n_2 + \ldots + n_k$. The maximum number of occasions that a unit remains in the sample is c.

If c is small, then estimates of the correlation parameters in Ω can be unstable, leading to instability in the estimates of interest $(\hat{\mu}_1, \hat{\mu}_2, \ldots, \hat{\mu}_k)^T$. Viewed another way, the total number of parameters is at least $k + 2$ and increases with time, *i.e.* with the addition of new occasions. Since the dimension of the parameter space increases with time, maximum likelihood estimates of parameters may be biased and inconsistent. The problem of the stability of estimates has been addressed in sampling on successive occasions, for example, by Blight

[1] D.R. Bellhouse, Department of Statistical and Actuarial Sciences, University of Western Ontario, London, Ontario, Canada N6A 5B9.

and Scott (1973), who assume that the elements of $(\mu_1, \mu_2, \ldots, \mu_k)^T$ follow a time series process. On using this assumption the dimension of the parameter space is fixed at a relatively small number so that the problems of instability, bias and inconsistency are resolved. In this paper, a different approach is taken. Here the fixed means assumption is retained and marginal and approximate conditional likelihoods are derived for the parameters in Ω, treating the fixed means as nuisance parameters.

Marginal likelihood estimation was introduced as a general method for eliminating nuisance parameters from the likelihood function (Fraser 1967; Kalbfleisch and Sprott 1970). Cox and Reid (1987) introduced approximate conditional likelihoods which also address this problem. They argued that the approximate conditional likelihood was preferable to the profile likelihood obtained by replacing the nuisance parameters in the likelihood by their maximum likelihood estimates when the parameters of interest are given. Bellhouse (1990) established the equivalence of marginal and approximate conditional likelihoods for correlation parameters under a normal model.

Following on the work of Cox and Reid, Cruddas *et al.* (1989) obtained an approximate conditional likelihood for the correlation parameter in several short series of autoregressive processes of order one with common variance and autocorrelation parameters. Based on a simulation study, Cruddas *et al.* (1989) showed that the estimate based on the approximate conditional likelihood had a much smaller bias and better coverage properties of the confidence interval than the maximum likelihood estimate from the profile likelihood. The situation described by Cruddas *et al.* (1989) applies directly to sampling on successive occasions in sample surveys. In order to reduce the response burden, individuals in a survey are retained in the sample for relatively short periods of time. It is expected that the use of marginal and approximate conditional likelihoods will improve the estimates of correlation parameters and consequently improve the estimates of the mean for each occasion.

Within a rotation group, the sample measurements on an individual are usually modelled by an autoregressive moving average process (ARMA), *i.e.* the parameters in Ω are comprised of the correlation parameters in the ARMA process. See Binder and Hidiroglou (1988) for a review of the application of time series models to sampling on successive occasions. Consequently, it is of interest to obtain marginal and approximate conditional likelihoods under ARMA models with application to rotation sampling. The marginal and approximate conditional likelihoods for the correlation parameters in a normal model are obtained in Section 2. The general results of Section 2 are illustrated in Section 3 by applying the results to sampling on successive occasions assuming simple random sampling of units in rotation groups. In Section 4, some methods are given to apply these likelihood methods to complex surveys.

2. MARGINAL AND APPROXIMATE CONDITIONAL LIKELIHOODS FOR CORRELATION PARAMETERS UNDER A NORMAL MODEL

Let y be a vector of sampled observations of dimension $m \times 1$ which follows the linear model

$$y = X\beta + \epsilon \tag{2}$$

with error vector $\epsilon \sim N(\mathbf{0},\sigma^2\Phi)$, where Φ is the $m \times m$ correlation matrix and where β is the $p \times 1$ vector of regression coefficients so that X is $m \times p$. The log-likelihood for β, σ^2 and Φ is given by

$$L(\beta,\sigma^2,\Phi) = -\left\{m \ln\sigma + (\ln|\Phi|)/2 + (y - X\beta)^T\Phi^{-1}(y - X\beta)/(2\sigma^2)\right\}. \tag{3}$$

For a given value of Φ,

$$\hat{\beta} = (X^T\Phi^{-1}X)X^T\Phi^{-1}y$$

and

$$s^2 = y^T\Phi^{-1}y - y^T\Phi^{-1}X\,(X^T\Phi^{-1}X)^{-1}\,X^T\Phi^{-1}y \tag{4}$$

are jointly sufficient for β and σ^2.

