TILBURG UNIVERSITY
KATHOLIEKE UNIVERSITEIT BRABANT
POSTBOX 90153
5000 LE TILBURG
THE NETHERLANDS

[redacted]

DEPARTMENT OF ECONOMICS

RESEARCH MEMORANDUM

[redacted]

ESTIMATION IN LINEAR REGRESSION UNDER THE PRESENCE OF HETEROSKEDASTICITY OF A COMPLETELY UNKNOWN FORM

B.B. van der Genugten

FEW 328

ESTIMATION IN LINEAR REGRESSION UNDER THE PRESENCE OF HETEROSKEDASTICITY OF A COMPLETELY UNKNOWN FORM

by

B.B. van der Genugten
Department of Econometrics
Tilburg University
Tilburg - The Netherlands

Abstract

A method is investigated for estimating the regression coefficients in a linear model with symmetrically distributed errors. The variances σ_t^2 of the errors are unknown; it is neither assumed that they are an unknown function of the explanatory variables nor that they are given in some parametric way.

The estimation is carried out in a bounded number of steps, the first one being OLS. In each step σ_t^2 is estimated with a weighted sum of m squared residuals in the neighbourhood of t and the coefficients are estimated using WLS. Furthermore an estimate of the covariance matrix is obtained. If in some sense this matrix decreases a new step is performed unless the last step has been reached.

Large sample properties of this estimation method are derived for fixed m. Some particular cases show that the asymptotic efficiency can be increased by allowing more than two steps. The asymptotic efficiency of the WLS-estimator with known variances can be approached by choosing m large.

Finally some finite sample properties are evaluated on base of simulation results.

1. Introduction

Consider the heteroskedastic linear regression model of the form

$$y_t = \beta' x_t + \varepsilon_t \qquad , \; E\{\varepsilon_t\} = 0, \; V\{\varepsilon_t\} = \sigma_t^2, \; t = 1,\dots,n$$

with $y_t \in \mathbb{R}$, $x_t \in \mathbb{R}^k$ and independent errors $\varepsilon_t \in \mathbb{R}$.
We concentrate on the asymptotic efficiency of estimators for β and the consistency of estimators for corresponding covariance matrices. Set $y = (y_1,\dots,y_n)' \in \mathbb{R}^n$ and $X = [x_1',\dots,x_n']' \in \mathbb{R}^{n\times k}$.

If the error variances σ_t^2 are known, we can use the WLS-estimator $\tilde{b}$ for β:

$$(1.1) \qquad \tilde{b} = (\sum_1^n x_t x_t'/\sigma_t^2)^{-1} \sum_1^n x_t y_t/\sigma_t^2 \; .$$

Under appropriate conditions we have asymptotic normality of the form

$$(1.2) \qquad \sqrt{n}(\tilde{b}-\beta) \to N_k(0,\tilde{\Phi})$$

(Strictly, an additional index n should be introduced when discussing the asymptotic behaviour as $n \to \infty$, e.g. $\tilde{b}_n$, $x_t(n)$, $\sigma_t^2(n)$. Because such notation becomes cumbersome we will omit this index.)

If the error variances σ_t^2 are completely unknown, the ordinary LS-estimator b_0 can be used:

$$(1.3) \qquad b_0 = (\sum_1^n x_t x_t')^{-1} \sum_1^n x_t y_t = (X'X)^{-1}X'y$$

with asymptotic normality of the form

$$(1.4) \qquad \sqrt{n}(b_0-\beta) \to N_k(0,\Phi_0)$$

However, its asymptotic efficiency $R_0 = \det(\tilde{\Phi})/\det(\Phi_0)$ with respect to $\tilde{b}$ can be low. This arises the main question: are alternatives available with a higher asymptotic efficiency for a broad class of heteroskedastic errors?

There is a large statistical and econometric literature available discussing all kind of alternatives.

A common approach is to specify a parametric form for the σ_t^2 and to estimate the underlying parameters together with β using some parametric method. This is a fruitful approach if the form of the heterskedasticity is of interest in itself. We refer to Nelder and Wedderburn (1972), Amemyia (1973), Jobson and Fuller (1980), Carroll and Ruppert (1982a), (1982b).

However, if the main problem is to estimate β other methods can be considered which are not optimal for a particular form of heteroskedasticity but are still good for a broad class of alternatives. We refer to Fuller and Rao (1978), White (1982), Carroll (1982). A possible approach to replace σ_t^2 in (1.1) by some weigthed sum $\hat{\sigma}_{t0}^2$ of squared LS-residuals in $(e_{10},\ldots,e_{n0})' = e_0 = y-Xb_0$. A particular approach used by Robinson (1987) is to assume that $\sigma_t^2 = \sigma^2(x_t)$ with $\sigma^2(.)$ some unknown function of the vector of explanatory variables x_t. Based on the work of Stone (1977), he has shown that there exist estimators for β which have the same asymptotic efficiency as the WLS-estimator (1.1). These estimators are called m-NN (Nearest Neighbour) because here σ_t^2 is estimated by a linear combination of m squared LS-residuals $\hat{\sigma}_{t0}^2$ corresponding to the m closest x_j to x_t.

The question arises if something like this can be done if it is not realistic to assume that σ_t^2 is a function of x_t alone. This may be expected in the context of time series where the index t has the natural interpretation of time. If we consider the errors to be the common effect of all variables not explicitely stated in the model, it is rather unlikely that the variance of the errors can be satisfactory explained by the explanatory variables. It is more likely that σ_t^2 is connected with the σ_j^2 for indices j in the neighbourhood of t. This suggests estimating σ_t^2 with a weighted sum of m squared LS-residuals $\hat{\sigma}_{j0}^2$ around t. We describe such a method which is suitable for symmetrically distributed errors. In the analysis no functional relationship of the errors is imposed.

Let $(w_j>0,\ j\in I_m)$ be a fixed set of m weighting coefficients (not depending on n). We propose an estimator $\hat{\beta}$ for β, which is the result of an iteration procedure stopped after $\hat{q}$ iterations not exceeding a (formal) fixed bound $Q > 0$. It is determined from the condition

$$(1.5) \quad \det(\hat{\Phi}_0) > \ldots > \det(\hat{\Phi}_{\hat{q}}) \le \det(\hat{\Phi}_{\hat{q}+1})$$

with the convention $\hat{q} = Q$ if no such $\hat{q} \le Q$ exists. Here, $\hat{\Phi}_0, \ldots, \hat{\Phi}_Q$ are covariance matrices calculated from the LS-residual vector e_0 and the weighting coefficients (w_j) with the interpretation that $\hat{\Phi}_{\hat{q}}$ estimates the covariance matrix of $\hat{\beta}$. The precise form is specified in section 2.

Let b_q denote the estimator for β at some intermediate step $q < Q$. Then new estimators for the σ_t^2, based on the squared residuals in $(e_{1q}, \ldots, e_{nq})' = e_q = y - Xb_q$ are calculated as

$$(1.6) \quad \hat{\sigma}_{tq}^2 = \sum_{j \in I_m} w_j e_{t+j,q}^2 .$$

(For the simulation results in section 4 we let $e_{tq} = e_{1q}$ for $t < 1$ and $e_{tq} = e_{nq}$ for $t > n$; this particular definition does not influence the asymptotic properties but has effect on the finite sample size results) Using the form (1.1) a new estimator b_{q+1} for the next step is calculated according to

$$(1.7) \quad b_{q+1} = \left(\sum_1^n x_t x_t' / \hat{\sigma}_{tq}^2\right)^{-1} \sum_1^n x_t y_t / \hat{\sigma}_{tq}^2 .$$

The final estimator $\hat{\beta}$ is defined by $\hat{\beta} = b_{\hat{q}}$.

Theorem 2.1 and 2.2 together show the asymptotic normality of b_q:

$$(1.8) \quad \sqrt{n}(b_q - \beta) \to N_k(0, \Phi_q)$$

where $\Phi_q = \text{plim } \hat{\Phi}_q$. Theorem 2.3 shows that

$$(1.9) \quad \sqrt{n}(\hat{\beta} - \beta) \to N_k(0, \Phi_q) \quad , \hat{q} \to q$$

for $0 \le q \le Q$ such that

$$(1.10) \quad \det(\Phi_0) > \ldots > \det(\Phi_q) < \det(\Phi_{q+1})$$

(with the obvious interpretation for $q = Q$).

From (1.9) it follows that the asymptotic efficiency of b_q (or $\hat{\beta}$) with respect to $\tilde{b}$, given by $R_q = \det(\tilde{\Phi})/\det(\Phi_q)$, is larger than that of b_0 unless $q = 0$.

