Discussiones Mathematicae
Probability and Statistics 20 (2000) 177–188

LIKELIHOOD AND PARAMETRIC HETEROSCEDASTICITY IN NORMAL CONNECTED LINEAR MODELS

João Tiago Mexia and Pedro Corte Real

Universidade Nova de Lisboa, Departamento de Matemática
da Faculdade de Ciências e Tecnologia,
Quinta da Torre, 2825 Monte de Caparica

e-mail: parcr@mail.fct.unl.pt

Abstract

A linear model in which the mean vector and covariance matrix depend on the same parameters is connected. Limit results for these models are presented. The characteristic function of the gradient of the score is obtained for normal connected models, thus, enabling the study of maximum likelihood estimators. A special case with diagonal covariance matrix is studied.

Keywords: linear model, connected model, normal model, maximum likelihood estimators, score function, Newton-Raphson method.

1991 Mathematics Subject Classification: 62F10; 62J99.

1 Introduction

A linear model

$$y = X\beta + e$$

with y-an n dimensional random vector, X-an $n \times k$ known fixed matrix of rank $k < n$, β-an k dimensional vector of unknown parameters and e-an n dimensional unobservable random vector with null mean value is connected when $V[y] = V[e] = V$ depends on X, β and, possibly, nuissance parameters. In this case we write $V = [g_{ij}(\lambda)]$, with $\lambda' = [\beta'|\tau']$-an s dimensional vector and τ-an h dimensional unknown vector. The $g_{ij}(\lambda)$, $i, j = 1, ..., n$ are real valued functions of λ and the observed constants, present in the matrix X.

The special case in which y is normal and V is diagonal with principal elements that depend on X through the corresponding line vectors was already studied (see [7]). We now obtain limit results that apply in the general case as well as the expression of the characteristic function of the score's gradient when errors are normal. This expression is useful in checking assumptions required in the limit results. We also show how to use the Newton-Raphson method to obtain maximum likelihood estimates in normal connected models. Lastly, we apply our results to a special case similar to the one considered in [7], so that comparisons can be made.

2 Limit Results

We will use the notation $N_l(\lambda)=\{a : ||a-\lambda|| < 1l\}$ and $N_l(\lambda)^*$ for the closure of N_l, and with $Z_n(y,\lambda) = Z_n(\lambda)$ a random variable, $\mu_n(\lambda) = E[Z_n(\lambda)]$ and $\sigma_n^2(\lambda) = V[Z_n(\lambda)]$. I_n will stand for the order n identity matrix, with J-a matrix, J^+ is the Moore-Penrose inverse, if J is square, $|J|$ will stand for the determinant and $i = \sqrt{-1}$. We will also use $\mathbb{R}^p$ to represent the p dimensional Euclidean space. $\overset{P}{\rightarrow}$ and $\overset{L}{\rightarrow}$ will denote convergence in probability and law, respectively. Unless otherwise specified, all limits will be taken as $n \rightarrow \infty$. The Appendix presents or at least outlines the proofs of:

Proposition 21. *If, with λ in an open set B, $\mu_n(\lambda) \rightarrow g(\lambda)$, $\sigma_n^2(\lambda) \rightarrow 0$, and $\forall \epsilon > 0, \exists N_l(\lambda) \subseteq B$, such that $P[\sup_{a \in N_l(\lambda)} |Z_n(a) - Z_n(\lambda)| < \epsilon] \rightarrow 1$, then $g(\lambda)$ is continuous in B.*

Proposition 22. *If $\sigma_n^2(\lambda) \rightarrow 0$, we have $Z_n(\lambda) \overset{P}{\rightarrow} g(\lambda)$ if and only if $\mu_n(\lambda) \rightarrow g(\lambda)$. If, whenever $\lambda \in B$, the conditions of Proposition 21 hold for $\{Z_n(\lambda)\}$, given a compact set $C \subset B$, then $\sup_{\lambda \in C} |Z_n(\lambda) - g(\lambda)| \overset{P}{\rightarrow} 0$. We write $Z_n(\lambda) \overset{P_u(C)}{\rightarrow} g(\lambda)$.*

Corollary 23. *If, whenever $\lambda \in B$, the conditions of Proposition 21 hold for $\{Z_n(\lambda)\}$ and if $\lambda_n \overset{P}{\rightarrow} \lambda$, then $Z_n(\lambda_n) \overset{P}{\rightarrow} g(\lambda)$.*

Propositions 21 and 22 togheter with Corollary 23, allow us to prove Proposition 24, which is a generalization of a result presented by Amemiya in [1, page 106].

