Ann. Inst. Statist. Math.
Vol. 47, No. 2, 351–369 (1995)

STOCHASTIC REGRESSION MODEL WITH HETEROSCEDASTIC DISTURBANCE*

DER-SHIN CHANG AND GUAN-CHYUN LIN

Institute of Statistics, National Tsing Hua University, Hsinchu, Taiwan 30043, R.O.C.

(Received February 28, 1994; revised September 12, 1994)

Abstract. This paper discusses some properties of stochastic regression model with continuous form of heteroscedastic disturbance. The strong consistency and asymptotic normality of a generalized weighted least squares estimate will be investigated under certain conditions on the stochastic regressors and errors. More, the linear hypothesis testing problem also be discussed and an example to be demonstrated to reestablish the results of Cheng and Chang (1990, Tech. Report, National Tsing Hua University).

Key words and phrases: Stochastic regressors, generalized weighted least squares, strong consistency, asymptotic normality, linear hypothesis, martingales, heteroscedastic disturbance.

1. Introduction

Consider the multiple regression model

$$y_n = \beta_1 x_{n1} + \beta_2 x_{n2} + \cdots + \beta_p x_{np} + \epsilon_n, \qquad n = 1, 2, \dots \tag{1.1}$$

where ϵ_n's are unobservable random errors, $\beta_1, \dots, \beta_p$ are unknown parameters, and y_n is the observed response corresponding to the design levels $x_{n1}, x_{n2}, \dots, x_{np}$. Let $x_n = (x_{n1}, \dots, x_{np})^T$, $X_n = (x_{ij})_{1 \le i \le n, 1 \le j \le p}$, $E_n = (\epsilon_1, \dots, \epsilon_n)^T$ and $Y_n = (y_1, \dots, y_n)^T$. The regression model (1.1) can be written as $Y_n = X_n\beta + E_n$ where $\beta = (\beta_1 \cdots \beta_p)^T$ and the ordinary least squares estimate of β based on the $x_1, y_1, \dots, x_n, y_n$ is $b_n = (b_{n1} \cdots b_{np})^T = (X_n^T X_n)^{-1} X_n^T Y_n$, assuming $X_n^T X_n$ nonsingular. Suppose that $\{\epsilon_n\}$ is a martingale difference sequence with respect to an increasing sequence of σ-fields $\{\mathcal{F}_n\}$ i.e. ϵ_n is $\mathcal{F}_n$-measurable and $E(\epsilon_n \mid \mathcal{F}_{n-1}) = 0$ for every n, more, assume that the design vector x_n at stage n depends on the previous observations $x_1, y_1, \dots, x_{n-1}, y_{n-1}$; i.e. x_n is $\mathcal{F}_{n-1}$-measurable. In the case of homoscedastic disturbance, i.e. $E(\epsilon_n^2 \mid \mathcal{F}_{n-1}) = \sigma^2$ for every n, the strong consistency and asymptotic normality of the least squares estimate b_n has

* Supported by the National Science Council Grant No. 810208M763 at National Tsing Hua University.

been established by Lai and Wei (1982), similar properties were also studied by Cheng and Chang (1990) under some regular conditions for the special case of linear heteroscedasticity.

In this paper we shall study the properties of a more general form, the continuous heterosedastic disturbance, which will generalize those of the above two cases. In Section 3 we try to establish the strong consistency and asymptotic normality of a generalized weighted least squares estimate under certain conditions on the stochastic regressors and errors. In Section 4 we will discuss the linear hypothesis testing problem in this model. In Section 5 we reestablish the results of Cheng and Chang (1990) from Corollary 5.1, Theorems 3.2, 3.4 and 4.1.

2. Some reviews

In this sections we review some important results in Lai and Wei (1982). In model (1.1), they assume:

$$\sup_n E\{|\epsilon_n|^\alpha \mid \mathcal{F}_{n-1}\} < \infty \qquad \text{a.s. for some } \alpha > 2 \tag{2.1}$$

and

$$\lim_{n\to\infty} E\{\epsilon_n^2 \mid \mathcal{F}_{n-1}\} = \sigma^2 \qquad \text{a.s. for some positive constant } \sigma^2. \tag{2.2}$$

A special case of (2.2) is $E(\epsilon_n^2 \mid \mathcal{F}_{n-1}) = \sigma^2$ for every n. Let $\lambda_{\max}(A)$, $\lambda_{\min}(A)$, $\lambda_{\max}(n)$ and $\lambda_{\min}(n)$, respectly, denotes the maximum and minimum eigenvalue of matrix A and $X_n^T X_n$. They established the strong consistency of $b_n = (X_n^T X_N)^{-1} X_n^T Y_n$ under the assumption that $\lambda_{\min}(n)$ tend to infinity faster than $\log \lambda_{\max}(n)$ in the following result.

THEOREM 2.1. *Suppose that in the regression model* (1.1), $\{\epsilon_n\}$ *satisfies* (2.1) *and*

$$\lambda_{\min}(n) \to \infty \qquad a.s.,$$
$$\log \lambda_{\max}(n) = o(\lambda_{\min}(n)) \qquad a.s.$$

then

$$b_n \to \beta \qquad a.s.$$

This result follows from the key

LEMMA 2.1. *Let* $\{\epsilon_n\}$ *be a martingale difference sequence with respect to an increasing sequence of* σ*-fields* $\{\mathcal{F}_n\}$ *such that* $\sup_n E(|\epsilon_n|^\alpha \mid \mathcal{F}_{n-1}\} < \infty$ *a.s. for some* $\alpha > 2$. *Let* $x_{n1}, \ldots, x_{np}$ *be* $\mathcal{F}_{n-1}$*-measurable random variable for every* n. *Define* $N = \inf\{n : X_n^T X_n$ *is nonsingular*$\}$ *and* $Q_n = E_n^T X_n (X_n^T X_n)^{-1} X_n^T E_n$, *assume* $N < \infty$ *a.s. Then for* $n \geq N$ *and on* $\{\lim_{n\to\infty} \lambda_{\max}(n) = \infty\}$ *we have*

$$Q_n = O(\log \lambda_{\max}(n)) \qquad a.s.$$

3. Estimation problem

In this section we will discuss the same estimation problem under that $E(\epsilon_n^2 \mid \mathcal{F}_{n-1}) = g(z_n, \theta)$, where z_n is observable and $\mathcal{F}_{n-1}$-measurable, z_n and θ belong to the k-dimensional Euclidean space, g is a real-valued known function of z_n and θ (θ unknown parameter vector). Let $s_n^2 = g(z_n, \theta)$, $\eta_n = y_n/s_n$, $w_n = x_n/s_n = (x_{n1}/s_n \cdots x_{np}/s_n)^T$ and $e_n = \epsilon_n/s_n$ for every n, then we can rewrite model (1.1) as

$$\eta_n = w_n^T\beta + e_n, \qquad n = 1, 2, \ldots. \tag{3.1}$$

Note that $E * (e_n^2 \mid \mathcal{F}_{n-1}) = 1$ for every n and the weighted least squares estimate of β, denoted by $\hat{\beta}_n$, in model (1.1) is just the ordinary least squares estimate of β in model (3.1). Since θ is unknown, we substitute the $\hat{\theta}_n$ described below into $\hat{\beta}_n$, and get a generalized weighted squares estimates $\hat{\beta}_n^*$ of β. In order to prove our main results, we need following Lemmas.

LEMMA 3.1. (Jennrich (1969)) *Let f be a real valued function on $\Theta \times Y$ where Θ be a compact subset of a Euclidean space and Y is a measurable space. For each θ in Θ, let $f(\theta, y)$ be a measurable function of y and a continuous function of θ for each y in Y. Then there exists a measurable function $\hat{\theta}$ from Y into Θ such that for all y in Y*

$$f(\hat{\theta}(y), y) = \inf_{\theta \in \Theta} f(\theta, y).$$

LEMMA 3.2. (Lai and Wei (1982), (4.15)) *Suppose that $\{\epsilon_n\}$ is a martingale difference sequence with respect to an increasing sequence of σ-fields $\{\mathcal{F}_n\}$ such that (2.1) holds, then*

$$\sum_{i=1}^{n} \epsilon_i^2 = \sum_{i=1}^{n} E(\epsilon_i^2 \mid \mathcal{F}_{i-1}) + o(n) \qquad a.s.$$

LEMMA 3.3. *Let Δ, Θ be subsets of k-dimensional Euclidean space, more, assume that Δ is a compact set. If $f(z, \rho)$ is real valued function which is continuous on $\Delta \times \Theta$, then $\sup_{z\in\Delta} f(z, \rho)$ and $\inf_{z\in\Delta} f(z, \rho)$ are continuous functions of ρ.*

PROOF. See Appendix.