In section 3 we impose a smoothing condition on the variances, which still permits all kinds of heteroskedasticity. Theorem 3.1 shows that the asymptotic efficiency R_q of $\hat{\beta}$ can be expressed as a simple function of R_0, the asymptotic efficiency of the LS-estimator b_0. From this it follows that the proposed method is often better for moderate values of R_0. Furthermore a guidance for choosing the weighting coefficients can be obtained. Roughly spoken, they should be high where the probability of the distribution of the errors is high. In particular it is important that $0 \in I_m$ (with high w_0) for common error distributions with a high probability around 0 (e.g. normal distributions). In that case we often have $q > 1$. This means that more than one iteration step increases the asymptotic efficiency. Only for rather curious (bimodal) error distributions with a low probability around 0 a choice $0 \notin I_m$ (or $0 \in I_m$ with w_0 low) seems to be appropriate. In that case mostly $q = 0$ or 1.

For large values of m the value of R_q often approaches 1. This raises the question of the finite sample behaviour of the method, where the choice of m will have to be made in relation to n. We touch this problem in a particular example of estimating a common level under the presence of heteroskedastic normal distributed errors. Here results are based on simulation.

Finally, section 4 contains the proofs of the theorems.

2. Statement of the main results

With respect to the errors ε_t we assume for the moments that for some $\varepsilon > 0$:

$$(2.1) \quad \inf_t \sigma_t^2 > 0 \qquad , \sup_t E|\varepsilon_t|^{2+\varepsilon} < \infty$$

$$(2.2) \quad \text{if } 0 \in I_m \text{ then } \sup_t E|\tilde{\sigma}_t^{-2}|^{1+\varepsilon} < \infty \text{ else } \sup_t E|\tilde{\sigma}_t^{-4}|^{1+\varepsilon} < \infty$$

where (compare (1.3)):

$$(2.3) \quad \tilde{\sigma}_t^2 = \sum_{j \in I_m} w_j \, \varepsilon_{t+j}^2 \, ,$$

and as conditions for symmetry that:

$$(2.4) \quad \begin{cases} E\{\varepsilon_t/\tilde{\sigma}_t^2\} = 0 \qquad , E\{\varepsilon_t \varepsilon_{t+i}/\tilde{\sigma}_t^4\} = 0 \qquad , i \neq 0 \\ E\{(\varepsilon_{t+i}/\tilde{\sigma}_{t+i}^2)(\varepsilon_{t+j}/\tilde{\sigma}_{t+j}^2)\} = 0 \qquad , i \neq j \, . \end{cases}$$

The conditions (2.1), (2.2) guarantee $0 < \inf E\{\tilde{\sigma}_t^{-2}\} \leq \sup E\{\tilde{\sigma}_t^{-2}| < \infty$ and that similar relations hold for $E\{\varepsilon_t^2/\hat{\sigma}_t^2\}$ and $E\{\varepsilon_t^2/\hat{\sigma}_t^4\}$. (see lemma 4.1) In particular this implies the existence of the moments in (2.4). For symmetrically distributed errors ε_t the conditions (2.4) are fulfilled for any set of weighting coefficients (w_j). However, (2.4) is just what is needed in the proofs of the theorems. For normal distributed errors (2.2) is fulfilled for $m \geq 3$ if $0 \in I_m$ and $m \geq 4$ if $0 \notin I_m$ (see example 1 of section 3).

The explanatory variables are assumed to be deterministic with

$$(2.5) \quad \sup_t |x_t| < \infty \qquad , C_0 = \lim_{n \to \infty} \frac{1}{n} \sum_1^n x_t x_t' > 0 \, .$$

Furthermore we assume also that the following limits exist:

$$(2.6)\quad \begin{cases} C_1 = \lim \frac{1}{n} \sum_1^n x_t x_t' \sigma_t^2 & , \; C_2 = \lim \frac{1}{n} \sum_1^n x_t x_t' / \sigma_t^2 \\ V_0 = \lim \frac{1}{n} \sum_1^n x_t x_t' \; E\{\tilde{\sigma}_t^{-2}\} & \\ V_1 = \lim \frac{1}{n} \sum_1^n x_t x_t' \; E\{\varepsilon_t^2 / \tilde{\sigma}_t^2\} & , \; V_2 = \lim \frac{1}{n} \sum_1^n x_t x_t' \; E\{\varepsilon_t^2 / \tilde{\sigma}_t^4\} \; . \end{cases}$$

With (2.1), (2.2) we see that the second relation in (2.5) implies that all limits in (2.6) are non-singular.

In the following theorems it is assumed that (2.1) - (2.6) hold.

Theorem 2.1

The matrices C_1, V_0, V_1, V_2 are consistently estimated by $\hat{C}_1$, $\hat{V}_0$, $\hat{V}_1$, $\hat{V}_2$, defined by

$$(2.7)\quad \begin{cases} \hat{C}_1 = \frac{1}{n} \sum_1^n x_t x_t' e_{t0}^2 & , \; \hat{V}_0 = \frac{1}{n} \sum_1^n x_t x_t' / \hat{\sigma}_{t0}^2 \\ \hat{V}_1 = \frac{1}{n} \sum_1^n x_t x_t' e_{t0}^2 / \hat{\sigma}_{t0}^2 & , \; \hat{V}_2 = \frac{1}{n} \sum_1^n x_t x_t' e_{t0}^2 / \hat{\sigma}_{t0}^4 \; . \end{cases}$$

Under (2.1), (2.5), (2.6) the asymptotic normality of the WLS-estimator $\tilde{b}$ and the LS-estimator b_0 is guaranteed: (1.2) holds with $\tilde{\Phi} = C_2^{-1}$ and (1.4) with $\Phi_0 = C_0^{-1} C_1 C_0^{-1}$. The following theorem shows that the same holds for b_q and specifies Φ_q in (1.8).

Theorem 2.2

The relation (1.8) holds with

$$(2.8)\quad \Phi_q = A_q V_2 A_q' + A_q V_1 B_q' + B_q V_1 A_q' + B_q C_1 B_q'$$

where

$$(2.9)\quad A_q = \sum_{j=0}^{q-1} (2V_0^{-1} W)^j V_0^{-1} \quad , \; B_q = (2V_0^{-1} W)^q C_0^{-1} \quad , \; W = w_0 V_2$$

(with $w_0 = 0$ if $0 \not\in I_m$)

Corollory

Let $\hat{\Phi}_q$ be defined by (2.8) with replacement of C_1, V_0, V_1, V_2, W by $\hat{C}_1$, $\hat{V}_0$, $\hat{V}_1$, $\hat{V}_2$, $\hat{W}$. Then theorem 2.1 implies that $\hat{\Phi}_q$ is a consistent estimator of Φ_q.

Theorem 2.3

If there exists a q for which (1.10) is fulfilled, then (1.9) holds.

The corollory of theorem 2.2 makes clear that for $0 \notin I_m$ a further iteration beyond $q = 1$ is useless. The asymptotic efficiency is constant for $q \geq 1$ since $w_0 = 0$. In fact the decision between OLS ($q=0$) and one further iteration step ($q=1$) is governed by $\Phi_0 = C_0^{-1}C_1C_0^{-1}$ and $\Phi_q = \Phi_1 = V_0^{-1}V_2V_0^{-1}$. However, for common error distributions it is better to choose $0 \in I_m$ and then Φ_q will depend on q. This will be explained more in detail in section 3.

3. A smoothing condition

Suppose that σ_t is a scale parameter of ε_t. We write E_1 in stead of E if $\sigma_t^2 = 1$ for all t. In the moment conditions (2.1), (2.2), (2.4) we may replace E and sup E by E_1 provided that we assume also that $\sup \sigma_t^2 < \infty$.

We introduce the following smoothing condition:

$$(3.1) \quad \max_{t \notin T_n} |\sigma_{t+1}^2/\sigma_t^2 - 1| \to 0$$

with T_n an exception set of $\{1,...,n\}$ such that $\#T_n$ is bounded in n. This condition is fulfilled for a broad class of heteroskedastic errors. A simple example satisfying (3.1) with $T_n = \emptyset$ is $\sigma_t^2 = \sigma^2(1+\lambda t/n)^2$, $\lambda > 0$. The exception set T_n is introduced to covor jumps: if the σ_t^2 fall into p classes of equal size with fixed levels then (3.1) is fulfilled with $\#T_n = p-1$.

