Ann. Inst. Statist. Math.
Vol. 44, No. 2, 213–227 (1992)

OPTIMAL ESTIMATION IN RANDOM COEFFICIENT REGRESSION MODELS

T. V. RAMANATHAN AND M. B. RAJARSHI

Department of Statistics, University of Poona, Pune, 411 007, India

(Received May 15, 1990; revised November 22, 1990)

Abstract. In linear regression models with random coefficients, the score function usually involves unknown nuisance parameters in the form of weights. Conditioning with respect to the sufficient statistics for the nuisance parameter, when the parameter of interest is held fixed, eliminates the nuisance parameters and is expected to give reasonably good estimating functions. The present paper adopts this approach to the problem of estimation of average slope in random coefficient regression models. Four sampling situations are discussed. Some asymptotic results are also obtained for a model where neither the regressors nor the random regression coefficients replicate. Simulation studies for normal as well as non-normal models show that the performance of the suggested estimating functions is quite satisfactory.

Key words and phrases: Conditional estimating function, random coefficient regression models, semi-parametric models, stratified data.

1. Introduction

Linear regression models with random regression coefficients have been studied by different authors (see Hildreth and Houck (1968), Swamy (1971), Anh (1988) and Dielman ((1989), Chapters V and VI)). For estimation of β (the average of the random regression coefficients), standard techniques involve fully-iterated maximum likelihood and weighted least squares. The former method requires the knowledge of underlying distributions. In the weighted least squares technique, the weights are functions of unknown parameters, estimates of which are usually obtained using the ordinary least square estimator. See, for example, Swamy (1971).

Our results are mainly concerned with situations where the data are in a stratified form. Let $\sigma_\beta^2(i)$ and $\sigma_\epsilon^2(i)$ respectively denote the variances of random regression coefficients and error terms in the i-th stratum, $i = 1, \dots, m$. Most of the work in this area assumes that $\sigma_\beta^2(i) = \sigma_\beta^2$ for all i and $\sigma_\epsilon^2(i) = \sigma_\epsilon^2$ for all i, though Swamy (1971) discussed the case where $\sigma_\beta^2(i) = \sigma_\beta^2$ for all i, but $\sigma_\epsilon^2(i)$'s may vary over the strata. We are primarily interested in situations where both $\sigma_\beta^2(i)$'s and $\sigma_\epsilon^2(i)$'s are allowed to vary over the strata.

We discuss four situations which are usually met in practice. For each situation, we consider the problem of estimation of the regression parameter β, from the point of view of the theory of estimating functions. We derive an optimal estimating function, which is free from the unknown (nuisance) parameters, through conditioning the data on the complete sufficient statistic for nuisance parameters, holding β fixed. For related discussions of such conditional estimating functions, the reader is referred to Lindsay (1982) and Mantel and Godambe (1989).

Let $\mathcal{G}$ denote the class of regular unbiased estimating functions (cf. Godambe (1985)).

DEFINITION 1.1. An estimating function $g^* \in \mathcal{G}$ is said to be locally optimal at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, if for any $g \in \mathcal{G}$,

$$E^2(\partial g^*/\partial\beta)E(g^2) \geq E^2(\partial g/\partial\beta)E(g^{*2}),$$

where β is the parameter of interest and $\boldsymbol{\theta}$ is the nuisance parameter (Mantel and Godambe (1989)).

DEFINITION 1.2. The standardized version g_s, of an estimating function g for the parameter β is defined by

$$g_s = g/E(\partial g/\partial\beta).$$

DEFINITION 1.3. With respect to a random variable $S(\beta)$ (free from $\boldsymbol{\theta}$), an estimating function g is said to be conditionally zero unbiased, if $E\{g \mid S(\beta)\} = 0$.

DEFINITION 1.4. The conditional information of a conditionally zero unbiased estimating function g is defined by $I\{g \mid S(\beta)\} = E^2\{(\partial g/\partial\beta) \mid S(\beta)\}/E\{g^2 \mid S(\beta)\}$.

Usually, we choose $S(\beta)$ to be a sufficient statistic for the nuisance parameter, holding the parameter of interest, β, fixed.

Let y_{ij}'s be random variables ($j = 1, \ldots, n_i$) with joint density function $f_i(y_{i1}, \ldots, y_{in_i}; \beta, \boldsymbol{\theta}) = f_i(\beta, \boldsymbol{\theta})$, which satisfies the Cramer regularity conditions, where β is the parameter of interest and $\boldsymbol{\theta}$ is the vector of nuisance parameters. Let U be the score function given by

$$U = \sum_{i=1}^{m} U_i = \sum_{i=1}^{m} \partial\{\log f_i(\beta, \boldsymbol{\theta})\}/\partial\beta. \tag{1.1}$$

The estimating function U is optimal if $\boldsymbol{\theta}$ were known or absent in (1.1). When U involves $\boldsymbol{\theta}$ (unknown), U cannot be directly used for estimation of β. Following Lindsay (1982), if for each fixed β, there exists a unique maximum likelihood estimator $\hat{\boldsymbol{\theta}}(\beta)$ of $\boldsymbol{\theta}$, we define the estimating function $\hat{U}$ as

$$\hat{U} = \sum_{i=1}^{m} \hat{U}_i = \sum_{i=1}^{m} U_i(\beta, \hat{\boldsymbol{\theta}}(\beta)). \tag{1.2}$$

But, $\hat{U}$ may fail to give a zero unbiased estimating function. To overcome this problem, Lindsay (1982) suggests using the conditional score function as described below. (Earlier work in this direction is due to Bartlett, Neyman and Scott; cf. Cox and Hinkley ((1974), p. 146) and Godambe (1976)).

