ASYMPTOTIC THEORY IN HETEROSCEDASTIC NONLINEAR MODELS *

by

Jun Shao
Purdue University

Technical Report #88-2

Department of Statistics
Purdue University

January 1988

* This research was supported in part by the NSF–AFOSR grant ISSA–860068.

⋆

ASYMPTOTIC THEORY IN HETEROSCEDASTIC NONLINEAR MODELS [1]

by

Jun Shao

Purdue University

Abstract

Under a nonlinear regression model with heteroscedastic errors, the consistency and asymptotic normality of the least squares estimator are proved and consistent estimators of the asymptotic covariance matrix of the least squares estimator are obtained. Statistical inference methods based on these results are then asymptotically valid in both homoscedastic and heteroscedastic models.

AMS 1980 *subject classifications.* Primary 62J02, 62F12.

Key word and phrases: Nonlinear least squares estimator, heteroscedasticity, weak and strong consistency, asymptotic normality, asymptotic covariance matrix, weighted and linear jack-knife.

[1] This research was supported in part by the NSF-AFOSR grant ISSA-860068.

1. Introduction

The purpose of this article is to establish an aymptotic theory for statistical inference by using least squares method in the following heteroscedastic nonlinear model:

(1.1) $$y_i = f(x_i, \theta) + e_i, \quad i=1,\ldots,n,$$

where $\theta \in \Theta \subset \mathbf{R}^p$ is a p-vector of unknown parameters of interest, x_i are known q-vectors, $f(x_i, \theta)$ are known nonlinear functions of $\theta \in \Theta$, and e_i are independently distributed errors with means zero and variances σ_i^2. The σ_i^2 are unknown and unequal (heteroscedastic).

The least squares estimator of θ based on data $\{y_i\}_{i=1}^n$ is defined to be any vector $\hat{\theta}_n \in \Theta$ which minimizes

(1.2) $$Q_n(\theta) = n^{-1}\sum_{i=1}^n (y_i - f(x_i, \theta))^2.$$

The existence and measurability of the least squares estimator $\hat{\theta}_n$ are ensured by the compactness of the parameter space Θ. In the general case where Θ is not compact, we shall follow Richardson and Bhattacharyya's (1986) approach, i.e., construct an increasing sequence $\{\mathbf{C}_n\}$ of compact subsets of Θ such that $\cup_n \mathbf{C}_n$ is dense in Θ and obtain the least squares estimator $\hat{\theta}_n$ by minimizing $Q_n(\theta)$ (1.2) over $\theta \in \mathbf{C}_n$ for each fixed n.

The method of least squares plays a central role in the statistical inference in the homoscedastic model, i.e., the model (1.1) with $\sigma_i^2 = \sigma^2$ for all i. In the heteroscedastic model (1.1), there may exist more efficient estimators of θ such as the weighted least squares estimator. See Fuller and Rao (1978), Carroll (1982) and Müller and Stadtmüller (1987) for some asymptotic theory of the weighted least squares estimators in linear models, i.e., $f(x_i, \theta) = x_i^\tau \theta$, where x^τ is the transpose of the vector x. However, we will concentrate on the (unweighted) least squares estimator for the following reasons:

(1) In an applied context, an investigator may overlook the error heteroscedasticity and use the unweighted least squares method because of its familiarity and simplicity. A relevant problem is then whether or not the statistical inference methods based on the least squares are robust against the error heteroscedasticity.

(2) The weighted least squares method involves a problem of choosing adequate weights and an additional computation of the weights.

(3) In many cases the σ_i^2 are unequal but close to each other and therefore the weighted least squares estimator may not be better than the unweighted least squares estimator (see Jacquez, Mather and Crawford (1968) and Shao (1987)).

When $f(x_i, \theta)$ are nonlinear in θ, the least squares estimator does not enjoy any tractable finite sample optimality property. Theoretical asymptotic properties of the nonlinear least squares estimator in the homoscedastic model ($\sigma_i^2 \equiv \sigma^2$) were established by Jennrich (1969), Malinvaud (1970), Wu (1981) and Richardson and Bhattacharyya (1986). We will prove, under the heteroscedastic model (1.1), the consistency and asymptotic normality of the least squares estimator $\hat{\theta}_n$ in Sections 2 and 3, respectively. Due to the error heteroscedasticity, the customary estimator of the asymptotic covariance matrix of $\hat{\theta}_n$ based on the "residual sum of squares" is *inconsistent*. Consistent estimators of the asymptotic covariance matrix of $\hat{\theta}_n$, which are essentially jackknife-based estimators, are obtained in Section 4. The conditions we impose on the functions $f(x_i, \theta)$ are the same as (or similar to) those in Jennrich (1969) and Wu (1981).