A marginal likelihood for Φ is obtained by making a transformation of the data y to the sufficient statistics $\hat{\beta}$ and s^2 and the ancillary statistic

$$a = \Phi^{-\frac{1}{2}}(y - X\,(X^T\Phi^{-1}X)^{-1}\,X^T\Phi^{-1}y)/s,$$

where $\Phi^{-\frac{1}{2}}$ is the $m \times m$ dimensional matrix such that $\Phi^{-1} = \Phi^{-\frac{1}{2}}\Phi^{-\frac{1}{2}}$. The marginal likelihood for Φ is the marginal distribution of the ancillary a times the product of the differentials da_i, $i = 1, \ldots, m$. See Kalbfleisch and Sprott (1970, eqs. 6 and 10) for a general discussion and a general expression for Πda_i. Bellhouse (1978) and, later independently Tunnicliffe Wilson (1989), showed that the marginal likelihood for Φ under the normal model is given by

$$L_M(\Phi) = \{|\ \Phi\ |^{\frac{1}{2}}\ |\ X^T\Phi^{-1}X\ |^{\frac{1}{2}}\ s^{m-p}\}^{-1}. \tag{5}$$

Note that (4) is proportional to the maximum likelihood estimate of σ^2 given Φ and that $s^2(X^T\Phi^{-1}X)^{-1}$ is proportional to the estimated variance-covariance matrix of the maximum likelihood estimate of β given Φ. Then (5) can be written as

$$L_M(\Phi) = \frac{|\ \text{est var}(\hat{\beta})\ |^{\frac{1}{2}}}{s^m\ |\ \Phi\ |^{\frac{1}{2}}}. \tag{6}$$

To obtain an approximate conditional likelihood, it is first necessary to transform the parameters to achieve parameter orthogonality between the parameters of interest and the nuisance parameters, which now may depend on the parameters of interest. Sets of parameters are orthogonal if the associated information matrix is block diagonal, with each block as the information matrix for each parameter set. The conditional likelihood is related to the distribution of the data y conditional on the maximum likelihood estimate of the nuisance parameters for fixed values of the parameters of interest. The approximate conditional likelihood is obtained by applying two approximations to this conditional distribution. See Cox and Reid (1987, Section 4.1) for a discussion of the derivation. For example, let Θ be the vector of parameters of interest and let Λ, possibly depending on Θ, be the vector of nuisance parameters orthogonal to Θ. The full likelihood of the data for parameters Θ and Λ is denoted by $L(\Theta,\Lambda)$ and the profile likelihood for $\Theta, L(\Theta,\hat{\Lambda})$ is the likelihood with Λ replaced by its maximum likelihood estimate. The approximate conditional likelihood for Θ is

$$L(\Theta,\hat{\Lambda})\ |\ I(\Theta,\hat{\Lambda})\ |^{\frac{1}{2}},$$

where $I(\Theta,\hat{\Lambda})$ is the observed information matrix for Λ at a fixed value of Θ. See Cox and Reid (1987, eq. 10).

Following Cruddas *et al.* (1989), Bellhouse (1990) suggested, for model (2), the parameter transformation $\lambda = \ln\sigma + (\ln \mid \Omega \mid)/(2m)$ leaving β the same. The log-likelihood under the new parameterization is denoted by $L(\beta, \lambda, \Phi)$ and can be obtained from (3). If the entries of Φ are functions of a parameter ϕ, then the nuisance parameters λ and β are each orthogonal to Φ, *i.e.*

$$-\frac{1}{m}E\left[\frac{\partial^2 L(\beta,\lambda,\Phi)}{\partial\phi\partial\lambda}\right] = 0$$

and

$$-\frac{1}{m}E\left[\frac{\partial^2 L(\beta,\lambda,\Phi)}{\partial\phi\partial\beta}\right] = 0,$$

when each entry of Φ is a continuous and differentiable function of ϕ. Moreover, in this case the approximate conditional likelihood for Φ, $L_C(\Phi)$ is the same as the marginal likelihood $L_M(\Phi)$, given by (5) or (6). See Bellhouse (1990) for details.