Proposition 24. *If, whenever* $\lambda \in B$*, the conditions of Proposition* 21 *hold for* $\{Z_n(\lambda)\}$*, if* $g(\lambda)$ *has a sole maximum;* $\lambda_0 \in B$ *and there is* $C \subset B$*, such that* $\lambda_0 \in C$ *and that* $P[Z_n(\lambda_0) > \sup_{\lambda \in B-C} Z_n(\lambda)] \to 1$*, then for any supremum* $\hat{\lambda}_n$ *of* $Z_n(\lambda)$ *in* B*,* $\hat{\lambda}_n \xrightarrow{P} \lambda_0$.

If, following Amemiya in [1, Chapter 4], we assume that $\partial Z_n(\lambda)/\partial\lambda$ and $\partial^2 Z_n(\lambda)/\partial\lambda\partial\lambda^{'}$ are continuous in λ, from Proposition 24 follows

Corollary 25. *If* $\sqrt{n}\partial Z_n(\lambda_0)/\partial\lambda \xrightarrow{L} N(0,W)$*,* $\partial Z_n(\tilde{\lambda}_n)/\partial\lambda = 0$*,* $\tilde{\lambda}_n \xrightarrow{P} \lambda_0$*, and if, whenever* $\hat{\lambda}_n \xrightarrow{P} \lambda_0$*,* $\partial^2 Z_n(\hat{\lambda}_n)/\partial\lambda\partial\lambda^{'} \xrightarrow{P} K$*, with* K *regular, then* $\sqrt{n}(\tilde{\lambda}_n - \lambda_0) \xrightarrow{L} N(0, K^{-1}WK^{-1})$.

The next corollary allows us to obtain the same asymptotic result for $\hat{\lambda}_n$, but now by checking the conditions of Proposition 21 over the second order partial derivative of $Z_n(\lambda)$ with respect to λ.

Corollary 26. *If*

i) *whenever* $\hat{\lambda}_n$ *is a maximum of* $Z_n(\lambda)$ *in* B*,* $\hat{\lambda}_n \xrightarrow{P} \lambda_0$,

ii) $\sqrt{n}\partial Z_n(\lambda_0)/\partial\lambda \xrightarrow{L} N(0,W)$,

iii) *there is a neighborhood* $N_l(\lambda_0)$ *such that if* $\lambda \in N_l(\lambda_0)$*, the conditions of Proposition* 21 *hold for* $\partial^2 Z_n(\lambda)/\partial\lambda\partial\lambda^{'}$,

iv) *the matrix* $K = \lim E[\partial^2 Z_n(\lambda_0)/\partial\lambda\partial\lambda^{'}]$ *is regular,*

then $\sqrt{n}(\hat{\lambda}_n - \lambda_0) \xrightarrow{L} N(0, K^{-1}WK^{-1})$.

Proposition 24 and its two corollaries will play a central part in our study.

3 Score function for normal models

We now assume the errors e_m, $m = 1, ..., n$, to be normal with null mean values. The score function divided by n will be

$$Z_n(\lambda) = -\ln(2\pi)/2 - \ln(|V|)/(2n) - (y - X\beta)^{'}V^{-1}(y - X\beta))/(2n) \tag{3.1}$$

the maximum likelihood estimates being derived from this function.