2. Consistency

Denote $f(x_i, \theta)$ by $f_i(\theta)$ and the unknown *true* parameter by θ_0. In the homoscedastic model, Jennrich (1969) proved the consistency of the least squares estimator under the following conditions.

(C1) Θ is a compact subset of $\mathbf{R}^p$.

(C2) $\lim_{n\to\infty} n^{-1}\sum_{i=1}^n f_i(t)f_i(s)$ exist uniformly for all t and s in Θ, and
$Q(\theta) = \lim_{n\to\infty} n^{-1}\sum_{i=1}^n (f_i(\theta)-f_i(\theta_0))^2$ has a unique minimum at $\theta=\theta_0$.

Theorem 1(i) is an extension of Jennrich's result to the heteroscedastic model. Some minor assumptions are added to handle the error heteroscedasticity.

Theorem 1. Assume (C1) and (C2). Let $\{ \hat{\theta}_n \}$ be a sequence of least squares estimators.

(i) If for a $\delta>0$,

(2.1) $$sup_i E\,|e_i|^{2+\delta} < \infty,$$

then

$$\hat{\theta}_n \to \theta_0 \quad a.s.$$

(ii) If the errors e_i in model (1.1) satisfy $e_i = \sigma_i^2 \varepsilon_i$, where ε_i are independently and identically distributed and $sup_i \sigma_i^2 < \infty$, then

$$\hat{\theta}_n \to_p \theta_0,$$

where $\to_p$ denotes convergence in probability.

Remark 1. By adopting Richardson and Bhattacharyya's approach (see Section 1), our results can be extended in a straightforward manner to the general case where Θ is not compact.

Proof. (i) Under Condition (2.1),

$$n^{-1}\sum_{i=1}^{n}(e_i^2 - \sigma_i^2) \to 0 \quad a.s.$$

(see Chung (1974, Theorem 5.4.1)). Let $\omega = \{e_i\}_{i=1}^{\infty}$ be fixed such that $n^{-1}\sum_{i=1}^{n}(e_i^2 - \sigma_i^2) \to 0$, and $\theta_n = \hat{\theta}_n(\omega)$. Suppose that θ' is a limit point of $\{\theta_n\}_{n=1}^{\infty}$. Then there is a subsequence $\{n_k\}_{k=1}^{\infty}$ such that

$$\lim_{k\to\infty} \theta_{n_k} = \theta'.$$

Denote $n^{-1}\sum_{i=1}^{n}\sigma_i^2$ by Δ_n. Since $\{\Delta_n\}_{n=1}^{\infty}$ is bounded under (2.1), there is a subsequence $\{n_j\}_{j=1}^{\infty} \subset \{n_k\}_{k=1}^{\infty}$ such that $\Delta = \lim_{j\to\infty}\Delta_{n_j}$ exists. Under (C1) and (C2),

$$\lim_{j\to\infty} Q_{n_j}(\theta) = Q(\theta) + \Delta$$

uniformly for all $\theta \in \Theta$ (see Jennrich (1969, Theorem 4)). Then

$$\lim_{j\to\infty} Q_{n_j}(\theta_{n_j}) = Q(\theta') + \Delta.$$

By the definition of the least squares estimator,

$$Q_{n_j}(\theta_{n_j}) \le Q_{n_j}(\theta_0) = n_j^{-1}\sum_{i=1}^{n_j} e_i^2 \to \Delta.$$

Hence $Q(\theta')=0$ and $\theta'=\theta_0$ follows from condition (C2). This proves $\hat{\theta}_n \to \theta_0$ *a.s.*

(ii) It suffices to show that for any subsequence $\{n_k\}_{k=1}^{\infty}$, there is a subsequence $\{n_j\}_{j=1}^{\infty} \subset \{n_k\}_{k=1}^{\infty}$ such that $\hat{\theta}_{n_j} \to \theta_0$ *a.s.*

By the boundedness of σ_i^2, for any $\{n_k\}_{k=1}^{\infty}$ there is a $\{n_l\}_{l=1}^{\infty} \subset \{n_k\}_{k=1}^{\infty}$ such that $\Delta = \lim_{l\to\infty} \Delta_{n_l}$ exists, where $\Delta_n = n^{-1}\sum_{i=1}^{n} \sigma_i^2$. Then from Jamison, Orey and Pruitt (1965),

$$n_l^{-1}\sum_{i=1}^{n_l}(e_i^2-\sigma_i^2) = n_l^{-1}\sum_{i=1}^{n_l}(\varepsilon_i^2-1)\sigma_i^2 \to_p 0.$$