Let $S_i(\beta)$ be a sufficient statistic for the nuisance parameter $\boldsymbol{\theta}$, when β is fixed. Consider

$$W = \sum_{i=1}^{m} W_i = \sum_{i=1}^{m} [U_i - E_\beta\{U_i \mid S_i(\beta)\}]. \tag{1.3}$$

Mantel and Godambe (1989) show that (1.3) is locally optimal at a specified value of $\boldsymbol{\theta}$. As noted by them, if W depends on $\boldsymbol{\theta}$, no optimal estimating function exists. Under certain assumptions, Lindsay (1982) has shown that W is a zero unbiased and information unbiased (i.e. $E\{(\partial W/\partial\beta) + W^2\} = 0$) estimating function and $\hat{W}$ is an unbiased estimating function, where

$$\hat{W}(\beta) = \sum_{i=1}^{m} W_i(\beta, \hat{\boldsymbol{\theta}}(\beta, S(\beta))). \tag{1.4}$$

Mantel and Godambe (1989), in a recent paper, adopt the following approach. Consider the functions $h_i(y_{i1}, \ldots, y_{in_i}; \beta) = h_i$ (say). Let $S(\beta) = (S_1(\beta), \ldots, S_m(\beta))$ be as described above. Suppose that the h_i's are conditionally zero unbiased and orthogonal; that is, $E\{h_i \mid S(\beta)\} = 0$ and $E\{h_i h_j \mid S(\beta)\} = 0$ for $i, j = 1, \ldots, m$, $i \neq j$. The conditionally optimal estimating function g^*, that is, one which maximizes $I\{g \mid S(\beta)\}$ for all β, in a semi-parametric set up, is given by

$$g^* = \sum_{i=1}^{m} a_i^* h_i, \tag{1.5}$$

where $a_i^* = E\{(\partial h_i/\partial\beta) \mid S(\beta)\}/E\{h_i^2 \mid S(\beta)\}$, $i = 1, \ldots, m$. The estimating function (1.5) is optimal in the class of conditionally zero unbiased estimating functions viz., $G^* = \{g = \sum_i a_i h_i,$ where the a_i's are functions of β and $S(\beta)$ only$\}$.

Various situations that we discuss in this paper are described below.

Model I. $\quad y_{ij} = \beta_i x_i + \epsilon_{ij}, \quad j = 1, \ldots, n_i; \quad i = 1, \ldots, m.$

ASSUMPTIONS. (i) ϵ_{ij}'s ($j = 1, \ldots, n_i$) are independently distributed with mean zero and variance $\sigma_\epsilon^2(i)$; $i = 1, \ldots, m$. (ii) β_i's are independently distributed with mean β and variance $\sigma_\beta^2(i)$ for $i = 1, \ldots, m$. (iii) β_i and ϵ_{jk} are independent for all i, j and k.

Notice that both the (random) regression coefficient and regressor are fixed throughout the i-th stratum.

Model II. $\quad y_{ij} = \beta_i x_{ij} + \epsilon_{ij}, \quad j = 1, \ldots, n_i; \quad i = 1, \ldots, m.$

Other assumptions are the same as those of Model I.

Model III. $$y_{ij} = \beta_{ij} x_i + \epsilon_{ij}, \quad j = 1, \ldots, n_i; \quad i = 1, \ldots, m.$$

ASSUMPTIONS. (i) ϵ_{ij}'s ($j = 1, \ldots, n_i$) are independently distributed with mean zero and variance $\sigma_\epsilon^2(i)$; $i = 1, \ldots, m$. (ii) β_{ij}'s ($j = 1, \ldots, n_i$) are independently distributed with mean β and variance $\sigma_\beta^2(i)$; $i = 1, \ldots, m$. (iii) ϵ_{ij} and β_{kl} are independent for all i, j, k and l.

These models are widely discussed in the literature. Hyde (1980) applied Model II, with an intercept factor to analyze "OUABAIN DATA" obtained from seventeen kidneys. He considers y_{ij} as the j-th observation of the oxygen consumption of the i-th kidney and x_{ij}, the corresponding sodium reabsorption, after changing the kidney equilibrium with the drug ouabain. The reciprocal of the regression coefficient β, is interpreted as the estimate of the pumping efficiency. In his statistical analysis, Hyde treats the β_i's as realizations of a random variable and notes that the variance of the error term changes from kidney to kidney. We refer to Dielman ((1989), Chapters V and VI), for more references regarding applications.

Section 2 discusses conditionally optimal estimating functions (1.5) for β, for each of the above heterogeneous models. The suggested estimating functions can very well be used in an appropriate semi-parametric set up. This is further supported by our simulation studies.

In Section 3, we deal with the above three models with equal variances over the strata. We also consider another regression model (Model IV) with a random coefficient, given by

Model IV. $$y_i = \beta_i x_i + \epsilon_i, \quad i = 1, \ldots, n;$$

with the same assumptions as that of the earlier models. In this model, there are no replications of x's or β's. The optimal estimating function depends on unknown parameters in the form of weights. Therefore, one usually proceeds to obtain an asymptotically optimal estimating function by replacing nuisance parameters by their consistent estimators. We have considered another estimating function where the nuisance parameters are substituted in terms of functions of the data and parameters of interest only. If β_i and ϵ_i are normally distributed, then this can be viewed as $\hat{W}$ of (1.4). Unlike the first three models, results of this subsection are asymptotic in nature. It is shown that the resulting estimating functions are asymptotically optimal in the sense that both the estimating functions have asymptotically the same variance as that of the estimating function with known weights. This extends a result due to Anh (1988), who assumes that $\sigma_\epsilon^2 = 0$. (It may be pointed out that Anh's results deal with estimators rather than estimating functions.) No distributional assumptions have been made here.

All our results can be easily extended to multiple regression models. The so called mixed effect models discussed in Dielman (1989) can also be seen as a special case of such a multiparameter extension.