To check the convergence in law of $\sqrt{n}\partial Z_n(\lambda_0)/\partial\lambda$, required in Corollaries 25 and 26, we will use later on the characteristic function $\phi_n(t)$ of the s dimensional vector $H = \sqrt{n}\partial Z_n(\lambda_0)/\partial\lambda$. It is easy to write the components of H as $-(h^{'}A_m h + 2b_m^{'}h + c_m)/\sqrt{n}$, where h is an n dimensional normal random vector, with null mean value and covariance matrix J^{-1}, $b_1, ..., b_s$ are fixed vectors, $c_1, ..., c_s$ are constants and $\sum_{j=1}^{s} t_j A_j$ is regular. Using the results on quadratic forms of normal vectors presented in [10, pages 16-18], we may write

$$\phi_n(t) =$$

$$\exp\left[-\frac{i}{n}t^{'}c - \frac{2}{n}t^{'}B^{'}G'\left(I_n + \frac{2i}{n}D\right)^{-1}GBt - \frac{1}{2}\ln\left(|I_n + \frac{2i}{n}D|\right)\right], \tag{3.2}$$

where $B^{'}$ is a matrix with line vectors $b_1, ..., b_s$ and G is such that $GJG^{'} = I_n$ and $G(\sum_{v=1}^{s} A_v t_v)G^{'} = D$, with D diagonal, about these matrices see [9, pages 6–7]. It can be shown that D does not depend on matrix G.

Usually, an explicit maximum likelihood estimator derived from (3.1) is very difficult to obtain. However, under mild conditions, we can use the Newton-Raphson method to obtain a root of the first order partial derivative with respect to λ, that is we can use the equation

$$\lambda_{v+1} = \lambda_v - \left[\frac{\partial^2 Z_n(\lambda_v)}{\partial\lambda'\partial\lambda}\right]^{-1}\frac{\partial Z_n(\lambda_v)}{\partial\lambda}, \quad v = 1, 2, ... \tag{3.3}$$

to generate a sequence convergent to a local maximum of $Z_n(\lambda)$.

To check on assumptions and how to solve some problems regarding the convergence of the method, see for instance [4, 5, 3] and [2].

4 A Special case

We will now add the assumptions:

Assumption 1. The parameter vector $\lambda \in S \subset \mathbb{R}^s$ an open bounded set (so, λ_0 will be an interior point of S).

Assumption 2. The rows of X, denoted by x_m, $m = 1, ..., n$, belong to $U \subset \mathbb{R}^p$ a compact set.

Assumption 3. $V[e] = D(g)$ with D a diagonal matrix and g an n dimensional vector with components, $g_m = \theta_0 x_t \beta_0$, $x_m^{'}\beta_0 > 0$ and $\theta_0 > 0$, $m = 1, ..., n$.

Assumption 4. The matrix limits $n^{-1}X'X$, $n^{-1}X'D(\mu)^{-1}X$ and $n^{-1}X'D(\mu)^{-2}X$ exist and are positive definite for all $\lambda \in S$. With $D(\mu)$ a diagonal matrix, whose main diagonal has components $x_m'\beta$, $m = 1, ..., n$.

In this case, we have for the limit of the natural logarithm of expression (3.2)

$$\lim \ln(\phi_n(t)) = -\frac{1}{2}t' \begin{bmatrix} \frac{1}{\theta_0}A(\mu_0) + \frac{1}{2}B(\mu_0) & 0 \\ 0 & \frac{1}{2\theta_0^2} \end{bmatrix} t$$

with $A(\mu_0) = \lim n^{-1}X'D(\mu_0)^{-1}X$ and $B(\mu_0) = \lim n^{-1}X'D(\mu_0)^{-2}X$, where $D(\mu_0)$ is a diagonal matrix, whose main diagonal has components $x_m'\beta_0$, $m = 1, ..., n$, and so, it's immediate that $\sqrt{n}\partial Z_n(\lambda_0)/\partial\lambda$ converges in distribution to a normal random vector with null mean vector.

Another nice property of this example is that we don't need to calculate the expected value of the second order partial derivatives of (3.1), because, we know that (see [1, pages 14–17] or [12, page 366]), $-E[\partial^2 Z_n(\lambda)/\partial\lambda\partial\lambda'] = E[(\partial Z_n(\lambda)/\partial\lambda)(\partial Z_n(\lambda)\partial/\lambda')]$, making obvious that $E[\partial^2 Z_n(\lambda)/\partial\lambda\partial\lambda']$ is negative definite, each is a necessary condition for the convergence of the sequence defined by (3.3). In this case, we can conclude that $\sqrt{n}(\hat{\lambda}_n - \lambda_0) \xrightarrow{L} N(0, W)$, with

$$W = \begin{bmatrix} (\frac{1}{\theta_0}A(\mu_0) + \frac{1}{2}B(\mu_0))^{-1} & 0 \\ 0 & 2\theta_0^2 \end{bmatrix}. \tag{4.1}$$

