This is to certify that the

dissertation entitled

Modified Cox Tests for Time Series and Panel Data

presented by

Donggeun Kim

has been accepted towards fulfillment
of the requirements for

Ph. D degree in Economics

Major professor

Date May 17, 2002

Modified Cox Tests for Time Series and Panel Data

By

Donggeun Kim

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Economics

2002

Abstract

Modified Cox Test for Time Series and Panel Data

By

Donggeun Kim

It has long been one of the main interests among econometricians to test nonnested models between two different families. But computational difficulties of the nonnested testing have restricted its application to rather simple linear or nonlinear regression models. This dissertation proposes a new approach based upon the conditional mean and the conditional variance specification in order to solve the computational difficulties and to extend its application to more complicated cases including time series and dynamic panel data. The first chapter of this dissertation proposes a modified Cox test under normality, examines its application to two different nonlinear error equation models with three different time series data sets, performs Monte Carlo experiments to investigate the potential applicability of our proposed test. Chapter two extends its applicability under nonnormality and develops a robust modified Cox test under nonnormality. Chapter three presents its application to the nonlinear dynamic panel data models with the U.S. patents and R&D expenditures data.

To my parents

ACKNOWLEDGEMENTS

During my time at MSU, I have been helped and supported by many good people. I thank you all.

I would like to thank my committee members, Richard Baillie, Anna-Maria Herrera, Christine Amsler, and Ernest Betts for their suggestions and support.

I must express my fullest respect to my mentor, Jeffrey M. Wooldridge. This dissertation could not have started without his insightful guidance, invaluable advice, constant encouragement, and extreme patience.

I would like to thank Ernest Betts for his support. I have been able to learn many important things while working for him as a tutorial coordinator.

I would like to thank Korean economics fellow students, especially my peer, Byungrae Choi and his family for their affection and special friendship. Also, I would like to thank my pastor Rick Erickson for his steadfast support and prayer.

I must thank my parents for their endless love and unlimited support. Without them, I would not be here today.

Even though my dissertation could not have been written without the helps and support of these special people, any remaining error is of my own responsibility.

Contents

List of Figures

List of Tables

Chapter 1

A Modified Cox Test for Dynamic Models of Conditional Means and Variances

1.1 Introduction

Since Cox (1961, 1962) devised a specification testing based upon a modification of the Neyman-Pearson maximum-likelihood ratio, testing nonnested models has been one of the main interests among econometricians. However, the application of the nonnested Cox test has been restricted to rather simple linear or nonlinear regression models mainly due to its complicated and, in many cases, intractable derivation of the pseudo-true value in the second part of the Cox test. [See, for example, Pesaran and Deaton (1978), Gourieroux, Monfort, and Trognon (1983), and Mizon and Richard (1986).] In general, the quasi-maximum likelihood estimate (QMLE) of a nonlinear model does not have a closed form, so it may not be possible to obtain the analytical derivation of its pseudo-true value and its finite sample estimation of the pseudo-true value in the Cox test.

To avoid these computational difficulties some authors developed alternative approaches. Davidson and Mackinnon (1981) combined the two nonnested models as an artificial nesting model and replaced the nuisance parameter under the null hypothesis with the estimated value under the alternative hypothesis to avoid the identification problems. For example, suppose there are two competing specifications M_1 and M_2,

$$M_1 : y_t = m_t(X_t, \gamma) + u_t, where \quad u_t \mid X_t \sim i.i.d(0, \sigma^2), t = 1, \cdots, T \tag{1.1}$$

$$M_2 : y_t = \mu_t(X_t, \delta) + v_t, where \quad v_t \mid X_t \sim i.i.d(0, \tau^2), t = 1, \cdots, T \tag{1.2}$$

then Davidson and Mackinnon (1981) transformed these two nonnested models as

$$y_t = (1-\lambda)m_t(X_t, \gamma) + \lambda\mu_t(X_t, \delta) + \omega_t, where \quad \omega_t \mid X_t \sim i.i.d(0, \eta^2), t = 1, \cdots, T \tag{1.3}$$

They replaced δ with an OLS estimate, $\hat{\delta}$, under M_2 instead of the pseudo-true value of δ underM_1 and tested if λ=0. The DM test can be written as

$$y_t = m_t(X_t, \gamma_0) + \lambda(\mu_t(X_t, \hat{\delta}) - m_t(X_t, \gamma_0)) + \omega_t \tag{1.4}$$

Now we can regard the DM test as an omitted variables test of $(\mu_t(X_t, \hat{\delta}) - m_t(X_t, \hat{\gamma}))$ in the nonlinear model $y_t = m_t(X_t^0) + e_t^0$. If there are nonnormality, heteroscedasticity, or serial correlation, their test statistics becomes invalid. Wooldridge (1990,1991) suggested a robust version of Davidson and Mackinnon test (DM test) by modifying the misspecification indicator of his conditional Mean Encompassing test (CME test). Under heteroskedasticity, a robust version of DM test is derived by simply setting the misspecification indicator $\hat{\lambda} \equiv (\mu_t(X_t, \hat{\delta}) - m_t(X_t, \hat{\gamma}))$ and applying the CME test procedure. For the weighted nonlinear least squares (WLNS) estimator, a robust DM test is obtained by setting $\hat{\lambda} \equiv (\hat{h_t}/\hat{\eta_t})(\mu_t(X_t, \hat{\delta}) - m_t(X_t, \hat{\gamma}))$ (See, Wooldridge (1990)). For possible nonzero correlation between the residuals $\hat{e_t} = y_t - m(X_t, \hat{\gamma})$ and a particular weighting of the difference in the estimated regression functions, set the misspecification indicator $\hat{\lambda} \equiv \hat{C_{t1}}\hat{C_{t2}^{-1}}(\mu_t(X_t, \hat{\delta}) - m_t(X_t, \hat{\gamma}))$, where $\hat{C_{t1}}$ is the estimated variance function for the model under the null and $\hat{C_{t2}}$ is the estimated variance function for the model under the alternative (See, Wooldridge 1991). On the other hand, Pesaran and Pesaran (1993) offered another approach to deal with the computational difficulties of obtaining the pseudo-true value of the Cox test by a method of stochastic simulation.

Let $H_f : f(y_t, \alpha \mid X_t)$ and $H_g : g(y_t, \beta \mid X_t)$ be the two nonnested competing models. Then the Cox test (1961,1962) is based upon

$$\begin{aligned} T_f &= \{L_f(\hat{\alpha}) - L_g(\hat{\beta})\} - E_{\hat{\alpha}}\{L_f(\hat{\alpha}) - L_g(\hat{\beta})\} \qquad (1.5) \\ &= L_f(\hat{\alpha}) - L_g(\hat{\beta}) + C(\hat{\alpha}, \hat{\beta}_*), \qquad (1.6) \\ & \quad where \quad C(\hat{\alpha}, \hat{\beta}_*) = E_{\hat{\alpha}}\{L_f(\hat{\alpha}) - L_g(\hat{\beta})\} \end{aligned}$$

$L_f(\hat{\alpha}) = T^{-1}\sum_{t=1}^{T} \log f(y_t, \alpha \mid X_t), L_g(\hat{\beta}) = T^{-1}\sum_{t=1}^{T} \log g(y_t, \beta \mid X_t)$ are the maximized log likelihood functions under H_f, H_g respectively, and $C(\hat{\alpha}, \hat{\beta}_*)$ is the unconditional expectation of the log likelihood ratio when the null is correctly specified. To obtain $\hat{\beta}_*$ by simulation method, a $T \times 1$ vector of independent observations of y_t is artificially generated under H_f and then the ML estimate of β is derived by using these artificially generated observations under H_g. This procedure is replicated R times to obtain

$$\hat{\beta}_* = \frac{1}{R}\sum_{i=1}^{R} \hat{\beta}_i \qquad (1.7)$$

Then, the same procedure is applied to obtain $C(\hat{\alpha}, \hat{\beta}_*)$ by the same simulation method

$$C(\hat{\alpha}, \hat{\beta}_*) = \frac{1}{R}\sum_{j=1}^{R} \{L_f(y_j, \hat{\alpha}) - L_g(y_j, \hat{\beta}_*)\} \qquad (1.8)$$

Even though Pesaran and Pesaran(1993) argues that these estimators obtained by the simulation method converge to the pseudo-true values consistently and fairly quickly with a relatively small number of replications, this simulated method is not such a favorable approach to the practitioners. Besides, it is very difficult to use the original Cox test if the given models contain the lagged dependent variables: $f_t(y_t, \alpha \mid x_t, y_{t-1}, x_{t-1}, \ldots)$ and $g_t(y_t, \beta \mid x_t, y_{t-1}, x_{t-1}, \ldots)$ because potentially very severe computational difficulties arise from computing the unconditional expectation of the differenced log likelihood functions. Bera and Higgins (1997) presented nonnested Cox test results using the stochastic simulation method proposed by Pesaran and Pesaran (1993) between two nonlinear equation error models, the autoregressive conditional heteroscedasticity(ARCH) and the bilinear models with three time series data sets.

The difficulty in applying the original Cox test in time series applications and possibly dynamic panel data is that it requires computing the unconditional expectation of the differenced log likelihood functions when the null is correctly specified. In this paper, we propose a new approach to solve the computational difficulties of the Cox statistic by using conditional mean and conditional variance. Our approach here is to compute, for each t, the conditional expectation. In some important applications including ARCH and GARCH models in time series, this approach leads to substantial simplifications. Another attractive feature of our approach is that we can test other distributional features because our approach uses the first two conditional moments while the DM test is for the conditional mean,

$E(y_t \mid x_t)$, only. In section 2, we describe our new modified Cox test procedure; in section 3, we present an empirical result with three time series data sets; in section 4, we provide simulation experiments of this modified Cox test and we draw conclusions in section 5.

1.2 A Modified Cox Test

1.2.1 Motivation and General Concepts

Suppose that there are T individually, identically distributed random variables $y_t, t = 1, \cdots, T$. $f(y_t, \alpha)$ is the probability density function under the null hypothesis, H_f, and $g(y_t, \beta)$ is the probability density function under the alternative hypothesis, H_g, where α, β are unknown parameters, and $f(y_t, \alpha)$ and $g(y_t, \beta)$ belong to separate families. If H_f is not nested in H_g, and H_g is not nested in H_f, then it is said that the two hypotheses, H_f and H_g, are nonnested with each other. If one model can account for the results from the other model, then the former is said to encompass the latter. [see Mizon and Richard (1986), and Hendry and Richard (1990).] This means that a correctly specified model can explain the results of its competing model and the pseudo-true value is the probability limit of the alternative model under the null hypothesis. Thus, the nonnested test statistic devised by Cox

(1961,1962) is an example of encompassing test. The Cox test, T_f,of H_f against H_g is based on

$$T_f = \{L_f(\hat{\alpha}) - L_g(\hat{\beta})\} - E_{\hat{\alpha}}\{L_f(\hat{\alpha}) - L_g(\hat{\beta})\} \tag{1.9}$$

where $L_f(\hat{\alpha}), L_g(\hat{\beta})$ are the maximized log likelihood functions under H_f and H_g respectively and $\hat{\alpha}, \hat{\beta}$ are maximum log likelihood estimators. The test statistic is based upon the difference between the log likelihood ratio and its expected estimate under the null hypothesis, H_f. If $E_{\hat{\alpha}}\{L_f(\hat{\alpha})-L_g(\hat{\beta})\} = 0$, then the Cox test statistic is just simplified to the form of log likelihood ratio statistic, but, in general, this term is nonzero under nonnested hypotheses. So the Cox test takes the deviation between the maximum log likelihood ratio and its expected value under the null hypothesis. Under the correctly specified null hypothesis, T_f should be close to zero while a large deviation from zero constitutes evidence against the null hypothesis. The standardized Cox test statistic, $\sqrt{T}\frac{T_f}{\hat{V}_f^{1/2}}$, where $\hat{V}_f$ is a consistent estimator of the asymptotic variance of T_f, is asymptotically distributed as unit normal. White (1982) provided general regularity conditions and the asymptotic normality of the Cox test statistic.

Despite its theoretically refined feature, the derivation of pseudo-true value of $E_{\hat{\alpha}}\{L_f(\hat{\alpha}) - L_g(\hat{\beta})\}$ is not straightforward, and even analytically intractable.

To solve these computational difficulties, we offer a new approach based upon the conditional mean and conditional variance method.

What makes difficult to apply the Cox test is that it requires computing the unconditional expectation of the differenced log likelihood functions that is not significantly analytical or tractable in many cases. The observations are assumed independent in case of Cox (1960,1961) and it reduces the computational difficulties in some degree but it still requires significant computational effort. Besides, for this reason, it becomes very challenging to apply the original Cox test to time series applications that contain the lagged dependant variables as the explanatory variables. But these difficulties can be avoided by computing the conditional expectation, for each t. And this leads to substantial simplifications in some important applications including ARCH and GARCH models. White (1994) showed the computational simplification of the second part of the Cox test using conditional densities f_t and g_t given I_{t-1} where I_{t-1} is the information set (σ-algebra generated by $\{x_t, y_{t-1}, x_{t-1}, \cdots\}$) available at time t.

Let $H_f : f_t(x_t, \alpha)$ and $H_g : g(x_t, \beta)$ be the two nonnested competing models, then their maximized log likelihood functions are, respectively,

$$L_n(\hat{\alpha}) = \frac{1}{T}\sum_{t=1}^{T} \log f_t(x_t, \hat{\alpha}) \tag{1.10}$$

$$L_n(\hat{\beta}) = \frac{1}{T}\sum_{t=1}^{T} \log g_t(x_t, \hat{\beta}) \tag{1.11}$$

and let $f : \Re^{vt} \times \alpha \to \Re^{+}$ and $g : \Re^{vt} \times \beta \to \Re^{+}$ be conditional densities and $\hat{\alpha}$ and $\hat{\beta}$ be the QMLEs under H_f and H_g respectively. An estimate of the expected value of the average log likelihood ratio when the null is correctly specified is

$$E_{\hat{f}}[\hat{L_{fn}} - \hat{L_{gn}}] \equiv \frac{1}{n}\int (\log f^n(x^n, \hat{\alpha}_n) - \log g^n(x^n, \hat{\beta}_n)) f^n(x^n, \hat{\alpha}_n) dv^n(x^n) \quad (1.12)$$

$$= \int (n^{-1} \sum_{t=1}^{n} (\log f_t(x^t, \hat{\alpha}_n) - \log g_t(x^t, \hat{\beta}_n)) \prod_{t=1}^{n} f_t(x^t, \hat{\alpha}_n) dv^n(x^n) \quad (1.13)$$

$$= n^{-1} \sum_{t=1}^{n} [\int (\log f_t(x^t, \hat{\alpha}_n) - \log g_t(x^t, \hat{\beta}_n)) f_t(x^t, \hat{\alpha}) dv^t(x^t)] \quad (1.14)$$

$$where f^n \equiv \prod_{t=1}^{n} f_t$$

The vt-fold integration in the equation(1.14) causes the severe difficulties of computing the unconditional expectation. It is assumed that the observations are independent in the case of Cox (1960,1961), so we can reduce the integral above equation as v-fold integral

$$= \frac{1}{n} \sum_{t=1}^{n} [\int (\log f_t(x_t, \hat{\alpha}) - \log g_t(x_t, \hat{\beta})) f_t(x_t, \hat{\alpha}) dv_t(x_t)] \quad (1.15)$$

White (1994) argues that it still requires computational effort, even though this is much more tractable than before. In addition, the analytical intractability still remains when we apply Cox test to time series applications that include the lagged dependent variables as explanatory variables. To avoid these difficulties, we suggest an approach using the conditional densities of f_t and g_t given I_{t-1}. By computing the conditional expectation for each t, we can achieve some substantial simplifications of the Cox test in some important applications. Now we can rewrite the second part of Cox test as

$$
\begin{aligned}
& E_{\hat{\alpha}}\left[L_f(\hat{\alpha}) - L_g(\hat{\beta})\right] \\
= & \frac{1}{T}\sum_{t=1}^{T}[\int(\log f_t(x_t,\hat{\alpha} \mid I_{t-1}) - \log g_t(x_t,\hat{\beta} \mid I_{t-1}))f_t(x_t,\hat{\alpha} \mid I_{t-1})dv_t(x_t)] \quad (1.16) \\
= & \frac{1}{T}\sum_{t=1}^{T}\left[E_{\hat{f}_t}[\log f_t(x_t,\hat{\alpha} \mid I_{t-1}) - \log g_t(x_t,\hat{\beta} \mid I_{t-1}) \mid I_{t-1}]\right] \quad (1.17)
\end{aligned}
$$

Equation (1.17) is the conditional expectation of the differenced log likelihood functions.