Hence there is a subsequence $\{n_j\}_{j=1}^{\infty} \subset \{n_l\}_{l=1}^{\infty}$ such that

$$n_j^{-1}\sum_{i=1}^{n_j}(e_i^2-\sigma_i^2) \to 0 \quad a.s.$$

Then the proof of $\hat{\theta}_{n_j} \to \theta_0$ *a.s.* is the same as that of (i). □

Under a different set of conditions Wu (1981) proved the consistency of $\hat{\theta}_n$. See his Theorems 2 and 3 and the remarks after them.

We state the following result from Wu (1981, Appendix), which will be used in the proofs of our main results in Sections 3 and 4.

Lemma 1. Let $\mathbf{C}$ be a compact subset of $\mathbf{R}^p$ and $\{h_i(t)\}_{i=1}^{\infty}$ be a sequence of continuous functions in $t \in \mathbf{C}$ satisfying

$$\text{(2.2)} \qquad sup_{s\neq t,\, s,\, t\in \mathbf{C}} \frac{|h_i(t)-h_i(s)|}{\|t-s\|} \le M\ sup_{t\in\mathbf{C}} |h_i(t)|,$$

where $0<M<\infty$ is independent of i and $\|\ \|$ is the Euclidean norm on $\mathbf{R}^p$. Let $\{\xi_i\}_{i=1}^{\infty}$ be a sequence of independent random variables with $E\xi_i=0$ and $sup_i E\xi_i^2<\infty$. If as $n\to\infty$,

$$d_n = \sum_{i=1}^{n} sup_{t\in\mathbf{C}} |h_i(t)|^2 \to \infty,$$

then

$$\lim_{n\to\infty} sup_{t\in\mathbf{C}} |\sum_{i=1}^{n} h_i(t)\xi_i| / d_n = 0 \quad a.s.$$

For differentiable functions $h_i(t)$, a sufficient condition for (2.2) is

$$sup_{t\in C}|h_i(t)| \geq c_1 > 0 \quad \text{and} \quad sup_{t\in C}\|h_i'(t)\| \leq c_2 < \infty,$$

where c_1 and c_2 are independent of i and $h_i'(t)$ is the gradient of $h_i(t)$.

3. Asymptotic normality

Assume that θ_0 is in the interior of Θ and the second order derivatives of $f_i(\theta)$ exist and are continuous in $\theta \in \{ \|\theta - \theta_0\| < 2\delta \}$, where $\delta>0$ is a constant. Denote

$$f_{ik}'(\theta) = \frac{\partial}{\partial\theta_k} f_i(\theta), \quad f_{ikl}''(\theta) = \frac{\partial^2}{\partial\theta_k \partial\theta_l} f_i(\theta),$$

(3.1) $$g_i(\theta) = (\, f_{i1}'(\theta) \, . \, . \, . \, f_{ip}'(\theta)\,)^\tau, \quad A_n(\theta) = \sum_{i=1}^n g_i(\theta) g_i^\tau(\theta),$$

where $g_i^\tau(\theta)$ is the transpose of $g_i(\theta)$, and

$$H_i(\theta) = [\, f_{ikl}''(\theta)\,]_{k,\, l=1}^p \, .$$

The asymptotic normality of the least squares estimator $\hat{\theta}_n$ is established under several of the following conditions:

(C3) There are positive constants ν and ρ such that $\nu \leq \sigma_i^2 \leq \rho$ for all i.

(C4) i. For sufficiently large n, the inverse of $A_n(\theta_0)$ exists, $A_n^{-1}(\theta_0) = O(n^{-1})$ and $A_n(\theta_0) = O(n)$.

ii. $A_n(\theta) A_n^{-1}(\theta_0)$ converges to the identity matrix I_p as $n \to \infty$ and $\theta \to \theta_0$.

iii. $\max_{i \leq n} g_i^\tau(\theta_0) A_n^{-1}(\theta_0) g_i(\theta_0) \to 0$ as $n \to \infty$.

(C5) $\lim_{n\to\infty} n^{-1} \sum_{i=1}^n f_i(t) f_i(s)$ and $\lim_{n\to\infty} n^{-1} \sum_{i=1}^n f_{ikl}''(t) f_{ikl}''(s)$ for all (k, l) exist uniformly for all t and $s \in C = \{ \|\theta - \theta_0\| \leq \delta \}$.