Now since $\hat{E}_n = Y_n - X_n b_n = (I_n - X_n(X_n^T X_n)^{-1} X_n^T) E_n$ where we denote $\hat{E}_n = [\hat{e}_1, \hat{e}_2, \ldots, \hat{e}_n]^T$. Define the least squares estimate $\hat{\theta}_n$ of θ to be the value of ρ that minimizing $Q_n(\rho) = \sum_{t=1}^{n} [\hat{e}_t^2 - g(z_t, \rho)]^2/n$, that is $Q_n(\hat{\theta}_n) = \inf_{\rho\in\Theta} Q_n(\rho)$.

THEOREM 3.1. *Suppose that in the regression model (1.1), $\{\epsilon_n)$ is a martingale difference sequence with respect to an increasing sequence of σ-fields $\{\mathcal{F}_n\}$*

such that (2.1) *holds and* $E(\epsilon_n^2 \mid \mathcal{F}_{n-1}) = g(z_n, \theta)$ *for every* n, *where* g *is a real valued known function of* z_n *and unknown parameter vector* θ. *We assume*

(i) *the parameter vector* θ *is contained in a bounded open sphere* S *where the closure of* S *is denoted by* Θ,

(ii) $\sup_n \|z_n\|_k < \infty$ *a.s., where* $\|\cdot\|_k$ *denotes the* k*-dimensional Euclidean norm,*

(iii) g *is continuous function on* $R^k \times \Theta$ *and* $g(z_n, \rho) > 0 \forall \rho \in \{\theta, \hat{\theta}_1, \ldots, \hat{\theta}_n, \ldots\}$ *a.s.,*

(iv) *the quantity*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} (g(z_t, \theta) - g(z_t, \rho))^2$$

has a unique minimum at $\rho = 0$ *a.s.,*

(v) $\log \lambda_{\max}(n)/n \to 0$ *a.s.*

Then we have the least squares estimate $\hat{\theta}_n \to \theta$ *a.s.*

PROOF. (a) Note first that the existence and measurability of the least squares estimate of θ would be followed from Lemma 3.1. Next let $\phi_m = \{w : \sup_n \|z_n(w)\|_k \le m\}$ and $\Delta_m = \{z \in R^k : \|z\|_k \le m\}$, then we have $P(\bigcup_{m=1}^{\infty} \phi_m) = 1$ and $z_n \in \Delta_m$ for all n on the set ϕ_m.

Since g is a continuous function on $\Delta_m \times \Theta$ and $\Delta_m \times \Theta$ is compact, there exists L_2 such that $g(z_n, \rho) \le L_2 < \infty \forall n$, $\rho \in \Theta$ on the set ϕ_m for any m, thus *we can assume without loss of generality in subsequence that* $\sup_n \|z_n(w)\|_k \le m$ *a.s.*

For every $\rho_n \in \Theta$, let $A_n = \operatorname{diag}(\sqrt{g(z_1, \rho_n)}, \ldots, \sqrt{g(z_n, \rho_n)})$. Then we have

$$(3.2)\quad \begin{aligned} \frac{1}{n}\sum_{t=1}^{n} \hat{e}_t^2 g(z_t, \rho_n) &= \frac{1}{n}(A_n \hat{E}_n)^T (A_n \hat{E}_n) \\ &= \frac{1}{n} E_n^T A_n^2 E_n - \frac{1}{n} E_n^T A_n^2 X_n (X_n^T X_n)^{-1} X_n^T E_n \\ &\quad - \frac{1}{n} E_n^T X_n (X_n^T X_n)^{-1} X_n^T A_n^2 E_n \\ &\quad + \frac{1}{n} E_n^T X_n (X_n^T X_n)^{-1} X_n^T A_n^2 X_n (X_n^T X_n)^{-1} X_n^T E_n. \end{aligned}$$

Since the 4th term of (3.2) equals to

$$(3.3)\quad \begin{aligned} \|A_n X_n (X_n^T X_n)^{-1} X_n^T E_n\|^2 &\le \frac{1}{n} \lambda_{\max}(A_n^2) \|X_n (X_n^T X_n)^{-1} X_n^T E_n\|^2 \\ &\le \frac{1}{n} L_2 \cdot O(\log \lambda_{\max}(n)) \qquad \text{(by Lemma 2.1)} \\ &= o(1) \qquad \text{a.s.} \end{aligned}$$

Then by Cauchy-Schwarz inequality, Lemma 3.2 and (3.3), we have the square of the 2nd term of (3.2)

$$\frac{1}{n^2} (E_n^T A_n^2 X_n (X_n^T X_n)^{-1} X_n^T E_n)^2$$

$$
\begin{aligned}
&\le \frac{1}{n^2}(E_n^T X_n (X_n^T X_n)^{-1} X_n^T A_n^2 X_n (X_n^T X_n)^{-1} X_n^T E_n) \cdot (E_n^T A_n^2 E_n) \\
&\le \left(L_2 \cdot \frac{1}{n} E_n^T E_n \right) \cdot \frac{1}{n} \| A_n X_n (X_n^T X_n)^{-1} X_n^T E_n \|^2 \\
&\le \left[\frac{L_2}{n} \sum_{t=1}^{n} g(z_t, \theta) + o(1) \right] \cdot o(1) \\
&\le [L_2^2 + o(1)] \cdot o(1) = o(1) \qquad \text{a.s.}
\end{aligned}
$$

Similarly, we have the squares of the 3rd term of (3.2)

$$
\frac{1}{n^2}(E_n^T X_n (X_n^T X_n)^{-1} X_n^T A_n^2 E_n)^2 = o(1) \qquad \text{a.s.}
$$

Hence from (3.2) we obtain

$$
\frac{1}{n} \sum_{t=1}^{n} \hat{e}_t^2 g(z_t, \rho_n) = \frac{1}{n} \sum_{t=1}^{n} \epsilon_t^2 g(z_t, \rho_n) + o(1) \qquad \text{a.s.} \tag{3.4}
$$

Take $\rho_n = \rho$, we have

$$
\frac{1}{n} \sum_{t=1}^{n} \hat{e}_t^2 g(z_t, \rho) = \frac{1}{n} \sum_{t=1}^{n} \epsilon_t^2 g(z_t, \rho) + o(1) \qquad \text{a.s.}
$$

Let $a_n = \epsilon_n \cdot \sqrt{g(z_n, \rho)}$ for every n, then $\{a_n\}$ is a martingale difference with respect to $\{\mathcal{F}_n\}$ and $\sup E(|a_n|^\alpha \mid \mathcal{F}_{n-1}) < \infty$ a.s. for some $\alpha > 2$. Then by (3.4) and Lemma 3.2, we have

$$
\begin{aligned}
\frac{1}{n} \sum_{t=1}^{n} \hat{e}_t^2 g(z_t, \rho) &= \frac{1}{n} \left[\sum_{t=1}^{n} E(a_t^2 \mid \mathcal{F}_{t-1}) + o(n) \right] + o(1) \\
&= \frac{1}{n} \sum_{t=1}^{n} g(z_t, \theta) g(z_t, \rho) + o(1) \qquad \text{a.s.} \quad \text{for all } \rho \in \Theta.
\end{aligned}
$$

In particular,

$$
\frac{1}{n} \sum_{t=1}^{n} \hat{e}_t^2 g(z_t, \theta) = \frac{1}{n} \sum_{t=1}^{n} g^2(z_t, \theta) + o(1) \qquad \text{a.s.} \tag{3.5}
$$