The marginal and approximate conditional likelihood in (5) or (6) can be evaluated at any Φ using state space models in the approach of Harvey and Phillips (1979). For any given Φ, once the recursions to estimate β and σ^2 are complete, the value of s^2 and $\mid \Phi \mid^{1/2}$ can be calculated from Harvey and Phillips (1979, eqs. 5.6 and 6.6, and 4.3 respectively). It is then necessary only to obtain $X^T\Phi^{-1}X$ and its determinant. The value of $X^T\Phi^{-1}X$ may be obtained from the final step in the recursive equations of Harvey and Phillips (1979, eq. 3.4).

3. SIMPLE RANDOM SAMPLING ON SUCCESSIVE OCCASIONS

3.1 Some General Results for Rotation Sampling

Suppose rotation group G_r first appears in the sample on occasion u and last appears on occasion v. Then u is either 1 or r and v is either $r + c - 1$ or k. The total number of occasions on which a unit in G_r is present in the sample is $b = v + 1 - u$. Let $\bar{y}_{u,r}, \ldots, \bar{y}_{v,r}$ be the sample means or elementary estimates for G_r on occasions $u, u + 1, \ldots, v - 1, v$ respectively. Then under model (1), the contribution of G_r to the log likelihood in (3) is

$$-\left\{bn_r \ln\sigma + (n_r/2)\ln(\mid \Omega_r \mid) + \left[n_r x_r^T \Omega_r^{-1} x_r + (n_r - 1)\,\mathrm{tr}(\Omega_r^{-1} S_r)\right]/(2\sigma^2)\right\}, \quad (7)$$

where x_r^T is the $1 \times b$ vector $(\bar{y}_{u,r} - \mu_u, \bar{y}_{u+1,r} - \mu_{u+1}, \ldots, \bar{y}_{v-1,r} - \mu_{v-1}, \bar{y}_{v,r} - \mu_v)$, where S_r is the $b \times b$ matrix of sample variances and covariances for observations within the rotation group, and where Ω_r is the $b \times b$ correlation matrix on the observations on a single unit within the rotation group. Note that the parameters in Ω as given in expression (1) will also be the parameters in Ω_r. The correlation matrix Ω is based on measurements from all occasions 1 through k; the correlation matrix Ω_r is from the subset of the data observed from occasions u through v. By the independence assumption, the full log likelihood is obtained by summing (7) over all rotation groups.

Given the parameters in Ω, or equivalently the parameters in $\Omega_1, \ldots, \Omega_{k+c-1}$, expressions for the maximum likelihood estimates $\hat{\mu}$ and $\hat{\sigma}^2$, for μ and σ^2 respectively, may be found. Likewise, $V(\hat{\mu})$, the estimated variance-covariance matrix of $\hat{\mu}$ may be obtained. This is illustrated for a first-order autoregressive process in Section 3.2. Then the marginal likelihood for the parameters in $\Omega_1, \ldots, \Omega_{k+c-1}$ is given by (5) with the expressions in (5) given by

$$| \Phi |^{1/2} = \prod_{r=1}^{k+c-1} \Omega_r,$$

$$| X^T \Phi^{-1} X |^{1/2} = V(\hat{\mu})/s^k, \tag{8}$$

$$s^2 = \sum_{r=1}^{k+c-1} \{n_r \hat{x}_r^T \Omega_r^{-1} \hat{x}_r + (n_r - 1) \operatorname{tr}(\Omega_r^{-1} S_r)\},$$

and $p = k$, where $\hat{x}_r$ is x_r with the μ's in x_r replaced by their maximum likelihood estimates.