2. Models with heterogeneity

2.1 *Analysis of Model I*

Let β_i and ϵ_{ij} be normally distributed with means β and zero respectively and variances $\sigma_\beta^2(i)$ and $\sigma_\epsilon^2(i)$ respectively. For notational convenience, we write σ_β^2 and σ_ϵ^2 instead of $\sigma_\beta^2(i)$ and $\sigma_\epsilon^2(i)$ when a stratum is fixed. The log likelihood function for the i-th stratum is given by

$$\begin{aligned}(2.1)\qquad \log L_i = &- (n_i/2)\log(2\pi) - \{(n_i-1)/2\}\log(\sigma_\epsilon^2)\\ &- (1/2)\log(n_i\sigma_\beta^2 x_i^2 + \sigma_\epsilon^2) - \{1/(2\sigma_\epsilon^2)\}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_i)^2\\ &- \frac{n_i(\bar{y}_i - \beta x_i)^2}{2(n_i\sigma_\beta^2 x_i^2 + \sigma_\epsilon^2)}.\end{aligned}$$

Here, $U = \sum_i n_i x_i(\bar{y}_i - \beta x_i)/(n_i\sigma_\beta^2 x_i^2 + \sigma_\epsilon^2)$ (which coincides with W) and it cannot be directly used for estimation purposes. Let $\hat{\boldsymbol{\theta}}(\beta)$ be the unique maximum likelihood estimator of $\boldsymbol{\theta} = (\sigma_\beta^2, \sigma_\epsilon^2)$ for fixed β. Lindsay (1982) suggests the use of $\hat{U}$ (of (1.2)), if it leads to an unbiased estimating function. Since, in this case, $\hat{U}_i = x_i/(\bar{y}_i - \beta x_i)$, this approach fails here.

Let $S_{i1} = \sum_j (y_{ij} - \bar{y}_i)^2$, $S_{i2} = (\bar{y}_i - \beta x_i)^2$. From (2.1), we see that $S_i(\beta) = (S_{i1}, S_{i2})$ is both complete and sufficient for the nuisance parameter $\boldsymbol{\theta} = (\sigma_\beta^2, \sigma_\epsilon^2)$, when β is fixed. It is clear from Rao ((1984), p. 197) that S_{i1} and S_{i2} are independent. The random variable S_{i1} is distributed as $\sigma_\epsilon^2\chi_{n_i-1}^2$, whereas S_{i2} is distributed as $(\sigma_\beta^2 x_i^2 + \sigma_\epsilon^2/n_i)\chi_1^2$.

Let $H_{ij} = y_{ij} - \beta x_i$. Here, in principle, it is possible to maximize the product of the conditional likelihoods of $(H_{ij}, j = 1, \ldots, n_i)$ given $S_i(\beta)$, $i = 1, \ldots, m$. However, this conditional distribution is too messy to handle, particularly since the range of the distribution depends upon β in a highly complex manner. Secondly, such an estimation procedure would be heavily influenced by the assumption of normality and would tend to be non-robust against departures from normality of $\beta_i - \beta$ and or ϵ_{ij}. We, therefore, adopt the following approach.

Let $\boldsymbol{H}_i = (H_{i1}, \ldots, H_{in_i})' = (y_{i1} - \beta x_i, \ldots, y_{in_i} - \beta x_i)'$. It is easy to see that $E\{H_{ij} \mid S_i(\beta)\} = 0$. Since for fixed β, $S_i(\beta)$ is both complete and sufficient for the family of distributions of $\boldsymbol{H}_i$, $E\{H_{ij}^2 \mid S_i(\beta)\}$ is the uniformly minimum variance unbiased estimator of $E(H_{ij}^2)$. Combined with the distributional properties of S_{i1} and S_{i2} described earlier, we have $E\{H_{ij}^2 \mid S_i(\beta)\} = \mathrm{Var}\{H_{ij} \mid S_i(\beta)\} = S_{i2} + S_{i1}/n_i$ almost surely, $E\{H_{ij}H_{ik} \mid S_i(\beta)\} = \mathrm{Cov}\{H_{ij}, H_{ik} \mid S_i(\beta)\} = S_{i2} - S_{i1}/\{n_i(n_i-1)\}$, almost surely. Since H_{ij}'s, $j = 1, \ldots, n_i$ are conditionally correlated, the estimating function (1.5) is no more optimal. However, it is easy to see that, in this situation, the estimating function $g^* = \sum_{i=1}^m \boldsymbol{E}_i' C_i^{-1}(\beta)\boldsymbol{H}_i$ is conditionally optimal, where $\boldsymbol{E}_i = -(x_i, \ldots, x_i)'$ and $C_i(\beta)$ is the conditional variance-covariance matrix of $\boldsymbol{H}_i$. Simplifying g^* further, we can see that it reduces to $\sum_i(\boldsymbol{E}_i'\boldsymbol{H}_i/n_i S_{i2})$, since $n_i S_{i2}$ is a characteristic root of $C_i(\beta)$ and $\boldsymbol{E}_i$ is the corresponding right characteristic vector. However, note that, $E|g^*| = \infty$. But, this difficulty can be easily dealt with by considering the conditional standardization of g^*, viz., $g_s^* = g^*/E\{(\partial g^*/\partial\beta) \mid S(\beta)\}$, i.e. we consider

$$g_s^* = \sum_{i=1}^{m} \boldsymbol{E}_i'\boldsymbol{H}_i \Big/ \Big\{ n_i S_{i2} \Big(\sum_{k=1}^{m} x_k^2/S_{k2} \Big) \Big\}. \tag{2.2}$$

It is easy to see that $E(g_s^*) = 0$ and $\mathrm{Var}(g_s^*) < \infty$.

The estimating equation (2.2) is solved by using the ordinary least square estimate obtained from

$$g_1 = \sum_{i=1}^{m} n_i x_i (\bar{y}_i - \beta x_i), \tag{2.3}$$

and using the iterative procedure

$$\beta_{n+1} = \frac{\sum_i \bar{y}_i x_i / S_{i2}(\beta_n)}{\sum_i x_i^2 / S_{i2}(\beta_n)},$$

where β_0 is the solution of (2.3). In our simulation study, we restrict to one step correction only. We note that the estimating function (2.2) is reasonable for semi-parametric models also. This follows from the fact that $S_{i2}(\beta)$ simply estimates the weight in the optimal estimating function U. We have obtained the mean square errors of estimates from U (for known values of nuisance parameters), g_s^* (of (2.2)) and g_1 (of (2.3)) for three different distributions of β_i and ϵ_{ij}. The results of simulations are reported in the table given below. Since U merely serves as a bench mark and is not used in practice, we report the ratios of the mean square errors which reflect the magnitude of the loss in efficiency when the nuisance parameters are unknown. (Throughout the Tables 1–4, $\mathrm{mse}(g_1)/\mathrm{mse}(g_2)$ denotes the ratio of the mean square error of the estimate obtained from g_1 to the mean square error of the estimate obtained from g_2.) From Table 1, we can see that the performance of g_s^* is much better than that of g_1, in almost all situations. Note that, even if the variances do not change over the strata, the estimating function (2.2) performs quite well. Further, we note that for distributions other than normal, U is not the score function and may not be a proper bench mark. Thus, in nonnormal situations, it is possible that the MSE ratio of suggested estimating functions be less than one. This is reflected in Table 1. Similar comments apply to the Table 2 also.