1.2.2 A Modified Cox Test

Let $\{y_t, Z_t; t = 1, \cdots, T\}$ be a sequence of observable random vectors with y_t $1 \times J$, and Z_t $1 \times K$; y_t is the vector of endogenous variables, and Z_t is the vector of explanatory variables. We assume the regularity conditions in White (1982) hold. Suppose the two competing nonnested parametric models f_t under the null and g_t under the alternative respectively, then

$$M_1 : \quad f_t(y_t \mid I_{t-1}, \theta_0), \quad \theta_0 \in \Theta \subseteq \Re^p, t = 1, 2, \cdots T \tag{1.18}$$

$$M_2 : \quad g_t(y_t \mid I_{t-1}, \delta_0), \quad \delta_0 \in \Delta \subseteq \Re^q, t = 1, 2, \cdots T \tag{1.19}$$

where I_{t-1} is the information set (σ-algebra generated by $\{y_{t-1}, Z_t, \cdots, Z_1\}$) available at time t.

If $\hat{\theta}$ is $\sqrt{T}$-consistent estimator of θ_0 under the null hypothesis when θ_0 is a true value of θ, and $\hat{\delta}$ is $\sqrt{T}$-consistent estimator of δ_0 under the alternative hypothesis when δ_0 is a true value of δ, then $\sqrt{T}(\hat{\theta}-\theta_0)$ and $\sqrt{T}(\hat{\delta}-\delta_0)$ are distributed as asymptotic normal. The null hypothesis is that M_1 is correctly specified and the alternative hypothesis is that M_2 is correctly specified. Now we write the modified Cox test as

$$\begin{aligned} T_{M_1} &= T^{-1}\sum_{t=1}^{T}\{\log f_t(y_t \mid I_{t-1};\theta_0) - \log g_t(y_t \mid I_{t-1};\delta^*)\} \\ &\quad -T^{-1}\sum_{t=1}^{T}\{E_{M_1}(\log f_t(y_t \mid I_{t-1};\theta_0) - \log g_t(y_t \mid I_{t-1};\delta^*) \mid I_{t-1})\} \end{aligned} \tag{1.20}$$

where $\delta^* = plim\hat{\delta}$ when M_1 is true. It is important to note that $\delta^* \neq \delta_0$ in general. θ_0 and δ_0 are the true parameters which are unknown, so we replace them with $\hat{\theta}$ and $\hat{\delta}$; the ML estimators of θ_0, and δ_0 respectively. This modified Cox test is composed of two parts, the averaged log likelihood ratio of the null hypothesis to the alternative hypothesis and its expected value under the null hypothesis. Now we analyze these two terms one after another.

i) The first term: Let the log likelihood functions of $\log f_t(y_t \mid I_{t-1};\hat{\theta})$ and $\log g(y_t \mid I_{t-1};\hat{\delta})$ be

$$\log f_t(y_t \mid I_{t-1};\hat{\theta}) = -\frac{1}{2}\log 2\pi - \frac{1}{2}\log h_t(\hat{\theta}) - \frac{1}{2}\frac{(y_t - m_t(\hat{\theta}))^2}{h_t(\hat{\theta})} \tag{1.21}$$

$$\log g_t(y_t \mid I_{t-1};\hat{\delta}) = -\frac{1}{2}\log 2\pi - \frac{1}{2}\log \eta_t(\hat{\delta}) - \frac{1}{2}\frac{(y_t - \mu_t(\hat{\delta}))^2}{\eta_t(\hat{\delta})} \tag{1.22}$$

Plug these two log likelihood functions into the first part of the Cox test and we can rewrite this term as

$$\sum_{t=1}^{T}\{\log f_t(y_t \mid I_{t-1};\hat{\theta}) - \log g_t(y_t \mid I_{t-1};\hat{\delta})\}$$
$$= -\frac{T}{2}\{\log h_t(\hat{\theta}) - \log \eta_t(\hat{\delta})\} - \frac{1}{2}\sum_{t=1}^{T}\left\{\frac{(y_t - m_t(\hat{\theta}))^2}{h_t(\hat{\theta})} - \frac{(y_t - \mu_t(\hat{\delta}))^2}{\eta_t(\hat{\delta})}\right\} \quad (1.23)$$

ii) The second term: The conditional expectation of the log likelihood ratio of $\log f_t(y_t \mid I_{t-1};\hat{\theta})$ to $\log g(y_t \mid I_{t-1};\hat{\delta})$ is

$$\sum_{t=1}^{T}\left[E_{M_1}\{\log f_t(y_t \mid I_{t-1};\hat{\theta}) - \log g_t(y_t \mid I_{t-1};\hat{\delta}) \mid I_{t-1}\}\right]$$
$$= \frac{T}{2}\log \eta_t(\hat{\delta}) - \frac{T}{2}\log h_t(\hat{\theta}) - \frac{T}{2} + \frac{1}{2}\sum_{t=1}^{T}\frac{h_t(\hat{\theta})}{\eta_t(\hat{\delta})} + \frac{1}{2}\sum_{t=1}^{T}\frac{m_t(\hat{\theta}) - \mu_t(\hat{\delta})}{\eta_t(\hat{\delta})} \quad (1.24)$$

Now we combine these two terms together and rewrite the modified Cox test as

$$\begin{aligned}\hat{T}_{M_1} &= \frac{1}{2} - \frac{T^{-1}}{2}\sum_{t=1}^{T}\left\{\frac{(y_t - m_t(\hat{\theta}))^2}{h_t(\hat{\theta})} - \frac{(y_t - \mu_t(\hat{\delta}))^2}{\eta_t(\hat{\delta})} - \frac{h_t(\hat{\theta})}{\eta_t(\hat{\delta})}\right. \\ &\quad \left. - \frac{m_t(\hat{\theta}) - \mu_t(\hat{\delta})}{\eta_t(\hat{\delta})}\right\} &(1.25)\\ &= T^{-1}\sum_{t=1}^{T}\left\{(y_t - m_t(\hat{\theta}))\frac{m_t(\hat{\theta}) - \mu_t(\hat{\delta})}{\eta_t(\hat{\delta})} + \frac{u_t(\hat{\theta})^2 - h_t(\hat{\theta})}{2}\right. \\ &\quad \left. \times\left(\frac{1}{\eta_t(\hat{\delta})} - \frac{1}{h_t(\hat{\theta})}\right)\right\} &(1.26)\end{aligned}$$

A more detailed derivation of this result is given in the appendix.

The equation (1.26) above is the modified Cox test statistic when we assume M_1 is correctly specified. If we exchange the role of the hypotheses, i.e. M_2 becomes the null hypothesis and M_1 becomes the alternative hypothesis, then under H_g, this modified Cox test has a different form as

$$\begin{aligned} T_{M_2} &= T^{-1}\{\log g_t(y_t \mid I_{t-1};\delta_0) - \log f_t(y_t \mid I_{t-1};\theta^*)\} \\ &\quad - \left[T^{-1}\sum_{t=1}^{T} E_{M_2}\{\log g_t(y_t \mid I_{t-1};\delta_0) - \log f_t(y_t \mid I_{t-1};\theta^*) \mid I_{t-1}\}\right] \quad (1.27) \\ &= T^{-1}\sum_{t=1}^{T}\left[(y_t - \mu_t(\delta_0))\frac{\mu_t(\delta_0) - m_t(\theta^*)}{h_t(\theta^*)} + \frac{\varepsilon_t(\delta_0)^2 - \eta_t(\delta_0)}{2}\right. \\ &\quad \left.\times\left(\frac{1}{h_t(\theta^*)} - \frac{1}{\eta_t(\delta_0)}\right)\right] \quad (1.28) \end{aligned}$$

Now we consider the asymptotic distribution of the modified Cox test under the null hypothesis that M_1 is correctly specified. Since we do not know the true values of parameters, θ_0 and δ^*, we use the consistent estimates, $\hat{\theta}$ and $\hat{\delta}$ instead, so the test statistic is based upon

$$\begin{aligned} \hat{T}_{M_1} &= T^{-1}\sum_{t=1}^{T}\left[(y_t - m_t(\hat{\theta}))\left\{\frac{m_t(\hat{\theta}) - \mu_t(\hat{\delta})}{\eta_t(\hat{\delta})}\right\}\right. \\ &\quad \left.+ \frac{u_t(\hat{\theta})^2 - h_t(\hat{\theta})}{2}\left(\frac{1}{\eta_t(\hat{\delta})} - \frac{1}{h_t(\hat{\theta})}\right)\right] \quad (1.29) \end{aligned}$$

We can expand $\hat{T}_{M_1}$ by the mean-value theorem as

$$\begin{aligned} &= T^{-1}\sum_{t=1}^{T}\left[(y_t - m_t(\theta_0)\left\{\frac{m_t(\theta_0) - \mu_t(\delta^*)}{\eta_t(\delta^*)}\right\} + \frac{u_t(\theta_0)^2 - h_t(\theta_0)}{2}\right. \\ &\quad \left.\times\left(\frac{1}{\eta_t(\delta^*)} - \frac{1}{h_t(\theta_0)}\right)\right] + T^{-1}\sum_{i=1}^{p}\sum_{t=1}^{T}\frac{\partial}{\partial\theta_i}\left[(y_t - m_t(\bar{\theta}))\left\{\frac{m_t(\bar{\theta}) - \mu_t(\bar{\delta})}{\eta_t(\bar{\delta})}\right\}\right. \\ &\quad \left.+\frac{u_t(\bar{\theta})^2 - h_t(\bar{\theta})}{2}\left(\frac{1}{\eta_t(\bar{\delta})} - \frac{1}{h_t(\bar{\theta})}\right)\right](\hat{\theta} - \theta_0) \quad (1.30) \end{aligned}$$

where $\bar{\theta}$ and $\bar{\delta}$ lie on the segment between $(\hat{\theta}, \theta_0)$ and $(\hat{\delta}, \delta^*)$.

Now we multiply $\sqrt{T}$ on both sides

$$
\begin{aligned}
= \quad & T^{-1/2}\sum_{t=1}^{T}\left[(y_t - m_t(\theta_0)\left\{\frac{m_t(\theta_0)-\mu_t(\delta^*)}{\eta_t(\delta^*)}\right\} + \frac{u_t(\theta_0)^2 - h_t(\theta_0)}{2}\right. \\
& \times\left.\left(\frac{1}{\eta_t(\delta^*)} - \frac{1}{h_t(\theta_0)}\right)\right] + T^{-1}\sum_{i=1}^{p}\sum_{t=1}^{T}\frac{\partial}{\partial\theta_i}\left[(y_t - m_t(\bar{\theta}))\left\{\frac{m_t(\bar{\theta})-\mu_t(\bar{\delta})}{\eta_t(\bar{\delta})}\right\}\right. \\
& \left.+\frac{u_t(\bar{\theta})^2 - h_t(\bar{\theta})}{2}\left(\frac{1}{\eta_t(\bar{\delta})} - \frac{1}{h_t(\bar{\theta})}\right)\right]\sqrt{T}(\hat{\theta}-\theta_0) \qquad (1.31)
\end{aligned}
$$

$$
\begin{aligned}
= \quad & T^{-1/2}\sum_{t=1}^{T}\left[(y_t - m_t(\theta_0)\left\{\frac{m_t(\theta_0)-\mu_t(\delta^*)}{\eta_t(\delta^*)}\right\} + \frac{u_t(\theta_0)^2 - h_t(\theta_0)}{2}\right. \\
& \times\left.\left(\frac{1}{\eta_t(\delta^*)} - \frac{1}{h_t(\theta_0)}\right)\right] + T^{-1}\sum_{t=1}^{T}\nabla_\theta\left[(y_t - m_t(\theta_0))\left\{\frac{m_t(\theta_0)-\mu_t(\delta^*)}{\eta_t(\delta^*)}\right\}\right. \\
& \left.+\frac{u_t(\theta_0)^2 - h_t(\theta_0)}{2}\left(\frac{1}{\eta_t(\delta^*)} - \frac{1}{h_t(\theta_0)}\right)\right]\sqrt{T}(\hat{\theta}-\theta_0) \qquad (1.32)
\end{aligned}
$$

under the null hypothesis and $(\bar{\theta} \to \theta_0)$ and $(\bar{\delta} \to \delta^*)$ by the mean-value property and ∇_θ is the gradient operator.

Define

$$
\begin{aligned}
\Psi_t(\theta_0, \delta^*) \quad \equiv \quad & T^{-1}\sum_{t=1}^{T}\nabla_\theta\left[(y_t - m_t(\theta_0))\left\{\frac{m_t(\theta_0)-\mu_t(\delta^*)}{\eta_t(\delta^*)}\right\}\right. \\
& \left.+\frac{u_t(\theta_0)^2 - h_t(\theta_0)}{2}\left(\frac{1}{\eta_t(\delta^*)} - \frac{1}{h_t(\theta_0)}\right)\right]
\end{aligned}
$$

Then

$$
\begin{aligned}
& T^{-1/2}\sum_{t=1}^{T}\left[(y_t - m_t(\hat{\theta}))\left\{\frac{m_t(\hat{\theta})-\mu_t(\hat{\delta})}{\eta_t(\hat{\delta})}\right\} + \frac{u_t(\hat{\theta})^2 - h_t(\hat{\theta})}{2}\left(\frac{1}{\eta_t(\hat{\delta})} - \frac{1}{h_t(\hat{\theta})}\right)\right] \\
& -T^{-1/2}\sum_{t=1}^{T}\left[(y_t - m_t(\theta_0)\left\{\frac{m_t(\theta_0)-\mu_t(\delta^*)}{\eta_t(\delta^*)}\right\}\right.
\end{aligned}
$$

$$+\frac{u_t(\theta_0)^2 - h_t(\theta_0)}{2}\left(\frac{1}{\eta_t(\delta^*)} - \frac{1}{h_t(\theta_0)}\right)\Bigg]$$

$$-\Psi_t(\theta_0, \delta^*)\sqrt{T}(\hat{\theta} - \theta_0) \xrightarrow{p} 0 \tag{1.33}$$

The asymptotic distribution of $\sqrt{T}\hat{T}_{M_1}$ is equivalent to the asymptotic distribution of $\sqrt{T}\left[(y_t - m_t(\theta_0)\left\{\frac{m_t(\theta_0)-\mu_t(\delta^*)}{\eta_t(\delta^*)}\right\} + \frac{u_t(\theta_0)^2-h_t(\theta_0)}{2}\left(\frac{1}{\eta_t(\delta^*)} - \frac{1}{h_t(\theta_0)}\right)\right] +$ $\Psi_t(\theta_0, \delta^*)\sqrt{T}(\hat{\theta}-\theta_0)$ and $\sqrt{T}(\hat{\theta}-\theta_0) = \left(-\frac{1}{T}\sum_{t=1}^{T} A_t(\theta_0)\right)^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\nabla_\theta \log f(y_t, \theta_0)$. Note that $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[(y_t - m_t(\theta_0)\left\{\frac{m_t(\theta_0)-\mu_t(\delta^*)}{\eta_t(\delta^*)}\right\} + \frac{u_t(\theta_0)^2-h_t(\theta_0)}{2}\left(\frac{1}{\eta_t(\delta^*)} - \frac{1}{h_t(\theta_0)}\right)\right]$ $-\Psi_t(\theta_0, \delta^*)\left(E\left[\frac{1}{T}\sum_{t=1}^{T} A_t(\theta_0)\right]\right)^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\nabla_\theta \log f(y_t, \theta_0)$ is martingale difference sequence random variable with mean zero and variance, $V(\theta_0, \delta^*)$, under the null hypothesis, where $V(\theta_0, \delta^*)$ is

$$V(\theta_0, \delta^*) = T^{-1}\sum_{t=1}^{T}\left[D_t - \left(\frac{1}{T}\sum_{t=1}^{T}\psi(\theta_0, \delta^*)\right)\left(\frac{1}{T}\sum_{t=1}^{T} A_t(\theta_0)\right)^{-1}\nabla_\theta \log f(y_t, \theta_0)\right]^2 \tag{1.34}$$