We now conduct a simple Monte Carlo study similar to the one presented in [7], making use of expression (3.3) to obtain estimates for the parameter vector $\lambda = [10, 4, -2, \theta]$ with θ taking values in $\{0.1, 0.2, 0.4, 1, 10\}$ (see tables below). We will use the same (40×3) X matrix as in [7] with the first column a column of ones, the second and third columns extracted from, respectively, Table E1 and Table E6 in [11], from this matrix we obtain a second (80×3) X matrix, by duplication of the first one. For each case, 200 multivariate normal e vectors were generated. We used as first element of the sequence (3.3)

$$\hat{\lambda}_1' = [\hat{\beta}', \hat{\theta}] = [((X'X)^{-1}X'y)', n^{-1}(y - \hat{y})'D^{-1}(\hat{\mu})(y - \hat{y})],$$

where $\hat{y} = X\hat{\beta}$ and $D(\hat{\mu})$ is a diagonal matrix, whose main diagonal has components $x_m'\hat{\beta}$, $m = 1, ..., n$.

The following tables summarize our study, where we used $Avrg$ and Var for shorthands of, respectively, the estimates' average and variance.

We obtain, for the sample size 40:

θ=0.1		β_1=10		β_2=4		β_3=-2	
Avrg	Var	Avrg	Var	Avrg	Var	Avrg	Var
.090254	.00056095	9.9616	1.6256	4.003	.0082831	-1.9972	.0048448

θ=0.2		β_1=10		β_2=4		β_3=- 2	
Avrg	Var	Avrg	Var	Avrg	Var	Avrg	Var
.18047	.0022451	9.945	3.2544	4.0043	.016556	-1.9961	.0096744

θ=0.4		β_1=10		β_2=4		β_3=- 2	
Avrg	Var	Avrg	Var	Avrg	Var	Avrg	Var
.36084	.0089928	9.9207	6.5178	4.0063	.033083	-1.9946	.019308

θ=1		β_1=10		β_2=4		β_3=- 2	
Avrg	Var	Avrg	Var	Avrg	Var	Avrg	Var
.90177	.05646	9.8704	16.338	4.0105	4.6914	-1.992	.048085

θ=10		β_1=10		β_2=4		β_3=- 2	
Avrg	Var	Avrg	Var	Avrg	Var	Avrg	Var
9.0456	6.102	9.5126	166.14	4.0462	.82552	-1.9891	.47324

For the sample size 80:

θ=0.1		β_1=10		β_2=4		β_3=-2	
Avrg	Var	Avrg	Var	Avrg	Var	Avrg	Var
.09498	.0002679	9.9705	.85565	4.0001	.0042058	-1.9963	.0026799

θ=0.2		β_1=10		β_2=4		β_3=-2	
Avrg	Var	Avrg	Var	Avrg	Var	Avrg	Var
.18993	.0010684	9.9567	1.7099	4.0002	.0083918	-1.9949	.0053614

θ=0.4		β_1=10		β_2=4		β_3=- 2	
Avrg	Var	Avrg	Var	Avrg	Var	Avrg	Var
.3798	.0042582	9.9358	3.4153	4.0006	.016724	-1.993	.010724

θ=1		β_1=10		β_2=4		β_3=- 2	
Avrg	Var	Avrg	Var	Avrg	Var	Avrg	Var
.94926	.026481	9.889	8.5104	4.0017	.041483	-1.9894	.026788

θ=10		β_1=10		β_2=4		β_3=- 2	
Avrg	Var	Avrg	Var	Avrg	Var	Avrg	Var
9.5088	2.7353	9.4738	82.607	4.023	.39147	-1.9799	.25972