2.2 *Analysis of Model II*

Let β_i and ϵ_{ij} be normally distributed with means β and zero respectively, and variances $\sigma_\beta^2(i)$ and $\sigma_\epsilon^2(i)$ respectively. As before, we write σ_β^2 and σ_ϵ^2 instead of $\sigma_\beta^2(i)$ and $\sigma_\epsilon^2(i)$ when a stratum is fixed. Let $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{in_i})'$. Then, $\boldsymbol{y}_i$ has an n_i-variate normal distribution with mean $\boldsymbol{\nu}_i$ and dispersion matrix V_i, where $\boldsymbol{\nu}_i = (\beta x_{i1}, \ldots, \beta x_{in_i})' = \beta \boldsymbol{X}_i'$ and $V_i = \sigma_\beta^2 \boldsymbol{X}_i \boldsymbol{X}_i' + \sigma_\epsilon^2 I$. Without loss of generality, we assume that $\boldsymbol{X}_i'\boldsymbol{X}_i = 1$ for all $i = 1, \ldots, m$. The log likelihood function L_i for

Table 1. Ratios of mse's of estimates obtained from g_s^* and g_1 to U. Number of simulations = 1000; $m = 6$; $n_i = 3, 4, 5, 6, 5, 4$.

Trial number	$\sigma_\beta^2(i)$	$\sigma_\epsilon^2(i)$	β
1	100, 200, 150, 50, 75, 20	10, 20, 15, 25, 10, 100	10
2	2, 4, 6, 8, 16, 32	1, 1, 1, 1, 1, 1	2
3	10, 10, 10, 10, 10, 10	10, 10, 10, 10, 10, 10	5

	$\beta_i \sim$ Normal	$\epsilon_{ij} \sim$ Normal
Trial	mse(g_s^*)/mse(U)	mse(g_1)/mse(U)
1	1.643829	2.855207
2	1.651286	4.429885
3	1.090402	1.547711

	$\beta_i \sim$ Logistic	$\epsilon_{ij} \sim$ Logistic
Trial	mse(g_s^*)/mse(U)	mse(g_1)/mse(U)
1	1.436874	1.169453
2	1.789639	5.481226
3	0.893838	1.372011

	$\beta_i - \beta \sim t_8$	$\epsilon_{ij} \sim t_8$
Trial	mse(g_s^*)/mse(U)	mse(g_1)/mse(U)
1	1.502885	1.472559
2	1.849164	2.339907
3	1.001800	1.325434

Note. The values of x_i's are arbitrarily chosen between 0 and 10. $t_8 \sim$ t-distribution with 8 degrees of freedom.

the i-th stratum is given by

$$\begin{aligned}(2.4)\qquad \log L_i = &-(n_i/2)\log(2\pi) - \{(n_i-1)/2\}\log(\sigma_\epsilon^2)\\ &-(1/2)\log(\sigma_\beta^2+\sigma_\epsilon^2)\\ &-\{(2\sigma_\epsilon^2)^{-1}(\boldsymbol{y}_i-\boldsymbol{\nu}_i)'\{I-(1+\lambda)^{-1}\boldsymbol{X}_i\boldsymbol{X}_i'\}(\boldsymbol{y}_i-\boldsymbol{\nu}_i)\},\end{aligned}$$

where $\lambda = \sigma_\epsilon^2/\sigma_\beta^2$. It follows that

$$U_i = \partial/\partial\beta\{\log L_i\} = \sum_{j=1}^{n_i}\frac{(y_{ij}-\beta x_{ij})x_{ij}}{\sigma_\beta^2(i)+\sigma_\epsilon^2(i)}.$$

It can be easily seen that $U = \sum_i U_i$, does not lead us to an estimating function which is free from the nuisance parameters. Further, even though $\hat{U}$ is free from nuisance parameters, it is not zero unbiased, since $E|\hat{U}_i| = \infty$.

Now, let $\boldsymbol{Z}_i = (Z_{i1}, \ldots, Z_{in_i})'$, where $Z_{ij} = y_{ij} - \beta x_{ij}$, $j = 1, \ldots, n_i$. Let $\boldsymbol{\theta} = (\sigma_\beta^2, \sigma_\epsilon^2)$. From (2.4), we see that $S_i^*(\beta) = (\boldsymbol{Z}_i'\boldsymbol{Z}_i, \boldsymbol{Z}_i'\boldsymbol{X}_i\boldsymbol{X}_i'\boldsymbol{Z}_i)$ is sufficient for the nuisance parameter $\boldsymbol{\theta}$, when β is fixed. Define $S_i(\beta) = (T_{i1}, T_{i2})$, where $T_{i1} = \boldsymbol{Z}_i'\boldsymbol{X}_i\boldsymbol{X}_i'\boldsymbol{Z}_i$ and $T_{i2} = \boldsymbol{Z}_i'(I - \boldsymbol{X}_i\boldsymbol{X}_i')\boldsymbol{Z}_i$. Then, we have the following: (i) T_{i1} and T_{i2} are independent; (ii) $S_i(\beta)$ is a complete sufficient statistic for $\boldsymbol{\theta}$; (iii) T_{i1} is distributed as $(\sigma_\beta^2 + \sigma_\epsilon^2)\chi_1^2$; and (iv) T_{i2} is distributed as $\sigma_\epsilon^2\chi_{n_i-1}^2$ (cf. Rao (1984), p. 188).

In this case, W coincides with U, whereas, $\hat{W}$, given by (1.4) reduces to $\hat{W} = \sum_i \hat{W}_i = \sum_i \sum_j (y_{ij} - \beta x_{ij})x_{ij}/T_{i1}$ is not zero unbiased. In fact $E(|\hat{W}_i|) = \infty$. We, therefore, proceed as follows.