Define

$$D_t \equiv \left[u_t(\theta_0)\left(\frac{m_t(\theta_0) - \mu_t(\delta^*)}{\eta_t(\delta^*)}\right) + \frac{u_t(\theta_0)^2 - h_t(\theta_0)}{2}\left(\frac{1}{\eta_t(\delta^*)} - \frac{1}{h_t(\theta_0)}\right)\right]$$

$$A_t(\theta_0) \equiv \partial^2 \log f(y_t, \theta_0)/\partial\theta_i\partial\theta_j$$

$$\psi(\theta_0, \delta^*) \equiv \nabla_\theta\left[u_t(\theta_0)\left(\frac{m_t(\theta_0) - \mu_t(\delta^*)}{\eta_t(\delta^*)}\right) + \frac{u_t(\theta_0)^2 - h_t(\theta_0)}{2}\left(\frac{1}{\eta_t(\delta^*)} - \frac{1}{h_t(\theta_0)}\right)\right]$$

Therefore, $\sqrt{T}\hat{T}_{M_1} \sim N(0, V(\theta_0, \delta^*))$ and $\frac{\sqrt{T}\hat{T}_{M_1}}{\hat{V}^{1/2}} \overset{a}{\sim} N(0,1)$ if $\hat{V}$ is a consistent estimator of $V(\theta_0, \delta^*)$. Under the regularity conditions, it is easily shown that

$$\hat{V}_T = T^{-1}\sum_{t=1}^{T}\left[\hat{D}_t - \left(\frac{1}{T}\sum_{t=1}^{T}\psi(\hat{\theta}, \hat{\delta})\right)\left(\frac{1}{T}\sum_{t=1}^{T} A_t(\hat{\theta})\right)^{-1}\nabla_\theta \log f_t(\hat{\theta})\right]^2 \tag{1.35}$$

$$\begin{aligned}
Thus, \quad & \hat{V}_T - V(\theta_0, \delta^*) \\
= \quad & \frac{1}{T}\sum_{t=1}^{T}\left[\hat{D}_t - \left(\frac{1}{T}\sum_{t=1}^{T}\psi(\hat{\theta},\hat{\delta})\right)\left(\frac{1}{T}\sum_{t=1}^{T}A_t(\hat{\theta})\right)^{-1}\nabla_\theta \log f_t(\hat{\theta})\right]^2 \\
& -\frac{1}{T}\sum_{t=1}^{T}\left[D_t - \left(\frac{1}{T}\sum_{t=1}^{T}\psi(\theta_0,\delta^*)\right)\left(\frac{1}{T}\sum_{t=1}^{T}A_t(\theta_0)\right)^{-1}\nabla_\theta \log f(y_t,\theta_0)\right]^2 \\
\xrightarrow{p} \quad & 0 \qquad (1.36)
\end{aligned}$$

Under the null hypothesis, the statistic of the modified Cox test is asymptotically normally distributed with mean zero and variance, $\hat{V}_{M_1}$. Thus, the standardized modified Cox test, $\frac{\sqrt{T}\hat{T}_{M_1}}{\hat{V}_{M_1}^{1/2}}$, is asymptotically distributed as unit normal $N(0,1)$ under the null hypothesis.

We now consider a time series application as an example. Suppose $y_t, t = 1, 2, \cdots, T$ is a sequence of $i.i.d$ observable random variables. Two competing models are given as

$$M_1 : y_t = m_t(\theta_0) + u_t, \quad where \quad u_t \sim N(0, h_t(\theta_0)), \qquad (1.37)$$

$$E(y_t \mid I_{t-1}) = m_t(\theta_0), \qquad (1.38)$$

$$Var(y_t \mid I_{t-1}) = h_t(\theta_0), \quad where \quad h_t(\theta_0) = \alpha_0 + \alpha_1 u_{t-1}^2 \cdots ARCH(1)(1.39)$$

$$and \quad M_2 : y_t = \mu_t(\delta_0) + \varepsilon_t, \qquad (1.40)$$

$$E(y_t \mid I_{t-1}) = \mu_t(\delta_0) + b_{11}\varepsilon_{t-1}\xi_{t-1} \qquad (1.41)$$

$$Var(y_t \mid I_{t-1}) = \sigma_\xi^2, \qquad (1.42)$$

$$where \quad \varepsilon_t = b_1\varepsilon_{t-1}\xi_{t-1} + \xi_t, \quad and \quad \xi_t \sim N(0, \sigma_\xi^2)$$

Under the null hypothesis that M_1 is correctly specified, we can write the modified Cox test as

$$\hat{T}_{M_1} = T^{-1} \sum_{t=1}^{T} \left[u_t(\hat{\theta}) \frac{m_t(\hat{\theta}) - \mu_t(\hat{\delta})}{\hat{\sigma}_\xi^2} + \frac{u_t^2(\hat{\theta}) - (\hat{\alpha_0} + \hat{\alpha_1} u_{t-1}^2)}{2} \times \left(\frac{1}{\hat{\sigma}_\xi^2} - \frac{1}{\hat{\alpha_0} + \hat{\alpha_1} u_{t-1}^2} \right) \right] \quad (1.43)$$

Define

$$D_{t1} \equiv \frac{m_t(\theta_0) - \mu_t(\delta^*)}{\sigma_\xi^2},$$

$$D_{t2} \equiv \frac{1}{\sigma_\xi^2} - \frac{1}{h_t(\theta_0)}, \quad \text{and}$$

$$D_t \equiv u_t(\theta_0) D_{t1} + \frac{u_t(\theta_0)^2 - h_t(\theta_0)}{2} D_{t2}$$

$$\text{then} \quad \sqrt{T}\hat{T}_{M1} = T^{-1/2} \sum_{t=1}^{T} \left\{ u_t(\hat{\theta}) \hat{D}_{t1} + \frac{u_t(\hat{\theta})^2 - h_t(\hat{\theta})}{2} \hat{D}_{t2} \right\} \quad (1.44)$$

$$\text{Therefore,} \quad T^{-1/2} \sum_{t=1}^{T} \left\{ u_t(\hat{\theta}) \hat{D}_{t1} + \frac{u_t(\hat{\theta})^2 - h_t(\hat{\theta})}{2} \hat{D}_{t2} \right\} - T^{-1/2} \sum_{t=1}^{T} \left\{ u_t(\theta_0) D_{t1} + \frac{u_t(\theta_0) - h_t(\theta_0)}{2} D_{t2} \right\} + \left(\frac{1}{T} \sum_{t=1}^{T} \psi_t(\theta_0, \delta^*) \right) \sqrt{T}(\hat{\theta} - \theta_0) \xrightarrow{p} 0 \quad (1.45)$$

$$\text{where} \quad T^{-1} \sum_{t=1}^{T} \psi_t(\theta_0, \delta^*) = T^{-1} \sum_{t=1}^{T} \nabla_\theta \left[u_t(\theta_0) D_{t1} + \frac{u_t(\theta_0)^2 - h_t(\theta_0)}{2} D_{t2} \right]$$

$$\text{and} \quad \sqrt{T}(\hat{\theta} - \theta_0) = -\left(\frac{1}{T} \sum_{t=1}^{T} A_t(\theta_0) \right)^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \nabla_\theta \log f(y_t, \theta_0)$$

The asymptotic distributions of the modified Cox test are as follows

$$\sqrt{T}\hat{T}_{M_1} = T^{-1/2}\sum_{t=1}^{T}\left[D_t - \left(p\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\psi_t(\theta_0,\delta^*)\right)\left(p\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}A_t(\theta_0)\right)^{-1}\right.$$
$$\left.\times\nabla_\theta \log f(y_t,\theta_0)\right] \quad (1.46)$$

$$Let \quad q_t = D_t - \left(\frac{1}{T}\sum_{t=1}^{T}\psi_t(\theta_0,\delta^*)\right)\left(\frac{1}{T}\sum_{t=1}^{T}A_t(\theta_0)\right)^{-1}\nabla_\theta \log f(y_t,\theta_0)$$

Then,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}q_t = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[D_t - \left(\frac{1}{T}\sum_{t=1}^{T}\psi_t(\theta_0,\delta^*)\right)\left(\frac{1}{T}\sum_{t=1}^{T}A_t(\theta_0)\right)^{-1}\nabla_\theta \log f(y_t,\theta_0)\right] \quad (1.47)$$

$$= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[D_t - \left(p\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\psi_t(\theta_0,\delta^*)\right)\left(p\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}A_t(\theta_0)\right)^{-1}\right.$$
$$\left.\times\nabla_\theta \log f(y_t,\theta_0)\right] + o_p(1) \quad (1.48)$$

$$= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}q_t^* \quad (1.49)$$

where

$$q_t^* = D_t - \left(p\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\psi_t(\theta_0,\delta^*)\right)$$
$$\times\left(p\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}A_t(\theta_0)\right)^{-1}\times\nabla_\theta \log f(y_t,\theta_0)$$

$$and \quad p\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\psi_t(\theta_0,\delta^*) = E\left[\psi_t(\theta_0,\delta^*)\right]$$
$$p\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}A_t(\theta_0) = E\left[A_t(\theta_0)\right]$$

Therefore,

$$E(q_t^* \mid I_{t-1}) = 0 \quad (1.50)$$

$$E(q_t^{*2} \mid I_{t-1}) = E\left[\left(D_t - \left(E\left[\psi_t(\theta_0,\delta^*)\right]\right)\left(E\left[A_t(\theta_0)\right]\right)^{-1}\right.\right.$$

$$\times \nabla_\theta \log f(y_t, \theta_0))^2 \mid I_{t-1}\Big] \tag{1.51}$$

$$= V_{M_1}(\theta_0, \delta^*) \tag{1.52}$$

Therefore,

$$\hat{V}_{M_1} = \frac{1}{T}\sum_{t=1}^{T}\left[\hat{D}_t^2 - 2\hat{D}_t\left(\frac{1}{T}\sum_{t=1}^{T}\psi_t(\hat{\theta},\hat{\delta})\right)\left(\frac{1}{T}\sum_{t=1}^{T}A_t(\hat{\theta})\right)^{-1}\nabla_\theta \log f(y_t,\hat{\theta})\right.$$
$$+\left(\frac{1}{T}\sum_{t=1}^{T}\psi_t(\hat{\theta},\hat{\delta})\right)\left(\frac{1}{T}\sum_{t=1}^{T}A_t(\hat{\theta})\right)^{-1}\nabla_\theta \log f(y_t,\hat{\theta})\nabla_\theta \log f(y_t,\hat{\theta})'$$
$$\left.\left(\frac{1}{T}\sum_{t=1}^{T}A_t(\hat{\theta})\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}\psi_t(\hat{\theta},\hat{\delta})\right)'\right] \tag{1.53}$$

The modified Cox test statistic under the conditional mean and variance, $\frac{\sqrt{T}\hat{T}_{M_1}}{\hat{V}_{M_1}^{1/2}}$, is standard normal, $N(0,1)$.

- Proposition

  Assume that the following conditions are satisfied under the null hypothesis,

  1. Regularity conditions[1] hold [see White (1982).]
  2. $T^{1/2}(\hat{\theta} - \theta_0) \rightarrow o_p(1)$ and $T^{1/2}(\hat{\delta} - \delta_0) \rightarrow o_p(1)$
  3. Conditional mean and conditional variance exist and are finite.

$$\textit{Then,}\quad \hat{T}_{M_1} = T^{-1}\sum_{t=1}^{T}\left[(y_t - m_t(\hat{\theta}))\frac{m_t(\hat{\theta}) - \mu_t(\hat{\delta})}{\eta_t(\hat{\delta})}\right.$$
$$\left.+\frac{u_t(\hat{\theta})^2 - h_t(\hat{\theta})}{2}\left(\frac{1}{\eta_t(\hat{\delta})} - \frac{1}{h_t(\hat{\theta})}\right)\right] \tag{1.54}$$

[1]The regularity conditions are given in the appendix.

and the standardized Cox test statistic, $\frac{\sqrt{T}\hat{T}_{M_1}}{\hat{V}_{M_1}^{1/2}}$ is asymptotically distributed as unit normal, $N(0,1)$, where $\hat{V}_{M_1}$ is the consistent asymptotic variance of $\sqrt{T}\hat{T}_{M_1}$.

Note that the equation (1.54) is a function of $\hat{\theta}$ and $\hat{\delta}$, the ML estimators of θ_0 and δ_0 respectively. Comparing to the Cox test (1961,1962) and the simulation method by Pesaran and Pesaran (1993), the modified Cox test does not require pseudo-true parameters or estimators from artificially generated data. This approach, based upon conditional mean and conditional variance specifications, is a more convenient method for a computational purpose. Now we apply this proposition as follows;

- Procedure 1.1

1. Obtain $\hat{\theta}$ and $\hat{\delta}$, the ML estimators of θ_0 and δ^*, save residuals, $u_t(\hat{\theta})$, and the conditional variance, $h_t(\hat{\theta})$ from the log likelihood function $\log f(y_t \mid I_{t-1};\hat{\theta})$ and $e_t(\hat{\delta})$, and $\eta(\hat{\delta})$ from the log likelihood function $\log g_t(y_t \mid I_{t-1};\hat{\delta})$.

2. Compute $\hat{D}_{t1}$, $\hat{D}_{t2}$, $\psi_t(\hat{\theta},\hat{\delta})$, and $\nabla_\theta \log f_t(\hat{\theta})$. Define $\hat{D}_{t1} \equiv \frac{m_t(\hat{\theta})-\mu_t(\hat{\delta})}{\eta_t(\hat{\delta})}$, $\hat{D}_{t2} \equiv \frac{1}{\eta_t(\hat{\delta})} - \frac{1}{h_t(\hat{\theta})}$, and $\frac{1}{T}\sum_{t=1}^T \psi_t(\hat{\theta},\hat{\delta}) \equiv \frac{1}{T}\sum_{t=1}^T \left[-\nabla_\theta \hat{m}_t \hat{D}_{t1} - \frac{\nabla_\theta \hat{h}_t}{2}\hat{D}_{t2}\right]$.

3. Compute $\sqrt{T}\hat{T}_{M_1} = T^{-1/2}\sum_{t=1}^T \Big[\hat{u}_t\hat{D}_{t1} + \frac{\hat{u}_t^2-\hat{h}_t}{2}\hat{D}_{t2} - (\frac{1}{T}\sum_{t=1}^T \psi(\hat{\theta},\hat{\delta}))$
$\times(\frac{1}{T}\sum_{t=1}^T A(\hat{\theta}))^{-1}\nabla_\theta \log f_t(\hat{\theta})\Big]$ and
$\hat{V}_{M_1} = \frac{1}{T}\sum_{t=1}^T \Big[\hat{D}_{t1}^2 h_t(\hat{\theta}) + \frac{h_t(\hat{\theta})^2}{4}\hat{D}_{t2}^2 + (\frac{1}{T}\sum_{t=1}^T \psi(\hat{\theta},\hat{\delta}))(\frac{1}{T}\sum_{t=1}^T A(\hat{\theta}))^{-1}$

$$\times\nabla_\theta \log f_t(\hat{\theta})\nabla_\theta \log f_t(\hat{\theta})'(\tfrac{1}{T}\textstyle\sum_{t=1}^T A(\hat{\theta}))^{-1}(\tfrac{1}{T}\sum_{t=1}^T \psi(\hat{\theta},\hat{\delta}))'$$

$$-2(\tfrac{1}{T}\textstyle\sum_{t=1}^T \psi(\hat{\theta},\hat{\delta}))(\tfrac{1}{T}\sum_{t=1}^T A(\hat{\theta}))^{-1}(\hat{D}_{t1}\nabla_\theta m_t(\hat{\theta}) + \frac{\hat{D}_{t2}}{4h_t(\hat{\theta})}\nabla_\theta h_t(\hat{\theta}))\Big]$$

and use the standardized Cox test statistic, $\frac{\sqrt{T}\hat{T}_{M_1}}{\hat{V}_{M_1}^{1/2}}$, as asymptotic unit normal under the null hypothesis.