It implies that

$$
\frac{1}{n} \sum_{t=1}^{n} (\hat{e}_t^2 - g(z_t, \theta))(g(z_t, \theta) - g(z_t, \rho)) = o(1) \qquad \text{a.s.}
$$

(b) Since

$$
\begin{aligned}
Q_n(\hat{\theta}_n) &= \inf_{\rho\in\Theta} \frac{1}{n}\sum_{t=1}^{n}(\hat{e}_t^2 - g(z_t,\rho))^2 \\
&= \inf_{\rho\in\Theta} \frac{1}{n}\sum_{t=1}^{n}(\hat{e}_t^2 - g(z_t,\theta) + g(z_t,\theta) - g(z_t,\rho))^2 \\
&= \inf_{\rho\in\Theta} \Bigg[\frac{1}{n}\sum_{t=1}^{n}(\hat{e}_t^2 - g(z_t,\theta))^2 \\
&\qquad + \frac{2}{n}\sum_{t=1}^{n}(\hat{e}_t^2 - g(z_t,\theta))(g(z_t,\theta) - g(z_t,\rho)) \\
&\qquad + \frac{1}{n}\sum_{t=1}^{n}(g(z_t,\theta) - g(z_t,\rho))^2\Bigg] \\
&= Q_n(\theta) + o(1) \quad \text{a.s.,}
\end{aligned}
\tag{3.6}
$$

by (a) and condition (iv).

(c) Since Θ is compact, let $\{\hat{\theta}_{n_k}\}$ be any convergent subsequence of $\{\hat{\theta}_n\}$ and say $\hat{\theta}_{n_k} \to \theta^*$. By (3.4) and (3.5), we have

$$
\begin{aligned}
&\frac{1}{n_k}\sum_{t=1}^{n_k}\hat{e}_t^2 g(z_t,\hat{\theta}_{n_k}) \\
&\quad = \frac{1}{n_k}\sum_{t=1}^{n_k}\epsilon_t^2 g(z_t,\hat{\theta}_{n_k}) + o(1) \quad \text{a.s.} \\
&\quad = \frac{1}{n_k}\sum_{t=1}^{n_k}\epsilon_t^2[g(z_t,\hat{\theta}_{n_k}) - g(z_t,\theta^*)] + \frac{1}{n_k}\sum_{t=1}^{n_k} g(z_t,\theta)g(z_t,\theta^*) + o(1) \quad \text{a.s.}
\end{aligned}
$$

By Lemmas 3.2 and 3.3, and $\hat{\theta}_{n_k} \to \theta^*$, we have

$$
\begin{aligned}
&\left|\frac{1}{n_k}\sum_{t=1}^{n_k}\epsilon_t^2[g(z_t,\hat{\theta}_{n_k}) - g(z_t,\theta^*)]\right| \\
&\quad \le \frac{1}{n_k}\sum_{t=1}^{n_k}\epsilon_t^2|g(z_t,\hat{\theta}_{n_k}) - g(z_t,\theta^*)| \\
&\quad \le \frac{1}{n_k}\sup_{z_t\in\Delta_m}|g(z_t,\hat{\theta}_{n_k}) - g(z_t,\theta^*)| \cdot \sum_{t=1}^{n_k}\epsilon_t^2 \\
&\quad = \sup_{z_t\in\Delta_m}|g(z_t,\hat{\theta}_{n_k}) - g(z_t,\theta^*)| \cdot \left(\frac{1}{n_k}\sum_{t=1}^{n_k} g(z_t,\theta) + o(1)\right) \quad \text{a.s.} \\
&\quad = \sup_{z_t\in\Delta_m}|g(z_t,\hat{\theta}_{n_k}) - g(z_t,\theta^*)| \cdot (L_2 + o(1)) \\
&\quad \to \sup_{z_t\in\Delta_m}|g(z_t,\theta^*) - g(z_t,\theta^*)| \cdot (L_2 + o(1)) = o(1) \quad \text{a.s.,}
\end{aligned}
$$

it implies that

$$\frac{1}{n_k}\sum_{t=1}^{n_k}\hat{e}_t^2 g(z_t,\hat{\theta}_{n_k}) = \frac{1}{n_k}\sum_{t=1}^{n_k} g(z_t,\theta)g(z_t,\theta^*) + o(1) \qquad \text{a.s.}$$

Similarly,

$$\frac{1}{n_k}\sum_{t=1}^{n_k} g(z_t,\theta)[g(z_t,\hat{\theta}_{n_k}) - g(z_t,\theta^*)] = o(1) \qquad \text{a.s.}$$

Thus

$$\frac{1}{n_k}\sum_{t=1}^{n_k}[\hat{e}_t^2 - g(z_t,\theta)][g(z_t,\theta) - g(z_t,\hat{\theta}_{n_k})] = o(1) \qquad \text{a.s.}$$

and by the same arguments of (3.6), we have

$$\frac{1}{n_k}\sum_{t=1}^{n_k}(g(z_t,\theta) - g(z_t,\hat{\theta}_{n_k}))^2 = o(1) \qquad \text{a.s.}$$

Further,

$$\begin{aligned}
&\lim_{n_k\to\infty}\frac{1}{n_k}\sum_{t=1}^{n_k}(g(z_t,\theta^*) - g(z_t,\hat{\theta}_{n_k}))^2 \\
&\quad\le \lim_{n_k\to\infty}\sup_{z_t\in\Delta_m}(g(z_t,\theta^*) - g(z_t,\hat{\theta}_{n_k}))^2 \\
&\quad= 0, \qquad \text{by Lemma 3.3 and } \hat{\theta}_{n_k}\to\theta^*.
\end{aligned}$$

Hence

$$\begin{aligned}
&\frac{1}{n_k}\sum_{t=1}^{n_k}(g(z_t,\theta^*) - g(z_t,\theta))^2 \\
&\quad= \frac{1}{n_k}\sum_{t=1}^{n_k}[(g(z_t,\theta^*) - g(z_t,\hat{\theta}_{n_k})) + (g(z_t,\hat{\theta}_{n_k}) - g(z_t,\theta))]^2 \\
&\quad= \frac{1}{n_k}\sum_{t=1}^{n_k}(g(z_t,\theta^*) - g(z_t,\hat{\theta}_{n_k}))^2 \\
&\qquad+ \frac{2}{n_k}\sum_{t=1}^{n_k}(g(z_t,\hat{\theta}_{n_k}) - g(z_t,\theta))(g(z_t,\theta^*) - g(z_t,\hat{\theta}_{n_k})) \\
&\qquad+ \frac{1}{n_k}\sum_{t=1}^{n_k}(g(z_t,\hat{\theta}_{n_k}) - g(z_t,\theta))^2 \\
&\quad\le 2\cdot 2L_2\cdot\frac{1}{n_k}\sum_{t=1}^{n_k}|g(z_t,\theta^*) - g(z_t,\hat{\theta}_{n_k})| + o(1) \\
&\quad\le 4L_2\cdot\sup_{z_t\in\Delta_m}|g(z_t,\theta^*) - g(z_t,\hat{\theta}_{n_k})| + o(1)\to 0 \qquad \text{as} \quad n_k\to\infty \qquad \text{a.s.}
\end{aligned}$$

thus $\theta^* = \theta$ a.s. by condition (iv). Since the null set for the subsequence $\{\hat{\theta}_{n_k}\}$ to be convergent on its complement can be chosen to be independent of $\{n_k\}_{k\in N}$, we obtain $\hat{\theta}_n = \theta + o(1)$ a.s. and then $\hat{\theta}_n \to \theta$ a.s. □

LEMMA 3.4. *Let x_n be the design vector at stage n. If $\sup_n \|x_n\|_p < \infty$ a.s., then $\lambda_{\max}(n) = O(n)$ a.s.*

PROOF. See Appendix.