(C6) i. $limsup_{n\to\infty} n^{-1} \sum_{i=1}^n sup_{\theta\in C} (f_{ikl}''(\theta))^2 < \infty$ for all (k, l).

ii. If for a pair (k, l), $\sum_{i=1}^\infty sup_{\theta \in C} (f_{ikl}''(\theta))^2 = \infty$, then the sequence $\{ f_{ikl}''(\theta) \}_{i=1}^\infty$ satisfies Condition (2.2) with $C = \{ \|\theta - \theta_0\| \leq \delta \}$.

Remark 2. (i) Condition (C4i) is equivalent to that the minimum and maximum eigenvalues of $n^{-1}A_n(\theta_0)$ are bounded away from 0 and ∞, and is much weaker than $n^{-1}A_n(\theta_0)$ converging to a positive definite matrix, a condition assumed in Jennrich (1969).

(ii) Condition (C4ii) implies that $A_n^{-1}(\theta)$ exists for sufficiently large n and θ near θ_0.

(iii) Condition (C4iii) is implied by either the existence of a positive definite matrix $A(\theta_0)=\lim_{n\to\infty} n^{-1}A_n(\theta_0)$ or $A_n^{-1}(\theta_0)\to 0$ and $sup_i \|g_i(\theta_0)\| \le c < \infty$.

(iv) Condition (C6i) is implied by the boundedness of $\{ sup_{\theta\in C}(f''_{ikl}(\theta))^2 \}_{i=1}^{\infty}$.

For any positive definite matrix P, there is a positive definite matrix A such that $P=A^2$. Define $P^{1/2}=A$ and $P^{-1/2}=A^{-1}$.

Theorem 2. Let $\{ \hat{\theta}_n \}_{n=1}^{\infty}$ be a sequence of (weakly) consistent least squares estimators and

$$D_n(\theta) = A_n^{-1}(\theta)\sum_{i=1}^n \sigma_i^2 g_i(\theta)g_i^{\tau}(\theta)A_n^{-1}(\theta).$$

(i) Assume Conditions (C3), (C4) and (C5). Then

$$D_n^{-1/2}(\theta_0)(\hat{\theta}_n-\theta_0) \to_d N(0, I_p), \tag{3.2}$$

where $\to_d$ denotes convergence in distribution.

(ii) Assume Conditions (C3), (C4) and (C6). Then (3.2) holds.

Remark 3. Under (C3) and (C4i), $D_n(\theta_0)=O(n^{-1})$ and $D_n^{-1}(\theta_0)=O(n)$. If $\lim_{n\to\infty} nD_n(\theta_0) =\Sigma(\theta_0)$ exists and is positive definite, which might be a too strong assumption, then $n^{-1}\Sigma(\theta_0)$ is the asymptotic covariance matrix of $\hat{\theta}_n$. For practical usrs, $D_n(\theta_0)$ can be treated as the asymptotic covariance matrix even if $nD_n(\theta_0)$ does not have a limit. Consistent estimators of $D_n(\theta_0)$ are given in the next section.

Proof. Let $Q'_n(\theta)$ and $Q''_n(\theta)$ be the gradient and hessian matrix of $Q_n(\theta)$, $\|\theta-\theta_0\| < 2\delta$. Then

$$Q'_n(\theta) = n^{-1}\sum_{i=1}^n (f_i(\theta)-f_i(\theta_0)-e_i)g_i(\theta)$$

and

$$Q_n''(\theta) = n^{-1}[\sum_{i=1}^n g_i(\theta)g_i^\tau(\theta) + \sum_{i=1}^n (f_i(\theta)-f_i(\theta_0)-e_i)H_i(\theta)].$$

On the set $\{\ \|\hat\theta_n-\theta_0\| \le\delta\ \}$, $Q_n'(\hat\theta_n)=0$. By the mean-value theorem,

$$Q_n'(\theta_0) = Q_n''(\theta_n^*)(\theta_0-\hat\theta_n),$$

where θ_n^* is a point on the line segment between θ_0 and $\hat\theta_n$. Thus,

$$\sum_{i=1}^n g_i(\theta_0)e_i = B_n A_n(\theta_0)(\hat\theta_n-\theta_0),$$

where

$$B_n = A_n(\theta_n^*)A_n^{-1}(\theta_0) + \sum_{i=1}^n (f_i(\theta_n^*)-f_i(\theta_0)-e_i)H_i(\theta_n^*)A_n^{-1}(\theta_0).$$