1.3 Empirical Application

Bera and Higgins (1997) took generalized autoregressive conditional heteroscedasticity (GARCH) by Bollerslev (1986) and bilinearity by Granger and Anderson (1978) as two competing models for nonlinear dependence in time series data and showed the nonnested Cox test results using a stochastically simulated method by Pesaran and Pesaran (1993) with three time series data sets; S&P 500 stock index, the daily pound/dollar exchange rate, and the rate of growth of the monthly U.S. index of industrial production. In this section we compare our modified Cox test results to those results from Bera and Higgins (1997).

1.3.1 GARCH and Bilinearity

Forecasting as well as estimating a model are very substantial components in econometrics. These components play a very important role in the analysis of time series data. If a series is assumed as a white noise (this is very common assumption in econometrics), the process is independent of its own past and it becomes very difficult to forecast this series because we cannot get any information from its own past. But most of the financial and macroeconomic data in time series shows evidence of a dependency upon the past. Granger and Anderson (1978) suggested that a white noise process could be forecastable from its own past in a nonlinear manner and introduced a bilinear process that allowed dependence on the past realization of the series.

Suppose $X_t = \beta\varepsilon_{t-1}X_{t-1} + \varepsilon_t$ where ε_t is white noise with mean zero and variance σ_ε^2 and X_t and ε_t are uncorrelated. The conditional mean X_t is $\beta X_{t-1}\varepsilon_{t-1}$ while the unconditional mean is zero because

$$E(X_t \mid I_{t-1}) = \beta X_{t-1}\varepsilon_{t-1} + E(\varepsilon_t \mid I_{t-1}) \tag{1.55}$$

$$= \beta X_{t-1}\varepsilon_{t-1} \tag{1.56}$$

$$\textit{while} \quad E(X_t) = E(\beta X_{t-1}\varepsilon_{t-1} + \varepsilon_t) \tag{1.57}$$

$$= \beta E(X_{t-1})E(\varepsilon_{t-1}) + E(\varepsilon_t) \tag{1.58}$$

$$= 0 \tag{1.59}$$

where I_{t-1} is an information set ($\sigma-$algebra) available at time t. Next, the conditional variance of y_t is σ_ε^2 while the unconditional variance is $\frac{\sigma_\varepsilon^2}{1-\beta^2\sigma_\varepsilon^2}$ because

$$Var(X_t \mid I_{t-1}) = Var(\varepsilon_t \mid I_{t-1}) \tag{1.60}$$

$$= \sigma_\varepsilon^2 \tag{1.61}$$

$$\begin{aligned} while \quad Var(X_t) &= \beta^2 Var(X_t) Var(\varepsilon_{t-1}) + Var(\varepsilon_t) & (1.62) \\ &= \beta^2 Var(X_t)\sigma_\varepsilon^2 + \sigma_\varepsilon^2 & (1.63) \\ Therefore, \quad Var(X_t) &= \frac{\sigma_\varepsilon^2}{1-\beta^2\sigma_\varepsilon^2} & (1.64) \end{aligned}$$

If $\beta^2\sigma_\varepsilon^2 < 1$, then this process is non-explosive and becomes stationary. So $\beta^2\sigma_\varepsilon^2 < 1$ is a very important condition for stationarity.

Engle (1982) further developed the idea of nonlinearity in his model, the autoregressive conditional heteroscedasticity (ARCH), which is very close to the bilinear process. Suppose $y_t = X_t'\beta + \varepsilon_t$, then $y_t \sim N(X_t'\beta, h_t)$ where $h_t = h(\varepsilon_{t-1}, \cdots, \varepsilon_{t-p}, \alpha)$. The conditional mean and the unconditional mean are both $X_t'\beta$:

$$\begin{aligned} E(y_t \mid I_{t-1}) &= X_t'\beta + E(\varepsilon_t \mid I_{t-1}) & (1.65) \\ &= X_t'\beta & (1.66) \\ and \quad E(y_t) &= E(X_t'\beta + \varepsilon_t) & (1.67) \\ &= X_t'\beta + E(\varepsilon_t) & (1.68) \\ &= X_t'\beta & (1.69) \end{aligned}$$

The conditional variance is $h_t = \alpha_0 + \alpha_1\varepsilon_{t-1} + \cdots + \alpha_p\varepsilon_{t-p}$ but the unconditional variance is σ_ε^2;

$$Var(y_t \mid I_{t-1}) = Var(\varepsilon_t \mid I_{t-1}) \qquad (1.70)$$

$$= h_t \tag{1.71}$$

$$and \quad Var(y_t) = E(\varepsilon_t^2) \tag{1.72}$$

$$= \sigma_\varepsilon^2 \tag{1.73}$$

Note that the conditional variance, h_t, contains the current and lagged values of independent variables through information set available at time t because $\varepsilon_t = y_t - X_t'\beta$. Thus, we can decompose the h_t as follows[2]

$$h_t = h_t(\varepsilon_{t-1}, \varepsilon_{t-2}, \cdots, \varepsilon_{t-p}, \alpha, X_t, X_{t-1}, \cdots, X_{t-p}) \tag{1.74}$$

$$= h_t(\varepsilon_{t-1}, \varepsilon_{t-2}, \cdots, \varepsilon_{t-p}, \alpha) h_t(X_t, X_{t-1}, \cdots, X_{t-p}) \tag{1.75}$$

Bollerslev (1986) extended the ARCH process to the generalized autoregressive conditional heteroscedasticity (GARCH) process allowing for a longer memory and a more flexible lag structure. The GARCH(p,q) process includes the lagged conditional variances as well as the linear function of past variances of the ARCH(q) process so, it corresponds to and forecasts from its own past in an adaptive expectation fashion . Suppose $y_t = X_t'\beta + \varepsilon_t$ where y_t is the dependant variable, X_t is a vector of independent variables, and β is a vector of unknown parameters, then the GARCH(p,q) process is given as

[2]See p.3 on ARCH selected reading, Engle, 1995

$$\varepsilon_t \mid I_{t-1} \sim N(0, h_t) \tag{1.76}$$

$$where \quad h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i h_{t-i} \tag{1.77}$$

$$and \qquad p \geq 0, \quad q > 0,$$

$$\alpha_0 > 0, \alpha_i \geq 0, \quad i = 1 \cdots, q,$$

$$\beta_i \geq 0, \quad i = 1 \cdots, p.$$

Therefore, $y_t \sim N(X_t'\beta, h_t)$ where $Var(y_t \mid I_{t-1}) = h_t$, while $Var(y_t) = Var(\varepsilon_t) = \sigma_\varepsilon^2$.

The bilinear process and the GARCH(p,q) process as well as the ARCH(q) process have forms of nonlinearity and provide more information for forecastability from their own past realization. Although it is hard to find the true specification between the bilinear and the GARCH processes due to the similarity between them, there are some remarkable differences between these two processes. The main and fundamental difference between the bilinear process and the GARCH (or ARCH) process is the conditional moments condition. The conditional distributions of a dependant variable between these two processes are pretty distinguishable. Suppose a dependant variable y_t is generated by $y_t = X_t'\beta + u_t$ where u_t is a stochastic error. Under the null hypothesis, u_t is specified as

$$M_1 : u_t \mid I_{t-1} \sim N(0, h_t) \quad (1.78)$$

$$where \quad h_t = \alpha_0 + \alpha_1 u_{t-1}^2 + \beta_1 h_{t-1} \cdots GARCH(1,1) \quad process \quad (1.79)$$

and under the alternative hypothesis, u_t is specified as

$$M_2 : u_t = b_{11} u_{t-1} \varepsilon_{t-1} + \varepsilon_t \quad (1.80)$$

$$where \quad \varepsilon_t \sim N(0, \sigma_\varepsilon^2) \cdots Bilinear \quad process \quad (1.81)$$

In the GARCH(1,1) model, $E(y_t \mid I_{t-1}) = X_t'\beta$ and $Var(y_t \mid I_{t-1}) = h_t$ and in the bilinear model, $E(y_t \mid I_{t-1}) = X_t'\beta + b_{11} u_{t-1} \varepsilon_{t-1}$ and $Var(y_t \mid I_{t-1}) = \sigma_\varepsilon^2$. The conditional mean of bilinearity shows that the bilinear process does augment the adaptive information between its past errors and innovations in a nonlinear manner while the conditional variance of the bilinear model is constant. This nonlinearity in the conditional mean of the bilinear model may increase the forecastability of the dependant variable while the GARCH(1,1) process does not bring any augmented information from its own past and innovation from the unconditional or conditional mean. On the other hand, the conditional variance of the GARCH(1,1) process provides augmented adaptive information from its own past realization while the conditional variance of the bilinear process is constant. Although the conditional distributions between the bilinear process and the GARCH process are fundamentally different, the unconditional distributions between these two are very similar, as shown earlier. Due to the nonlinearity and similar unconditional distributions between the bilinear process and the GARCH process, it is more difficult to find the true specification. In the next section we do the modified nonnested Cox test between these two nonlinear specifications with three time series data sets.

1.3.2 Empirical Application

In this application, we consider three time series data sets: the daily percentage changes of the S&P 500 stock index, the daily log price changes of the British pound in terms of the U.S. dollar (£/$), and the annualized growth rate of the U.S. monthly index of industrial production (IP). Note that the first two data sets are high frequency financial time series and the third data set is a non-financial time series. These three data sets are the same ones that Bera and Higgins (1997) used.[3]

We consider that the stochastic error equation follows the nonlinearity and specify the GARCH model as the null hypothesis and the bilinear model as the alternative hypothesis. The exogenous variables are considered as autoregressive

[3]They retained the last 10 per cent of the observations to compute root mean squared errors for the one-step-ahead forecastability from each of models and we used the same data samples as they did for nonnested test between GARCH and bilinear models.

Table 1.1: Summary statistics: S&P 500

	Mean	s.d.	Skew	Kurt	Max	Min	sample size
Bera&Higgins	.060	.820	-.651	8.759	3.468	-5.877	1138
Kim	.042	.925	-.711	8.796	3.455	-7.008	1138

Table 1.2: Summary statistics: British pound

	Mean	s.d.	Skew	Kurt	Max	Min	sample size
Bera&Higgins	-.023	.477	.032	4.758	1.959	-2.252	1210
Kim	.0260	.692	-.202	4.632	2.990	-2.784	1210

(AR) models[4]. Now the model specifications are given as

$$M1 : y_t = X_t'\beta + u_t \tag{1.82}$$

$$u_t \mid I_{t-1} \sim i.i.d(0, h_t) \tag{1.83}$$

$$where \quad h_t = \alpha_0 + \alpha_1 u_{t-1}^2 + \delta h_{t-1} \tag{1.84}$$

$$M2 : y_t = X_t'\beta + u_t \tag{1.85}$$

$$where \quad u_t = b_{11} u_{t-1} \varepsilon_{t-1} + \varepsilon_t, \tag{1.86}$$

$$and \quad \varepsilon_t \sim i.i.d(0, \sigma_\varepsilon^2) \tag{1.87}$$

First, we take the daily S&P 500 stock index (SP) from January 4, 1978 to May 28, 1993 and compare our statistics summary to that of Bera and Higgins (1997)

[4]In modeling of exogenous variables, we take autoregressive (AR) models and the order of the autoregression following Bera and Higgins (1997).

Table 1.3: Summary statistics: IP

	Mean	s.d.	Skew	Kurt	Max	Min	sample size
Bera&Higgins	3.357	10.604	-.645	5.653	37.699	-51.732	359
Kim	2.728	8.823	-.623	5.649	33.483	-42.364	359

in Table 1.1. Next, we take the daily log exchange rate of the British pound to the U.S. dollar (£/$) in a sample period from December 12, 1985 to February 28, 1991 and present the statistics results in Table 1.2. As Bera and Higgins (1997) considered in their paper, we also take the annualized growth rate of the U.S. monthly index of industrial production (IP), a non-financial time series data set, from January, 1960 to March, 1993 for the third empirical application and present the summary statistics in Table 1.3. Note that the summary statistics between Bera and Higgins (1997) and our findings given in Table 1.1 to 1.3, are similar but not exactly the same, even though we used the same data sets with the same sample periods that Bera and Higgins (1997) considered. There are a couple of things to be noted from the summary statistics. First, as given in Table 1.1 through 1.3, all the series are of high kurtosis, especially S&P 500 stock index series. Another is that we have different signs of the mean values in the British Pound series; -0.023 of Bera & Higgins and 0.026 of us.

Table 1.4 through 1.6 present the estimation results of the GARCH(1,1) model. Again our estimation results, using the British Pound series, reveal the

Table 1.4: Estimated GARCH Models: S&P 500

Bera & Higgins			Kim		
y_t=.052	+.066y_{t-1}+	u_t	y_t=.041	+.010y_{t-1}+	u_t
(.025)	(.031)		(.023)	(.030)	
h_t =.011	+.013u_{t-1}^2+	.968h_{t-1}	h_t=.013	+.012u_{t-1}^2+	.972h_{t-1}
(.006)	(.005)	(.013)	(.004)	(.004)	(.006)
$l(\hat{\theta})$=	-1367.67		$l(\hat{\theta})$=	-1511.091	

Table 1.5: Estimated GARCH Models: British Pound

Bera & Higgins			Kim		
y_t=-.024	+u_t		y_t=.032	+u_t	
(.014)			(.018)		
h_t =.010	+.059u_{t-1}^2+	.897h_{t-1}	h_t=.017	+.065u_{t-1}^2+	.902h_{t-1}
(.004)	(.002)	(.017)	(.009)	(.022)	(.036)
$l(\hat{\theta})$=	-785.72		$l(\hat{\theta})$=	-1231.208	

Table 1.6: Estimated GARCH Models: IP

Bera & Higgins			Kim		
y_t=2.68	+.279y_{t-1}	+.114y_{t-2}+u_t	y_t=2.135	+.270y_{t-1}	+.122y_{t-2}+u_t
(.303)	(.033)	(.013)	(.766)	(.100)	(.054)
h_t =60.4	+.235u_{t-1}^2	+.101h_{t-1}	h_t=47.114	+.233u_{t-1}^2	+.040h_{t-1}
(5.42)	(.034)	(.021)	(35.128)	(.116)	(.524)
$l(\hat{\theta})$=	-1301.327		$l(\hat{\theta})$=	-1247.462	

Table 1.7: Estimated Bilinear Models: S&P 500

Bera & Higgins		Kim	
y_t=.017	$+.102y_{t-1} + u_t$	y_t=-.0004	$+.045y_{t-1} + u_t$
(.030)	(.017)	(.029)	(.031)
u_t=	$.053u_{t-1}\varepsilon_{t-1}+\varepsilon_t$	u_t	$=.047u_{t-1}\varepsilon_{t-1}$
	(.017)		(.011)
$l(\hat{\theta})$=	-1368.884	$l(\hat{\theta})$=	-1517.299
$\hat{\sigma}_\varepsilon^2$=	.651	$\hat{\sigma}_\varepsilon^2$=	.844

difference in sign; -0.024 of Bera & Higgins vs. 0.032 of our estimation in Table 1.5. Table 1.6 shows that the two estimations results are very close and the GARCH effects in the IP series are rather small in both estimations (0.101 from Bera and Higgins vs. 0.040 from out estimation) compared to the two other GARCH effects in S & P and £/$ data sets.

Table 1.7 through 1.9 present the estimation results of the bilinear model and there are some significant differences between Bera & Higgins and our estimation results. First, the bilinear effects in the British Pound series are different in sign; (0.039, -0.083) from Bera & Higgins vs. (-0.016, 0.025) from our estimation. Second, the estimated variances of the bilinear model($\hat{\sigma}_\varepsilon$) are also different between Bera & Higgins and our results.