For technical reasons we introduce also the condition that for all $i \in I_m$:

$$(3.2) \quad \text{if } 0 \in I_m \text{ then } E_1|\varepsilon_t^2\varepsilon_{t+i}^2/\tilde{\sigma}_t^6| < \infty \text{ else } E_1|\varepsilon_{t+i}^2/\tilde{\sigma}_t^6| < \infty$$

For normal distributed errors (3.2) is satisfied again for $m \geq 3$ if $0 \in I_m$ and $m \geq 4$ if $0 \notin I_m$ (see example 1).

The corollory of the following theorem shows that R_q can be expressed as a simple function of R_0.

Theorem 3.1

Under the additional conditions (3.1), (3.2) we have

$$(3.3) \quad V_0 = v_0 C_2 \quad , V_1 = v_1 C_0 \quad , V_2 = v_2 C_2$$

where

$$(3.4) \quad v_0 = E_1\{\tilde{\sigma}_t^{-2}\} \quad , v_1 = E_1\{\varepsilon_t^2/\tilde{\sigma}_t^2\} \quad , v_2 = E_1\{\varepsilon_t^2/\tilde{\sigma}_t^4\} .$$

Corollory

Substitution of (3.3) in (2.8) leads to

(3.5) $\Phi_q = a_q\tilde{\Phi} + b_q\Phi_0$

(3.6) $R_q^{-1} = a_q^k + b_q^k R_0^{-1}$

where

(3.7) $$\begin{cases} a_q = c_q^2 v_2/v_0^2 + 2\tau^q c_q v_1/v_0 & , \; b_q = \tau^{2q} \\ c_q = (1-\tau^q)/(1-\tau) & , \; \tau = 2w_0v_2/v_0 \, . \end{cases}$$

From its definition it follows that $0 \le \tau \le 2$.
If $0 \notin I_m$ then $\tau = 0$ and so $R_q = \bar{R}_\infty = (v_0^2/v_2)^k$, $q \ge 1$. Therefore $\hat{\beta}$ is asymptotically equivalent to OLS if $\bar{R}_\infty < R_0$ and to WLS with q = 1 if $\bar{R}_\infty > R_0$.
If $0 \in I_m$ and if the weighting coefficients are choosen such that $\tau < 1$ then $R_q \to R_\infty = \{(1-\tau)^2 v_0{}^2/v_2\}^k$ if $q < \infty$. Therefore we choose $0 \in I_m$ and make iteration steps if $R_q > R_0$. This can be wise even if $R_\infty < R_0$ depending on the maximum value of R_q.

Example 1

Let ε_t have a symmetric distribution such that $\varepsilon_t^2 \sim \Gamma(\rho/\sigma_t^2, \rho)$, where $\Gamma(\lambda,\rho)$ stands for the gamma-distribution with scale-parameter $1/\lambda$ and shape-parameter $\rho > 0$. So $\varepsilon_t \sim N(0,\sigma_t^2)$ for $\rho = \frac{1}{2}$. For $\rho > \frac{1}{2}$ we get an error distribution with even higher probabilities around 0 and for $0 < \rho < \frac{1}{2}$ a bimodal distribution with low probabilities around 0.

In this case the WLS-estimator $\tilde{b}$ for β is MVUE. Therefore the asymptotic efficiency of any other (regular) estimator for β cannot exceed 1.

Simple analytic results can only be obtained for equal weighting coefficients. We restrict ourselves to this case. Then all necessary expectations follow from

$$E_1\{\prod_{j=1}^{m} \varepsilon_j^{2\alpha_j}/(\sum_{j=1}^{m} \varepsilon_j^2)^{\beta}\} = \rho^{\beta-\Sigma\alpha_j}(\Gamma(\Sigma\alpha_j+m\rho-\beta)/\Gamma(\Sigma\alpha_j+m\rho))\Gamma^{-m}(\rho)\prod_{j=1}^{m}\Gamma(\alpha_j+\rho)$$

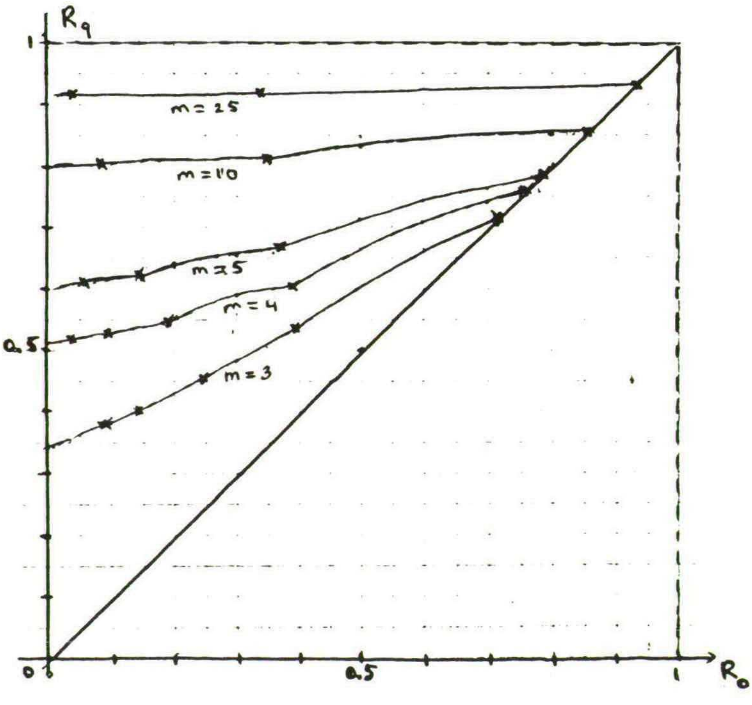
(for $\beta,\alpha_j \geq 0$ and $\beta < \Sigma\alpha_j + m\rho$).

For $0 \notin I_m$ we get $v_0 = 1/(m-1/\rho)$, $v_2 = 1/\{(m-1/\rho)(m-2/\rho)\}$, $\bar{R}_\infty^{1/k} = \{1-2/(m\rho)\}/\{1-1/(m\rho)\}$ provided that $m > 2/\rho$.

For $0 \in I_m$ we get under the weaker condition $m > 1/\rho$ that $v_0 = 1/(m-1/\rho)$, $v_1 = 1/m$, $v_2 = 1/\{m(m-1/\rho)\}$, $\tau = 2/m$, $a_q = c_q(c_q+2\tau^q)(1-1/(m\rho))$. This leads for $m \geq 3$ to $R_\infty^{1/k} = (1-2/m)^2/\{1-1/(m\rho)\}$. Hence, $\bar{R}_\infty > R_\infty$ iff $\rho > \frac{1}{2}$ and $m \geq 2\rho/(2\rho-1)$. This illustrates that for error distributions with high probabilities around 0 (e.g. $\rho \leq \frac{1}{2}$) it is wise to include 0 in I_m.

Example 2

We continue example 1 for $0 \in I_m$ and take the special values $\rho = \frac{1}{2}$ (normal distributed errors) and $k = 1$ (one explanatory variable). The figure below shows for $m = 3,4,5,10,25$ the asymptotic efficiency R_q of $\hat{\beta}$ as a function of R_0. Points at which the number of iteration steps q changes are indicated by stars (e.g. for $m = 3$ the value $R_0 = 0.60$ leads to $R_q = 0.66$ with $q = 1$ and $R_0 = 0.20$ gives $R_q = 0.43$ with $q = 3$). It is seen that $\hat{\beta}$ is better than b_0 for moderate values of R_0.



Example 3

The results of example 2 describe the behaviour of the estimators for large n. In this example we give some simulation results for the (finite) efficiencies $R_0(n) = V\{b_0\}/V\{\tilde{b}\}$ and $R_{\hat{q}}(n) = V\{\hat{\beta}\}/V\{\tilde{b}\}$ of $b_0, \hat{\beta}$ for some particular values of n. Here we take the special case that the explanatory variable is the constant term ($x_t=1$). So the problem becomes to estimate the constant level β_1 of independent normal distributed $y_t \sim N(\beta_1, \sigma_t^2)$, $t = 1,\dots,n$. The estimator $\hat{\beta}$ will be based on equal weighting coefficients with $I_m = \{-[m/2] + 1,\dots,[m/2]\}$.

As a model for the variances we take $\sigma_t^2 = \sigma^2(1+\lambda t/n)^2$, $\lambda > 0$ (see section 3). Then $R_0 = (1+\lambda)/(1+\lambda+\lambda^2/3)$ decreases with increasing λ.

The following table is based on simulation. It gives the values of $R_0(n)$ and $R_{\hat{q}}(n)$ for some interesting values of n, m, and λ based on N = 5000 runs. In each run a sample $(y_1,\dots,y_n)$ was drawn and the estimates b_0, $\tilde{b}$ and $\hat{\beta}$ were calculated. The variances $V\{b_0\}$, $V\{\tilde{b}\}$, $V\{\hat{\beta}\}$ were calculated from those N estimates. The values $\lambda = 0$ and $\lambda = 10$ correspond resp. with homoskedasticity and a large increasing heteroskedasticity. The random generator was that of Logitech's Modula-2 compiler (version 3.0).