Then (3.2) follows from

(3.3) $\qquad B_n \to_p I_p,$

and

(3.4) $\qquad D_n^{-1/2}(\theta_0)A_n^{-1}(\theta_0)\sum_{i=1}^n g_i(\theta_0)e_i \to_d N(0, I_p).$

Proof of (3.3). From (C4ii) and the consistency of $\hat\theta_n$, $A_n(\theta_n^*)A_n^{-1}(\theta_0)\to_p I_p$. By the Cauchy-Schwarz inequality,

$$\sum_{i=1}^n (f_i(\theta_n^*)-f_i(\theta_0))H_i(\theta_n^*)A_n^{-1}(\theta_0) \to_p 0$$

follows from $\theta_n^*\to_p \theta_0$, (C4) and either (C5) or (C6i). Then (3.3) follows from

$$n^{-1}\sum_{i=1}^n H_i(\theta)e_i \to_p 0$$

uniformly on $C=\{\ \|\theta-\theta_0\|\le\delta\ \}$, which is implied by either (C5) or Lemma 1 under (C6ii).

Proof of (3.4). Let λ be a fixed p-vector and

$$c_i = \lambda^\tau D_n^{-1/2}(\theta_0)A_n^{-1}(\theta_0)g_i(\theta_0).$$

Then by the Cauchy-Schwarz inequality and (C3),

$$\begin{aligned} c_i^2 &\le \lambda^\tau\lambda\ g_i^\tau(\theta_0)A_n^{-1}(\theta_0)D_n^{-1}(\theta_0)A_n^{-1}(\theta_0)g_i(\theta_0) \\ &\le \nu^{-1}\lambda^\tau\lambda\ g_i^\tau(\theta_0)A_n^{-1}(\theta_0)g_i(\theta_0). \end{aligned}$$

Also,

$$\begin{aligned}\sum_{i=1}^{n} c_i^2 &= \lambda^\tau D_n^{-1/2}(\theta_0)A_n^{-1}(\theta_0)\sum_{i=1}^{n} g_i(\theta_0)g_i^\tau(\theta_0)A_n^{-1}(\theta_0)D_n^{-1/2}(\theta_0)\lambda \\ &= \lambda^\tau D_n^{-1/2}(\theta_0)A_n^{-1}(\theta_0)D_n^{-1/2}(\theta_0)\lambda \\ &\geq \rho^{-1}\lambda^\tau D_n^{-1/2}(\theta_0)D_n(\theta_0)D_n^{-1/2}(\theta_0)\lambda \\ &= \rho^{-1}\lambda^\tau\lambda.\end{aligned}$$

Hence

$$\max_{i\leq n}(c_i^2/\sum_{i=1}^{n} c_i^2) \leq \nu^{-1}\rho \max_{i\leq n} g_i^\tau(\theta_0)A_n^{-1}(\theta_0)g_i(\theta_0) \to 0$$

by (C4iii). Thus, the Lindeberg's condition holds and therefore

$$\lambda^\tau D_n^{-1/2}(\theta_0)A_n^{-1}(\theta_0)\sum_{i=1}^{n} g_i(\theta_0)e_i \to_d N(0, \lambda^\tau\lambda).$$

This shows (3.4) and completes the proof. □

4. Consistent estimators of the asymptotic covariance matrix

For the purpose of inference, the result in Theorem 2 is useful only if we can find a consistent estimator of $D_n(\theta_0)$, the aymptotic covariance matrix of $\hat{\theta}_n$. That is, an estimator $\hat{D}_n$ such that

$$n[\hat{D}_n - D_n(\theta_0)] \to 0 \;\; a.s. \;\; (\text{ or in probability }).$$

The estimator based on the residual sum of squares, $Q_n(\hat{\theta}_n)A_n^{-1}(\hat{\theta}_n)$, is consistent in the homoscedastic model but *inconsistent* in the heteroscedastic model.

In linear models (i.e., $f(x_i, \theta) = x_i^\tau\theta$), the following estimators are shown to be consistent in both homoscedastic and heteroscedastic models (Shao and Wu (1987)):

(4.1) $$(\sum_{i=1}^{n} x_i x_i^\tau)^{-1}\sum_{i=1}^{n}(1-w_i)^{-1}(y_i - x_i^\tau\hat{\theta}_n)^2 x_i x_i^\tau(\sum_{i=1}^{n} x_i x_i^\tau)^{-1},$$

which is equal to Hinkley's (1977) weighted jackknife estimator if $w_i = p/n$, and is equal to Wu's (1986) weighted jackknife estimator if $w_i = x_i^\tau(\sum_{i=1}^{n} x_i x_i^\tau)^{-1}x_i$.