3.2 First-Order Autoregressive Processes

When specific forms of the correlation matrices $\Omega_1, \ldots, \Omega_{k+c-1}$ are used, some simplifications to the general form of the marginal likelihood for correlation parameters, given by (5) and (6), may be obtained. For example, assume the first-order autoregressive model

$$y_{tj} = \mu_t + \phi (y_{t-1,j} - \mu_{t-1}) + \epsilon_{tj}, \tag{9}$$

where $\epsilon_{tj} \sim N(0,\sigma^2)$ for $t = 1, \ldots, k$ and $j = 1, \ldots, N$, and where the ϵ's are mutually independent. Model (9), essentially Patterson's (1950) model, is a special case of (1). As in Section 3.1, the vector of regression parameters $\beta = (\mu_1, \ldots, \mu_k)^T$. When the data vector y contains the measurements on each unit grouped by all the occasions on which it was sampled, as in the rotation sampling description of Section 3.1, the correlation matrix Φ, now a function of the autoregressive parameter ϕ, can be written as a direct sum of matrices, each of which are the correlation matrices of a first-order autoregressive process.

The following notation, similar to Patterson (1950), is used to denote various sample sizes, means and sums of squares and cross products (corrected for the appropriate mean) for occasion t:

- π_t = the proportion of units on occasion t that are matched with units from the previous occasion $(t - 1)$;
- n_t = the number of units sampled on occasion t;
- $\bar{y}_t'$ = the mean of the units on occasion t that are matched with units from the previous occasion $(t - 1)$;
- $\bar{y}_t''$ = the mean of the units on occasion t that are unmatched with units from the previous occasion $(t - 1)$;
- $\bar{y}_t$ = the mean of all the units on occasion t;
- $\bar{x}_t'$ = the mean of the units on occasion t that are matched with units from the following occasion $(t + 1)$;
- syy_t' = the sum of squares among units on occasion t which are matched with units from the previous occasion $(t - 1)$;
- syy_t'' = the sum of squares among units on occasion t which are unmatched with units from the previous occasion $(t - 1)$;
- sxx_t' = the sum of squares among units on occasion t which are matched with units from the following occasion $(t + 1)$;
- syy_t = the sum of squares among all the units on occasion t;
- sxy_t' = the sum of cross products for measurements on sample units from occasion t matched with sample units from $(t - 1)$.

Under the special case of model (9), and after much algebra, it may be shown that (7) summed over all rotation groups r, the log-likelihood for the data reduces to

$$L(\mu_1, \ldots, \mu_k, \sigma^2, \phi) = -m \ln\sigma + (d/2)\ln(1 - \phi^2) - \{A(\mu,\phi) + B(\phi)\}/(2\sigma^2), \quad (10)$$

where d is the distinct number of units sampled (irrespective of the number of occasions on which a unit is sampled) and m is the total sample size $(n_1 + \ldots + n_k)$. Further in (10),

$$A(\mu,\phi) = (1 - \phi^2)n_1(\bar{y}_1 - \mu_1)^2 + \sum_{t=2}^{k} \left[\pi_t n_t\{\bar{y}_t' - \mu_t - \phi(\bar{x}_{t-1}' - \mu_{t-1})\}^2 + (1 - \pi_\tau)n_t(1 - \phi^2)(\bar{y}_t'' - \mu_t)^2\right] \quad (11)$$

and

$$B(\phi) = (1 - \phi^2)\, syy_1 + \sum_{t=2}^{k} \left\{\phi^2\, sxx_{t-1}' - 2\phi\, sxy_t' + syy_t' + (1 - \phi^2)\, syy_t''\right\}. \quad (12)$$

For any given value of ϕ the maximum likelihood estimator is $\hat{\mu} = \boldsymbol{G}^{-1}\boldsymbol{z}$ and $\hat{\sigma}^2 = \{A(\hat{\mu},\phi) + B(\phi)\}/m$, where $A(\hat{\mu},\phi)$ is (11) with μ replaced with its maximum likelihood estimate and where $\boldsymbol{G}$ is a symmetric $k \times k$ band matrix of band width 3 and $\boldsymbol{z}$ is a $k \times 1$ vector. The nonzero entries of $\boldsymbol{G}$ are