COROLLARY 3.1. *The result of Theorem 3.1 still holds when the condition (v) is replaced by $\sup_n \|x_n\|_p < \infty$ a.s.*

After getting the strong consistent estimate $\hat{\theta}_n$ of θ, we can define the generalized weighted least squares estimate of β for (1.1) as

$$\hat{\beta}_n^* = (X_n^T G_n^* X_n)^{-1} X_n^T G_n^* Y_n \tag{3.7}$$

where $G_n^* = \operatorname{diag}[\frac{1}{g(z_1,\hat{\theta}_n)}, \ldots, \frac{1}{g(z_n,\hat{\theta}_n)}]$. Note that the weighted least squares estimate of β for (1.1) is

$$\hat{\beta}_n = (X_n^T G_n X_n)^{-1} X_n^T G_n Y_n$$

with $G_n = \operatorname{diag}[\frac{1}{g(z_1,\theta)}, \ldots, \frac{1}{g(z_n,\theta)}]$.

In order to investigate the strong consistency of $\hat{\beta}_n^*$ under some regular conditions on the regression matrix X_n, we need the following Lemmas.

LEMMA 3.5. *Let B is $n \times p$ matrix. If C is $n \times n$ symmetric matrix, then $\lambda_i((\lambda_{\min}C) \cdot B^T B) \le \lambda_i(B^T C B) \le \lambda_i((\lambda_{\max}C) \cdot B^T B)$, where $\lambda_1(M) \ge \cdots \ge \lambda_p(M)$, M is any $p\times p$ matrix and $\lambda_i(M)$ is the i-th eigenvalue of M. In particular, if C is nonnegative definite, then for $i = 1, 2, \ldots, p$ we have*

$$(\lambda_{\min}C) \cdot \lambda_i(B^T B) \le \lambda_i(B^T C B) \le (\lambda_{\max}C)\lambda_i(B^T B).$$

PROOF. See Appendix.

LEMMA 3.6. *Let A_n, B_n be invertible symmetric $p \times p$ matrices and $A_n = B_n + o(1)$. If $\liminf_n \lambda_{\min}(A_n) > 0$, then $A_n^{-1} = B_n^{-1} + o(1)$.*

PROOF. See Appendix.

LEMMA 3.7. *If $0 < L_1 \le g(z_n, \rho) \le L_2 < \infty \forall n$, $\forall \rho \in \{\theta, \hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_n, \ldots\}$ and $\sup_n \|x_n\|_p < \infty$ a.s., then we have*

$$(1)\quad \frac{1}{n} X_n^T G_n^* E_n = \frac{1}{n} X_n^T G_n E_n + o(1) \quad a.s. \quad and$$

$$(2)\quad \frac{1}{n} X_n^T G_n^* X_n = \frac{1}{n} X_n^T G_n X_n + o(1) \quad a.s.$$

Furthermore, $\liminf_n \lambda_{\min}(n)/n > 0$, *then*

$$\left(\frac{1}{n}X_n^T G_n^* X_n\right)^{-1} = \left(\frac{1}{n}X_n^T G_n X_n\right)^{-1} + o(1) \qquad a.s.$$

PROOF. See Appendix.

THEOREM 3.2. *Suppose that in the regression model* (1.1), $\{\epsilon_n\}$ *is a martingale difference sequence with respect to an increasing sequence of* σ*-fields* $\{\mathcal{F}_n\}$ *such that* (2.1) *holds and* $E(\epsilon_n^2 \mid \mathcal{F}_{n-1}) = g(z_n, \theta)$, *with* g *a known function and* θ *unknown parameter vector. If the regression matrix* X_n *and* $g(z_n, \theta)$ *satisfy conditions* (i), (ii), (iii), (iv) *of Theorem* 3.1 *and the more extra conditions*

(v) $\sup_n \|x_n\|_p < \infty$ *a.s. and*

(vi) $\liminf_n \lambda_{\min}(n)/n > 0$ *a.s.*

then $\hat{\beta}_n^* \to \beta$ *a.s.*

PROOF. Let ϕ_m, L_2 as in the proof of Theorem 3.1. By Lemma 3.3 and $\hat{\theta}_n \to \theta$, we have $\inf_t g(z_t, \hat{\theta}_n) \to \inf_t g(z_t, \theta)$ a.s. Since $\inf_t g(z_t, \rho) > 0 \forall \rho \in \{\theta, \hat{\theta}_1, \ldots, \hat{\theta}_n, \ldots\}$ a.s. we take $0 < \epsilon < \inf_t g(z_t, \theta)$. For this ϵ, there exists N such that

$$\inf_t g(z_t, \hat{\theta}_n) > \inf_t g(z_t, \theta) - \epsilon \qquad \forall n > N,$$

take $L_1 = \min\{\inf_t g(z_t, \hat{\theta}_1), \inf_t g(z_t, \hat{\theta}_2), \ldots, \inf_t g(z_t, \hat{\theta}_N), \inf_t g(z_t, \theta) - \epsilon\}$, then $0 < L_1 \le g(z_n, \rho) \le L_2 < \infty \forall n$ and $\forall \rho \in \{\theta, \hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_n, \ldots\}$ a.s. By Lemma 3.4, condition (vi) and Lemma 3.5, we have

$$\begin{aligned}
&\lambda_{\min}(n) \to \infty \qquad \text{a.s.},\\
&\limsup_n \frac{\log \lambda_{\max}(n)}{\lambda_{\min}(n)} \times \frac{n}{n}\\
&\qquad \le \frac{1}{\liminf_n(\lambda_{\min}(n)/n)} \cdot \frac{\log \lambda_{\max}(n)}{n} = o(1) \qquad \text{a.s.},\\
&L_1 \lambda_i(X_n^T G_n X_n) \le \lambda_i(X_n^T X_n)\\
&\qquad\qquad \le L_2 \lambda_i(X_n^T G_n X_n) \forall i = 1, 2, \ldots, p \qquad \text{a.s.},
\end{aligned}$$

and

$$\lambda_{\min}(X_n^T G_n X_n) \to \infty, \qquad \log \lambda_{\max}(X_n^T G_n X_n) = o(\lambda_{\min}(X_n^T G_n X_n)) \qquad \text{a.s.}$$

Let e_n as in (3.1), since $\sup_n E(|e_n|^\alpha \mid \mathcal{F}_{n-1}) < \infty$ a.s. and by Theorem 2.1 we have

$$\hat{\beta}_n = (X_n^T G_n X_n)^{-1} X_n^T G_n Y_n = \beta + o(1) \qquad \text{a.s.}$$

Now we rewrite

$$\hat{\beta}_n^* - \beta = (X_n^T G_n^* X_n)^{-1} X_n^T G_n^* Y_n - \beta = (X_n^T G_n^* X_n)^{-1} X_n^T G_n^* E_n.$$

Since the elements of G_n are bounded a.s., by condition (V), Lemma 3.7 and (2.9) of Lai and Wei (1982), we have

$$\hat{\beta}_n^* - \beta = \left[\frac{1}{n} X_n^T G_n X_n\right]^{-1} \left[\frac{1}{n} X_n^T G_n E_n\right] + o(1) = (\hat{\beta}_n - \beta) + o(1) = o(1)$$

and

$$\hat{\beta}_n^* = \beta + o(1) \quad \text{a.s.}$$

Thus we conclude the result

$$\hat{\beta}_n^* \to \beta \quad \text{a.s.}$$

Before proving the asymptotic normality of $\hat{\beta}_n^*$, we would like first to give a weaker form of asymptotic normality of b_n than that of Lai and Wei (1982), and try to study the asymptotic normality of $\hat{\beta}_n^*$ to a more general result. Denote $\xrightarrow{p}$ and $\xrightarrow{d}$, respectively, the convergence in probability and in distribution.