Natural extensions of estimators (4.1) to the nonlinear model are

$$(4.2)\qquad \hat{D}_n = A_n^{-1}(\hat{\theta}_n)\sum_{i=1}^n (1-w_i)^{-1} r_i^2 g_i(\hat{\theta}_n) g_i^\tau(\hat{\theta}_n) A_n^{-1}(\hat{\theta}_n),$$

where $A_n(\theta)$ and $g_i(\theta)$ are defined in (3.1), $r_i = y_i - f_i(\hat{\theta}_n)$ is the ith residual, and w_i equals either p/n or $g_i^\tau(\hat{\theta}_n)A_n^{-1}(\hat{\theta}_n)g_i(\hat{\theta}_n)$. From Remark 2(ii), $A_n^{-1}(\hat{\theta}_n)$ exists for large n if (C4) holds and $\hat{\theta}_n$ is consistent for θ_0. If we assume

$$(4.3)\qquad \| g_i(\theta) \| \le c, \quad \text{for } \theta \in \{ \| \theta - \theta_0 \| \le \delta \},$$

then

$$\max_{i \le n} g_i^\tau(\hat{\theta}_n) A_n^{-1}(\hat{\theta}_n) g_i(\hat{\theta}_n) \to_p 0.$$

Thus, the estimators defined in (4.2) for two choices of w_i are asymptotically equivalent. We will show in Theorem 3 that $\hat{D}_n$ (4.2) are consistent estimators of $D_n(\theta_0)$ for any choice of w_i satisfying $0 \le w_i < 1$ and

$$(4.4)\qquad \max_{i \le n} w_i \to 0 \ \ a.s. \text{ (or in probability) if } \hat{\theta}_n \to \theta_0 \ \ a.s. \text{ (or in probability).}$$

When w_i in (4.2) are chosen to be identically equal to zero, $\hat{D}_n$ is the variance estimator obtained by the linear jackknife (Fox, Hinkley and Larntz (1980)). There exist other estimators of $D_n(\theta_0)$ such as the estimator obtained by the exact jackknife (Duncan (1978)) and its modifications (Simonoff and Tsai (1986)). However, the consistency of these estimators has not been justified thus far.

Some of the following conditions will be used in proving the consistency of $\hat{D}_n$. For a fixed pair (k, l), let $h_i(\theta) = f'_{ik}(\theta) f'_{il}(\theta)$.

(C7) $\quad sup_i Ee_i^4 < \infty$.

(C8) $\quad$ i. $sup_i \| g_i(\theta_0) \| \le c < \infty$;

ii. $\{ g_i(\theta) \}_{i=1}^{\infty}$ are equicontinuous in $\theta \in C = \{ \| \theta - \theta_0 \| \le \delta \}$.

(C9) $\quad$ The sequence $\{ h_i(\theta) \}_{i=1}^{\infty}$ satisfies condition (2.2) with $C = \{ \| \theta - \theta_0 \| \le \delta \}$, if

$$\sum_{i=1}^{\infty} sup_{\theta\in \mathbf{C}}[h_i(\theta)]^2 = \infty.$$

(C10) The limits of $n^{-1}\sum_{i=1}^{n} h_i(t)h_i(s)$ exist uniformly for t and s in **C**.

Remark 4. Under (C8), (4.3) is satisfied. Condition (C8) is implied by that $g_i(\theta)=g(x_i, \theta)$ is continuous on $X\times\mathbf{C}$, where X is a compact subset of $\mathbf{R}^q$ and $x_i\in X$ for all i.

We establish the following lemma first.

Lemma 2. (i) Assume (4.3), (C7) and (C9). Then

(4.5) $$sup_{t\in \mathbf{C}}|n^{-1}\sum_{i=1}^{n}(e_i^2-\sigma_i^2)h_i(t)| \to 0 \quad a.s.$$

(ii) Assume (C7) and (C10). Then (4.5) holds.

Proof. (i) Let $d_n=\sum_{i=1}^{n} sup_{t\in \mathbf{C}}[h_i(t)]^2$ and $\zeta_n = sup_{t\in \mathbf{C}}|n^{-1}\sum_{i=1}^{n}(e_i^2-\sigma_i^2)h_i(t)|$. Since d_n is increasing in n, we have either $d_n\to\infty$ or $d_n\to d<\infty$.

Case 1: $d_n\to\infty$. By (C7), (C9) and Lemma 1,

$$n\zeta_n/d_n \to 0 \quad a.s.$$

Under (4.3), $\{ n^{-1}d_n \}_{n=1}^{\infty}$ is bounded. Hence (4.5) holds.