$$g_{tt} = \pi_t n_t + (1 - \pi_t)n_t(1 - \phi^2) + \pi_{t+1}n_{t+1}\phi^2, \quad \text{for} \quad t = 1, \ldots, k$$

and

$$g_{t,t+1} = -\pi_{t+1}n_{t+1}\phi, \quad \text{for} \quad t = 1, \ldots, k - 1,$$

where $\pi_1 = \pi_{k+1} = 0$. The entries of $\boldsymbol{z}$ are

$$z_t = \pi_t n_t(\bar{y}_t' - \phi\bar{x}_{t-1}') + (1 - \pi_t)n_t\bar{y}_t''(1 - \phi^2) - \pi_{t+1}n_{t+1}(\bar{y}_{t+1}' - \phi\bar{x}_t'),$$

for $t = 1, \ldots, k$, where $\pi_1 = \pi_{k+1} = 0$ and $\bar{y}_1'' = \bar{y}_1$. The vector of estimated means $\hat{\mu}$ is unbiased for μ under model (9) and its variance-covariance matrix is $\sigma^2\boldsymbol{G}^{-1}$. It follows from (5) or (6) that the marginal and approximate conditional likelihood for ϕ is

$$L_M(\phi) = \frac{(1 - \phi^2)^{d/2}}{\{A(\hat{\mu},\phi) + B(\phi)\}^{(m-k)/2}\, |\boldsymbol{G}|^{1/2}}. \quad (13)$$

3.3 Example

The data for this example are forestry data taken from Cunia and Chevrou (1969, p. 220). The data are the merchantable volume of timber per plot measured on three occasions with partial replacement of the sample units. In rotation sampling it is assumed that once a unit

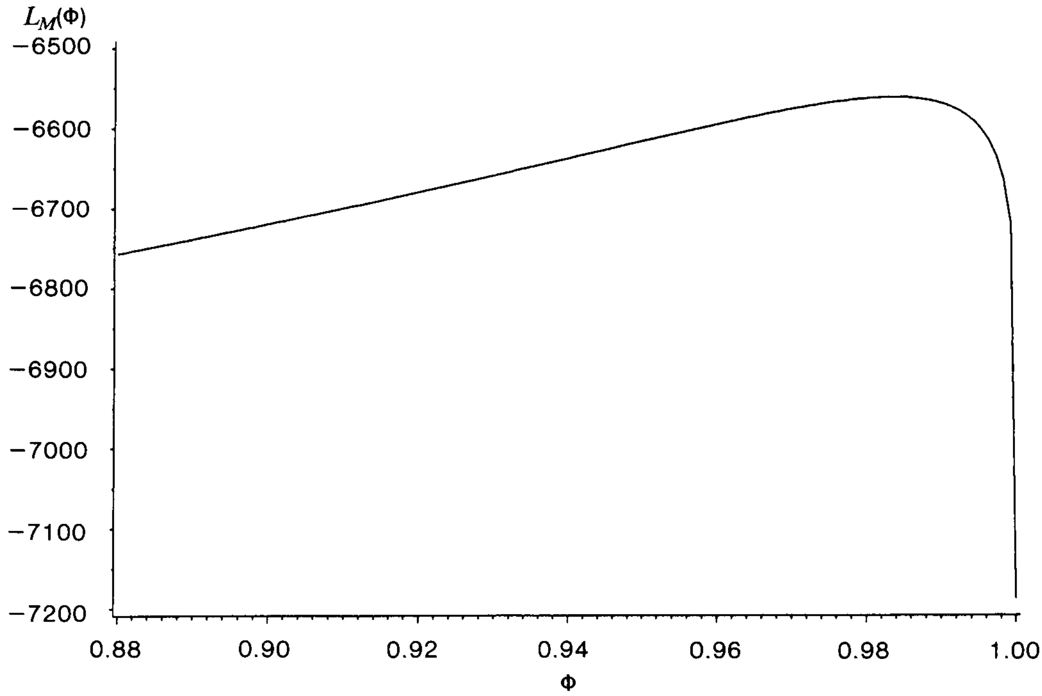

Figure 1. Marginal Likehood for the AR(1) Parameter

is dropped from the sample it is not selected again. In view of this assumption an adjustment to the data in Cunia and Chevrou was made. In particular, the measurements from sample units matched on the first and third occasions without matching units on the second occasion were dropped from the current example. From the remaining data the following calculations may be made:

$\pi_2 = 86/139$, $\pi_3 = 76/100$, $n_1 = 104$, $n_2 = 139$, $n_3 = 100$, $\bar{y}_2' = 161.5581$, $\bar{y}_3' = 179.9211$, $\bar{y}_1'' = 154.0673$, $\bar{y}_2'' = 167.2075$, $\hat{y}_3'' = 181.125$, $\bar{x}_1' = 147.6512$, $\bar{x}_2' = 163.4342$, $syy_2' = 864129.2$, $syy_3' = 555369.5$, $syy_1'' = 943948.5$, $syy_2'' = 266820.7$, $syy_3'' = 271762.6$, $sxx_1' = 800753.5$, $sxx_2' = 559850.7$, $sxy_2' = 812435.7$, $sxy_3' = 550943.6$, $d = 181$, and $m - k = 340$.

On substituting these data into (13) the marginal and approximate conditional likelihood of the data for the autoregressive order one parameter ϕ may be obtained. This is shown in Figure 1.

4. COMPLEX SURVEYS

There are several ways in which one may proceed to analyze time series data from complex surveys. Each method that can be put forward will depend upon the sample information that is available.

If data are available at the micro level, then variance-covariance matrices based on the complex design can be computed for the elementary estimates for each rotation group. A pseudo marginal likelihood is obtained by replacing $\hat{x}_r$ and S_r in (5) and (8) by their complex survey counterparts. A similar approach is taken, for example, by Roberts, Rao and Kumar (1987) in logistic regression analysis for complex surveys: obtain a likelihood or a set of likelihood equations and replace the usual statistics by their complex survey counterparts.

Under simple random sampling, S_r estimates the finite population variance-covariance matrix for measurements on the occasions covered by rotation group r. Consequently, in a complex design, S_r is replaced by a design-consistent estimate of the corresponding finite population variance-covariance matrix. For example, Kilpatrick (1981) looked at a stratified sampling design on two occasions for evaluation of the standing volume of state forests in Northern Ireland; the strata were based on the times, beginning in the 1920's, at which the forests were planted. In order to calculate the stratified sampling equivalent to S_r, it is necessary to have the estimates of the means on each occasion, strata means, strata variances, and strata covariances for the unmatched and matched samples from the two occasions. For a stratified population, the finite population variance (or covariance) may be decomposed into terms comprising the variation (or covariation) between strata and the variation (or covariation) within strata; see, for example, Cochran (1977, eq. 5.32). Estimates of the means and strata means would be used to obtain a consistent estimates of the between strata variation or covariation component and estimates of the strata variances and covariances would be used to obtain estimates of the within strata variation and covariation. Unfortunately, only certain strata variance and covariance estimates were relevant to Kilpatrick's study, so that there is insufficient published data in the article to calculate a maximum marginal likelihood estimate for the correlation between timber volumes on the two occasions.

In many cases the data at the micro level will not be available. The estimation procedure then depends upon the data that are available. One scenario is considered here; others could be formulated. Suppose that only the elementary estimates and their design effects are available. Let $\bar{y}_{t,r}$ be the estimate from rotation group G_r on occasion t based on a sample of size m_r. Let $\text{deff}_{t,r}$ be the design effect associated with $\bar{y}_{t,r}$. If σ^2/m_r is the variance of $\bar{y}_{t,r}$ under simple random sampling, then on appealing to the Central Limit Theorem,

$$(\bar{y}_{t,r} - \mu_t)/(\text{deff}_{t,r})^{1/2} \sim N(0,\sigma^2/m_r) \tag{14}$$

approximately. The modelling may proceed by assuming, within G_r, an ARMA-type process such as

$$(\bar{y}_{t,r} - \mu_t)/(\text{deff}_{t,r})^{1/2} = \phi(\bar{y}_{t-1,r} - \mu_{t-1})/(\text{deff}_{t-1,r})^{1/2} + \epsilon_t, \tag{15}$$