Table 1.10 and 1.11 present the modified Cox test results. Table 1.10 reports the test results when the GARCH model is the null hypothesis and Table 1.11 reports the test results when the bilinear model is the null hypothesis. Note that

Table 1.8: Estimated Bilinear Models: British Pound

Bera & Higgins		Kim	
y_t=.024	$+u_t$	y_t=.034	$+u_t$
(.020)		(.022)	
u_t=.039$u_{t-1}\varepsilon_{t-1}$	-.083$u_{t-2}\varepsilon_{t-1} + \varepsilon_t$	u_t=-.016$u_{t-1}\varepsilon_{t-1}$	+.025$u_{t-2}\varepsilon_{t-1} + \varepsilon_t$
(.021)	(.024)	(.031)	(.040)
$l(\hat{\theta})=$	-819.62	$l(\hat{\theta})=$	-1271.357
$\hat{\sigma}_\varepsilon^2=$	.226	$\hat{\sigma}_\varepsilon^2=$	.478

Table 1.9: Estimated Bilinear Models: IP

Bera & Higgins			Kim		
y_t=2.34+	.321y_{t-1}+	.125$y_{t-2} + u_t$	y_t=2.057+	.307y_{t-1}+	.133$y_{t-2} + u_t$
(.634)	(.053)	(.030)	(.566)	(.069)	(.055)
$u_t=$	-.006$u_{t-1}\varepsilon_{t-1}$	$+\varepsilon_t$	$u_t=$	-.008$u_{t-1}\varepsilon_{t-1}$	$+\varepsilon_t$
	(.003)			(.005)	
$l(\hat{\theta})=$	-1311.15		$l(\hat{\theta})=$	-1255.231	
$\hat{\sigma}_\varepsilon^2=$	90.69		$\hat{\sigma}_\varepsilon^2=$	63.651	

Table 1.10: Test results: H_0:GARCH vs. H_1: Bilinear

	Bera & Higgins	modified Cox test
S&P 500	.023	.322
British Pound	.196	-.033
Industrial Production	.533	.021

the absolute values of our test results are bigger than those of Bera & Higgins when the bilinear model is the null hypothesis in Table 1.11. When the GARCH model is the null hypothesis, our test results are close to zero for all three series, so we cannot reject the null hypothesis in those three data sets at any significance levels. In Table 1.11, Bera & Higgins reject the British Pound series as the null at 1 % of significance level and reject the IP series at 10 % of significance level when the null hypothesis is the bilinear model, but all three series are rejected in our test results, which produces much greater test values in absolute value than those from Bera and Higgins (1997). In Table 1.9, the estimated bilinear effect is -0.008 and the standard deviation is 0.005 in our estimation results, it is marginally significant and indicates the bilinear effect is very trivial for the IP series. For the IP series the test value is -23.724, which is almost 15 times bigger than that of Bera and Higgins and rejects the bilinear model as the null hypothesis at any significance levels. It is shown that the error equation does not follow the bilinear model in the IP series and this is consistent with the test result in Table 1.11 for the IP series.

Table 1.11: Test results: H_0:Bilinear *vs.* H_1: GARCH

	Bera & Higgins	modified Cox test
S&P 500	-.910	-6.775
British Pound	-2.797	-8.300
Industrial Production	-1.643	-23.724

1.4 Simulation Experiments

In this section we perform some simulation experiments to investigate the potential applicability of the modified Cox test.

We consider a linear regression model with two different nonlinear error equations as competing models. We specify an AR(1)-GARCH(1,1) model as the null hypothesis and a AR(1)-first order bilinear model as the alternative hypothesis. Thus, the nonnested model specifications are

$$M_1 : y_{1,t} = \alpha_0 + \alpha_1 y_{1,t-1} + u_t, \tag{1.88}$$

$$u_t \mid I_{t-1} \sim N(0, h_t), \tag{1.89}$$

$$h_t = \kappa + \gamma u_{t-1}^2 + \delta h_{t-1}, \tag{1.90}$$

$$\textit{and} \quad u_t = \sqrt{h_t} v_t, \qquad v_t \sim N(0,1) \tag{1.91}$$

$$M_2 : y_{2,t} = \beta_0 + \beta_1 y_{2,t-1} + \varepsilon_t, \tag{1.92}$$

$$\varepsilon_t = b_{11} \varepsilon_{t-1} \xi_{t-1} + \xi_t, \tag{1.93}$$

$$and \quad \xi_t \sim N(0,1)$$

We generate the artificial data in the following way. First,we generate the normally distributed random variables from RNDN GAUSS program to calculate the AR(1)-GARCH(1,1) model, $y_{1,t}$. Then again we generate the normally distributed random variables from RNDN GAUSS program for the AR(1)-first order bilinear model, $y_{2,t}$. The pseudo-true population parameters for M_1 are given as $y_{1,t} = 0.15 + 0.85y_{1,t-1} + u_t$ with a strong GARCH effect; $h_t = 0.1 + 0.2u_{t-1}^2 + 0.75h_{t-1}$. For M_2, the pseudo-true population parameters are given as $y_{2,t} = 0.19 + 0.8y_{2,t-1} + \varepsilon_t$ where $\varepsilon_t = 0.385\varepsilon_{t-1}\xi_{t-1} + \xi_t$. The parameter values chosen for both models correspond to the empirical estimates of the time series. Next we combine these two data sets with weight λ to generate a new data set $y_t = \lambda y_{1,t} + (1-\lambda)y_{2,t}$. Using this new generated data, we perform the testing experiments by setting different values of λ; λ=0, and 1. If λ=1, then $y_t = y_{1,t}$, so M_1 becomes the correctly specified one, while M_2 is correctly specified if λ=0. The QMLEs of these two specifications are calculated based on BHHH algorithm and the simulation results are calculated from 200 replications and with a sample size of 1000, 2000, 3000, and 5000 and 250, 500, and 750 for the small sample size properties. T1 is the modified Cox test when M_1 is correctly specified and T2 is the modified Cox test when M_2 is correctly specified. When the null is true, the test value(T) should be approximately zero.

Table 1.12: Simulation results when GARCH(1,1) is true

sample size	N=1000		N=2000		N=3000		N=5000	
	T1	T2	T1	T2	T1	T2	T1	T2
mean	0.056	-19.498	0.129	-34.577	-0.027	-44.458	0.054	-59.661
s.d	0.846	7.045	0.986	3.986	0.945	2.353	1.019	0.678
skew	-0.062	1.770	-0.03	5.692	-0.369	6.129	0.449	-0.276
kurt	2.866	4.423	2.368	39.269	3.782	44.359	3.630	3.138
R.F.(α=.05)	0.020	0.960	0.040	0.995	0.045	1.000	0.055	1.000
toohigh	0.005	0.000	0.030	0.000	0.010	0.000	0.040	0.000
toolow	0.015	0.960	0.010	0.995	0.035	1.000	0.015	1.000

two-tailed test with $\alpha = 0.05$ and $\lambda = 1$

In Table 1.12, we report the simulation results when the null is the GARCH(1,1) model with $\lambda = 1$. The four moments of the unconditional probability distribution of the simulated test are very close to normal for all four sample sizes. The actual size is very close to the nominal size for all sample sizes except for N=1000, in which the actual size is little bit understated.

Table 1.13 reports the simulation results when the null is the bilinear model for N=1000, 2000, 3000, and 5000. The distribution of the simulation results appear to be very close to the standard normal distribution for all sample sizes. And the actual size is very close to the nominal size.

Figure 1.1 and 1.2 present the empirical density functions (edfs) of the modified Cox test against the edf of N(0,1) for N= 1000, 2000, 3000, and 5000 and 200 replications. Figure 1.1 shows that the edfs of the simulation results of T1 appear

Table 1.13: Simulation results when Bilinear(1,1) is true

sample size	N=1000		N=2000		N=3000		N=5000	
	T1	T2	T1	T2	T1	T2	T1	T2
mean	-9.236	-0.063	-13.045	-0.145	-15.883	-0.039	-20.811	-0.136
s.d	3.125	1.029	4.050	0.923	4.631	0.969	5.394	0.905
skew	2.002	0.022	2.203	-0.118	2.369	0.022	2.548	-0.241
kurt	5.949	2.794	6.851	2.439	8.069	2.745	8.720	2.625
R.F.($\alpha = .05$)	0.945	0.050	0.945	0.045	0.995	0.050	0.980	0.030
toohigh	0.000	0.020	0.000	0.005	0.000	0.025	0.000	0.000
toolow	0.945	0.030	0.945	0.040	0.995	0.025	0.980	0.030

two-tailed test with $\alpha = 0.05$ and $\lambda = 0$

to be normal for all sample sizes. Figure 1.2 shows that the edfs of the simulation results of T2 appear to be approximately normal.

Figure1.1 The edfs of the modified Cox test vs. the cdf of N(0,1)

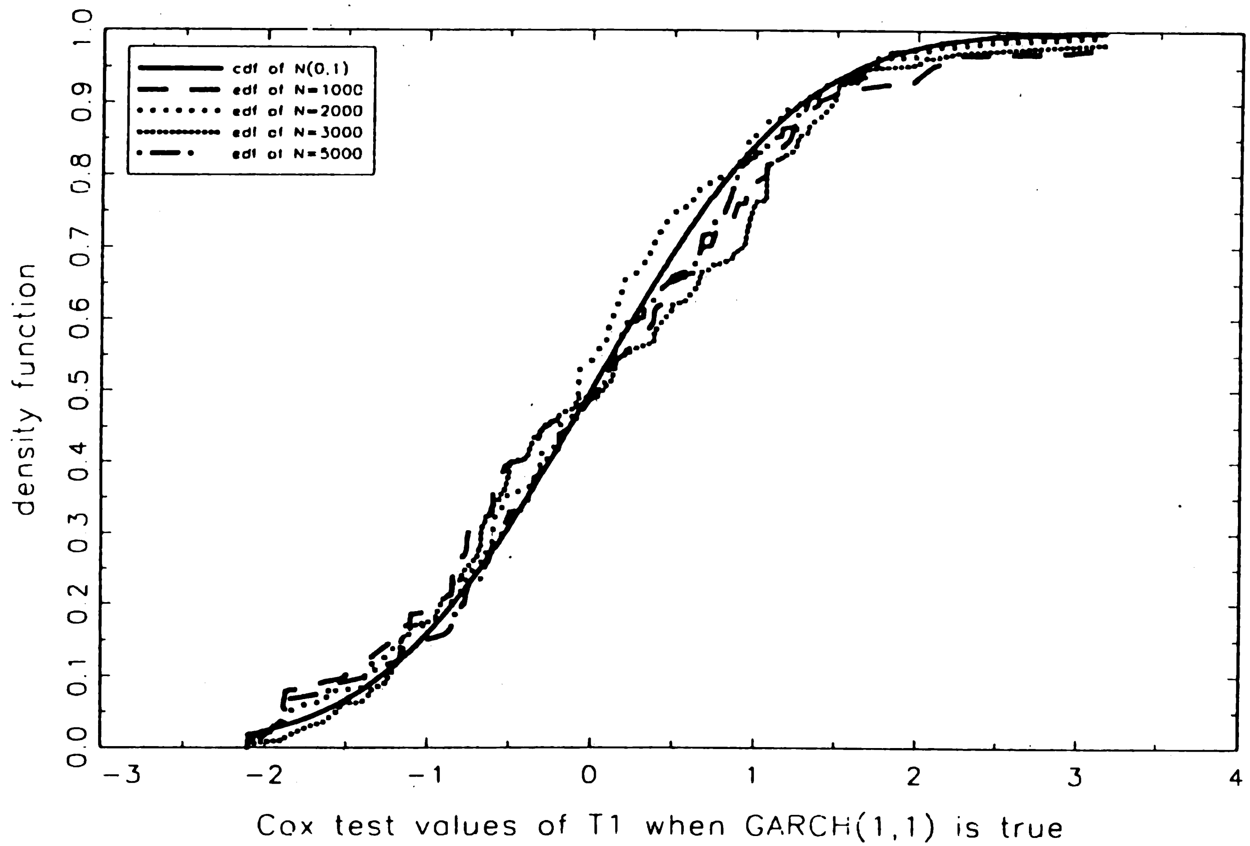


Figure1.2 The edfs of the modified Cox test vs the cdf of N(0,1)

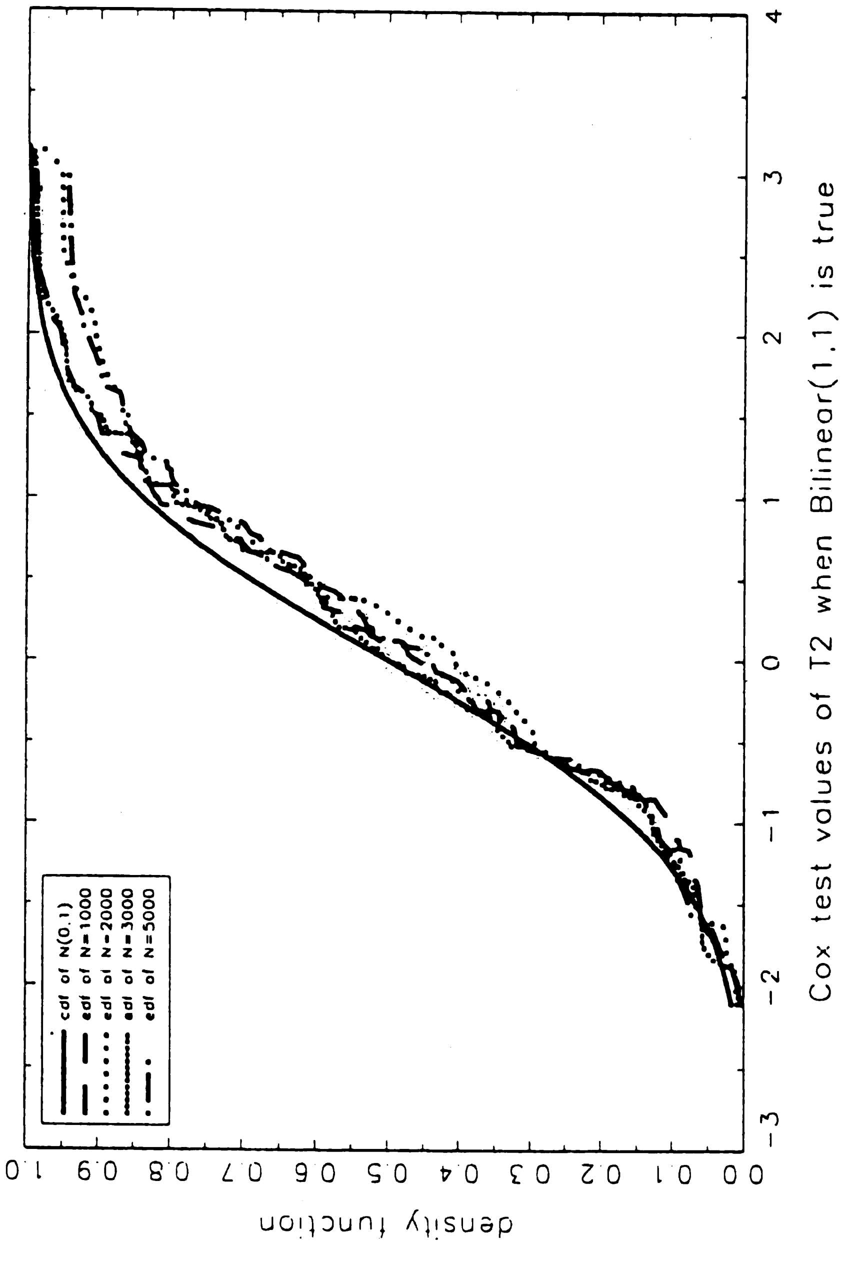


Table 1.14: Simulation results when GARCH(1,1) is true

sample size	N=250		N=500		N=750	
	T1	T2	T1	T2	T1	T2
mean	-0.277	-5.507	0.024	-11.027	-0.007	-15.509
s.d	0.760	3.466	0.875	4.973	0.936	6.106
skew	-0.446	0.411	-0.223	1.107	-0.177	1.519
kurt	3.015	1.485	2.871	2.667	3.104	3.736
R.F.(α=.05)	0.030	0.725	0.020	0.880	0.040	0.935
toohigh	0.000	0.000	0.005	0.000	0.020	0.000
toolow	0.030	0.725	0.015	0.880	0.020	0.935

two-tailed test with $\alpha = 0.05$ and $\lambda = 1$

Table 1.14 reports the simulation results with small sizes for N = 250, 500, and 750. The mean and standard deviation for N = 250 slightly deviate from the standard normal N(0,1) but close to normal for other sample sizes. The simulation results undersize for all three sample sizes and the rejection frequency of T2 is lower than 0.95 for N = 250 and 500.