We can immediately check that, even for large values of θ, the Monte Carlo study is consistent with the large-sample theory, while in [7] large variances are not considered. These conclusions continue to hold when the asymptotic covariance matrices are considered. For instance, in the extremes cases of our study; $\theta = 0.1$ and $\theta = 10$, we have for expression (4.1), respectively:

$$\begin{bmatrix} 68.8505 & -3.95221 & 0.66972 & 0 \\ -3.95221 & 0.331564 & -0.167516 & 0 \\ 0.66972 & -0.167516 & 0.195028 & 0 \\ 0 & 0 & 0 & 0.002 \end{bmatrix}$$

and

$$\begin{bmatrix} 6297.5 & -364.48 & 63.1942 & 0 \\ -364.48 & 30.7949 & -15.556 & 0 \\ 63.1942 & -15.556 & 17.9186 & 0 \\ 0 & 0 & 0 & 200 \end{bmatrix}.$$

5 Appendix

***Proof of Proposition* 21.** Given $a \in N_l(\lambda)$ there exists $m \in \aleph$ such that for $n > m$

$$|\mu_n(a) - g(a)| < \epsilon,$$

so that

$$P[|Z_n(a) - g(a)| < 2\epsilon] \geq P[|Z_n(a) - \mu_n(a)| < \epsilon] \geq 1 - \sigma_n^2(a)/\epsilon^2 \to 1,$$

and

$$P[|Z_n(\lambda) - g(\lambda)| < 2\epsilon] \to 1,$$

thus

$$P[|g(a) - Z_n(a)| < 2\epsilon, |Z_n(a) - Z_n(\lambda)| < \epsilon, |Z_n(\lambda) - g(\lambda)| < 2\epsilon] \to 1.$$

The last three inequalities, taken simultaneously, imply

$$|g(a) - g(\lambda)| < 5\epsilon,$$

making $g(\lambda)$ continuous in B. ■

***Proof of Proposition* 22.** If $\sigma_n^2(\lambda) \to 0$, we will have

$$P[|Z_n(\lambda) - \mu_n(\lambda)| \leq \epsilon] \geq 1 - \sigma_n^2(\lambda)/\epsilon^2 \to 1$$

and the first part of the thesis is established.

According to Proposition 21, $g(\lambda)$ will be continuous in B and so, given $\epsilon > 0$, and $\lambda \in B$, there will be $N_t(\lambda)$ such that

$$\sup_{a \in N_t(\lambda)} |g(a) - g(\lambda)| < \epsilon.$$

Thus, with $N_{l^0}(\lambda) = N_l(\lambda) \bigcap N_t(\lambda)$, we will have

$$\sup_{a \in N_{l^0}(\lambda)} |g(a) - g(\lambda)| < \epsilon,$$

as well as

$$P\left[\sup_{a \in N_{l^0}(\lambda)} |Z_n(a) - Z_n(\lambda)| < \epsilon\right] \to 1,$$

and so

$$P\left[\sup_{a \in N_{l^0}(\lambda)} |Z_n(a) - g(a)| < 3\epsilon\right] \geq$$

$$P\left[\sup_{a \in N_{l^0}(\lambda)} |Z_n(a) - Z_n(\lambda)| < \epsilon, |Z_n(\lambda) - g(\lambda)| < \epsilon,\right.$$

$$\left.\sup_{a \in N_{l^0}(\lambda)} |g(\lambda) - g(a)| < \epsilon\right] \to 1$$

since, according to the first part of the thesis, $Z_n(\lambda) \xrightarrow{P} g(\lambda)$.

We can find points $\lambda_j \in B$ such that $B \subseteq \bigcup_{j=1}^{J} N_{l_j}(\lambda_j)$ with λ_j and l_j chosen such that $P[\sup_{a \in N_{l_j}(\lambda_j)} |Z_n(a) - g(a)| < \epsilon] \to 1$ (see [8, Chapter 5]). Now C is a compact contained in B and so, from the open cover $\bigcup_{j=1}^{J} N_{l_j}(\lambda_j)$, we can extract a finite subcover, such that $C \subseteq \bigcup_{j=1}^{K} N_{l_j}(\lambda_j)$. Since

$$\sup_{\lambda \in C} |Z_n(\lambda) - g(\lambda)| \leq \max_{j=1,\ldots,K} \left\{ \sup_{\lambda \in N_{l_j}(\lambda_j)} |Z_n(\lambda) - g(\lambda)| \right\}$$

the thesis follows from ϵ being arbitrary and from

$$P\left[\sup_{\lambda \in C} |Z_n(\lambda) - g(\lambda)| < \epsilon\right] \geq P\left[\bigcap_{j=1}^{K} \sup_{a \in N_{l_j}(\lambda_j)} |Z_n(a) - g(a)| < \epsilon\right] \to 1.$$