THEOREM 3.3. (Chang and Chang (1990)) *Suppose that in the regression model* (1.1), $\{\epsilon_n\}$ *is a martingale difference sequence with respect to an increasing sequence of* σ*-fields* $\{\mathcal{F}_n\}$ *such that* (2.1) *holds and* $\lim_{n\to\infty} E(\epsilon_n^2 \mid \mathcal{F}_{n-1}\} = \sigma^2$ *a.s. Moreover assume that there exists a sequence of non-random nonsingular matrix* $\{B_n\}$ *such that*

$$B_n^{-1}(X_n^T X_n)(B_n^T)^{-1} \xrightarrow{p} \Gamma$$

where Γ *is a positive definite matrix and*

$$\max_{1\le i\le n} x_i^T (B_n B_n^T)^{-1} x_i \xrightarrow{p} 0$$

then

$$\text{(i)} \quad (B_n^{-1}(X_n^T X_n)(B_n^T)^{-1}, B_n^{-1}(X_n^T X_n)(b_n - \beta)) \xrightarrow{d} (\Gamma, \Gamma^{1/2} N)$$

$$\text{(ii)} \quad (b_n - \beta)^T (X_n^T X_n)(b_n - \beta) \xrightarrow{d} \sigma^2 \chi_p^2 \qquad \text{as} \quad n \to \infty$$

where $N \sim N(0, \sigma^2 I_p)$ *and* χ_p^2*, the chi-squared distribution with p degrees of freedom.*

Now we establish the asymptotic normality of $\hat{\beta}_n^*$ under certain conditions on the regression matrix X_n as following

THEOREM 3.4. *Suppose that in the regression model* (1.1), $\{\epsilon_n\}$ *is a martingale difference sequence with respect to an increasing sequence of* σ*-fields* $\{\mathcal{F}_n\}$ *such that* (2.1) *holds and* $E(\epsilon_n^2 \mid \mathcal{F}_{n-1}) = g(z_n, \theta)$*, where g is a known function and* θ *is unknown parameter vector. Moreover, assume that conditions* (i), (ii), (iii), (iv) *and* (v) *in Theorem 3.2 hold and there exists a sequence of nonrandom nonsingular matrix* $\{B_n\}$ *such that*

$$B_n^{-1}(X_n^T G_n X_n)(B_n^T)^{-1} \xrightarrow{p} \Gamma$$

where Γ is a positive definite matrix and

$$\max_{1\le i\le n} x_i^T(B_nB_n^T)^{-1}x_i \xrightarrow{p} 0$$

then

$$\begin{aligned}
&\text{(i)}\quad B_n^{-1}(X_n^TG_nX_n)(\hat\beta_n-\beta)\xrightarrow{d}\Gamma^{1/2}\cdot N,\\
&\qquad (\hat\beta_n-\beta)^T(X_n^TG_nX_n)(\hat\beta_n-\beta)\xrightarrow{d}\chi_p^2.\\
&\text{(ii)}\quad B_n^{-1}(X_n^TG_n^*X_n)(\hat\beta_n^*-\beta)\xrightarrow{d}\Gamma^{1/2}\cdot N\\
&\qquad (\hat\beta_n^*-\beta)^T(X_n^TG_n^*X_n)(\hat\beta_n^*-\beta)\xrightarrow{d}\chi_p^2 \qquad \textit{as}\quad n\to\infty
\end{aligned}$$

where $N\sim N(O,I_p)$.

PROOF. (i) Let ϕ_m, L_1, L_2 as in the proof of Theorem 3.2, then by Lemmas 3.3 and 3.5 and $\hat\theta_n\to\theta$, we have

$$\begin{aligned}
&\|B_n^{-1}(X_n^TG_nX_n)(B_n^T)^{-1}-B_n^{-1}(X_n^TG_n^*X_n)(B_n^T)^{-1}\|\\
&\quad=\|B_n^{-1}(X_n^TT_nX_n)(B_n^T)^{-1}\|\\
&\quad=o(1)\|B_n^{-1}(X_n^TX_n)(B_n^T)^{-1}\|\\
&\quad\le o(1)\cdot L_2\cdot\|B_n^{-1}(X_n^TG_n^*X_n)(B_n^T)^{-1}\|\xrightarrow{p}0
\end{aligned}$$

where $T_n=G_n-G_n^*$, we obtain

$$B_n^{-1}(X_n^TG_nX_n)(B_n^T)^{-1}-B_n^{-1}(X_n^TG_n^*X_n)(B_n^T)^{-1}\xrightarrow{p}0$$

and then

$$B_n^{-1}(X_n^TG_n^*X_n)(B_n^T)^{-1}\xrightarrow{p}\Gamma. \tag{3.8}$$

Next since

$$\max_{1\le i\le n} w_i^T(B_nB_n^T)^{-1}w_i\le\frac{1}{L_1}\max_{1\le i\le n}x_i^T(B_nB_n^T)^{-1}x_i \qquad \text{a.s.},$$

we have

$$\max_{1\le i\le n} w_i^T(B_nB_n^T)^{-1}w_i\xrightarrow{p}0 \qquad \text{a.s.} \tag{3.9}$$

where w_i's are stated in (3.1). And since $\hat\beta_n$ is the ordinary least squares estimate of the rewritten model (3.1) with w_i's replacing x_i's in Theorem 3.3, we have

$$B_n^{-1}(X_n^TG_nX_n)(\hat\beta_n-\beta)\xrightarrow{d}\Gamma^{1/2}\cdot N \tag{3.10}$$

and

$$(\hat\beta_n-\beta)^T(X_n^TG_nX_n)(\hat\beta_n-\beta)\xrightarrow{d}\chi_p^2.$$

(ii) By conditions (i), (ii), (iii), (iv), (v) in Theorem 3.2 and Corollary 3.1, we have

$$\begin{aligned}
&B_n^{-1}(X_n^T G_n^* X_n)(\hat{\beta}_n^* - \beta) \\
&\quad = B^{-1}(X_n^T G_n^* X_n)[(X_n^T G_n^* X_n)^{-1} X_n^T G_n^* (X_n\beta + E_n) - \beta] \\
&\quad = B_n^{-1} X_n^T G_n^* E_n \\
&\quad = B_n^{-1} X_n^T G_n E_n - B_n^{-1} X_n^T T_n E_n \\
&\quad = B_n^{-1}(X_n^T G_n X_n)(\hat{\beta}_n - \beta) - B_n^{-1} X_n^T T_n E_n,
\end{aligned}$$

where

$$T_n = G_n - G_n^* = \text{diag}[t_1, \ldots, t_n].$$

Now consider the second term $B_n^{-1} X_n^T T_n E_n$, we have, for any p-vector c, $c^T B_n^{-1} X_n^T E_n = \sum_{i=1}^n c^T B_n^{-1} x_i \epsilon_i$ is a martingale transform, then by Lemma 3.3, (2.9) of Lai and Wei (1982) and (3.8), we have

$$\begin{aligned}
|c^T B_n^{-1} X_n T_n E_n| &= \left| \sum_{i=1}^n c^T B_n^{-1} x_i \epsilon_i t_i \right| \\
&= o(1) \left| \sum_{i=1}^n c^T B_n^{-1} x_i \epsilon_i \right| \qquad \text{a.s.} \\
&= o(1) \left| o(1) \cdot \sum_{i=1}^n (c^T B_n^{-1} x_i)^2 + O(1) \right| \qquad \text{a.s.} \\
&= o(1) \cdot \sum_{i=1}^n (c^T B_n^{-1} x_i)^2 \qquad \text{a.s.} \\
&\le o(1) \cdot L_2 \cdot \sum_{i=1}^n c^T B_n^{-1} x_i \left[\frac{1}{g(z_i, \theta)} \right] x_i^T (B_n^T)^{-1} c \\
&= o(1) \cdot L_2 c^T B_n^{-1} (X_n^T G_n X_n)(B_n^T)^{-1} c \xrightarrow{p} 0.
\end{aligned}$$

Thus

$$c^T B_n^{-1}(X_n^T G_n^* X_n)(\hat{\beta}_n^* - \beta) - c^T B_n^{-1}(X_n^T G_n X_n)(\hat{\beta}_n - \beta) \xrightarrow{p} 0.$$

By Cramér-Wold Theorem and (3.10), we have

$$B_n^{-1}(X_n^T G_n^* X_n)(\hat{\beta}_n^* - \beta) \xrightarrow{d} \Gamma^{1/2} \cdot N.$$

Finally, since

$$\begin{aligned}
&(\hat{\beta}_n^* - \beta)^T (X_n^T G_n^* X_n)(\hat{\beta}_n^* - \beta) \\
&\quad = [B_n^{-1}(X_n^T G_n^* X_n)(\hat{\beta}_n^* - \beta)]^T B_n^T (X_n^T G_n^* X_n)^{-1} B_n [B_n^{-1}(X_n^T G_n^* X_n)(\hat{\beta}_n^* - \beta)] \\
&\quad \xrightarrow{d} \chi_p^2.
\end{aligned}$$

□

4. Hypothesis testing problem

Suppose that in model (1.1), $E(\epsilon_n^2 \mid \mathcal{F}_{n-1}) = g(z_n, \theta)$ where g is a known function of z_n and unknown parameter θ. For the null hypothesis $H_0 : H^T\beta = h$ against the alternative hypothesis $H_1 : H^T\beta \neq h$ where H^T is a $k \times p$ $(k < p)$ matrix with rank k and h is a $k \times 1$ known vector, we will discuss this hypothesis as follows.