		$\lambda = 0$		$\lambda = 10$			
m		n = 25	n = ∞	n = 10	n = 25	n = 100	n = ∞
-	$R_0(n)$	1.00	1.00	0.34	0.28	0.26	0.25
3	$R_{\hat{q}}(n)$	0.97	1.00	0.38	0.34	0.39	0.45
5		0.97	1.00	0.41	0.44	0.62	0.65
10		0.97	1.00	0.43	0.56	0.76	0.81
15		0.98	1.00	-	0.57	0.81	0.87
20		0.99	1.00	-	0.58	0.83	0.90
25		0.99	1.00	-	0.52	0.86	0.92
50		-	1.00	-	-	0.82	0.96
100		-	1.00	-	-	0.60	0.98

In the homoskedastic case of course b_0 is better than $\hat{\beta}$ but not much is lost. In the heteroskedastic case $\hat{\beta}$ is much better than b_0 although the effect is less for small sample sizes. Note that for fixed n efficiency of $R_{\hat{q}}(n)$ attains a maximum in m.

4. Proofs of the theorems

In these proofs c, c_i denote general positive constants not depending on n.

Lemma 4.1

a) $E\{\tilde{\sigma}_t^{-2}\}$, $E\{\varepsilon_t^2/\tilde{\sigma}_t^2\}$, $E\{\varepsilon_t^2/\tilde{\sigma}_t^4\}$ are bounded away from 0.

b) for some $\varepsilon > 0$ $E|\varepsilon_t/\tilde{\sigma}_t|^{1+\varepsilon}$, $E|\varepsilon_t/\tilde{\sigma}_t^3|^{1+\varepsilon}$, $E|\varepsilon_t^2/\tilde{\sigma}_t^2|^{1+\varepsilon}$, $E|\varepsilon_t^2/\tilde{\sigma}_t^4|^{1+\varepsilon}$, $E|\varepsilon_t\varepsilon_{t+i}/\tilde{\sigma}_t^4|^{1+\varepsilon}$ $(i \in I_m)$ are bounded.

Proof

The assertions follow from (2.1), (2.2).

a) For some $i \in I_m$ we have $E\{\tilde{\sigma}_t^{-2}\} \geq cE\{1/\varepsilon_{t+i}^2\} \geq c/E\{\varepsilon_{t+i}^2\} \geq 1/\sup \sigma_t^2$. If $0 \in I_m$ then $\varepsilon_t^2/\tilde{\sigma}_t^2 \geq c$, $\varepsilon_t^2/\tilde{\sigma}_t^4 \geq c/\tilde{\sigma}_t^{-2}$ and if $0 \notin I_m$ then $E\{\varepsilon_t^2/\tilde{\sigma}_t^2\} = E\{\varepsilon_t^2\}E\{\tilde{\sigma}_t^{-2}\}$, $E\{\varepsilon_t^2/\tilde{\sigma}_t^4\} = E\{\varepsilon_t^2\}E\{\tilde{\sigma}_t^{-4}\} \geq E\{\varepsilon_t^2\}(E\{\sigma_t^{-2}\})^2$.

b) If $0 \in I_m$ then $\varepsilon_t \leq c$, $\eta_t \leq c\tilde{\sigma}_t$ and so $|\varepsilon_t/\tilde{\sigma}_t| \leq c$, $|\varepsilon_t/\tilde{\sigma}_t^3| \leq c\tilde{\sigma}_t^{-2}$, $\varepsilon_t^2/\sigma_t^2 \leq c$, $\varepsilon_t^2/\tilde{\sigma}_t^4 \leq c\tilde{\sigma}_t^{-2}$, $|\varepsilon_t\varepsilon_{t+i}|/\tilde{\sigma}_t^4 \leq c\tilde{\sigma}_t^{-2}$.

If $0 \notin I_m$ then $E|\tilde{\sigma}_t^{-2}|^{1+\varepsilon} \leq (E|\tilde{\sigma}_t^{-4}|^{1+\varepsilon})^{1/2}$, $E|\varepsilon_t/\tilde{\sigma}_t^3|^{1+\varepsilon} = E|\varepsilon_t|^{1+\varepsilon}E|\tilde{\sigma}_t^{-3}|^{1+\varepsilon} \leq (E|\varepsilon_t^2|^{1+\varepsilon})^{1/2}(E|\tilde{\sigma}_t^{-4}|^{1+\varepsilon})^{3/4}$, $E|\varepsilon_t^2/\tilde{\sigma}_t^2)^{1+\varepsilon} = E|\varepsilon_t^2|^{1+\varepsilon}E|\tilde{\sigma}_t^{-2}|^{1+\varepsilon}$, $E|\varepsilon_t^2/\tilde{\sigma}_t^4|^{1+\varepsilon} = E|\varepsilon_t^2|^{1+\varepsilon}E|\tilde{\sigma}_t^{-4}|^{1+\varepsilon}$, $E|\varepsilon_t\varepsilon_{t+i}/\tilde{\sigma}_t^4|^{1+\varepsilon} \leq cE|\varepsilon_t/\tilde{\sigma}_t^3|^{1+\varepsilon}$. □

For estimators for β and related variables at some iteration stage some preparatory lemma's are needed.

Let $b = b_n$ be any estimator for β such that $b_n - \beta = O(1/\sqrt{n})$ (in probability). Let $e = y - Xb$, $f = \varepsilon - e$ with corresponding components $e_t = \varepsilon_t - f_t$, $f_t = x_t'(b_n - \beta)$. According to (1.6) set $\hat{\sigma}_t^2 = \sum_j w_j e_{t+j}^2$ and let

$$\eta_t = \max_{j \in I_m} |\varepsilon_{t+j}|, \quad g_n = |b_n - \beta| / \min_{1 \leq t \leq n} \tilde{\sigma}_t .$$

Note that $\eta_t \leq c_1\tilde{\sigma}_t \leq c_2\eta_t$.

Lemma 4.2

(4.1) $g_n \xrightarrow{P} 0$

(4.2) $\forall\delta > 0$: $P\{\min_{1\le t\le n} \tilde{\sigma}_t \ge \delta \max_{1\le t\le n} |f_t|\} \to 1$.

Proof

For $\delta > 0$, $M > 0$ we have:

$$P\{|g_n| \ge \delta\} = P\{\sqrt{n}(n_n-\beta) \ge \delta\sqrt{n} \min \tilde{\sigma}_t\} \le$$

$$\le P\{\sqrt{n}|b_n-\beta| \ge M\} + P\{\min \tilde{\sigma}_t^2 \le M^2/(n\delta^2)\} .$$

Since $b_n-\beta = O(1/\sqrt{n})$ the first term can be made arbitrary small by taking M large enough. For such M the second term tends to 0 as follows from

$$P\{\min \tilde{\sigma}_t^2 \le c/n\} \le \sum_1^n P\{\tilde{\sigma}_t^2 \le c/n\} = \sum_1^n P\{\tilde{\sigma}_t^{-2} \ge n/c\} \le$$

$$\le \sum_1^n (\tfrac{c}{n})^{1+\varepsilon} E|\tilde{\sigma}_t^{-2}|^{1+\varepsilon} \le c_1 n^{-\varepsilon} \sup E|\tilde{\sigma}_t^{-2}|^{1+\varepsilon} \to 0,$$

using lemma 4.1,b. This proves (4.1). Furthermore, using (2.5):

$$\max|f_t| \le |b_n-\beta|\max|x_t| \le c_1 g_n \min \tilde{\sigma}_t \le c_2 g_n \min \eta_t .$$

Hence, $\max|f_t|/\min \eta_t \le c_2 g_n \xrightarrow{P} 0$, proving (4.2). □

Choose δ such that $0 < \delta < 1$ and set $A_n = \{\min \eta_t \ge \delta \max|f_t|\}$. Then $P(A_n') \to 0$ according to (4.2). The following lemma shows that results about convergence in probability can be obtained by restricting the analysis to the set A_n.

Lemma 4.3

Let $|h_n| \le r_n$ or A_n. Then:

$$r_n = o(\alpha_n) \Rightarrow h_n = o(\alpha_n), \qquad r_n = O(\alpha_n) \Rightarrow h_n = O(\alpha_n) .$$

Proof

For $\Delta > 0$ we have:

$$P\{\alpha_n^{-1}|h_n| \geq \Delta\} = P\{\alpha_n^{-1}|h_n| \geq \Delta,\ A_n\} + P\{\alpha_n^{-1}|h_n| \geq \Delta,\ A_n'\}$$

$$\leq P\{\alpha_n^{-1}|r_n| \geq \Delta\} + P\{A_n'\}$$

and so the result follows from $P\{A_n'\} \to 0$. □

The following inequalities on A_n will be used.