■

***Proof of Corollary* 23.** If $\lambda_n \xrightarrow{P} \lambda$, we have $P[\lambda_n \in N_l^*(\lambda)] \to 1$. If $N_l^*(\lambda) \subset B$, which can be achieved with l small enough, it will be a compact contained in B and so, according to Proposition 22

$$Z_n(\lambda) \overset{P_u(N_l^*(\lambda))}{\longrightarrow} g(\lambda),$$

and

$$\begin{aligned} &P[|Z_n(\lambda_n) - g(\lambda_n)| < \epsilon] \geq \\ &P\left[\lambda_n \in N_l^*(\lambda), \sup_{a \in N_l^*(\lambda)} |Z_n(a) - g(a)| < \epsilon\right] \to 1. \end{aligned}$$

Since, according to Proposition 21, $g(\lambda)$ is continuous in B the Slutsky theorem enables us to write $g(\lambda_n) \xrightarrow{P} g(\lambda)$. The thesis follows from ϵ being arbitrary and from

$$\begin{aligned} &P[|Z_n(\lambda_n) - g(\lambda)| < 2\epsilon] \geq \\ &P[|Z_n(\lambda_n) - g(\lambda_n)| < \epsilon, |g(\lambda_n) - g(\lambda)| < \epsilon] \to 1. \end{aligned}$$

■

***Proof of Proposition* 24.** First, note that the existence of a supremum, $\tilde{\lambda}_n$ of $Z_n(\lambda)$, which depends on the random vector y in a measurable way, is guaranted by [6, page 637].

Let $Z_n(\tilde{\lambda}_n) = \sup_{\lambda\in C} Z_n(\lambda)$ with

$$\epsilon = g(\lambda_0) - \max_{\lambda\in C-N_l(\lambda_0)} g(\lambda) \tag{5.1}$$

and define the event

$$E_n = |Z_n(\lambda) - g(\lambda)| < \epsilon/2, \quad \text{for all } \lambda.$$

Then

$$E_n \Rightarrow g(\tilde{\lambda}_n) > Z_n(\tilde{\lambda}_n) - \epsilon/2 \quad \text{and} \quad E_n \Rightarrow Z_n(\lambda_0) > g(\lambda_0) - \epsilon/2. \tag{5.2}$$

But, due to the definition of $\tilde{\lambda}_n$, $Z_n(\tilde{\lambda}_n) \geq Z_n(\lambda_0)$, being so, the first of the implications in (5.2) allows us to write

$$E_n \Rightarrow g(\tilde{\lambda}_n) \geq Z_n(\lambda_0) - \epsilon/2 \tag{5.3}$$

by adding both sides of the inequalities in the second implication of (5.2) and (5.3) we obtain

$$E_n \Rightarrow g(\tilde{\lambda}_n) > g(\lambda_0) - \epsilon. \tag{5.4}$$

Therefore, from (5.1) and (5.4) we can conclude that $E_n \Rightarrow \tilde{\lambda}_n \in N_l(\lambda_0)$, which implies $P[\tilde{\lambda}_n \in N_l(\lambda_0)] \geq P[E_n] \stackrel{P}{\rightarrow} 1$, due to Proposition 21. Since ϵ is arbitrary we can conclude that $\tilde{\lambda}_n \stackrel{P}{\rightarrow} \lambda_0$. Finally, define $Z_n(\hat{\lambda}_n) = \sup_{\lambda\in B} Z_n(\lambda)$, since

$$\left\{Z_n(\lambda_0) > \sup_{\lambda\in B-C} Z_n(\lambda)\right\} \Rightarrow \{\hat{\lambda}_n \in C\}$$

we have

$$P\left[Z_n(\lambda_0) > \sup_{\lambda\in B-C} Z_n(\lambda)\right] \leq P[\hat{\lambda}_n \in C].$$