Consider the rewritten model (3.1), $\eta_n = w_n^t\beta + e_n$ and $E(e_n^2 \mid \mathcal{F}_{n-1}) = 1$ for every n. Let

$$R_0^2 = (\psi_n - W_n\hat{\beta}_n)^T(\psi_n - W_n\hat{\beta}_n),$$
$$R_1^2 = (\psi_n - W_n\hat{\beta}_n^*)^T(\psi_n - W_n\hat{\beta}_n^*)$$

where $W_n = (w_1, \ldots, w_n)^T$, $\psi_n = (\eta_1, \ldots, \eta_n)^T$, $\hat{\beta}_n$ is the weighted least squares estimate of β, and $\hat{\beta}_n^*$ is the constrained weighted least squares estimate of β under H_0. It can be shown that $\hat{\beta}_n = (W_n^TW_n)^{-1}W_n^T\psi_n$ and $\hat{\beta}_n^* = \hat{\beta}_n - (W_n^TW_n)^{-1} \cdot H[H^T(W_n^TW_n)^{-1}H]^{-1}(H^T\hat{\beta}_n - h)$. Let $P_n = W_n(W_n^TW_n)^{-1}W_n^T$ and $P_n^* = W_n(W_n^TW_n)^{-1}H[H^T(W_n^TW_n)^{-1}H]^{-1}H^T(W_n^TW_n)^{-1}W_n^T$, then P_n and $P_n - P_n^*$ are the projection operators of $\mathcal{M}(W_n)$, the space generated by W_n, and the space $\{W_n\beta \mid H^T\beta = h\}$ respectively. Furthermore, $(P_n - P_n^*)$ is orthogonal to P_n^* (see Chang and Chang (1990)), denote $P_w = P_n - P_n^*$, then we have

$$\begin{aligned}
R_1^2 - R_0^2 &= (\psi_n - W_n\hat{\beta}_n^*)^T(\psi_n - W_n\hat{\beta}_n^*) - (\psi_n - W_n\hat{\beta}_n)^T(\psi_n - W_n\hat{\beta}_n) \\
&= \|(\psi_n - P_w\psi_n)\|^2 - \|\psi_n - P_n\psi_n\|^2 \\
&= \|P_n\psi_n - P_w\psi_n\|^2 \\
&= (H^T\hat{\beta}_n - h)^T[H^T(W_n^TW_n)^{-1}H]^{-1}(H^T\hat{\beta}_n - h) \\
&= (H^T\hat{\beta}_n - H^T\beta + H^T\beta - h)^T[H^T(W_n^TW_n)^{-1}H]^{-1} \\
&\quad \cdot (H^T\hat{\beta}_n - H^T\beta + H^T\beta - h).
\end{aligned}$$

Thus under H_0, $R_1^2 - R_0^2 = (\hat{\beta}_n - \beta)^TH[H^T(W_n^TW_n)^{-1}H]^{-1}H^T(\hat{\beta}_n - \beta)$.

Let

$$A_n = P_n^*W_n \tag{4.1}$$

then

$$A_n^TA_n = H[H^T(W_n^TW_n)^{-1}H]^{-1}H^T,$$

and

$$R_1^2 - R_0^2 = (\hat{\beta}_n - \beta)^TA_n^TA_n(\hat{\beta}_n - \beta) \tag{4.2}$$

Now we give a test statistic of linear hypothesis $H_0 : H^T\beta = h$ and construct an asymptotic distribution as follows.

THEOREM 4.1. *Suppose that in the regression model* (1.1), $\{\epsilon_n\}$ *is a martingale difference sequence with respect to an increasing sequence of* σ*-fields* $\{\mathcal{F}_n\}$

such that (2.1) *holds and* $E(\epsilon_n^2 \mid \mathcal{F}_{n-1}) = g(z_n, \theta)$ *where* g *is a known function of* z_n *and unknown parameter vector* θ. *Moreover, assume that conditions in Theorem* 3.2 *hold and there exists a sequence of nonrandom nonsingular matrix* $\{B_n\}$ *such that*

$$B_n^{-1}(X_n^T G_n X_n)(B_n^T)^{-1} \xrightarrow{p} \Gamma,$$

where Γ *is a positive definite matrix with*

$$\max_{1\le i\le n} x_i^T (B_n B_n^T)^{-1} x_i \xrightarrow{p} 0$$

and

$$B_n^{-1}(A_n^T A_n)(B_n^T)^{-1} \xrightarrow{p} \Gamma^{1/2} \begin{bmatrix} U_{k\times k} & 0 \\ 0 & 0 \end{bmatrix} \Gamma^{1/2} \tag{4.3}$$

where $U_{k\times k}$ *is an idempotent matrix with rank* k. *Then under* $H_0 : H^T\beta = h$, *not only* $R_1^{*2} - R_0^{*2} = (\hat\beta_n^* - \beta)^T H[H^T(X_n^T G_n^* X_n)^{-1}H]^{-1}H^T(\hat\beta_n^* - \beta)$ *but also* $R_1^2 - R_0^2$ *converges in distribution to* χ_k^2 *as* $n \to \infty$, *where* $\hat\beta_n^*$ *and* G_n^* *as in* (3.7), R_i^{*2} *as in* (4.2) *except replacing* $\hat\beta_n$, θ *by* $\hat\beta_n^*$ *and* $\hat\theta_n$, *respectively for* $i = 0, 1$.

Proof. First from (3.9)

$$\max_{1\le i\le n} w_i^T (B_n B_n^T)^{-1} w_i \xrightarrow{p} \text{a.s.}$$

then from (4.2) by substituting $\hat\beta_n^*$ for $\hat\beta_n$ and A_n^* for A_n, where $A_n^* = P_n^* W_n^*$ as in (4.1) except substituting $\hat\theta_n$ for θ in the entries of W_n, we have

$$\begin{aligned} R_1^{*2} - R_0^{*2} &= (\hat\beta_n^* - \beta)(A_n^{*T} A_n^*)(\hat\beta_n^* - \beta) \\ &= [B_n^{-1}(X_n^T G_n^* X_n)(\hat\beta_h^* - \beta)]^T [B_n^T (X_n^T G_n^* X_n)^{-1}(B_n)] \\ &\quad \cdot [B_n^{-1}(A_n^{*T} A_n^*)(B_n^T)^{-1}](B_n^T(X_n^T G_n^* X_n)^{-1} B_n) \\ &\quad \cdot [B_n^{-1}(X_n^T G_n^* X_n)(\hat\beta_n^* - \beta)]. \end{aligned}$$

Using Theorem 3.4 and (4.3) we have

$$R_1^{*2} - R_0^{*2} \xrightarrow{d} N^T \Gamma^{1/2}\Gamma^{-1}\Gamma^{1/2} \begin{bmatrix} U_{k\times k} & 0 \\ 0 & 0 \end{bmatrix} \Gamma^{1/2}\Gamma^{-1}\Gamma^{1/2} N = \chi_k^2,$$

as $n \to \infty$.