Case 2: $d_n\to d<\infty$. From

$$\zeta_n \le d_n^{1/2} n^{-1}[\sum_{i=1}^{n}(e_i^2-\sigma_i^2)^2]^{1/2},$$

the result follows if

(4.6) $$n^{-2}\sum_{i=1}^{n}(e_i^2-\sigma_i^2)^2 \to 0 \quad a.s.$$

Let $z_i=(e_i^2-\sigma_i^2)^2-E(e_i^2-\sigma_i^2)^2$. Then by (C7), $sup_i E|z_i|<\infty$ and therefore

$$\sum_{i=1}^{\infty} E|z_i|/i^2 < \infty.$$

From Theorem 5.4.1 of Chung (1974),

$$n^{-2}\sum_{i=1}^{n} z_i \to 0 \quad a.s.,$$

which implies (4.6). This completes the proof of (i).

(ii) The result follows from Theorem 4 of Jennrich (1969). □

Theorem 3. Suppose that $\hat{\theta}_n \to \theta_0$ *a.s.* Let $\hat{D}_n$ be defined in (4.2) with w_i satisfying (4.4).

(i) Assume Conditions (C4), (C7), (C8) and (C9). Then

$$n[\hat{D}_n - D_n(\theta_0)] \to 0 \quad a.s. \tag{4.7}$$

(ii) Assume Conditions (C4), (C7), (C8) and (C10). Then (4.7) holds.

Remark 5. (i) Condition (C8) implies that $w_i = g_i^\tau(\hat{\theta}_n) A_n^{-1}(\hat{\theta}_n) g_i(\hat{\theta}_n)$ satisfies (4.4). Thus, the class of estimators (4.2) with w_i satisfying (4.4) includes the linear jackknife estimator ($w_i \equiv 0$), the extensions of Hinkley's weighted jackknife estimator ($w_i \equiv p/n$) and Wu's weighted jackknife estimator ($w_i = g_i^\tau(\hat{\theta}_n) A_n^{-1}(\hat{\theta}_n) g_i(\hat{\theta}_n)$). They are all consistent according to the theorem.

(ii) If $\hat{\theta}_n$ is only weakly consistent, i.e., $\hat{\theta}_n \to_p \theta_0$, then $\hat{D}_n$ is weakly consistent, i.e., $n[\hat{D}_n - D_n(\theta_0)] \to_p 0$. The proof remains the same except that $\to$ *a.s.* should be replaced by $\to_p$.

Proof. Under Conditions (C7), (C8) and either (C9) or (C10), the conditions of Lemma 2 are satisfied. Hence by the consistency of $\hat{\theta}_n$ and Lemma 2,

$$n^{-1} \sum_{i=1}^n (e_i^2 - \sigma_i^2) g_i(\hat{\theta}_n) g_i^\tau(\hat{\theta}_n) \to 0 \quad a.s.$$

Then from Condition (C4),

$$nA_n^{-1}(\hat{\theta}_n) \sum_{i=1}^n (e_i^2 - \sigma_i^2) g_i(\hat{\theta}_n) g_i^\tau(\hat{\theta}_n) A_n^{-1}(\hat{\theta}_n) \to 0 \quad a.s. \tag{4.8}$$

Let $u_i = f_i(\theta_0) - f_i(\hat{\theta}_n)$. Then $r_i = e_i + u_i$. From the mean-value theorem,

$$u_i = g_i^\tau(\theta_n^*)(\theta_0 - \hat{\theta}_n),$$

where θ_n^* is on the line segment between θ_0 and $\hat{\theta}_n$. Under (C8), there is a constant $M>0$

such that

$$\max_{i\le n} u_i^2 \le \max_{i\le n}[g_i^\tau(\theta_n^*)g_i(\theta_n^*)]\,\|\hat\theta_n-\theta_0\|^2 \le M\ \|\hat\theta_n-\theta_0\|^2 \to 0 \quad a.s.$$

and therefore

$$nA_n^{-1}(\hat\theta_n)\sum_{i=1}^n u_i^2 g_i(\hat\theta_n)g_i^\tau(\hat\theta_n)A_n^{-1}(\hat\theta_n) \to 0 \quad a.s.$$

Thus, by (4.8) and the Cauchy-Schwarz inequality,

$$nA_n^{-1}(\hat\theta_n)\sum_{i=1}^n (r_i^2-\sigma_i^2)g_i(\hat\theta_n)g_i^\tau(\hat\theta_n)A_n^{-1}(\hat\theta_n) \to 0 \quad a.s. \tag{4.9}$$