In Table 1.15, the means are little bit greater than zero in absolute value for all three sample sizes but this deviation is getting smaller as the sample size increases. The actual size and the rejection frequency are approximately equivalent to the nominal levels.

Table 1.15: Simulation results when Bilinear(1,1) is true

sample size	N=250		N=500		N=750	
	T1	T2	T1	T2	T1	T2
mean	-4.952	-0.417	-6.806	-0.348	-8.274	-0.256
s.d	1.493	1.047	2.017	1.028	2.530	0.968
skew	2.406	-0.015	2.198	-0.325	2.076	-0.004
kurt	9.508	2.730	7.411	2.573	6.253	2.560
R.F(α=.05)	0.935	0.070	0.940	0.070	0.945	0.055
toohigh	0.005	0.010	0.000	0.000	0.000	0.005
toolow	0.930	0.060	0.940	0.070	0.945	0.050

two-tailed test with $\alpha = 0.05$ and $\lambda = 0$

1.5 Conclusions

A new approach based upon the conditional mean and the conditional variance specifications has been proposed in order to solve the computational difficulties of the Cox test. This modified Cox test has some attractive features. The major attraction of the modified Cox test is its computational conveniency because it does not require computing the pseudo-true values. As this proposed test is based upon the specification of the first two conditional moments, we can also test other distributional features unlike the DM test is for the conditional mean property only. Furthermore, it can be easily extended to the more complicated nonlinear models. Monte Carlo experiments indicate that this proposed test seems to perform well for all different sample sizes. The actual size from this proposed test is almost always close to the nominal size but the actual size is slightly different from the nominal size for N = 250, and 500. Further study needs to be done to examine the applicability to the finite-sample properties.

Chapter 2

A Robust Version of the Modified Cox Test

2.1 Introduction

In the previous chapter we proposed a modified version of Cox test under specification of the first two conditional moments. We examined its applicability with three different data sets: S&P 500 stock index, the £/ $ exchange rate, and U.S monthly IP data sets, and we also did some Monte Carlo simulation experiments. Both, empirical and simulated, test results are quite convincing the applicability of the modified Cox test and the actual size from the simulation results is very close to the nominal size regardless of sample size. But these empirical test results and simulation performances are derived under normality assumption. In this chapter we relax this normality assumption and extend our model in the univariate case to the robust version of the modified Cox test under nonnormality. In section 2, we reexamine our modified Cox test under nonnormality and derive the robust and nonrobust versions of the modified Cox test. Section 3 summarizes some Monte Carlo simulation experiments under nonnormality. In section 4, we compare the test results in the previous chapter assuming normality to the test results from the robust modified Cox test under nonnormality. Then we follow with a summary and conclusion in section 5.

2.2 A Robust Modified Cox Test

Assume there are two competing nonnested parametric models under the conditional mean and variance specifications.

$$M_1 : y_t = m_t(\theta_0) + u_t \tag{2.1}$$

$$E(y_t \mid I_{t-1}) = m_t(\theta_0) \tag{2.2}$$

$$V(y_t \mid I_{t-1}) = h_t(\theta_0) \tag{2.3}$$

and

$$M_2 : y_t = \mu_t(\delta_0) + \varepsilon_t \tag{2.4}$$

$$E(y_t \mid I_{t-1}) = \mu_t(\delta_0) \tag{2.5}$$

$$V(y_t \mid I_{t-1}) = \eta_t(\delta_0) \tag{2.6}$$

Following the previous chapter when M_1 is correctly specified, the modified Cox test is

$$\begin{aligned} \hat{T}_{M_1} &= \frac{1}{T}\sum_{t=1}^{T}\Bigg[(y_t - m_t(\hat{\theta}))\left\{\frac{m_t(\hat{\theta}) - \mu_t(\hat{\delta})}{\eta_t(\hat{\delta})}\right\} \\ &\quad + \frac{u_t(\hat{\theta})^2 - h_t(\hat{\theta})}{2}\left(\frac{1}{\eta_t(\hat{\delta})} - \frac{1}{h_t(\hat{\theta})}\right)\Bigg] \end{aligned} \tag{2.7}$$

And the asymptotic distributions of the modified Cox test are as follows[1]

$$\begin{aligned} \sqrt{T}\hat{T}_{M_1} &= \frac{1}{T}\sum_{t=1}^{T}\Bigg[(y_t - m_t(\hat{\theta}))\left\{\frac{m_t(\hat{\theta}) - \mu_t(\hat{\delta})}{\eta_t(\hat{\delta})}\right\} \\ &\quad + \frac{u_t(\hat{\theta})^2 - h_t(\hat{\theta})}{2}\left(\frac{1}{\eta_t(\hat{\delta})} - \frac{1}{h_t(\hat{\theta})}\right) \\ &\quad - (p\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\psi_t(\hat{\theta},\delta))(p\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}A_t(\hat{\theta}))^{-1} \\ &\quad \times\nabla_\theta \log f_t(\hat{\theta})\Big] + o_p(1) \end{aligned} \tag{2.8}$$

$$\begin{aligned} Define \qquad q_t^* &\equiv D_t - (p\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\psi_t(\theta_0,\delta^*))(p\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}A_t(\theta_0))^{-1} \\ &\quad \times\nabla_\theta \log f_t(\theta_0) \end{aligned}$$

$$Then, \qquad E(q_t^* \mid I_{t-1}) = 0, \tag{2.9}$$

$$\begin{aligned} E(q_t^{*2} \mid I_{t-1}) &= E\left[\left(D_t - \left(E[\frac{1}{T}\sum_{t=1}^{T}\psi_t(\theta_0,\delta^*)]\right)\left(E[\frac{1}{T}\sum_{t=1}^{T}A_t(\theta_0)]\right)^{-1}\right.\right. \\ &\quad \left.\times\nabla_\theta \log f_t(\theta_0))^2 \mid I_{t-1}\right] \end{aligned} \tag{2.10}$$

[1]We follow these from the previous chapter. See section 1.2.2 for more details.

$$= \quad V(q_t^* \mid I_{t-1}) \tag{2.11}$$

Under conditional normality when M_1 is correctly specified,

$$E(u_t(u_t^2 - h_t) \mid I_{t-1}) \quad = \quad 0, \tag{2.12}$$

$$and \qquad E(u_t^4 \mid I_{t-1}) \quad = \quad 2h_t(\theta_0)^2, \tag{2.13}$$

$$\begin{aligned} So, \qquad E(q_t^{*2} \mid I_{t-1}) \quad = \quad & D_{t1}^2 h_t(\theta_0) + \frac{h_t(\theta_0)^2}{4} D_{t2}^2 \\ & +(\frac{1}{T}\sum_{t=1}^{T} \psi_t(\theta_0, \delta^*))(\frac{1}{T}\sum_{t=1}^{T} A_t(\theta_0))^{-1}(\frac{1}{T}\sum_{t=1}^{T} \psi_t(\theta_0, \delta^*))' \\ & -2(\frac{1}{T}\sum_{t=1}^{T} \psi_t(\theta_0, \delta^*))(\frac{1}{T}\sum_{t=1}^{T} A_t(\theta_0))^{-1} \\ & \times(D_{t1}\nabla_\theta m_t(\theta_0) + \frac{D_{t2}}{4h_t(\theta_0)}\nabla_\theta h_t(\theta_0)) \end{aligned} \tag{2.14}$$

$$\begin{aligned} Thus, \qquad \hat{V}_{T_1} \quad = \quad & \frac{1}{T}\sum_{t=1}^{T}\Big[\hat{D}_{t1}^2 h_t(\hat{\theta}) + \frac{h_t(\hat{\theta})^2}{4}\hat{D}_{t2}^2 \\ & +(\frac{1}{T}\sum_{t=1}^{T} \psi_t(\hat{\theta}, \hat{\delta}))(\frac{1}{T}\sum_{t=1}^{T} A_t(\hat{\theta}))^{-1}(\frac{1}{T}\sum_{t=1}^{T} \psi_t(\hat{\theta}, \hat{\delta}))' \\ & -2(\frac{1}{T}\sum_{t=1}^{T} \psi_t(\hat{\theta}, \hat{\delta}))(\frac{1}{T}\sum_{t=1}^{T} A_t(\hat{\theta}))^{-1} \\ & \times(\hat{D}_{t1}\nabla_\theta m_t(\hat{\theta}) + \frac{\hat{D}_{t2}}{4h_t(\hat{\theta})}\nabla_\theta h_t(\hat{\theta}))\Big] \end{aligned} \tag{2.15}$$

Under conditional normality, the modified Cox test, $\frac{\sqrt{T}\hat{T}_1}{\hat{V}_{T_1}}^{1/2} \sim N(0,1)$, but if the conditional normality does not hold, then the limiting distribution of $\frac{\sqrt{T}\hat{T}_1}{\hat{V}_{T_1}}^{1/2}$ is not standard normal in general. Under nonnormality the modified Cox test derived from the previous chapter is not valid and the actual size from this nonrobust modified Cox test can be different from the nominal size. The robust modified Cox test under nonnormality is extended from equation(2.10)and (2.11). Under nonnormality,

$$E(u_t(u_t^2 - h_t) \mid I_{t-1}) = E(u_t^3 - u_t h_t \mid I_{t-1})$$
$$= E(u_t^3 \mid I_{t-1}) \quad (2.16)$$
$$\textit{and} \quad E((u_t^2 - h_t)^2 \mid I_{t-1}) = E(u_t^4 - 2u_t^2 h_t + h_t^2 \mid I_{t-1})$$
$$= E(u_t^4 \mid I_{t-1}) - h_t^2 \quad (2.17)$$

Under nonnormality, $E(u_t^3 \mid I_{t-1})$ and $E(u_t^4 \mid I_{t-1})$ are generally unspecified but we can derive the conditional variance of the robust modified Cox test using the Law of Iterated Expectation (L.I.E):

$$E\left[\frac{D_{t1}D_{t2}}{2}E\left(u_t(\theta_0)^3 \mid I_{t-1}\right)\right] = E\left[\frac{D_{t1}D_{t2}}{2}u_t(\theta_0)^3\right] \quad (2.18)$$
$$\textit{and} \quad E\left[\frac{D_{t2}^2}{4}E\left(u_t(\theta_0)^4 \mid I_{t-1}\right)\right] = E\left[\frac{D_{t2}^2}{4}u_t(\theta_0)^4\right] \quad (2.19)$$

So the conditional variance of the robust modified Cox test under nonnormality is

$$\hat{V}_{T_1}^R = \frac{1}{T}\sum_{t=1}^{T}\left[\hat{D}_{t1}^2 h_t(\hat{\theta}) + \frac{\hat{D}_{t1}\hat{D}_{t2}}{2}u_t(\hat{\theta})^3 + \frac{\hat{D}_{t2}^2}{4}(u_t(\hat{\theta})^4 - h_t(\hat{\theta})^2)\right.$$
$$+(\frac{1}{T}\sum_{t=1}^{T}\psi_t(\hat{\theta},\hat{\delta}))(\frac{1}{T}\sum_{t=1}^{T}A_t(\hat{\theta}))^{-1}(\frac{1}{T}\sum_{t=1}^{T}\psi_t(\hat{\theta},\hat{\delta}))'$$

$$-2(\frac{1}{T}\sum_{t=1}^{T}\psi_t(\hat{\theta},\hat{\delta}))(\frac{1}{T}\sum_{t=1}^{T}A_t(\hat{\theta}))^{-1}$$
$$\times(\hat{D}_{t1}\nabla_\theta m_t(\hat{\theta})+\frac{\hat{D}_{t2}}{4h_t(\hat{\theta})}(1-\frac{1}{h_t(\hat{\theta})})\nabla_\theta h_t(\hat{\theta}))\Big] \tag{2.20}$$

Now we apply these properties as follows:

- Procedure 2.1

1. Obtain $\hat{\theta}$ and $\hat{\delta}$, the QML estimators of θ_0 and δ_0, save residuals, $u_t(\hat{\theta})$, and the conditional variance, $h_t(\hat{\theta})$, from the quasi-log likelihood function $\log f(y_t \mid I_{t-1};\hat{\theta})$ and $e_t(\hat{\delta})$ and $\eta_t(\hat{\delta})$ from the quasi-log likelihood function $\log g(y_t, \mid I_{t-1};\hat{\delta})$.

2. Compute $\hat{D}_{t1}$, $\hat{D}_{t2}$, $\frac{1}{T}\sum_{t=1}^{T}\psi_t(\hat{\theta},\hat{\delta})$, $(\frac{1}{T}\sum_{t=1}^{T}A_t(\hat{\theta}))^{-1}$, $\nabla_\theta \log f_t(\hat{\theta})$, $u_t(\hat{\theta})^3$, $u_t(\hat{\theta})^4$, and $h_t(\hat{\theta})^2$. Define $D_{t1} \equiv \frac{m_t(\hat{\theta})-\mu_t(\hat{\delta})}{\eta_t(\hat{\delta})}$, $D_{t2} \equiv \frac{1}{\eta_t(\hat{\delta})}-\frac{1}{h_t(\hat{\theta})}$, $A_t(\hat{\theta}) \equiv \frac{\partial^2 \log f_t(\hat{\theta})}{\partial\hat{\theta}\partial\hat{\theta}'}$, and $\psi_t(\hat{\theta},\hat{\delta}) \equiv \nabla_\theta(\hat{u}_t\hat{D}_{t1}-\frac{\hat{u}_t^2-\hat{h}_t}{2}\hat{D}_{t2})$.

3. Compute $\sqrt{T}\hat{T}_{M_1} = T^{-1/2}\sum_{t=1}^{T}\Big[\hat{u}_t\hat{D}_{t1}+\frac{\hat{u}_t^2-\hat{h}_t}{2}\hat{D}_{t2}-(\frac{1}{T}\sum_{t=1}^{T}\psi_t(\hat{\theta},\hat{\delta}))$ $(\frac{1}{T}\sum_{t=1}^{T}A_t(\hat{\theta}))^{-1}\nabla_\theta \log f_t(\hat{\theta})\Big]$ and $\hat{V}_{T_1}^R = \frac{1}{T}\sum_{t=1}^{T}\Big[\hat{D}_{t1}^2 h_t(\hat{\theta})+\frac{\hat{D}_{t1}\hat{D}_{t2}}{2}u_t(\hat{\theta})^3+\frac{\hat{D}_{t2}^2}{4}(u_t(\hat{\theta})^4-h_t(\hat{\theta})^2)$ $+(\frac{1}{T}\sum_{t=1}^{T}\psi_t(\hat{\theta},\hat{\delta}))(\frac{1}{T}\sum_{t=1}^{T}A_t(\hat{\theta}))^{-1}(\frac{1}{T}\sum_{t=1}^{T}\psi_t(\hat{\theta},\hat{\delta}))'$ $-2(\frac{1}{T}\sum_{t=1}^{T}\psi_t(\hat{\theta},\hat{\delta}))(\frac{1}{T}\sum_{t=1}^{T}A_t(\hat{\theta}))^{-1}$ $\times(\hat{D}_{t1}\nabla_\theta m_t(\hat{\theta})+\frac{\hat{D}_{t2}}{4h_t(\hat{\theta})}\nabla_\theta h_t(\hat{\theta}))\Big]$ and use the robust modified Cox test statistic, $\frac{\sqrt{T}\hat{T}_{T_1}}{\hat{V}_{T_1}^{R1/2}}$ as asymptotically unit normal under the null hypothesis.

This robust modified Cox test has some appealing characteristics. First, as Bollerslev and Wooldridge (1992) showed in their robust version of LM test, this approach is also valid under normality and can be applied to the case where normality assumption does hold. Second, this procedure requires only the first derivatives of the conditional mean and variance functions, it is relatively easy to compute. Finally, even though this robust inference procedure requires the conditional third and fourth moments, this is not a restrictive condition and we can calculate the third and fourth moments for the robust modified Cox test using L.I.E.