where ϵ_t has constant variance. This may be easily cast into the framework of model (2), where the data vector y contains data of the form $\bar{y}_{t,r}/(\text{deff}_{t,r})^{1/2}$, where β is $(\mu_1, \mu_2, \ldots, \mu_k)^T$, and where X contains entries of the form $1/(\text{deff}_{t-1,r})^{1/2}$. The marginal likelihood, obtained as a special case of (5) or (6), may be evaluated using the state space models of Harvey and Phillips (1979) as noted in Section 2. Marginal and approximate conditional likelihood estimation is especially desirable under the model given by (14) and (15). The estimate of ϕ in this case is based on the variation between elementary estimates within each rotation group; the variation within elementary estimates is not available. The length of time a rotation group remains in the sample is short so that the problems of bias and inconsistency in the maximum likelihood estimates will be applicable here.

5. DISCUSSION

Binder and Dick (1990) have also suggested the use of marginal likelihood estimation techniques for sampling on successive occasions. In their framework, suppose that the survey estimates of the means, say $\bar{y}_t$, are available for each occasion $t = 1, \ldots, k$. Also, the matrix, say S, of variances and covariances of the surveys estimates is available. As in Binder and Dick (1989, 1990), among several others, the $\bar{y}_t$'s may be modelled by

$$\bar{y}_t = \mu_t + e_t, \tag{16}$$

where e_t is the survey error at time t with variance-covariance matrix estimated by S. The means on each occasion, μ_t for occasion t, follow an ARMA process. Model (16) is a special case of the random coefficients regression model, so that the appropriate marginal likelihood is different from (5).

A marginal or approximate conditional likelihood for correlation parameters in a random coefficients regression model is obtained as follows. Suppose in model (2) that β is a random vector modelled by $\beta = W\delta + u$, where W is a $p \times q$ matrix of known values, δ is a $q \times 1$ vector of parameters, and $u \sim N(\mathbf{0}, \gamma^2\Gamma)$, independent of ϵ. Under the composite model $y = XW\delta + Xu + \epsilon$, the log-likelihood for δ, Ω, Γ, γ^2, and $\kappa = \sigma^2/\gamma^2$, denoted by $L(\delta,\kappa,\gamma^2,\Gamma,\Omega)$, is given by (3), with Ω replaced by $\kappa\Omega + X\Gamma X^T$ and $X\beta$ replaced by $XW\delta$. Likewise, the marginal likelihood, denoted by $L_M(\kappa,\Gamma,\Omega)$, is given by (5), with X replaced by XW and Ω replaced by $\kappa\Omega + X\Gamma X^T$. This yields

$$L_M(\kappa,\Gamma,\Omega) = \left\{ | \kappa\Omega + X\Gamma X^T |^{1/2} \; | (XW)^T(\kappa\Omega + X\Gamma X^T)^{-1}XW |^{1/2} g^{m-q} \right\}^{-1}, \tag{17}$$

where

$$g = y^T(\kappa\Omega + X\Gamma X^T)^{-1}y$$

$$- y^T(\kappa\Omega + X\Gamma X^T)^{-1}XW((XW)^T(\kappa\Omega + X\Gamma X^T)^{-1}XW)^{-1}(XW)^T(\kappa\Omega + X\Gamma X^T)^{-1}y.$$

Now the dimension of Ω may be large in comparison to Γ; this can be the case in sampling on successive occasions. As an alternate approach, one could take the likelihood implied by (3), multiply it by the distribution for β, and integrate over β to obtain the likelihood for the parameters under the random coefficient model. This will yield matrices of the same dimension as Γ.