From (C8), for any pair (k, l),

$$\max_{i\le n}|f'_{ik}(\hat\theta_n)f'_{il}(\hat\theta_n) - f'_{ik}(\theta_0)f'_{il}(\theta_0)| \to 0 \quad a.s.$$

Under (C7), $\{\sigma_i^2\}_{i=1}^\infty$ is bounded. Hence

$$nA_n^{-1}(\hat\theta_n)\sum_{i=1}^n \sigma_i^2[g_i(\hat\theta_n)g_i^\tau(\hat\theta_n) - g_i(\theta_0)g_i^\tau(\theta_0)]A_n^{-1}(\hat\theta_n) \to 0 \quad a.s. \tag{4.10}$$

From (C4),

$$nA_n^{-1}(\hat\theta_n)\sum_{i=1}^n \sigma_i^2 g_i(\theta_0)g_i^\tau(\theta_0)A_n^{-1}(\hat\theta_n) - nD_n(\theta_0) \to 0 \quad a.s. \tag{4.11}$$

Then the result follows from (4.9)-(4.11) and that w_i satisfy (4.4). □

Thus, the following result can be used in making statistical inferences about θ.

Corollary. Assume that the conditions in Theorems 2 and 3 hold and $\hat\theta_n$ is weakly consistent for θ_0. Then

$$\hat D_n^{-1/2}(\hat\theta_n-\theta_0) \to_d N(0, I_p),$$

for any $\hat D_n$ defined in (4.2) with w_i satisfying (4.4).

5. Summary

Under the heteroscedastic model (1.1), the least squares estimator of θ is consistent and asymptotically normal. The estimators defined in (4.2) are consistent for the asymptotic covariance matrix of the least squares estimator. These results hold, of course, in the special case of $\sigma_i^2=\sigma^2$ for all i. Thus, any statistical inference method based on the least squares and Corollary in Section 4 is asymptotically valid and is robust against error heteroscedasticity, if the model is assumed to be homoscedastic.

REFERENCES

Carroll, R. J. (1982). Adapting for heteroscedasticity in linear models. *Ann. Statist.* **10**, 1224-1233.

Chung, K. L. (1974). *A Course in Probability Theory*, 2nd ed., Academic Press, New York.

Duncan, G. T. (1978). An empirical study of Jackknife-constructed confidence regions in non-linear regression. *Technometrics* **20**, 123-129.

Fox, T., Hinkley, D. and Larntz, K. (1980). Jackknifing in nonlinear regression. *Technometrics* **22**, 29-33.

Fuller, W. A. and Rao, J. N. K. (1978). Estimation for a linear regression model with unknown diagonal covariance matrix. *Ann. Statist.* **6**, 1149-1158.

Hinkley, D. V. (1977). Jackknifing in unbalanced situations. *Technometrics* **19**, 285-292.

Jacquez, J. A., Mather, F. J. and Crawford, C. R. (1968). Linear regression with nonconstant, unknown error variances: sampling experiments with least squares, weighted least squares and maximum likelihood estimators. *Biometrics* **24**, 607-626.

Jamison, B, Orey, S. and Pruitt, W. (1965). Convergence of weighted average of independent random variables. *Z. Wahrscheinlichkeitstheorie verw. Geb.* **4**, 40-44.

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

Malinvaud, E. (1970). The consistency of nonlinear regressions. *Ann. Math. Statist.* **41**, 956-

969.

Müller, H.-G. and Stadtmüller, U. (1987). Estimation of heteroscedasticity in regression analysis. *Ann. Statist.* **15**, 610-625.

Richardson, G. D. and Bhattacharyya, B. B. (1986). Consistent estimators in nonlinear regression for a noncompact parameter space. *Ann. Statist.* **14**, 1591-1596.

Shao, J. (1987). Estimating heteroscedastic variances in linear models I and II. Tech. Reports #87-42 and #87-43, Department of Statistics, Purdue University.

Shao, J. and Wu, C. F. J. (1987). Heteroscedasticity-robustness of jackknife variance estimators in linear models. *Ann. Statist.* **15**, to appear.

Simonoff, J. S. and Tsai, C. L. (1986). Jackknife-based estimators and confidence regions in nonlinear regression. *Technometrics* **28**, 103-112.

Wu, C. F. J. (1981). Asymptotic theory of nonlinear least squares estimation. *Ann. Statist.* **9**, 501-513.

Wu, C. F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis (with discussion). *Ann. Statist.* **14**, 1261-1350.

Department of Statistics
Purdue University
West Lafayette, IN 47907