■

***Proof of Corollary* 25.** We have by a Taylor expansion

$$0 = \partial Z_n(\tilde{\lambda}_n)/\partial\lambda = \partial Z_n(\lambda_0)/\partial\lambda + (\partial^2 Z_n(a)/\partial\lambda\partial\lambda')(\tilde{\lambda}_n - \lambda_0),$$

where $a = \lambda_0 + h(\tilde{\lambda}_n - \lambda_0)$, with $h \in]0,1[$. Now, since $\tilde{\lambda}_n \stackrel{P}{\rightarrow} \lambda_0$ implies $a \stackrel{P}{\rightarrow} \lambda_0$, $\partial^2 Z_n(a)/\partial\lambda\partial\lambda' \stackrel{P}{\rightarrow} K$ thus, see [1, page 111] and [12, page 24], $[\partial^2 Z_n(a)/\partial\lambda\partial\lambda']^+ \stackrel{P}{\rightarrow} K^{-1}$,

$$\begin{aligned}\sqrt{n}(\tilde{\lambda}_n - \lambda_0) &= [\partial^2 Z_n(a)/\partial\lambda\partial\lambda^{'}]^{+}\sqrt{n}\partial Z_n(\lambda_0)/\partial\lambda \\ &\overset{L}{\to} N(0, K^{-1}WK^{-1}),\end{aligned}$$

where the convergence in law is due to $\sqrt{n}\partial Z_n(\lambda_0)/\partial\lambda \overset{L}{\to} N(0,W)$ and to repetead applications of Slutsky's theorem (see, for instance, [12, page 122]). ■

***Proof of Corollary* 26.** From Proposition 21 and Corollary 23 we see that if $\tilde{\lambda}_n \overset{P}{\to} \lambda_0$, we have

$$[\partial^2 Z_n(\tilde{\lambda})/\partial\lambda\partial\lambda^{'}] \overset{P}{\to} K$$

thus the thesis will follow from Corollary 25, since

$$\partial Z_n(\hat{\lambda})/\partial\lambda = 0 = \partial Z_n(\lambda_0)/\partial\lambda + (\partial^2 Z_n(a)/\partial\lambda\partial\lambda^{'})(\hat{\lambda}_n - \lambda_0),$$

where $a = \lambda_0 + h(\hat{\lambda}_n - \lambda_0)$, with $h \in]0,1[$. ■

References

[1] T. Amemiya, *Advanced Econometrics*, Harvard University Press, Harvard 1985.

[2] E.R. Berndt B.H. Hall and J.A. Hausman, *Estimation and Inference in Nonlinear Structural Models*, Annals of Economic and Social Measurement **3** (1974), 653–666.

[3] W.C. Davidson, *Variable Metric Methods for Minimization*, Atomic Energy Commision, Research Development Report ANL-5990, Washington, D.C. 1959.

[4] S.M. Goldfeld, R.E. Quandt and H.F. Trotter, *Maximization by Quadratic Hill-Climbing*, Econometrica **34** (1966), 541–551.

[5] S.M. Goldfeld and R.E. Quandt, *Nonlinear Methods in Econometrics*, North-Holland Publishing, Amsterdam 1972.

[6] R.I. Jennrich, *Asymptotic Properties of Non-Linear Least-Squares Estimators*, The Annals of Mathematical Statistics **40** (1969), 633–643.

[7] J.D. Jobson and W.A. Fuller, *Least Squares Estimation When the Covariance Matrix and Parameter Vector Are Functionally Related*, Journal of the American Statistical Association **75** (1980), 176–181.

[8] J.L. Kelley, *General Topology*, Princeton, New Jersey 1961.

[9] J.T. Mexia, *Linear Models with Partially Controlled Heteroscedasticity*, Trabalhos de Investigação 2. Departamento de Matemática - Faculdade de Ciências e Tecnologia/Universidade Nova de Lisboa 1993.

[10] R.L. Plackett, *Principles of Regression Analysis*, Oxford University Press, Oxford 1960.

[11] S.J. Prais and H.S. Houthakker, *The Analysis of Family Budgets*, Cambridge University Press, Cambridge 1955.

[12] C.R. Rao, *Linear Statistical Inference and Its Applications (Second Edition)*, John Wiley & Sons 1973.

Received 19 April 1999
Revised 15 March 2000