Arguing as before, we have $E\{Z_{ij} \mid S_i(\beta)\} = 0$, almost surely. Further,

$$\begin{aligned} E\{Z_{ij}^2 \mid S_i(\beta)\} &= \operatorname{Var}\{Z_{ij} \mid S_i(\beta)\} \\ &= x_{ij}^2\left(T_{i1} - \frac{T_{i2}}{(n_i-1)}\right) + \frac{T_{i2}}{(n_i-1)}, \end{aligned}$$

almost surely, and

$$\begin{aligned} E\{Z_{ij}Z_{ik} \mid S_i(\beta)\} &= \operatorname{Cov}\{(Z_{ij}, Z_{ik}) \mid S_i(\beta)\} \\ &= \left(T_{i1} - \frac{T_{i2}}{(n_i-1)}\right) x_{ij}x_{ik}, \qquad j \neq k, \end{aligned}$$

almost surely. Thus, the conditionally optimal estimating function is given by $g_1^* = \sum_{i=1}^m \boldsymbol{X}_i'\Omega_i^{-1}(\beta)\boldsymbol{Z}_i$, which simplifies to $g_1^* = \sum_{i=1}^m (\boldsymbol{X}_i'\boldsymbol{Z}_i)/T_{i1}$ (this follows since T_{i1} is a characteristic root of $\Omega_i(\beta)$ and $\boldsymbol{X}_i$ is the corresponding vector). But, again, $E|g_1^*| = \infty$. As before, standardization leads to

$$g_{1s}^* = \sum_{i=1}^m \boldsymbol{X}_i'\boldsymbol{Z}_i \Big/ \left\{T_{i1}\left(\sum_{k=1}^m \boldsymbol{X}_k'\boldsymbol{X}_k/T_{k1}\right)\right\}, \tag{2.5}$$

which is zero unbiased, further $\operatorname{Var}(g_{1s}^*) < \infty$. To solve (2.5), we consider

$$g = \sum_{i=1}^m \sum_{j=1}^{n_i} (y_{ij} - \beta x_{ij})x_{ij}. \tag{2.6}$$

The procedure is the same as described in Subsection 2.1.

Here also, a similar type of simulation study is carried out as explained in Subsection 2.1. The parameters remain the same. From Table 2 it is clear that the performance of the estimating functions (2.5) and (2.6) is quite good for all the three combinations of distributions of β_i and ϵ_{ij}. It is also seen that the estimating function g_{1s}^* has an edge over (2.6). Further, note that even if the variances do not change over the strata, (2.5) performs reasonably well.

Table 2. Ratios of mse's of estimates obtained from g^*_{1s} and g to U. Number of simulations = 1000; $m = 6$; $n_i = 3, 4, 5, 6, 5, 4$.

Trial number	$\sigma^2_\beta(i)$	$\sigma^2_\epsilon(i)$	β
1	100, 200, 150, 50, 75, 20	10, 20, 15, 25, 10, 100	10
2	2, 4, 6, 8, 16, 32	1, 1, 1, 1, 1, 1	2
3	10, 10, 10, 10, 10, 10	10, 10, 10, 10, 10, 10	5

Trial	$\beta_i \sim$ Normal mse(g^*_{1s})/mse(U)	$\epsilon_{ij} \sim$ Normal mse(g)/mse(U)
1	1.237109	1.355899
2	1.200304	1.397028
3	1.074006	1

Trial	$\beta_i \sim$ Logistic mse(g^*_{1s})/mse(U)	$\epsilon_{ij} \sim$ Logistic mse(g)/mse(U)
1	0.858224	1.461172
2	1.248242	1.549602
3	0.856324	1

Trial	$\beta_i - \beta \sim t_8$ mse(g^*_{1s})/mse(U)	$\epsilon_{ij} \sim t_8$ mse(g)/mse(U)
1	1.023723	1.267342
2	1.362298	1.745427
3	1.049131	1

Note. The values of x_{ij}'s are arbitrarily chosen between 0 and 10.

2.3 *Analysis of Model III*

Under the assumption of normality for ϵ_{ij} and β_{ij}, (1.1) and (1.3) give the same estimating function, which is not free from nuisance parameters. Thus, we are naturally led to the conditional procedure. With the natural choice of $h_i = \bar{y}_i - \beta x_i = \sum_j (y_{ij}/n_i) - \beta x_i$, $i = 1, \ldots, m$, it is easy to see that the conditionally optimal estimating function in the class G^* (with respect to $S(\beta) = (S_1(\beta), \ldots, S_m(\beta))$, where $S_i(\beta) = \sum_j (y_{ij} - \beta x_i)^2$) is given by

$$g^*_1 = \sum_{i=1}^{m} \frac{n_i^2 x_i (\bar{y}_i - \beta x_i)}{S_i(\beta)}. \tag{2.7}$$

Note that (2.7) is nothing but $\sum \hat{U}_i(\beta)$. (Here, it is understood that strata for which $n_i = 1$ are ignored.)

One can also use the replications to suggest another estimating function as follows: the sample variance of the i-th stratum is an unbiased estimator of the

Table 3. Ratios of mse's of estimates obtained from g_1^*, $\hat{\beta}_{ls}$ and g_2^* to U. Number of simulations $= 1000$; $m = 6$; For trial 1, $n_i = 3$ for all i, for trial 2, $n_i = 4, 5, 6, 7, 6, 7$ and $n_i = 6, 7, 8, 6, 7, 8$, for trial 3.