2.3 Monte Carlo Experiments

To investigate the applicability of the robust modified Cox test, we perform some simulation experiments for different sample sizes. Following the previous chapter, we consider a linear regression model with two different nonlinear error equations: we specify the AR(1)-GARCH(1,1) model as the null hypothesis and the AR(1)-the first order bilinear model as the alternative hypothesis. Thus, these two nonnested model specifications are

$$M_1 : y_t = \alpha_0 + \alpha_1 y_{t-1} + u_t, \tag{2.21}$$

$$u_t \mid I_{t-1} \sim i.i.d(0, h_t), \tag{2.22}$$

$$h_t = \kappa + \gamma u_{t-1}^2 + \delta h_{t-1}, \quad (2.23)$$

$$and \quad u_t = \sqrt{h_t}\nu_t \quad (2.24)$$

$$M_2 : y_t = \beta_0 + \beta_1 y_{t-1} + \varepsilon_t, \quad (2.25)$$

$$and \quad \varepsilon_t \mid I_{t-1} = b_{11}\varepsilon_{t-1}\xi_{t-1} + \xi_t \quad (2.26)$$

As seen often in time series analysis, high frequency financial time series are of leptokurtosis and the unconditional distribution of many finantial time series typically shows fatter tails than a normal distribution. But as shown in Engle (1982) and Bollerslev (1986), unconditional error distribution could be leptokurtic even though the conditional error distribution is normal. Bollerslev (1987) proposed that if the error distribution is not normal, for example the conditionally t-distributed errors, then it permits a conditional leptokurtic distribution and it also accounts for the unconditional kurtosis. To investigate the valid inference from the robust modified Cox test under nonnormality, we generate the error terms from two different nonnormal distributions. First, we have considered that ν_t (ξ_t as well) is conditionally distributed as a Student's t-distribution with 5 and 10 degrees of freedom. The mean and variance of t-distribution are 0 and $\frac{v}{v-2}, if \quad v \geq 3$, respectively, where v is degree of freedom, so the variance from t_v distributed random variables is bigger than the variance from the random variables generated by the standard normal distribution, if v is relatively small number. But the t-distributed random variables still contain symmetricity on the distribution. Next, in order to examine the effect of asymmetric error distribution, we generate the error terms from two i.i.d χ_1^2 distribution, i.e. ν_t (ξ_t as well) is formed from $\frac{x_{t1}-x_{t2}}{2}$ where x_{t1}, x_{t2} are i.i.d χ_1^2 variates, respectively. Thus, the distribution of ν_t (ξ_t as well) is i.i.d(0,1) pertaining to asymmetric property. The t_v distributed random variables were formed as $(v-2)$ times a N(0,1) random variable divided by the square root of χ_v^2 variate. The normal variate was generated by the RNDN GAUSS program and χ_v^2 with v df by the RNDCHI GAUSS program. Beside the error generating procedures under nonnormality, we proceed in a similar way for the pseudo-true population parameters for M_1 and M_2[2]and for the data generating procedure in the previous chapter (see 1.4 Simulation Experiments). The QMLEs for both models are found through BHHH algorithm and the simulation results are based upon 200 replications and a sample size of 500, 1000, 2000, and 3000.

Table 2.1 and 2.2 report the simulation results under nonnormality: the error terms were generated from χ_1^2 distribution. In Table 2.1, the four moments of the unconditional probability distribution of T1 are approximately close to normal but the means are a little bit larger than zero in absolute value for N = 500, and 1000. The actual size is very close to the nominal size for all sample sizes.

[2]We change the bilinear parameter value to $b_{11} = 0.085$ for a computational conveniency.

Table 2.1: Robust Cox test results when GARCH(1,1) is true

sample size	N=500		N=1000		N=2000		N=3000	
	T1	T2	T1	T2	T1	T2	T1	T2
mean	-0.373	-14.269	-0.207	-23.380	-0.075	-35.692	0.105	-46.313
s.d	0.966	3.341	1.069	4.390	1.008	5.043	1.050	2.436
skew	-0.272	2.023	-0.031	3.080	0.003	4.011	0.170	10.404
kurt	4.868	6.470	4.361	11.810	3.148	19.776	2.599	127.053
R.F.(α=.05)	0.050	1.000	0.060	1.000	0.060	1.000	0.055	1.000
toohigh	0.010	0.000	0.015	0.000	0.020	0.000	0.045	0.000
toolow	0.040	1.000	0.045	1.000	0.040	1.000	0.010	1.000
max[3]	0.148		0.076		0.015		0.060	
min[4]	-0.096		-0.075		-0.056		-0.047	
mean[5]	0.004		0.001		0.001		-0.000	

Data are generated from $\chi^2_{(1)}$ distribution and R=200

Table 2.2 reports the simulation results from the nonrobust modified Cox test under nonnormality: the error terms were formed from the χ^2_1 distribution. Nonrobustness indicates that we apply the modified Cox test derived from the normality assumption to the situation where this normality assumption does not hold any more. The means and standard deviations are overstated in absolute value and the actual size is more than twice as large as the nominal size varying from 0.095 to

[3]The maximum value of correlations for a sample size N = 500, 1000, 2000, and 3000, and with 200 replications

[4]The minimum value of correlations for a sample size N = 500, 1000, 2000, and 3000, and with 200 replications

[5]The mean of correlations

Table 2.2: Nonrobust Cox test results when GARCH(1,1) is true

sample size	N=500		N=1000		N=2000		N=3000	
	T1	T2	T1	T2	T1	T2	T1	T2
mean	-0.495	-13.930	-0.439	-22.590	0.059	-36.517	0.008	-45.741
s.d	1.211	3.821	1.262	5.108	1.383	3.139	1.240	3.822
skew	0.165	1.669	0.029	2.337	0.168	6.017	0.106	6.561
kurt	3.582	4.636	2.875	7.194	2.879	48.729	3.354	54.734
R.F.(α=.05)	0.120	0.995	0.160	1.000	0.135	1.000	0.095	1.000
toohigh	0.035	0.000	0.035	0.000	0.085	0.000	0.045	0.000
toolow	0.085	0.995	0.125	1.000	0.050	1.000	0.050	1.000
max	0.117		0.083		0.072		0.045	
min	-0.102		-0.069		-0.052		-0.055	
mean	-0.001		-0.004		-0.003		0.003	

Data are generated from $\chi^2_{(1)}$ distribution and R=200

0.160. Under nonnormality, the robust modified Cox test performs far much better than the nonrobust modified Cox test.

Figure 2.1 and 2.2 show the empirical density functions (the edfs) of the robust and the nonrobust modified Cox tests against the cdf of N(0,1). In Figure 2.1, the empirical density functions appear to be equivalent to the cdf of N(0,1) for all the sample sizes except for N = 500. In Figure 2.2, the edfs are distorted and are far from normal for all four sample sizes.

Table 2.3 reports the simulation results of the robust modified Cox test under χ^2_1 distribution when the bilinear model is correctly specified. The actual size is a little bit overstated than the nominal size but approaches the nominal size as sample

Table 2.3: Robust Cox test results when Bilinear(1,1) is true

sample size	N=500		N=1000		N=2000		N=3000	
	T1	T2	T1	T2	T1	T2	T1	T2
mean	-3.736	-0.404	-4.898	-0.090	-6.865	-0.106	-8.278	0.152
s.d	1.930	1.326	2.711	1.108	3.563	1.182	4.448	1.060
skew	0.215	-0.811	0.405	-0.294	0.556	-0.328	0.589	-0.469
kurt	2.689	4.066	2.389	2.479	2.268	2.549	2.182	3.276
R.F.(α=.05)	0.840	0.120	0.820	0.090	0.865	0.080	0.845	0.060
toohigh	0.000	0.005	0.000	0.020	0.000	0.010	0.000	0.030
toolow	0.840	0.115	0.820	0.070	0.865	0.070	0.845	0.030
max	0.151		0.101		0.101		0.088	
min	-0.126		-0.077		-0.051		-0.044	
mean	-0.002		-0.000		-0.002		0.001	

Data are generated from $\chi^2_{(1)}$ distribution and R=200

size becomes larger. The four moments of the unconditional probability distribution of T2 are close to normal but, again, the mean and standard deviation are slightly different from N(0,1) for N = 500.

Table 2.4 reports the simulation results of the nonrobust modified Cox test under χ^2_1 distribution. As expected, the four moments of the unconditional probability distribution of T2 are far from normal and the actual size is very different from the nominal size and overstated varying from 0.150 to 0.260.

Figure 2.3 and 2.4 show the edfs of the robust and nonrobust modified Cox test of T2. In Figure 2.3, the edfs are very close to the cdf of N(0,1) for all four sample sizes but the edfs are severely distorted and far from the cdf of N(0,1) in

Table 2.4: Nonrobust Cox test results when Bilinear(1,1) is true

sample size	N=500		N=1000		N=2000		N=3000	
	T1	T2	T1	T2	T1	T2	T1	T2
mean	-3.736	0.172	-4.898	0.309	-6.865	0.358	-8.278	0.322
s.d	1.930	2.080	2.711	2.820	3.563	2.912	4.448	2.877
skew	0.215	1.138	0.405	1.610	0.556	0.450	0.589	0.337
kurt	2.689	6.646	2.389	9.842	2.268	4.106	2.182	3.376
R.F.(α=.05)	0.840	0.290	0.820	0.350	0.865	0.450	0.845	0.475
toohigh	0.000	0.150	0.000	0.220	0.000	0.260	0.000	0.255
toolow	0.840	0.140	0.820	0.130	0.865	0.190	0.845	0.220
max	0.135		0.074		0.079		0.049	
min	-0.107		-0.068		-0.054		-0.045	
mean	-0.002		0.001		0.001		0.000	

Data are generated from the $\chi^2_{(1)}$ distribution and R=200

Figure 2.4.

Table 2.5 and Table 2.6 report the simulation results for the robust and the nonrobust modified Cox test under t_5 distribution. The means and standard deviations are approximately normal but the actual size is understated than the nominal size for N = 500 and slightly overstated for other three sample sizes in Table 2.5.

Table 2.6 reports the simulation results of the nonrobust modified Cox test under t_5 distribution. The actual size and the unconditional four moments of the probability distribution of T1 are far from normal for all sample sizes.

Table 2.5: Robust Cox test results when GARCH(1,1) is true

sample size	N=500		N=1000		N=2000		N=3000	
	T1	T2	T1	T2	T1	T2	T1	T2
mean	-0.112	-13.739	0.101	-23.109	0.026	-35.247	0.147	-45.876
s.d	0.965	3.003	1.065	3.504	1.115	5.193	1.105	2.170
skew	0.693	1.657	0.277	3.435	-0.016	3.866	0.469	8.670
kurt	3.748	4.852	2.937	15.485	3.410	18.060	3.251	96.825
R.F.	0.025	1.000	0.070	1.000	0.075	1.000	0.075	1.000
toohigh	0.020	0.000	0.055	0.000	0.055	0.000	0.050	0.000
toolow	0.005	1.000	0.015	1.000	0.020	1.000	0.025	1.000

Data are generated from the t-distribution with 5 degrees of freedom and R=200.

Table 2.6: Nonrobust Cox test results when GARCH(1,1) is true

sample size	N=500		N=1000		N=2000		N=3000	
	T1	T2	T1	T2	T1	T2	T1	T2
mean	-0.754	-13.255	-0.407	-22.793	-0.102	-35.999	0.019	-45.765
s.d	1.912	3.338	2.309	4.134	2.228	3.470	2.232	2.439
skew	-1.327	1.352	-0.971	2.741	-0.531	5.113	-0.354	5.532
kurt	7.495	3.704	4.700	10.334	2.963	31.989	2.981	39.032
R.F.	0.250	1.000	0.325	1.000	0.370	1.000	0.385	1.000
toohigh	0.035	0.000	0.125	0.000	0.180	0.000	0.200	0.000
toolow	0.215	1.000	0.200	1.000	0.190	1.000	0.185	1.000

Data are generated from the t-distribution with 5 degrees of freedom and R=200.

Figure 2.5 and 2.6 show the edfs of the robust and the nonrobust modified Cox test. The edfs in Figure 2.5 are very close to the cdf of N(0,1) while the edfs in Figure 2.6 are far from the cdf of N(0.1).

Table 2.7 and 2.8 report the simulation results of the robust and the nonrobust modified Cox test under t_5 distribution when the null is the bilinear model. The simulation results show that the robust modified Cox test performs far better than the nonrobust modified Cox test generally under nonnormality except for the simulation results from the robust modified Cox test for N = 500. They are very similar to the results from the nonrobust modified Cox test for the same sample size. The edfs in Figure 2.7 are slightly deviated from the cdf of N(0,1). In Figure 2.8, the edfs are more distorted from the cdf of N(0,1). We suspect that this relatively poor performance may be mainly due to the parameter value chosen for the bilinear effect.[6]

Table 2.9 and 2.10 report the simulation results under t_{10} distribution. In Table 2.9, the simulation results are very close to normal and the actual size is also

[6] We could not compute the Hessian matrix for the GARCH(1,1) model when we used the same parameter value(0.385) for the bilinear effect in the previous chapter, so we chose a different parameter value(0.085) that managed to fit in the GARCH(1,1) model. But this parameter value chosen for the bilinear effect is rather small and dose not provide a strong bilinear effect. Thus, the error terms generated from this bilinear parameter value do not fit in well enough to perform the simulation experiments and the performance is particularly worse in the small sample size.

Table 2.7: Robust Cox test results when Bilinear(1,1) is true

sample size	N=500		N=1000		N=2000		N=3000	
	T1	T2	T1	T2	T1	T2	T1	T2
mean	-4.052	-0.569	-5.386	-0.375	-7.823	-0.151	-8.496	-0.209
s.d	2.208	1.529	2.828	1.193	4.206	0.932	4.693	0.967
skew	-0.429	-1.332	0.167	-0.678	0.275	-0.323	-0.011	-0.277
kurt	4.425	5.217	2.316	3.979	1.931	2.752	1.908	3.141
R.F.(α=.05)	0.880	0.145	0.865	0.080	0.900	0.040	0.905	0.050
toohigh	0.000	0.005	0.000	0.000	0.000	0.000	0.000	0.005
toolow	0.880	0.140	0.865	0.080	0.900	0.040	0.905	0.045

Data are generated from the t-distribution with 5 degrees of freedom and R=200

Table 2.8: Nonrobust Cox test results when Bilinear(1,1) is true

sample size	N=500		N=1000		N=2000		N=3000	
	T1	T2	T1	T2	T1	T2	T1	T2
mean	-4.052	-0.101	-5.386	-0.222	-7.823	-0.174	-8.496	-0.320
s.d	2.208	1.587	2.828	1.577	4.206	1.630	4.693	1.840
skew	-0.429	1.126	0.167	0.824	0.275	0.050	-0.011	-0.783
kurt	4.425	7.683	2.316	4.716	1.931	4.170	1.908	7.390
R.F.(α=.05)	0.880	0.145	0.865	0.175	0.900	0.210	0.905	0.265
toohigh	0.000	0.080	0.000	0.070	0.000	0.080	0.000	0.105
toolow	0.880	0.065	0.865	0.105	0.900	0.130	0.905	0.160

Data are generated from the t-distribution with 5 degrees of freedom and R=200

Table 2.9: Robust Cox test results when GARCH(1,1) is true

sample size	N=500		N=1000		N=2000		N=3000	
	T1	T2	T1	T2	T1	T2	T1	T2
mean	-0.083	-13.378	0.041	-22.550	0.118	-35.794	0.079	-45.379
s.d	0.882	2.643	0.967	2.523	1.050	1.394	1.021	0.709
skew	0.130	1.826	0.246	3.332	0.085	6.267	0.051	0.630
kurt	2.532	5.476	3.438	16.512	2.723	62.431	3.496	5.046
R.F.	0.015	1.000	0.050	1.000	0.055	1.000	0.060	1.000
toohigh	0.005	0.000	0.035	0.000	0.030	0.000	0.045	0.000
toolow	0.010	1.000	0.015	1.000	0.025	1.000	0.015	1.000

Data are generated from the t-distribution with 10 degrees of freedom and R=200.

very close to the nominal size for all four sample sizes. The simulation results of the nonrobust modified Cox test in Table 2.10 are far from normal. Figure 2.9 and 2.10 show evidence that robust modified Cox test performs better than the nonrobust modified Cox test under nonnormality and that the robust modified Cox test is also very accurate.