Since S is available, an estimate of Ω, the correlation matrix of the survey error, may be easily obtained. An estimate of $\kappa = \sigma^2/\gamma^2$, may also be obtained. From assumptions which lead to the marginal likelihood in (17), it is necessary to assume that e_t in (16) is a stationary random variable. Then an estimate of σ^2 is the average of the diagonal elements in S. If γ^2 is the variance of the μ's then the variation between $\bar{y}_t$, $t = 1, \ldots, k$ provides an estimate of $\sigma^2 + \gamma^2$. From these two estimates, an estimate of κ may be obtained. Under model (16), X in (17) is the $k \times k$ identity matrix, while W is a $k \times 1$ column vector of 1's. The resulting marginal likelihood is a pseudo likelihood since some of the parameters have been replaced by estimates. In this case, the pseudo marginal likelihood for the parameters in Γ (pseudo since κ and Ω have been replaced by their estimates) and is given by (17) with the appropriate substitutions. The parameters in Γ are the correlation parameters in the ARMA process on μ_t. If k, the number of occasions, is relatively large in comparison to the number of parameters in Γ, then the marginal and approximate conditional likelihood estimates should be similar to the maximum likelihood estimator. For ease of computation, it seems that the full likelihood approach using the state space models as outlined by Binder and Dick (1989a, Section 3) appears to be the simplest approach to use in this situation.

ACKNOWLEDGEMENTS

This work was supported by a grant from the Natural Sciences and Engineering Research Council of Canada. The author would like to thank the referee for his comments on an earlier draft of the paper.

REFERENCES

BELLHOUSE, D.R. (1978). Marginal Likelihoods for distributed lag models. *Statistische Hefte*, 19, 2-14.

BELLHOUSE, D.R. (1989). Optimal estimation of linear functions of finite population means in rotation sampling. *Journal of Statistical Planning and Inference*, 21, 69-74.

BELLHOUSE, D.R. (1990). On the equivalence of marginal and approximate conditional likelihoods for correlation parameters under a normal model. *Biometrika*, 77, 743-746.

BINDER, D.A., and HIDIROGLOU, M.A. (1988). Sampling in time. *Handbook of Statistics, Volume 6 (Sampling)* (Eds. P.R. Krishnaiah and C.R. Rao). Amsterdam: North-Holland, 187-211.

BINDER, D.A., and DICK, J.P. (1989). Modelling and estimation for repeated surveys. *Survey Methodology*, 15, 29-45.

BINDER, D.A., and DICK, J.P. (1990). A method for the analysis of seasonal ARIMA models. *Survey Methodology*, 16, 239-253.

BLIGHT, B.J.N., and SCOTT, A.J. (1973). A stochastic model for repeated surveys. *Journal of the Royal Statistical Society*, Series B, 35, 61-68.

COCHRAN, W.G. (1977). *Sampling Techniques*. 3rd Ed. New York: Wiley.

COX, D.R., and REID, N. (1987). Parameter orthogonality and approximate conditional inference (with discussion). *Journal of the Royal Statistical Society*, Series B, 49, 1-39.

CRUDDAS, A.M., REID, N., and COX, D.R. (1989). A time series illustration of approximate conditional likelihood. *Biometrika*, 76, 231-237.

CUNIA, T., and CHEVROU, R.B. (1969). Sampling with partial replacement on three or more occasions. *Forest Science*, 15, 204-224.

FRASER, D.A.S. (1967). Data transformations and the linear model. *The Annals of Mathematical Statistics*, 38, 1456-1465.

HARVEY, A.C., and PHILLIPS, G.D.A. (1979). Maximum likelihood estimates of regression models with autoregressive-moving average disturbances. *Biometrika*, 66, 49-58.

KALBFLEISCH, J.D., and SPROTT, D.A. (1970). Application of likelihood methods to models involving large numbers of parameters. *Journal of the Royal Statistical Society*, Series B, 32, 175-208.

KILPATRICK, D.J. (1981). Optimum allocation in stratified sampling of forest populations on successive occasions. *Forest Science*, 27, 730-738.

PATTERSON, H.D. (1950). Sampling on successive occasions with partial replacement of units. *Journal of the Royal Statistical Society*, Series B, 12, 241-255.

ROBERTS, G., RAO, J.N.K., and KUMAR, S. (1987). Logistic regression analysis of sample survey data. *Biometrika*, 74, 1-12.

TUNNICLIFFE WILSON, G. (1989). On the use of marginal likelihood in time series model estimation. *Journal of the Royal Statistical Society*, Series B, 51, 15-27.