Lemma 4.4

On A_n we have for all $t = 1,\ldots,n$:

(4.3) $\quad |e_{t+j}^2-\varepsilon_{t+j}^2| \leq c|b_n-\beta|(\eta_t+|\varepsilon_{t+j}|),\ j \in I_m$ or $j = 0$

(4.4) $\quad |\hat{\sigma}_t^2-\tilde{\sigma}_t^2| \leq c|b_n-\beta|\eta_t$

(4.5) $\quad \hat{\sigma}_t^2 \geq c\tilde{\sigma}_t^2$

(4.6) $\quad |\hat{\sigma}_t^{-2}-\tilde{\sigma}_t^{-2}| \leq cg_n/\tilde{\sigma}_t^2$

(4.7) $\quad |\varepsilon_t^2/\tilde{\sigma}_t^2-e_t^2/\hat{\sigma}_t^2| \leq cg_n(1+|\varepsilon_t|/\tilde{\sigma}_t+\varepsilon_t^2/\tilde{\sigma}_t^2)$

(4.8) $\quad |\varepsilon_t^2/\tilde{\sigma}_t^4-e_t^2/\hat{\sigma}_t^4| \leq cg_n(\tilde{\sigma}_t^{-2}+|\varepsilon_t|/\tilde{\sigma}_t^3+\varepsilon_t^2/\tilde{\sigma}_t^4)$

(4.9) $\quad |\varepsilon_t/\hat{\sigma}_t^2-\varepsilon_t/\tilde{\sigma}_t^2-2(b_n-\beta)'\sum_j w_jx_{t+j}\varepsilon_{t+j}\varepsilon_t/\tilde{\sigma}_t^4| \leq c|b_n-\beta|\varepsilon_t|/\tilde{\sigma}_t^3$.

Proof

Note that $e_t^2 - \varepsilon_t^2 = f_t(f_t-2\varepsilon_t)$.

(4.3): With (2.5) we get

$$|e_{t+j}^2-\varepsilon_{t+j}^2| \leq |b_n-\beta||x_t|(|f_{t+j}|+2|\varepsilon_{t+j}|) \leq c|b_n-\beta|(\eta_t+|\varepsilon_{t+j}|)$$

(4.4): Since $|\varepsilon_{t+j}| \le \eta_t$ for $j \in I_m$ we have

$$|\hat{\sigma}_t^2-\tilde{\sigma}_t^2| \le \Sigma\, w_j|e_{t+j}^2-\varepsilon_{t+j}^2| \le c|b_n-\beta|\eta_t \ .$$

(4.5): Let j' be such that $|\varepsilon_{t+j'}| = \eta_t$. Then

$$\hat{\sigma}_t^2 \ge w_{j'}e_{t+j'}^2 = w_{j'}(\varepsilon_{t+j'}-f_{t+j'})^2 \ge w_{j'}(1-\delta)^2\varepsilon_{t+j'}^2 =$$

$$= w_{j'}(1-\delta)^2\eta_t \ge c\tilde{\sigma}_t^2 \ .$$

(4.6): With (4.4) and (4.5) we get

$$|\hat{\sigma}_t^{-2}-\tilde{\sigma}_t^{-2}| = |\hat{\sigma}_t^2-\tilde{\sigma}_t^2|\hat{\sigma}_t^{-2}\tilde{\sigma}_t^{-2} \le c_1|\hat{\sigma}_t^2-\tilde{\sigma}_t^2|/\tilde{\sigma}_t^4 \le$$

$$\le c_2|b_n-\beta|\eta_t/\tilde{\sigma}_t^4 \le c_2 g_n\eta_t/\tilde{\sigma}_t^3 \le c_3 g_n/\tilde{\sigma}_t^2 \ .$$

(4.7): With (4.3), (4.5), (4.6) we get

$$|\varepsilon_t^2/\tilde{\sigma}_t^2-e_t^2/\hat{\sigma}_t^2| \le |\varepsilon_t^2-e_t^2|/\hat{\sigma}_t^2 + \varepsilon_t^2|\hat{\sigma}_t^{-2}-\tilde{\sigma}_t^{-2}| \le$$

$$\le c_1|b_n-\beta|(\eta_t+|\varepsilon_t|)/\tilde{\sigma}_t^2 + c_2 g_n\varepsilon_t^2/\tilde{\sigma}_t^2 \le$$

$$\le cg_n(1+|\varepsilon_t|)/\tilde{\sigma}_t + cg_n\varepsilon_t^2/\tilde{\sigma}_t^2 \ .$$

(4.8): In the same way as (4.7) it follows

$$|\varepsilon_t^2/\tilde{\sigma}_t^4-e_t^2/\hat{\sigma}_t^4| \le |\varepsilon_t^2-e_t^2|/\hat{\sigma}_t^4 + \varepsilon_t^2|\hat{\sigma}_t^{-2}-\tilde{\sigma}_t^{-2}|\,|\hat{\sigma}_t^{-2}+\tilde{\sigma}_t^{-2}| \le$$

$$\le cg_n(1+|\varepsilon_t|)/\tilde{\sigma}_t^3 + cg_n\varepsilon_t^2/\tilde{\sigma}_t^4 \ .$$

(4.9): Substitution of

$$\hat{\sigma}_t^{-2} = \tilde{\sigma}_t^{-2} - \tilde{\sigma}_t^{-4}(\hat{\sigma}_t^2-\tilde{\sigma}_t^2) + \tilde{\sigma}_t^{-4}\hat{\sigma}_t^2(\hat{\sigma}_t^2-\tilde{\sigma}_t^2)^2 =$$

$$= \tilde{\sigma}_t^{-2} + \tilde{\sigma}_t^{-4}\sum_j w_j f_{t+j}(2\varepsilon_{t+j}-f_{t+j}) + \tilde{\sigma}_t^{-4}\hat{\sigma}_t^2(\hat{\sigma}_t^2-\tilde{\sigma}_t^2)^2$$

leads with (4.4), (4.5) to

$$|\varepsilon_t/\hat{\sigma}_t^2 - \varepsilon_t/\tilde{\sigma}_t^2 - 2\tilde{\sigma}_t^{-4} \sum_j w_j f_{t+j}\varepsilon_{t+j}\varepsilon_t| \le$$

$$\le (\tilde{\sigma}_t^{-4} \sum_j w_j f_{t+j}^2 + c_1\tilde{\sigma}_t^{-6}|b_n-\beta|^2\eta_t^2)|\underline{\varepsilon}_t| \le$$

$$\le c_2|b_n-\beta|^2\tilde{\sigma}_t^{-4}(1+\eta_t^2/\tilde{\sigma}_t^2)|\varepsilon_t| \le c_3|b_n-\beta|g_n|\varepsilon_t|/\tilde{\sigma}_t^3. \quad \square$$

In agreement with (2.7) set $\hat{C}_1 = \frac{1}{n}\sum x_t x_t' e_t^2$, $\tilde{C}_1 = \frac{1}{n}\sum x_t x_t' \varepsilon_t^2$. Let $\hat{V}_i$, $\tilde{V}_i$ for i = 0,1,2 be defined in the same way, e.g. $\hat{V}_2 = \frac{1}{n}\sum x_t x_t' e_t^2/\hat{\sigma}_t^2$, $\tilde{V}_2 = \frac{1}{n}\sum x_t x_t' \varepsilon_t^2/\tilde{\sigma}_t^2$. We have:

<u>Lemma 4.5</u>

(4.10) $\hat{C}_1 - \tilde{C}_1 \overset{P}{\to} 0 \quad , \hat{V}_i - \tilde{V}_i \overset{P}{\to} 0 \quad (i = 0,1,2)$.

<u>Proof</u>

Let $\|.\| = \|.\|_2$.

Ad C_1: On A_n we get with (2.5), (4.3)

$$\|\hat{C}_1-\tilde{C}_1\| \le c_1.\frac{1}{n}\sum|\varepsilon_t^2-e_t^2| \le c_2|b_n-\beta|\frac{1}{n}\sum_1^n (\eta_t+|\varepsilon_t|) .$$

So lemma 4.3 gives $\hat{C}_1 - C_1 \overset{P}{\to} 0$ if the right hand side tends to 0 in probability. From (2.1) we get $E|\varepsilon_{t+j}| \le \sigma_{t+j} \le \sup \sigma_t \le c$ or $\frac{1}{n}\sum \eta_t = O(1)$, $\frac{1}{n}\sum|\varepsilon_t| = O(1)$ and so this follows from $b_n-\beta = O(1/\sqrt{n})$.