Trial number	$\sigma_\beta^2(i)$	$\sigma_\epsilon^2(i)$	β
1	$50, 70, 100, 150, 200, 100$	$10, 20, 30, 40, 50, 20$	25
2	$10, 20, 30, 40, 50, 20$	$100, 70, 50, 200, 150, 100$	25
3	$100, 100, 100, 100, 100, 100$	$10, 10, 10, 10, 10, 10$	25

Trial	$\beta_i \sim$ Normal mse(g_1^*)/mse(U)	$\epsilon_{ij} \sim$ Normal mse($\hat{\beta}_{ls}$)/mse(U)	mse(g_2^*)/mse(U)
1	2.514146	3.003451	2.712723
2	1.299891	1.139620	1.517157
3	1.436803	1.504831	1.333610

Trial	$\beta_i \sim$ Logistic mse(g_1^*)/mse(U)	$\epsilon_{ij} \sim$ Logistic mse($\hat{\beta}_{ls}$)/mse(U)	mse(g_s^*)/mse(U)
1	1.146593	1.205535	1.083448
2	1.422302	1.612142	1.136712
3	1.243003	1.514950	1.002894

Trial	$\beta_i \sim t_8$ mse(g_1^*)/mse(U)	$\epsilon_{ij} \sim t_8$ mse($\hat{\beta}_{ls}$)/mse(U)	mse(g_2^*)/mse(U)
1	2.013145	2.132145	2.190246
2	1.214482	1.120904	1.212345
3	1.761266	2.161171	1.311367

Note. The values of x_i's are arbitrarily chosen between 0 and 10.

nuisance parameter $\sigma_\beta^2(i)x_i^2 + \sigma_\epsilon^2(i)$. Let $S_i^2 = \sum_j (y_{ij} - \bar{y}_i)^2/(n_i - 1)$. The corresponding estimating function is given by

$$g_2^* = \sum_{i=1}^{m} \frac{n_i x_i(\bar{y}_i - \beta x_i)}{S_i^2}. \tag{2.8}$$

This is an unbiased estimating function and its unconditional variance exists if $n_i > 5$. Variances of the standardized versions of the above estimating functions can be easily obtained and are not reported here.

The estimating function (2.8) can be very well applied in a semi-parametric set up also. The function g_1^* (of (2.7)), is optimal for the models, which have $E\{\bar{y}_i \mid S_i(\beta)\} = \beta x_i$ and $\text{Var}\{\bar{y}_i \mid S_i(\beta)\} = S_i(\beta)/n_i$.

In Table 3, we give some simulation results for Model III. The estimating function g_1^* is solved by searching a zero in the confidence interval based on the ordinary least square estimator viz., $\hat{\beta}_{ls} = \sum_i \bar{y}_i x_i / \sum_i x_i^2$. Variance of this estimator is estimated by Wu's weighted jackknife procedure for heteroscedastic regression models (see Wu (1987)). The ratio of the mean square error of $\hat{\beta}_{ls}$ to the mean square error of U is also included in Table 3. It can be noticed that the performance of the conditional estimating function g_1^* as well as g_2^* is quite good.

3. Models with homogeneity

3.1 *The first three models*

Even when the variances do not change over the strata, it is fruitful to follow the approach of conditional estimating functions to avoid estimation of σ_β^2 and σ_ϵ^2. For example, in Model I, an estimating function optimal in an appropriate class is given by

$$g = \sum_{i=1}^{m} \frac{n_i x_i(\bar{y}_i - \beta x_i)}{\{T_1/(N-m)\}\{1 - x_i^2(\sum_i x_i^2)^{-1}\sum_i(1/n_i)\} + T_2 x_i^2(\sum_i x_i^2)^{-1}}$$

where $T_1 = \sum_i \sum_j (y_{ij} - \bar{y}_i)^2$, $T_2 = \sum_i(\bar{y}_i - \beta x_i)^2$ and $N = \sum_i n_i$. The above estimating function is obtained as follows. We have $E(T_1) = \sigma_\epsilon^2(N-m)$, $E(T_2) = \sum_i(\sigma_\beta^2 x_i^2 + \sigma_\epsilon^2/n_i)$. Consider, $h_i = \bar{y}_i - \beta x_i$ and $S(\beta) = (T_1, T_2)$, the sufficient and complete statistic for the model holding β fixed. Note that $E\{h_i \mid S(\beta)\} = 0$, $E\{h_i h_j \mid S(\beta)\} = 0$ for $i \neq j$. Further, $E\{h_i^2 \mid S(\beta)\}$ is given by

$$\frac{1}{n_i}\left\{\frac{T_1}{(N-m)}\left\{1 - \left(\sum_i x_i^2\right)^{-1}\sum_i(1/n_i)x_i^2\right\} + T_2 x_i^2\left(\sum_i x_i^2\right)^{-1}\right\}.$$

Thus, the optimal estimating function g is as given by (1.5). Similarly, an optimal estimating function can be obtained for Model III also. (In the case of Model II, it is clear from the form of the score function that the ordinary least square estimator itself is the optimal estimator).

3.2 *Regression Model IV*

In this section, we restrict ourselves to Model IV. The ordinary least square estimator of β in this case is

$$\hat{\beta}_{ls} = \sum_{i=1}^{n} x_i y_i \left(\sum_{i=1}^{n} x_i^2\right)^{-1}. \tag{3.1}$$

Consider the estimating function $g = \sum_i a_i h_i$, with $h_i = y_i - \beta x_i$. In view of Godambe (1985), the estimating function g, optimal within $\mathcal{G}$, the class of regular unbiased estimating functions which are of the form $\sum_i a_i h_i$, is given by

$$g = \sum_{i=1}^{n} \frac{(y_i - \beta x_i)x_i}{(\sigma_\beta^2 x_i^2 + \sigma_\epsilon^2)}, \tag{3.2}$$

provided $\lambda = \sigma_\beta^2/\sigma_\epsilon^2$ is known. Note that, under the assumption of normality on both β_i and ϵ_i, (3.2) is nothing but the score function. We start with the following lemma.

LEMMA 3.1. *Under the assumption that* $\sum_i(\lambda x_i^2+1)^{-1}x_i^2 \to \infty$ *as* $n \to \infty$, *for some* $\lambda \in (0, \Lambda)$, $\Lambda < \infty$, *the following results hold.*

(i) $b_n^{-1/2} g$ *has an asymptotic normal distribution with mean zero and variance one, where* $b_n = \sum_i \{\sigma_\epsilon^2(\lambda x_i^2+1)\}^{-1} x_i^2$.

(ii) $\hat{\beta}_{\omega l}$, *the estimate obtained by solving* $g = 0$, *is consistent for* β.

PROOF. Part (i) follows from the assumptions and Feller-Lindeberg central limit theorem. Part (ii) follows in a routine manner.