Secondly, from Lemmas 3.4, 3.6 and 3.7, we have

$$(X_n^T G_n^* X_n)^{-1} = (X_n^T G_n X_n)^{-1} + o\left(\frac{1}{n}\right)$$

and

$$(4.4)\qquad \begin{aligned}(A_n^{*T}A_n^*) &= H[H^T(X_n^TG_n^*X_n)^{-1}H]^{-1}H^T\\ &= H\left[H^T(X_n^TG_nX_n)^{-1}H+o\left(\frac{1}{n}\right)\right]^{-1}H^T\\ &= n\cdot H[n\cdot H^T(X_n^TG_nX_n)^{-1}H+o(1)]^{-1}H^T\\ &= n\cdot H\{[n\cdot H^T(X_n^TG_nX_n)^{-1}H]^{-1}+o(1)\}H^T\\ &= H[H^T(X_n^TG_nX_n)^{-1}H]^{-1}H^T+o(n)\\ &= (A_n^TA_n)+o(n) \qquad \text{a.s.}\end{aligned}$$

Then from Lemma 3.5, condition (vi), (3.8) and (4.4), we obtain

$$\begin{aligned}\|B_n^{-1}[(A_n^{*T}A_n^*)-(A_n^TA_n)](B_n^T)^{-1}\| &\le o(n)\cdot\|B_n^{-1}(B_n^T)^{-1}\|\\ &\le o(n)\cdot\frac{1}{\lambda_{\min}(n)}\|B_n^{-1}X_n^TX_n(B_n^T)^{-1}\|\xrightarrow{p}0\end{aligned}$$

i.e.

$$B_n^{-1}(A_n^{*T}A_n^*)(B_n^T)^{-1}\xrightarrow{p}\Gamma^{1/2}\begin{bmatrix}U_{k\times k} & 0\\ 0 & 0\end{bmatrix}\Gamma^{1/2}.$$

Using Theorem 3.4 we obtain

$$R_1^2-R_0^2\xrightarrow{d}\chi_k^2. \qquad \square$$

By this result, we can conclude that, in practical, the null hypothesis H_0 would be rejected if $R_i^{*2}-R_0^{*2}>\chi^2_{k,1-\alpha}$ for significance level α and in large n.

5. Example

In this section, we apply our method to the following linear heteroscedastic model. We demonstrate the linear heteroscedastic model of the regression model (1.1) with $E(\epsilon_n^2 \mid \mathcal{F}_{n-1})=z_n^T\cdot\theta$, where $z_n=(z_{n1},z_{n2},\ldots,z_{nk})^T$ is an observable random vector, $\mathcal{F}_{n-1}$-measurable and $\theta=(\theta_1,\ldots,\theta_k)^T$ is a parameter vector. In this model, $\hat\theta_n$ is equal to $(Z_n^TZ_n)^{-1}Z_n^T\hat e_n^2$ with $Z_n=(z_{ij})_{1\le i\le n,1\le j\le k}$.

COROLLARY 5.1. *Suppose that in linear heteroscedastic model, $\{\epsilon_n\}$ is a martingale difference sequence with respect to an increasing sequence of σ-fields $\{\mathcal{F}_n\}$ such that* (2.1) *holds. We assume the conditions* (i), (ii) *and* (v) *of Theorem* 3.1 *are satisfied and* $\liminf_n \frac{\lambda_{\min}(Z_n^TZ_n)}{n}>0$ *a.s., then $\hat\theta_n\to\theta$ a.s.*

PROOF. Since the condition (ii) of Theorem 3.1 clearly holds, we only need to check the condition (iv) of Theorem 3.1 whether or not holds. Assume $\lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{1}(z_n^T(\theta-\rho))^2=0$, from

$$\begin{aligned}0=\lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}(z_n^T(\theta-\rho))^2 &= \lim_{n\to\infty}\frac{1}{n}(\theta-\rho)^TZ_n^TZ_n(\theta-\rho)\\ \ge \liminf_n\frac{\lambda_{\min}(Z_n^TZ_n)}{n}\|\theta-\rho\|_k^2\ge 0 &\qquad \text{(by Lemma 3.5)}\end{aligned}$$

we have $\theta = \rho$ a.s.

Thus, by Theorem 3.1, we have $\hat{\theta}_n \to \theta$ a.s. □

Now we consider a special case of linear heteroscedasticity model. Let $z_n = (1, u_n)^T$ and $\theta = (d, \sigma^2)^T$, then $g(z_n, \theta) = \sigma^2 u_n + d$ and

$$Z_n^T Z_n/n = \begin{pmatrix} 1 & \bar{u} \\ \bar{u} & \bar{u}^2 \end{pmatrix}$$

where $\bar{u} = \sum_1^n u_i/n$, $\bar{u}^2 = \sum_1^n u_i^2/n$.

Let $\lambda = \lambda_{\min}(Z_n^T Z_n/n) = \{1 + \bar{u}^2 - [(1-\bar{u}^2)^2 + (2\bar{u}^2)]^{1/2}\}/2$, we have

$$\begin{aligned}(1-\bar{u}^2)^2 + (2\bar{u})^2 &= 4\lambda^2 - 4(1+\bar{u}^2)\lambda + (1+\bar{u}^2)^2 \\ &\geq -4(1+\bar{u}^2)\lambda + (1+\bar{u}^2)^2.\end{aligned}$$

It implies

$$\begin{aligned}\lambda &\geq [\bar{u}^2 - (\bar{u})^2]/[4(1+\bar{u}^2)] \\ &= \left[n\sum_1^n u_i^2 - \left(\sum_1^n u_i\right)^2\right] \Big/ \left(n^2 + n\sum_1^n u_i^2\right) \\ &\geq (n-1)\sum_1^n u_i^2 \Big/ \left(n^2 + n\sum_1^n u_i^2\right).\end{aligned}$$

If $\inf_n u_n = c > 0$ then $\sum_{i=1}^n u_i^2 \geq nc$ and

$$\liminf_n \frac{(n-1)\sum_{i=1}^n u_i^2}{n^2 + n\sum_{i=1}^n u_i^2} \geq \liminf_n \frac{nc - c}{n(1+c)} > 0$$

i.e.

$$\liminf_n \lambda_{\min}(Z_n^T Z_n)/n > 0.$$

Therefore, we reestablish the results of Cheng and Chang (1990) from Corollary 5.1, Theorems 3.2, 3.4 and 4.1.

6. Discussion

Note that in many simulations the convergent rate of $\hat{\theta}_n$ converging to θ in Lemma 3.1 (Jennrich (1969)) are sometime slow, so it is not efficient in application, thus to find a better method to substitute it is necessary in practice. If θ is permitted to contain elements of β, i.e. $\theta = [\alpha_1, \ldots, \alpha_r, \beta_1, \ldots, \beta_p]^T$, we can find all results still hold under those conditions mentioned before.

Acknowledgements

We would like to thank the referee and associate editor for their many helpful comments and clarifications.

Appendix

In this appendix, we prove Lemmas 3.3, 3.4, 3.5, 3.6 and 3.7.

PROOF OF LEMMA 3.3. Since f is uniformly continuous on compact subsets of $\Delta \times \Theta$, for any $\rho_0 \in \Theta$ and any $\epsilon > 0$, there exists $\delta > 0$ such that $f(z_t, \rho_0) - \epsilon < f(z_t, \rho) < f(z_t, \rho_0) + \epsilon$ for all $z_t \in \Delta$ and $\|\rho - \rho_0\|_k < \delta$. Thus, for $\|\rho - \rho_0\|_k < \delta$

$$\sup_{z_t \in \Delta} f(z_t, \rho_0) - \epsilon \leq \sup_{z_t \in \Delta} f(z_t, \rho) \leq \sup_{z_t \in \Delta} f(z_t, \rho_0) + \epsilon.$$

Since ϵ is arbitrary $\sup_{z_t \in \Delta} f(z_t, \rho)$ is continuous at $\rho = \rho_0$, and ρ_0 is arbitrary, so $\sup_{z_t \in \Delta} f(z_t, \rho)$ is continuous for all $\rho \in \Theta$. Similarly, $\inf_{z_t \in \Delta} f(z_t, \rho)$ is continuous for all $\rho \in \Theta$. □