Ad V_i: On A_n we get with (4.6), (4.7), (4.8):

$$\|\hat{V}_0-\tilde{V}_0\| \le c.g_n.\frac{1}{n}\sum \tilde{\sigma}_t^{-2} ,$$

$$\|\hat{V}_1-\tilde{V}_1\| \le c.g_n(1+ \frac{1}{n}\sum|\varepsilon_t|/\tilde{\sigma}_t + \frac{1}{n}\sum \varepsilon_t^2/\tilde{\sigma}_t^2) ,$$

$$\|\hat{V}_2-\tilde{V}_2\| \le c.g_n(\frac{1}{n}\sum \tilde{\sigma}_t^{-2} + \frac{1}{n}\sum|\varepsilon_t|/\tilde{\sigma}_t^3 + \frac{1}{n}\sum \varepsilon_t^2/\tilde{\sigma}_t^4) .$$

So with (4.1) and lemma 4.3 it follows in the same way that $\hat{V}_i - \tilde{V}_i \xrightarrow{P} 0$ provided that $E\{\tilde{\sigma}_t^{-2}\}$, $E|\varepsilon_t/\tilde{\sigma}_t|$, $E|\varepsilon_t^2/\tilde{\sigma}_t^2|$, $E|\varepsilon_t/\tilde{\sigma}_t^3|$, $E|\varepsilon_t^2/\tilde{\sigma}_t^4|$ are bounded in t. However, this follows from lemma 4.1.b. □

The sequence of random variables $U_1, U_2, \ldots$ is called p-dependent ($p \geq 0$) if $(U_1, \ldots, U_t)$ and $(U_{t+p+1}, U_{t+p+2}, \ldots)$ are independent for all t. We have the following weak law of large numbers:

Lemma 4.6

Let $U_1, U_2, \ldots$ be p-dependent with $E\{U_t\} = 0$ and $\sup E|U_t|^{1+\varepsilon} < \infty$ for some $\varepsilon > 0$. Then

$$\frac{1}{n} \sum_1^n a_t U_t \xrightarrow{P} 0$$

for any bounded sequence $a_1, a_2, \ldots$

Proof

The case p = 0 is implied by Rao [1973], excercise 4.5, p. 146. The general case follows easily from this by splitting up the sum in independent parts. □

Lemma 4.7

$$(4.11) \quad \tilde{C}_1 - C_1 \xrightarrow{P} 0 \quad , \tilde{V}_i - V_i \xrightarrow{P} 0 \qquad (i = 0,1,2) \ .$$

Proof

Ad C_1: Since

$$\tilde{C}_1 - C_1 = \frac{1}{n} \Sigma\, x_t x_t' (\varepsilon_t^2 - \sigma_t^2)$$

this follows from (2.5) and lemma 4.6 for p = 0 provided that $\sup E|\varepsilon_t^2 - \sigma_t^2|^{1+\varepsilon/2} < \infty$ for some $\varepsilon > 0$. However, this is implied by the condition $\sup E|\varepsilon_t|^{2+\varepsilon} < \infty$ in (2.1).

Ad V_2: Since

$$\tilde{V}_2 - V_2 = \frac{1}{n} \Sigma\, x_t x_t' (\varepsilon_t^2/\tilde{\sigma}_t^4 - E\{\varepsilon_t^2/\tilde{\sigma}_t^4\})$$

this follows in the same way from lemma 4.5 for $p > \mathrm{dist}(I_m)$ provided that $\sup E|\varepsilon_t^2/\tilde{\sigma}_t^4|^{1+\varepsilon} < \infty$. However, this is guaranteed by lemma 4.1.b.

Ad V_0, V_1: Compare the proof of V_2. □

Corollory

From (4.10), (4.11) we get:

$$(4.12) \quad \hat{C}_1 \xrightarrow{P} C_1 \quad , \hat{V}_i \xrightarrow{P} V_i \qquad (i = 0,1,2).$$

Proof of theorem 2.1

The result follows immediately from (4.12) provided that we can show that $b_0 - \beta = O(1/\sqrt{n})$. However, this follows immediately from $E\{b_0\} = \beta$ and $nV\{b_0\} \to C_0^{-1} C_1 C_0^{-1}$, where C_0, C_1 are specified in (2.5), (2.6). □

Remark

Note that theorem 1.1 is not only true for the OLS-estimator b_0 as starting point but also for any estimator b_0 for which $b_0 - \beta = O(1/\sqrt{n})$.

Lemma 4.8

$$(4.13) \quad \frac{1}{n} \sum_t \sum_j w_j x_t x_{t+j}' \varepsilon_{t+j} \varepsilon_t / \tilde{\sigma}_t^4 \xrightarrow{P} W .$$

Proof

For fixed $j \in I_m$ it follows from lemma 4.6 for some $p \geq m+1$ that

$$\frac{1}{n} \sum_t x_t x_{t+j}' (\varepsilon_{t+j} \varepsilon_t / \tilde{\sigma}_t^4 - E\{\varepsilon_{t+j} \varepsilon_t / \tilde{\sigma}_t^4\}) \xrightarrow{P} 0$$

provided that $\sup E|\varepsilon_{t+j} \varepsilon_t / \tilde{\sigma}_t^4|^{1+\varepsilon} < \infty$. This is guaranteed by lemma 4.1.b. However, by (2.4) we have $E\{\varepsilon_{t+j} \varepsilon_t / \tilde{\sigma}_t^4\} = 0$ for $j \neq 0$. Since $w_0 = 0$ for $0 \notin I_m$ this gives

$$\frac{1}{n}\sum_t\sum_j w_j x_t x'_{t+j}\varepsilon_{t+j}\varepsilon_t/\tilde{\sigma}_t^4 - \frac{1}{n}\sum x_t x'_t w_0 E\{\varepsilon_t^2/\tilde{\sigma}_t^4\} \xrightarrow{P} 0 .$$

Then (4.13) follows with (2.6), (2.9). □

Lemma 4.9

$$(4.14)\quad \frac{1}{\sqrt{n}}\sum_t x_t\varepsilon_t/\hat{\sigma}_t^2 - \frac{1}{\sqrt{n}}\sum_t x_t\varepsilon_t/\tilde{\sigma}_t^2 - 2W.\sqrt{n}(b_n-\beta) \xrightarrow{P} 0 .$$

Proof

On A_n we get with (4.9):

$$\left|\frac{1}{\sqrt{n}}\sum x_t\varepsilon_t/\hat{\sigma}_t^2 - \frac{1}{\sqrt{n}}\sum x_t\varepsilon_t/\tilde{\sigma}_t^2 - 2\left(\frac{1}{n}\sum_t\sum_j w_j x_t x'_{t+j}\varepsilon_{t+j}\varepsilon_t/\tilde{\sigma}_t^4\right).\sqrt{n}(b_n-\beta)\right| \le$$

$$\le c\ \sqrt{n}(b_n-\beta)g_n.\frac{1}{n}\sum_t x_t|\varepsilon_t|/\tilde{\sigma}_t^3 .$$

So with (4.1), lemma 4.3 and $\sup E|\varepsilon_t|/\tilde{\sigma}_t^3 < \infty$ it follows that the left hand side of this inequality tends to 0 in probability. With (4.13) and $b_n-\beta = O(1/\sqrt{n})$ this leads to (4.14). □

Lemma 4.10

$$(4.15)\quad E\{\frac{1}{\sqrt{n}}\sum x_t\varepsilon_t/\tilde{\sigma}_t^2\} = 0$$

$$(4.16)\quad \mathrm{Cov}\{\frac{1}{\sqrt{n}}\sum x_t\varepsilon_t/\tilde{\sigma}_t^2\} \to V_2 .$$

Proof

Both relations follow from (2.4). In particular the left hand side of (4.16) equals

$$E\{\frac{1}{n}\sum_t\sum_s x_t x'_s\varepsilon_t\varepsilon_s/\tilde{\sigma}_t^2\tilde{\sigma}_s^2\} = \frac{1}{n}\sum_t x_t x'_t E\{\varepsilon_t^2/\tilde{\sigma}_t^4\}$$

and this tends to V_2 according to (2.6). □

Corollory

From (4.15), (4.16) and (4.14) we get:

(4.17) $\frac{1}{\sqrt{n}} \sum_t x_t \varepsilon_t / \tilde{\sigma}_t^2$ and $\frac{1}{\sqrt{n}} \sum x_t \varepsilon_t / \hat{\sigma}_t^2$ are P-bounded.