Here, the usual procedure consists of estimating σ_β^2 and σ_ϵ^2 and then substituting these estimates in (3.2). The following discussion shows that, under appropriate conditions on the regressors, this procedure is asymptotically as good as using (3.2) with known $\lambda = \sigma_\beta^2/\sigma_\epsilon^2$.

To estimate σ_β^2 and σ_ϵ^2, consider the residual sum of squares after estimating β by the ordinary least square estimator (3.1). This leads to a regression equation

$$E\left((y_i - \hat{\beta}_{ls}x_i)^2 \left\{1 - \left(\sum_{i=1}^{n} x_i^2\right)^{-1} x_i^2\right\}^{-1}\right) = \sigma_\beta^2 x_i^2 + \sigma_\epsilon^2,$$

which further yields

$$\begin{aligned} \hat{\sigma}_\beta^2 &= \left\{\sum_{i=1}^{n}(Z_{2i} - \bar{Z}_{2n})^2\right\}^{-1} \sum_{i=1}^{n}(Z_{1i} - \bar{Z}_{1n})(Z_{2i} - \bar{Z}_{2n}); \\ \hat{\sigma}_\epsilon^2 &= \bar{Z}_{1n} - \hat{\sigma}_\beta^2 \bar{Z}_{2n}, \end{aligned} \tag{3.3}$$

where $Z_{1i} = (y_i - \hat{\beta}_{ls}x_i)^2\{1 - (\sum x_i^2)^{-1}x_i^2\}^{-1}$, $\bar{Z}_{1n} = n^{-1}\sum Z_{1i}$, $Z_{2i} = x_i^2$ and $\bar{Z}_{2n} = (1/n)\sum_i Z_{2i}$. It is possible that either of $\hat{\sigma}_\beta^2$ and $\hat{\sigma}_\epsilon^2$ (or both) is negative. The simplest remedy would be to use alternative estimator $\hat{\sigma}_\beta^{2*}$ defined by, $\hat{\sigma}_\beta^{2*} = \max(\hat{\sigma}_\beta^2, 0)$, as suggested by Hildreth and Houck (1968). The estimator $\hat{\sigma}_\epsilon^2$ can be modified in a similar manner. If, in a given sample, both the modified estimators happen to be zero, we estimate β by ordinary least square estimator $\hat{\beta}_{ls}$.

The following result presents the relevant properties of (3.3).

LEMMA 3.2. *Suppose that the following hold.*

(i) $E(Y_i^4) < \infty$,

(ii) $\left(\sum_{i=1}^n x_i^2\right)^{-1} x_i^2 \to 0$ *uniformly in* i,

(iii) $\limsup_n \left\{n^{-1}\sum_{i=1}^n (Z_{2i} - \bar{Z}_{2n})^2\right\}^{-1} \bar{Z}_{2n} < \infty$.

Then, each of the estimates $\hat{\sigma}_\beta^2$ *and* $\hat{\sigma}_\epsilon^2$ *defined by* (3.3) *is mean square consistent.*

PROOF. Proof is based on the direct evaluation of the mean square errors.

The following result generalizes a result due to Anh (1988), who discusses the case $\sigma_\epsilon^2 = 0$. It may be pointed out that Anh (1988) discusses properties of the estimator itself rather than those of the estimating function.

THEOREM 3.1. *In addition to the assumptions of Lemma 3.2, suppose that* $\limsup_n \sum x_i^{12}(\lambda^* x_i^2 + 1)^{-8} < \infty$, λ^* *varying in a neighbourhood of* λ. *Consider the estimating function*

$$\hat{g} = \sum_{i=1}^{n} \frac{(y_i - \beta x_i)x_i}{\hat{\sigma}_\beta^2 x_i^2 + \hat{\sigma}_\epsilon^2}. \tag{3.4}$$

Then, (i) $n^{-1/2}(\hat{g} - g) = o_p(1)$, (ii) $\mathrm{Var}(\hat{g}) = \mathrm{Var}(g) + O(n^{-1})$.

PROOF. We have,

$$\begin{aligned} &n^{-1/2}(\hat{g} - g) \\ &\quad = -n^{-1/2}(\hat{\sigma}_\beta^2 - \sigma_\beta^2) \sum_{i=1}^{n} \frac{\delta_i x_i^3}{(\hat{\sigma}_\beta^2 x_i^2 + \hat{\sigma}_\epsilon^2)(\sigma_\beta^2 x_i^2 + \sigma_\epsilon^2)} \\ &\quad\quad - n^{-1/2}(\hat{\sigma}_\epsilon^2 - \sigma_\epsilon^2) \sum_{i=1}^{n} \frac{\delta_i x_i}{(\hat{\sigma}_\beta^2 x_i^2 + \hat{\sigma}_\epsilon^2)(\sigma_\beta^2 x_i^2 + \sigma_\epsilon^2)}, \end{aligned}$$

where $\delta_i = y_i - \beta x_i$. The asymptotic distributions of

$$n^{-1/2} \sum_{i=1}^{n} \frac{\delta_i x_i^3}{(\hat{\sigma}_\beta^2 x_i^2 + \hat{\sigma}_\epsilon^2)(\sigma_\beta^2 x_i^2 + \sigma_\epsilon^2)} \quad \text{and} \quad n^{-1/2} \sum_{i=1}^{n} \frac{\delta_i x_i^3}{(\sigma_\beta^2 x_i^2 + \sigma_\epsilon^2)^2}$$

are the same. Using Liapounov's lemma (see Anh (1988)), $n^{-1/2} \sum_i (\sigma_\beta^2 x_i^2 + \sigma_\epsilon^2)^{-2} \delta_i x_i^3$ converges in distribution to a random variable having normal distribution with mean zero and variance $\lim_{n\to\infty} (\sigma_\epsilon^2 n^{1/3})^{-3} \sum_i (\lambda x_i^2 + 1)^{-3} x_i^6$. The same argument along with the consistency of $\hat{\sigma}_\beta^2$ and $\hat{\sigma}_\epsilon^2$ proved in Lemma 3.2 proves part (i).