PROOF OF LEMMA 3.4. Since

$$\lambda_{\max}(n) \leq \sum_{i=1}^{p} \lambda_i(n) = t_r X_n^T X_n = \sum_{i=1}^{n} \|x_n\|_p^2 = O(n). \qquad \square$$

PROOF OF LEMMA 3.5. Since C is symmetric, there exists orthogonal matrix U such that $C = U^T \Lambda U$ where $\Lambda = \operatorname{diag}(\lambda_1(C), \lambda_2(C), \ldots, \lambda_n(C))$, we have

$$\begin{aligned} B^T C B &= (UB)^T \Lambda (UB) \\ &= (UB)^T [\Lambda - \lambda_n(C) \times I_{n\times n}](UB) + (UB)^T [\lambda_n(C) \times I_{n\times n}](UB). \end{aligned}$$

Further since $(UB)^T[\Lambda - \lambda_n(C) \times I_{n\times n}](UB)$ is nonnegative definite and by Weyl's inequalities, for $i = 1, 2, \ldots, p$, we have

$$\begin{aligned} \lambda_i(B^T C B) &\geq \lambda_i\{(UB)^T [\lambda_n(C) \times I_{n\times n}](UB)\} \\ &= \lambda_i(\lambda_n(C) \cdot B^T B). \end{aligned}$$

On the other hand,

$$\begin{aligned} &(UB)^T [\lambda_1(C) \times I_{n\times n}](UB) \\ &\quad = (UB)^T [\lambda_1(C) \times I_{n\times n} - \Lambda](UB) + (UB)^T \Lambda (UB). \end{aligned}$$

Thus

$$\lambda_i(B^T C B) \leq \lambda_i(\lambda_1(C) B^T B). \qquad \square$$

PROOF OF LEMMA 3.6. From, $\|A_n^{-1} - B_n^{-1}\| \leq \|B_n^{-1}\| \|A_n - B_n\| \|A_n^{-1}\|$ and $\liminf_n \lambda_{\min}(A_n) > 0$, we have $A_n^{-1} = B_n^{-1} + o(1)$ a.s. □

PROOF OF LEMMA 3.7. Let $\varphi_m = \{w : \sup_n \|x_n(w)\|_p \leq m\}$, then $P(\bigcup_{m=1}^{\infty} \varphi_m) = 1$, and let

$$h(z_t,\rho) = \frac{1}{g(z_t,\rho)} - \frac{1}{g(z_t,\theta)} = \frac{g(z_t,\theta)-g(z_t,\rho)}{g(z_t,\theta)g(z_t-\rho)}$$

we have by assumption $\frac{1}{L_2} \leq \frac{1}{g(z_t,\rho)} \leq \frac{1}{L_1} \forall\rho \in \{\theta, \hat{\theta}_1, \ldots, \hat{\theta}_n, \ldots\}$. Thus

$$\begin{aligned}\frac{1}{n}\sum_{t=1}^{n} h^2(z_t,\rho) &= \frac{1}{n}\sum_{t=1}^{n}\left(\frac{g(z_t,\theta)-g(z_t,\rho)}{g(z_t,\theta)g(z_t-\rho)}\right)^2 \\ &\leq \frac{1}{L_1^4}\cdot\frac{1}{n}\sum_{t=1}^{n}(g(z_t,\theta)-g(z_t,\rho))^2 \forall\rho \in \{\theta, \hat{\theta}_1, \ldots\}\end{aligned}$$

from the proof of Theorem 3.1, we have

$$\frac{1}{n}\sum_{t=1}^{n}(g(z_t,\theta)-g(z_t,\hat{\theta}_n))^2 = o(1) \qquad \text{a.s.}$$

and

$$\frac{1}{n}\sum_{t=1}^{n} h^2(z_t,\hat{\theta}_n) = o(1) \qquad \text{a.s.}$$

Now

$$\begin{aligned}\frac{1}{n}X_n^T G_n^* E_n - \frac{1}{n}X_n^T G_n E_n &= \frac{1}{n}X_n^T(G_n^* - G_n)E_n \\ &= \begin{pmatrix} \frac{1}{n}\sum_{t=1}^{n} x_{t1}\cdot h(z_t,\hat{\theta}_n)\epsilon_t \\ \vdots \\ \frac{1}{n}\sum_{t=1}^{n} x_{tk}\cdot h(z_t,\hat{\theta}_n)\epsilon_t \\ \vdots \\ \frac{1}{n}\sum_{t=1}^{n} x_{tp}\cdot h(z_t,\hat{\theta}_n)\epsilon_t \end{pmatrix}_{p\times 1}\end{aligned}$$

for $k = 1, 2, \ldots, p$, and

$$\frac{1}{n}\sum_{t=1}^{n} x_{tk}^2 h^2(z_t,\hat{\theta}_n) \leq m\frac{1}{n}\sum_{t=1}^{n} h^2(z_t,\hat{\theta}_n) = o(1) \qquad \text{a.s.}$$

By Lai and Wei (1982) (2.9), we have

$$\frac{1}{n}\sum_{t=1}^{n} x_{tk}h(z_t,\hat{\theta}_n)\epsilon_t = \frac{1}{n}o\left(\sum_{t=1}^{n} x_{tk}^2 h^2(z_t,\hat{\theta}_n)\right) + \frac{1}{n}O(1) = o(1) \qquad \text{a.s.}$$

and

$$\frac{1}{n}X_n^T G_n^* E_n = \frac{1}{n}X_n^T G_n E_n + o(1) \qquad \text{a.s.}$$

Hence

$$\frac{1}{n}X_n^T G_n^* E_n = \frac{1}{n}X_n^T G_n E_n + o(1) \qquad \text{a.s.}$$

Next

$$\frac{1}{n}X_n^T G_n^* X_n - \frac{1}{n}X_n^T G_n X_n = \begin{pmatrix} \frac{1}{n}\sum_{t=1}^{n} x_{t1}h(z_t,\hat{\theta}_n)x_{t1}, \ldots, \frac{1}{n}\sum_{t=1}^{n} x_{t1}h(z_t,\hat{\theta}_n)x_{tp} \\ \vdots \\ \frac{1}{n}\sum_{t=1}^{n} x_{tp}h(z_t,\hat{\theta}_n)x_{t1}, \ldots, \frac{1}{n}\sum_{t=1}^{n} x_{tp}h(z_t,\hat{\theta}_n)x_{tp} \end{pmatrix}$$

for $1 \leq h,\ k \leq p$, and by Cauchy-Schwarz inequality

$$\left|\frac{1}{n}\sum_{t=1}^{n} x_{th}\cdot h(z_t,\hat{\theta}_n)x_{tk}\right| \leq m^2 \left|\frac{1}{n}\sum_{t=1}^{n} h(z_t,\hat{\theta}_n)\right| \leq m^2 \sqrt{\frac{1}{n}\sum_{t=1}^{n} h^2(z_t,\hat{\theta}_n)} = o(1) \qquad \text{a.s.,}$$

we have

$$\frac{1}{n}X_n^T G_n^* X_n = \frac{1}{n}X_n^T G_n X_n + o(1) \qquad \text{a.s.}$$

Hence

$$\frac{1}{n}X_n^T G_n^* X_n = \frac{1}{n}X_n^T G_n X_n + o(1) \qquad \text{a.s.}$$

Finally, from Lemmas 3.5 and 3.6, we have

$$\left(\frac{1}{n}X_n^T G_n^* X_n\right)^{-1} = \left(\frac{1}{n}X_n^T G_n X_n\right)^{-1} + o(1) \qquad \text{a.s.} \qquad \square$$

References

Chang, D. S. and Chang, M. R. (1990). Hypothesis testing in stochastic regression models, Tech. Report, National Tsing Hua University.

Cheng, M. I. and Chang, D. S. (1990). A note on stochastic regression model with heteroscedastic disturbances, Tech. Report, National Tsing Hua University.

Jennrich, R. I. (1969). Asymptotic properties of non-linear least squares estimators, *Ann. Math. Statist.*, **40**, 633–643.

Lai, T. L. and Wei, C. Z. (1982). Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems, *Ann. Statist.*, **10**, 154–166.