Lemma 4.10

Let

$$(4.18) \quad \hat{b}_n = (\sum_t x_t x_t' / \hat{\sigma}_t^2)^{-1} \sum_t x_t y_t / \hat{\sigma}_t^2,$$

then

$$(4.19) \quad \sqrt{n}(\hat{b}_n - \beta) = V_0^{-1} \frac{1}{\sqrt{n}} \sum_t x_t \varepsilon_t / \tilde{\sigma}_t^2 + 2V_0^{-1} W.\sqrt{n}(b_n - \beta) + o(1)$$

Proof

We have with (4.12), (4.17), (4.14)

$$\sqrt{n}(\hat{b}_n - \beta) = (\frac{1}{n} \sum_t x_t x_t' / \hat{\sigma}_t^2)^{-1} . \frac{1}{\sqrt{n}} \sum_t x_t \varepsilon_t / \hat{\sigma}_t^2 = \hat{V}_0^{-1} \frac{1}{\sqrt{n}} \sum x_t \varepsilon_t / \hat{\sigma}_t^2 =$$

$$= V_0^{-1} \frac{1}{\sqrt{n}} \sum_t x_t \varepsilon_t / \hat{\sigma}_t^2 + o(1) = V_0^{-1} \frac{1}{\sqrt{n}} \sum_t x_t \varepsilon_t / \tilde{\sigma}_t^2 + 2V_0^{-1} W.\sqrt{n}(b_n - \beta) + o(1)$$

and this gives (4.16). ▫

Lemma 4.11

$$(4.20) \quad \sqrt{n}(\underline{b}_q - \beta) = A_q \frac{1}{\sqrt{n}} \sum_t x_t \varepsilon_t / \tilde{\sigma}_t^2 + B_q \frac{1}{\sqrt{n}} \sum x_t \varepsilon_t + o(1) \ .$$

Proof

From (1.7), (4.15), (4.16) it follows that

$$\sqrt{n}(b_{q+1} - \beta) = V_0^{-1} \frac{1}{\sqrt{n}} \sum_t x_t \varepsilon_t / \tilde{\sigma}_t^2 + 2V_0^{-1} W.\sqrt{n}(b_q - \beta) + o(1),$$

provided that $b_q - \beta = O(1/\sqrt{n})$. With (4.17) we see that then also $b_{q+1} - \beta = O(1/\sqrt{n})$. Iteration and substitution of (2.9) leads to

$$\sqrt{n}(b_q-\beta) = A_q \frac{1}{\sqrt{n}} \sum_t x_t\varepsilon_t/\tilde{\sigma}_t^2 + B_q\sqrt{n}(b_0-\beta) + o(1)$$

provided that $b_0-\beta = O(1/\sqrt{n})$. However, this has been shown in the proof of theorem 2.1. Then (4.20) follows with (1.3). □

Lemma 4.12

Let $U_1, U_2, \ldots$ be p-dependent with $E\{U_t\} = 0$ and $\sup E|U_t|^{2+\varepsilon} < \infty$ for some $\varepsilon > 0$. Then

$$\frac{1}{\sqrt{n}} \sum_1^n a_t U_t \overset{L}{\to} N(0,\sigma^2)$$

for any bounded sequence $a_1, a_2, \ldots$ for which

$$\sigma^2 := \lim_{n\to\infty} \frac{1}{n} V\{\sum_1^n a_t U_t\}$$

exists.

Proof

See Anderson (1971), theorem 7.7.9, p. 431. □

Remark

By considering linear combinations the theorem is easily extended to random vectors and a bounded sequence of matrices.

Proof of theorem 2.2

Since $E|\varepsilon_t/\tilde{\sigma}_t^2|^{2+\varepsilon}$ and $E|\varepsilon_t|^{2+\varepsilon}$ are bounded for some $\varepsilon > 0$, we can apply lemma 4.12, remark to the right hand side of (4.20) by taking $p > \text{dist}(I_m)$ It remains to calculate the covariance matrix of the limit distribution. Using (2.4) we get

$$\text{Cov}\{\frac{1}{\sqrt{n}} \sum_t x_t\varepsilon_t/\tilde{\sigma}_t^2\} = \frac{1}{n} \sum_t x_t x_t' E\{\varepsilon_t^2/\tilde{\sigma}_t^4\} \to V_2$$

$$\text{Cov}\{\frac{1}{\sqrt{n}} \sum_t x_t\varepsilon_t\} = \frac{1}{n} \sum_t x_t x_t' \sigma_t^2\} \to C_1$$

$$\mathrm{Cov}\{\frac{1}{\sqrt{n}} \sum_t x_t\varepsilon_t/\tilde{\sigma}_t^2, \frac{1}{\sqrt{n}} \sum_t x_t\varepsilon_t\} = \frac{1}{n} \sum_t x_t x_t' E\{\varepsilon_t^2/\tilde{\sigma}_t^2\} \to V_1$$

and so this covariance matrix equals

$$A_q V_2 A_q' + A_q V_1 B_q + B_q V_1 A_q' + B_q C_1 B_q'.$$

However, this is Φ_q in (2.8) and so (1.8) follows. This completes the proof. □

Proof of theorem 2.3

Under (1.10) we can take $\varepsilon > 0$ less than $\det(\Phi_{q+1}) - \det(\Phi_q)$ and $\det(\Phi_j) - \det(\Phi_{j+1})$ for all $j = 0,\ldots,q-1$. Then $P\{\hat{q}=q\} = P\{\det(\hat{\Phi}_j) > \ldots > \det(\hat{\Phi}_q) \le \det(\hat{\Phi}_{q+1})\} \ge P\{|\det(\hat{\Phi}_j) - \det(\Phi_j)| < \varepsilon/2,\ j = 0,\ldots,Q\} \to 1$ according to theorem 2.2, corollory. This implies $P\{\hat{\beta}=b_q\} \to 1$. With (1.8) this completes the proof of (1.9). □

Proof of theorem 3.1

Let $(\alpha,\beta) = (0,1)$, $(1,1)$ or $(1,2)$. If we can show that

$$\frac{1}{n} \sum x_t x_t' E\{\varepsilon_t^{2\alpha}/\tilde{\sigma}_t^{2\beta}\} - \frac{1}{n} \sum x_t x_t' \sigma_t^{2(\alpha-\beta)} E_1\{\varepsilon_t^{2\alpha}/\tilde{\sigma}_t^{2\beta}\} \to 0 ,$$

then (3.3) follows from (2.6) and (3.4). Let η_0, η_j for $j \in I_m$ have the same distribution as ε_t/σ_t, $\varepsilon_{t+j}/\sigma_{t+j}$ for $j \in I_m$. Then it suffices to prove that

$$\frac{1}{n} \Sigma' |E\{\eta_0^{2\alpha}/(\Sigma\, w_{tj}\eta_j^2)^\beta\} - E\{\eta_0^{2\alpha}/(\Sigma\, w_j\eta_j^2)^\beta\}| \to 0 ,$$

where $w_{tj} = w_j\sigma_{t+j}^2/\sigma_t^2$. Here Σ' denotes summation outside the exception set T_n of (3.1). Note that (3.1) implies $\sup|w_{tj}-w_t| \to 0$. The partial derivative of $f(\eta,v) = \eta_0^{2\alpha}/(\Sigma\, v_j\eta_j^2)^\beta$ is given by $\partial f/\partial v_j = \eta_0^{2\alpha}\eta_j^2/(\Sigma\, v_j\eta_j^2)^{\beta+1}$. So the result follows easily with the mean value theorem provided that

$$E\{\eta_0^{2\alpha}\eta_j^2/(\Sigma\, v_j\eta_j^2)^{\beta+1}\} = E_1\{\varepsilon_t^{2\alpha}\varepsilon_{t+j}^2/\tilde{\sigma}_t^{2(\beta+1)}\} < \infty .$$

However, this is implied by (2.2), lemma 4.1,b and (3.2). □

References

Amemyia, T. (1973) - Regression analysis when the variance of the dependent variable is proportional to the square of its expectation - JASA 68, 928-934.

Anderson, T.W. (1971) - The Statistical Analysis of Time Series - Wiley, New York.

Carroll, R.J. (1982) - Adapting for heteroskedasticity in linear models - Ann.Stat. 10, 1224-1233.

Carroll, R.J. and Ruppert, D. (1982a) - Robust estimation in heteroskedastic linear models, Ann.Stat. 10, 429-441.

Carroll, R.J. and Ruppert, D. (1982b) - A comparison between maximum likelihood and generalized least squares in a heteroskedastic linear model - JASA 77, 878-882.