Expanding (3.4) at σ_β^2 and σ_ϵ^2 by a Taylor series expansion and making use of the assumptions, it can be shown that $\mathrm{Var}(\hat{g}) = \mathrm{Var}(g) + O(n^{-1})$. This completes the proof.

Another useful estimating function, which avoids the two stage estimation, can be obtained as follows. Consider the substitution of nuisance parameters σ_β^2 and σ_ϵ^2, by appropriate functions of the data and β. More precisely, let

$$\hat{\sigma}_\beta^2(\beta) = \frac{\sum_{i=1}^{n} \left\{ (y_i - \beta x_i)^2 - \left[\sum_{j=1}^{n} (y_j - \beta x_j)^2 / n \right] \right\} \{Z_{2i} - \bar{Z}_{2n}\}}{\sum_{i=1}^{n} (Z_{2i} - \bar{Z}_{2n})^2}$$

and $\hat{\sigma}_\epsilon^2(\beta) = \left\{ \sum_{i=1}^{n} (y_i - \beta x_i)^2 / n \right\} - \hat{\sigma}_\beta^2(\beta) \bar{Z}_{2n}$. Substituting σ_β^2 and σ_ϵ^2 by $\hat{\sigma}_\beta^2(\beta)$ and $\hat{\sigma}_\epsilon^2(\beta)$ respectively in (3.2), we get

$$\hat{g}_1 = \sum_{i=1}^{n} \frac{(y_i - \beta x_i)x_i}{\frac{(Z_{2i} - \bar{Z}_{2n})}{\sum_{i=1}^{n} (Z_{2i} - \bar{Z}_{2n})^2} \{ \sum_{k=1}^{n} (y_k - \beta x_k)^2 (Z_{2k} - \bar{Z}_{2n}) \} + \{ \sum_{j=1}^{n} (y_j - \beta x_j)^2 / n \}}$$

It can be seen that, if β_i and ϵ_i follow a normal distribution, then $\hat{g}_1$ is nothing but $\hat{W}$ defined in (1.4). Proof of the following result is similar to that of Lemma 3.2.

LEMMA 3.3. *Under the assumptions* (a) *and* (b), *viz.*, (a) *assumption* (i) *of Lemma* 3.2,

(b)
$$\lim_{n\to\infty} \frac{\sum_{i=1}^{n} Z_{2i}^2 (Z_{2i} - \bar{Z}_{2n})^2}{\left[\sum_{i=1}^{n} (Z_{2i} - \bar{Z}_{2n})^2\right]^2} = 0,$$

$\hat{\sigma}_\beta^2(\beta)$ and $\hat{\sigma}_\epsilon^2(\beta)$ are mean square consistent for σ_β^2 and σ_ϵ^2 respectively.

A similar result holds for the standardized estimating functions also.

Result similar to Theorem 3.1 establishing the optimality of $\hat{g}_1$ is immediate, as the mean square error consistency of $\hat{\sigma}_\beta^2(\beta)$ and $\hat{\sigma}_\epsilon^2(\beta)$ is assured by the above lemma.

Below, we give Table 4 which gives the ratio of mean square errors of three estimates obtained by solving corresponding estimating equations. It is seen that the mean square errors of the estimates obtained by solving $\hat{g} = 0$ as well as $\hat{g}_1 = 0$ are considerably less than the ordinary least square estimator, irrespective of the distributions of the error term and the random coefficient. Further, both $\hat{g}$ and $\hat{g}_1$ perform reasonably well in comparison with (3.2) which assumes the value of λ to be known.

Table 4. Ratios of mse's of estimates obtained from $\hat{g}$, $\hat{g}_1$ and $\hat{\beta}_{ls}$ to g (based on 1000 simulations with sample size = 50).

Distribution of ϵ_i	Distribution of β_i	mse($\hat{g}$)/mse(g)	mse($\hat{g}_1$)/mse(g)	mse($\hat{\beta}_{ls}$)/mse(g)
Normal$(0, 36)$	Normal$(15, 225)$	1.586911	1.717785	1.781797
Logistic$(0, 36)$	Logistic$(15, 225)$	1.034116	1.073874	1.133878
$(\epsilon_i/6) - t_8$	$(\beta_i - 15)/15 - t_8$	1.023369	1.088428	1.221408

Note. x_i's are chosen in such a way that $0 < x_i < 1$, for all i.

Acknowledgements

The first author wishes to thank the University Grants Commission, India, for an award of a junior research fellowship. We are thankful to both the referees whose comments improved the presentation of the paper.

REFERENCES

Anh, V. V. (1988). On the Hildreth-Houck estimator for a random coefficient regression model, *Austral. J. Statist.*, **30**, 189–195.

Cox, D. R. and Hinkley, D. V. (1974). *Theoretical Statistics*, Chapman and Hall, London.

Dielman, T. E. (1989). *Pooled Cross-sectional and Time Series Data Analysis*, Dekker, New York.

Godambe, V. P. (1976). Conditional likelihood and unconditional optimum estimating equations, *Biometrika*, **63**, 277–284.

Godambe, V. P. (1985). The foundations of finite sample estimation in stochastic processes, *Biometrika*, **72**, 419–428.

Hildreth, C. and Houck, J. P. (1968). Some estimators for a linear model with random coefficients, *J. Amer. Statist. Assoc.*, **63**, 585–595.

Hyde, J. (1980). Determining an average slope, *Biostatistics Casebook* (eds. R. G. Miller, Jr., B. Efron, B. W. Brown, Jr. and L. E. Moses), 171–189, Wiley, New York.

Lindsay, B. G. (1982). Conditional score functions: some optimality results, *Biometrika*, **69**, 503–512.

Mantel, H. J. and Godambe, V. P. (1989). Estimating functions for conditional inference: many nuisance parameter case (submitted for publication).

Rao, C. R. (1984). *Linear Statistical Inference and Its Applications*, Second ed., Wiley Eastern, New Delhi.

Swamy, P. A. V. B. (1971). *Statistical Inference in Random Coefficient Regression Models*, Springer, Berlin.

Wu, C. F. J. (1987). Jackknife, bootstrap and other resampling methods in regression analysis, *Ann. Statist.*, **14**, 1261–1295.