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

Jobson, J.D. and Fuller, W.A. (1980) - Least squares estimation when the covariance matrix and parameter vector are functionally related - JASA 78, 176-181.

Nelder, F.D. and Wedderburn, R.W.M. (1972) - Generalized linear models - J.Royal Stat. Soc. B. 135, 370-384.

Rao, C.R. (1973) - Linear Statistical Inference and Its Applications, 2nd edit. - Wiley, New York.

Robinson, P.M. (1987) - Asymptotically efficient estimation in the presence of heteroskedasticity of unknown form - Econometrica 55, 875-891.

Stone, C.J. (1977) - Consistent nonparametric regression - Ann.Stat. 5, 595-645.

White, H. (1982) - Instrumental variable regression with independent observations - Econometrica 50, 483-499.

IN 1987 REEDS VERSCHENEN

242 Gerard van den Berg
Nonstationarity in job search theory

243 Annie Cuyt, Brigitte Verdonk
Block-tridiagonal linear systems and branched continued fractions

244 J.C. de Vos, W. Vervaat
Local Times of Bernoulli Walk

245 Arie Kapteyn, Peter Kooreman, Rob Willemse
Some methodological issues in the implementation of subjective poverty definitions

246 J.P.C. Kleijnen, J. Kriens, M.C.H.M. Lafleur, J.H.F. Pardoel
Sampling for Quality Inspection and Correction: AOQL Performance Criteria

247 D.B.J. Schouten
Algemene theorie van de internationale conjuncturele en strukturele afhankelijkheden

248 F.C. Bussemaker, W.H. Haemers, J.J. Seidel, E. Spence
On (v,k,λ) graphs and designs with trivial automorphism group

249 Peter M. Kort
The Influence of a Stochastic Environment on the Firm's Optimal Dynamic Investment Policy

250 R.H.J.M. Gradus
Preliminary version
The reaction of the firm on governmental policy: a game-theoretical approach

251 J.G. de Gooijer, R.M.J. Heuts
Higher order moments of bilinear time series processes with symmetrically distributed errors

252 P.H. Stevers, P.A.M. Versteijne
Evaluatie van marketing-activiteiten

253 H.P.A. Mulders, A.J. van Reeken
DATAAL - een hulpmiddel voor onderhoud van gegevensverzamelingen

254 P. Kooreman, A. Kapteyn
On the identifiability of household production functions with joint products: A comment

255 B. van Riel
Was er een profit-squeeze in de Nederlandse industrie?

256 R.P. Gilles
Economies with coalitional structures and core-like equilibrium concepts

257 P.H.M. Ruys, G. van der Laan
Computation of an industrial equilibrium

258 W.H. Haemers, A.E. Brouwer
Association schemes

259 G.J.M. van den Boom
Some modifications and applications of Rubinstein's perfect equilibrium model of bargaining

260 A.W.A. Boot, A.V. Thakor, G.F. Udell
Competition, Risk Neutrality and Loan Commitments

261 A.W.A. Boot, A.V. Thakor, G.F. Udell
Collateral and Borrower Risk

262 A. Kapteyn, I. Woittiez
Preference Interdependence and Habit Formation in Family Labor Supply

263 B. Bettonvil
A formal description of discrete event dynamic systems including perturbation analysis

264 Sylvester C.W. Eijffinger
A monthly model for the monetary policy in the Netherlands

265 F. van der Ploeg, A.J. de Zeeuw
Conflict over arms accumulation in market and command economies

266 F. van der Ploeg, A.J. de Zeeuw
Perfect equilibrium in a model of competitive arms accumulation

267 Aart de Zeeuw
Inflation and reputation: comment

268 A.J. de Zeeuw, F. van der Ploeg
Difference games and policy evaluation: a conceptual framework

269 Frederick van der Ploeg
Rationing in open economy and dynamic macroeconomics: a survey

270 G. van der Laan and A.J.J. Talman
Computing economic equilibria by variable dimension algorithms: state of the art

271 C.A.J.M. Dirven and A.J.J. Talman
A simplicial algorithm for finding equilibria in economies with linear production technologies

272 Th.E. Nijman and F.C. Palm
Consistent estimation of regression models with incompletely observed exogenous variables

273 Th.E. Nijman and F.C. Palm
Predictive accuracy gain from disaggregate sampling in arima - models

274 Raymond H.J.M. Gradus
The net present value of governmental policy: a possible way to find the Stackelberg solutions

275 Jack P.C. Kleijnen
A DSS for production planning: a case study including simulation and optimization

276 A.M.H. Gerards
A short proof of Tutte's characterization of totally unimodular matrices

277 Th. van de Klundert and F. van der Ploeg
Wage rigidity and capital mobility in an optimizing model of a small open economy

278 Peter M. Kort
The net present value in dynamic models of the firm

279 Th. van de Klundert
A Macroeconomic Two-Country Model with Price-Discriminating Monopolists

280 Arnoud Boot and Anjan V. Thakor
Dynamic equilibrium in a competitive credit market: intertemporal contracting as insurance against rationing

281 Arnoud Boot and Anjan V. Thakor
Appendix: "Dynamic equilibrium in a competitive credit market: intertemporal contracting as insurance against rationing

282 Arnoud Boot, Anjan V. Thakor and Gregory F. Udell
Credible commitments, contract enforcement problems and banks: intermediation as credibility assurance

283 Eduard Ponds
Wage bargaining and business cycles a Goodwin-Nash model

284 Prof.Dr. hab. Stefan Mynarski
The mechanism of restoring equilibrium and stability in polish market

285 P. Meulendijks
An exercise in welfare economics (II)

286 S. Jørgensen, P.M. Kort, G.J.C.Th. van Schijndel
Optimal investment, financing and dividends: a Stackelberg differential game

287 E. Nijssen, W. Reijnders
Privatisering en commercialisering; een oriëntatie ten aanzien van verzelfstandiging

288 C.B. Mulder
Inefficiency of automatically linking unemployment benefits to private sector wage rates

289 M.H.C. Paardekooper
A Quadratically convergent parallel Jacobi process for almost diagonal matrices with distinct eigenvalues

290 Pieter H.M. Ruys
Industries with private and public enterprises

291 J.J.A. Moors & J.C. van Houwelingen
Estimation of linear models with inequality restrictions

292 Arthur van Soest, Peter Kooreman
Vakantiebestemming en -bestedingen

293 Rob Alessie, Raymond Gradus, Bertrand Melenberg
The problem of not observing small expenditures in a consumer expenditure survey

294 F. Boekema, L. Oerlemans, A.J. Hendriks
Kansrijkheid en economische potentie: Top-down en bottom-up analyses

295 Rob Alessie, Bertrand Melenberg, Guglielmo Weber
Consumption, Leisure and Earnings-Related Liquidity Constraints: A Note

296 Arthur van Soest, Peter Kooreman
Estimation of the indirect translog demand system with binding non-negativity constraints

IN 1988 REEDS VERSCHENEN

297 Bert Bettonvil
Factor screening by sequential bifurcation

298 Robert P. Gilles
On perfect competition in an economy with a coalitional structure

299 Willem Selen, Ruud M. Heuts
Capacitated Lot-Size Production Planning in Process Industry

300 J. Kriens, J.Th. van Lieshout
Notes on the Markowitz portfolio selection method

301 Bert Bettonvil, Jack P.C. Kleijnen
Measurement scales and resolution IV designs: a note

302 Theo Nijman, Marno Verbeek
Estimation of time dependent parameters in lineair models using cross sections, panels or both

303 Raymond H.J.M. Gradus
A differential game between government and firms: a non-cooperative approach

304 Leo W.G. Strijbosch, Ronald J.M.M. Does
Comparison of bias-reducing methods for estimating the parameter in dilution series

305 Drs. W.J. Reijnders, Drs. W.F. Verstappen
Strategische bespiegelingen betreffende het Nederlandse kwaliteits-concept

306 J.P.C. Kleijnen, J. Kriens, H. Timmermans and H. Van den Wildenberg
Regression sampling in statistical auditing

307 Isolde Woittiez, Arie Kapteyn
A Model of Job Choice, Labour Supply and Wages

308 Jack P.C. Kleijnen
Simulation and optimization in production planning: A case study

309 Robert P. Gilles and Pieter H.M. Ruys
Relational constraints in coalition formation

310 Drs. H. Leo Theuns
Determinanten van de vraag naar vakantiereizen: een verkenning van materiële en immateriële factoren

311 Peter M. Kort
Dynamic Firm Behaviour within an Uncertain Environment

312 J.P.C. Blanc
A numerical approach to cyclic-service queueing models