Table 2.11 and 2.12 report the simulation results of the robust and the nonrobust modified Cox tests under t_{10} distribution. As expected, the simulation results are very similar to those from Table 2.7 and 2.8.

Figure 2.11 and 2.12 show the edfs of the robust and the nonrobust modified Cox test. Again, the edfs in Figure 2.11 slightly deviate from the cdf of N(0,1) but they are relatively close to the cdf of N(0,1) compared to the edfs in Figure 2.12.

So far, we have illustrated the applicability of the robust modified Cox test

Table 2.10: Nonrobust Cox test results when GARCH(1,1) is true

sample size	N=500		N=1000		N=2000		N=3000	
	T1	T2	T1	T2	T1	T2	T1	T2
mean	-0.030	-22.592	-0.096	-22.347	0.120	-35.860	0.129	-45.420
s.d	1.309	2.622	1.315	2.969	1.443	1.179	1.523	0.755
skew	-0.158	3.182	0.226	2.914	-0.117	4.036	0.269	0.377
kurt	3.131	14.538	2.921	12.381	2.859	29.950	2.813	2.816
R.F.	0.130	1.000	0.110	1.000	0.160	1.000	0.210	1.000
toohigh	0.050	0.000	0.030	0.000	0.090	0.000	0.125	0.000
toolow	0.080	1.000	0.080	1.000	0.070	1.000	0.085	1.000

Data are generated from the t-distribution with 10 degrees of freedom and R=200.

Table 2.11: Robust Cox test results when Bilinear(1,1) is true

sample size	N=500		N=1000		N=2000		N=3000	
	T1	T2	T1	T2	T1	T2	T1	T2
mean	-3.885	-0.510	-6.280	-0.632	-8.182	-0.261	-10.227	-0.189
s.d	2.195	1.407	3.550	1.232	4.962	0.992	6.202	1.005
skew	-0.596	-1.532	0.060	-0.453	0.024	-0.293	-0.072	-0.204
kurt	2.951	6.808	2.097	3.329	1.865	2.914	1.700	2.499
R.F.(α=.05)	0.830	0.120	0.850	0.160	0.880	0.070	0.875	0.065
toohigh	0.000	0.000	0.000	0.010	0.000	0.010	0.000	0.010
toolow	0.830	0.120	0.850	0.150	0.880	0.060	0.875	0.055

Data are generated from the t-distribution with 10 degrees of freedom and R=200

Table 2.12: Nonrobust Cox test results when Bilinear(1,1) is true

sample size	N=500		N=1000		N=2000		N=3000	
	T1	T2	T1	T2	T1	T2	T1	T2
mean	-4.052	-0.101	-5.864	-0.285	-8.324	-0.153	-9.712	0.167
s.d	2.208	1.587	3.370	1.215	4.763	1.219	6.019	1.420
skew	-0.429	1.126	-0.075	0.619	0.043	0.343	-0.063	0.356
kurt	2.076	4.195	2.076	4.195	1.778	2.880	1.783	4.015
R.F.(α=.05)	0.870	0.120	0.870	0.120	0.915	0.120	0.885	0.125
toohigh	0.000	0.080	0.000	0.045	0.000	0.065	0.000	0.050
toolow	0.880	0.065	0.870	0.075	0.915	0.055	0.885	0.075

Data are generated from the t-distribution with 10 degrees of freedom and R=200

from the procedure 2.1 but in order to use this proposed test we have to calculate each term in the conditional variance in equation (2.15). This might be a little cumbersome. As an alternative way, we suggest $E[q_t^{*2} \mid I_{t-1}]$ from equation (2.10) as the conditional variance of the robust modified Cox test. Both the robust and nonrobust modified Cox test are originally derived from the equations (2.10) and (2.11). An attractive feature of this robust modified Cox test is that it does not require computing every term in equation (2.15). If the error terms do not follow normality, then the third and fourth moments are automatically calculated in equation (2.10). We did the simulation experiments with some selective cases. But these simulation results strongly suggest that this alternative robust modified Cox test would perform properly for other cases as well.

Table 2.13 reports the simulation results under nonnormality when we use

Table 2.13: Robust Cox test results when GARCH(1,1) is true

sample size	N=500		N=1000		N=2000		N=3000	
	T1	T2	T1	T2	T1	T2	T1	T2
mean	-0.355	-14.130	-0.192	-22.715	-0.059	-35.198	0.043	-45.845
s.d	0.899	3.597	0.927	4.937	0.951	6.516	0.9745	3.761
skew	0.313	1.840	0.285	2.222	-0.668	3.498	-0.043	5.766
kurt	3.848	5.427	2.791	6.731	5.060	14.637	2.834	37.900
R.F.(α=.05)	0.030	0.995	0.025	1.000	0.040	1.000	0.045	1.000
toohigh	0.010	0.000	0.010	0.000	0.010	0.000	0.020	0.000
toolow	0.020	0.995	0.015	1.000	0.030	1.000	0.025	1.000
max	0.096		0.061		0.047		0.060	
min	-0.096		-0.077		-0.051		-0.056	
mean	0.004		0.001		-0.000		-0.002	

Data are generated from $\chi^2_{(1)}$ distribution, R=200, and the conditional variance from equation (2.10) is used equation (2.10) as the conditional variance for the robust modified Cox test. These simulation results are very similar to those in Table 2.1. Note that the actual size in Table 2.13 is slightly understated than the nominal size while the actual size in Table 2.1 is slightly overstated than the nominal size. But these differences from both cases are trivial and very close to the nominal size for all sample sizes.

Figure 2.13 shows the edfs of the alternatively proposed robust modified Cox test. The edfs are also very close to those in Figure 2.1 and they approach the cdf of N(0,1) as sample size increases.

Table 2.14 reports the simulation results under χ_1^2 distribution when the bilinear model is correctly specified. The means and standard deviations of this simulated results appear to be approximately normal and the actual size is lower than that in Table 2.3 for N = 500, and 1000.

Figure 2.14 shows the edfs of this proposed test. The edfs are very close to the cdf of N(0,1) for all sample sizes as in Figure 2.3.

Table 2.15 reports the simulation results of the robust modified Cox test using the conditional variance from equation (2.10) under t_{10} distribution. The simulation statistics are outperformed compared to Table 2.7. The four moments of the probability distribution of T2 are very close to normal and the actual size is also very close to the nominal size for all sample sizes. Note that the actual size

Table 2.14: Robust Cox test results when Bilinear(1,1) is true

sample size	N=500		N=1000		N=2000		N=3000	
	T1	T2	T1	T2	T1	T2	T1	T2
mean	-3.736	-0.250	-4.898	-0.076	-6.865	0.091	-8.278	0.078
s.d	1.930	1.139	2.711	1.026	3.563	1.196	4.448	1.063
skew	0.215	-0.433	0.405	-0.380	0.556	-0.473	0.589	-0.074
kurt	2.689	2.651	2.390	2.900	2.268	3.068	2.182	2.944
R.F.(α=.05)	0.840	0.085	0.820	0.060	0.865	0.100	0.845	0.060
toohigh	0.000	0.005	0.000	0.010	0.000	0.040	0.000	0.025
toolow	0.840	0.080	0.820	0.050	0.865	0.060	0.845	0.035
max	0.094		0.099		0.071		0.066	
min	-0.122		-0.088		-0.063		-0.044	
mean	-0.001		-0.001		-0.001		0.000	

Data are generated from $\chi^2_{(1)}$ distribution, R=200, and the conditional variance from equation (2.10) is used

Table 2.15: Robust Cox test results when Bilinear(1,1) is true

sample size	N=500		N=1000		N=2000		N=3000	
	T1	T2	T1	T2	T1	T2	T1	T2
mean	-4.052	-0.235	-5.386	-0.254	-7.823	-0.119	-8.496	-0.197
s.d	2.208	0.920	2.828	1.021	4.206	0.914	4.693	0.914
skew	-0.429	-0.095	0.167	-0.051	0.275	-0.143	-0.011	-0.180
kurt	4.425	2.538	2.316	2.609	1.931	2.553	1.908	3.065
R.F.(α=.05)	0.880	0.040	0.865	0.050	0.900	0.015	0.905	0.045
toohigh	0.000	0.000	0.000	0.010	0.000	0.000	0.000	0.005
toolow	0.880	0.040	0.865	0.040	0.900	0.015	0.905	0.040

Data are generated from t_{10} distribution, R=200, and the conditional variance from equation (2.10) is used

for N = 500 in Table 2.15 is 0.040 which is almost equivalent to the nominal size of 0.05 and it is more than three times lower than the actual size of 0.145 from Table 2.7 for N = 500.

Figure 2.15 shows the edfs of the robust modified Cox test. The edfs are almost equivalent to those in Figure in 2.7 and they appear approximately to be the cdf of N(0,1).

These simulation results, in Table 2.13 through 2.15, exhibit that the actual size is very close to the nominal size and that the simulation statistics are very similar to those from Table 2.1, 2.3, and 2.7. But the main difference between this type of robust modified Cox test and the previously proposed one is that the actual size of the alternative way of the robust modified Cox test is slightly understated while the actual size of the robust modified Cox test is slightly overstated.

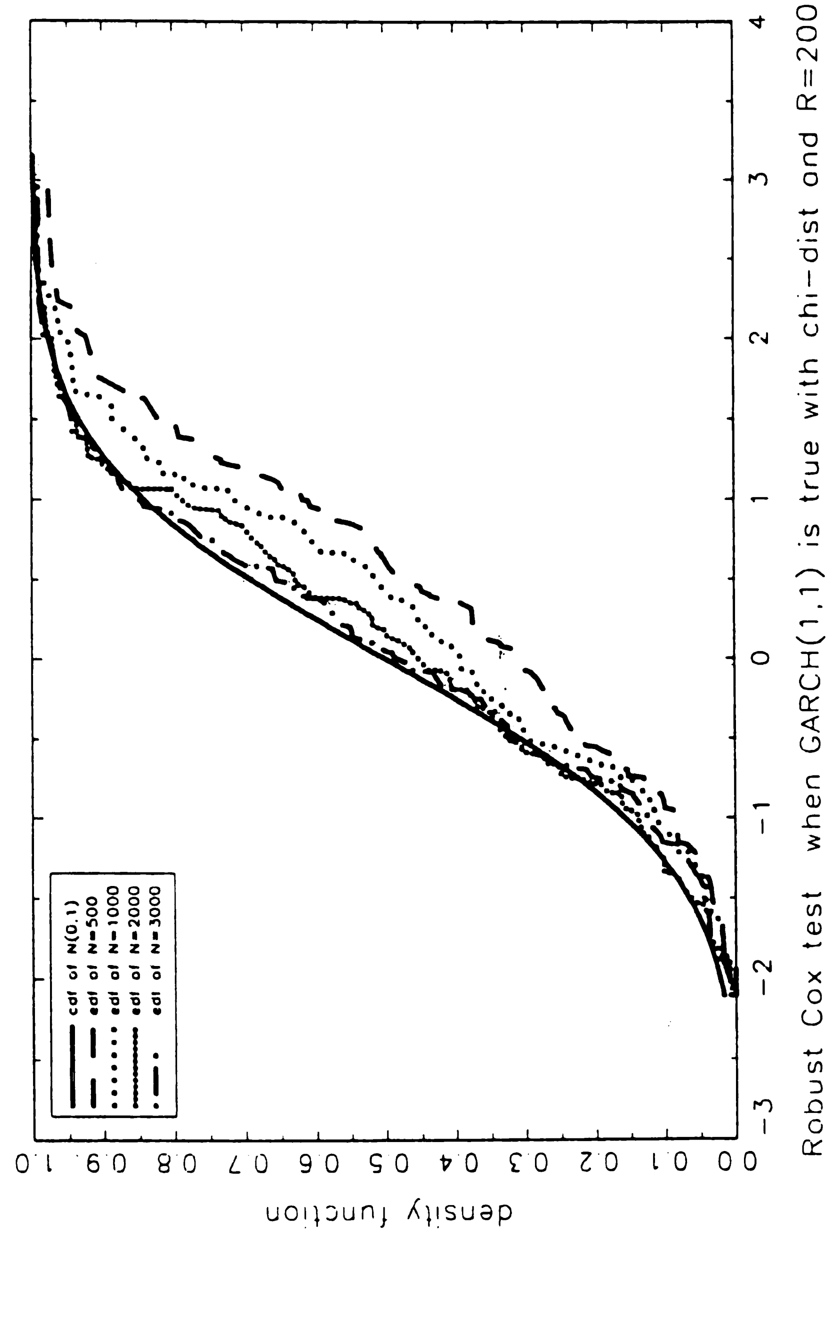
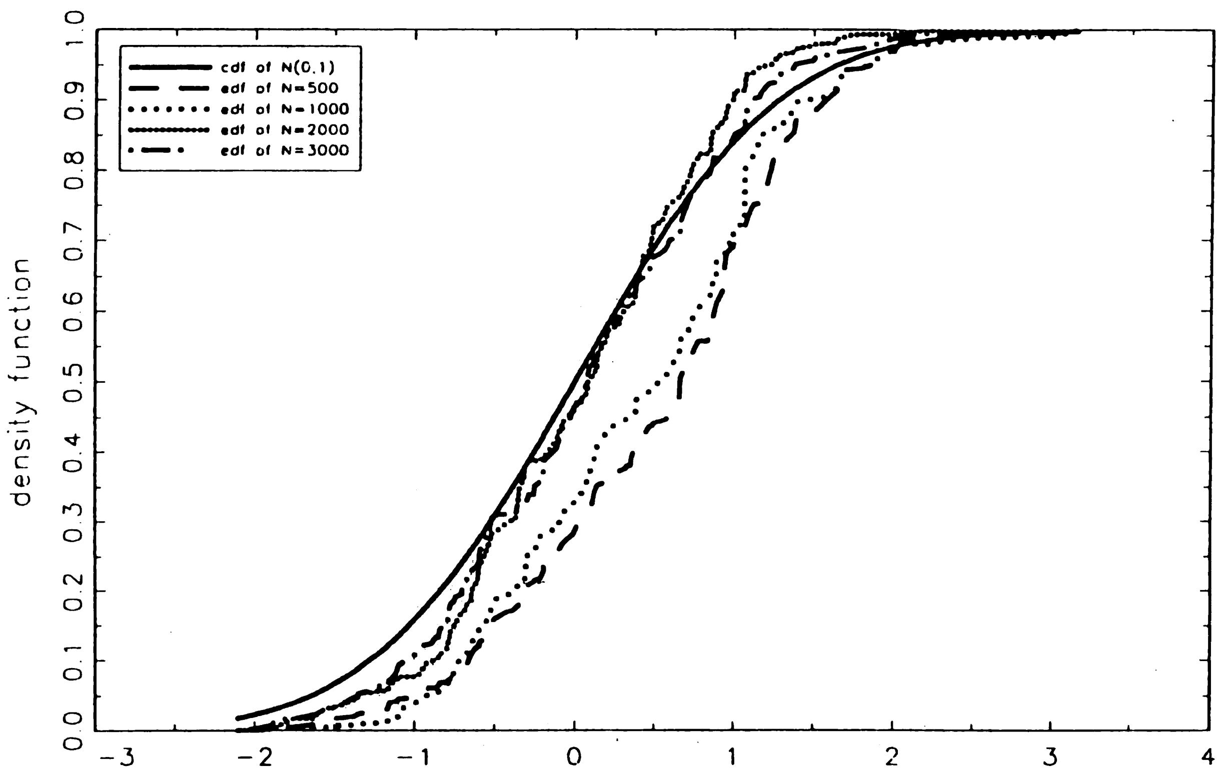


Figure2.1 The edfs of the modified Cox test vs. the cdf of N(0,1)

Figure2.2 The edfs of the modified Cox test vs. the cdf of N(0,1)



Nonrobust Cox test when GARCH(1,1) is true with chi-dist and R=200

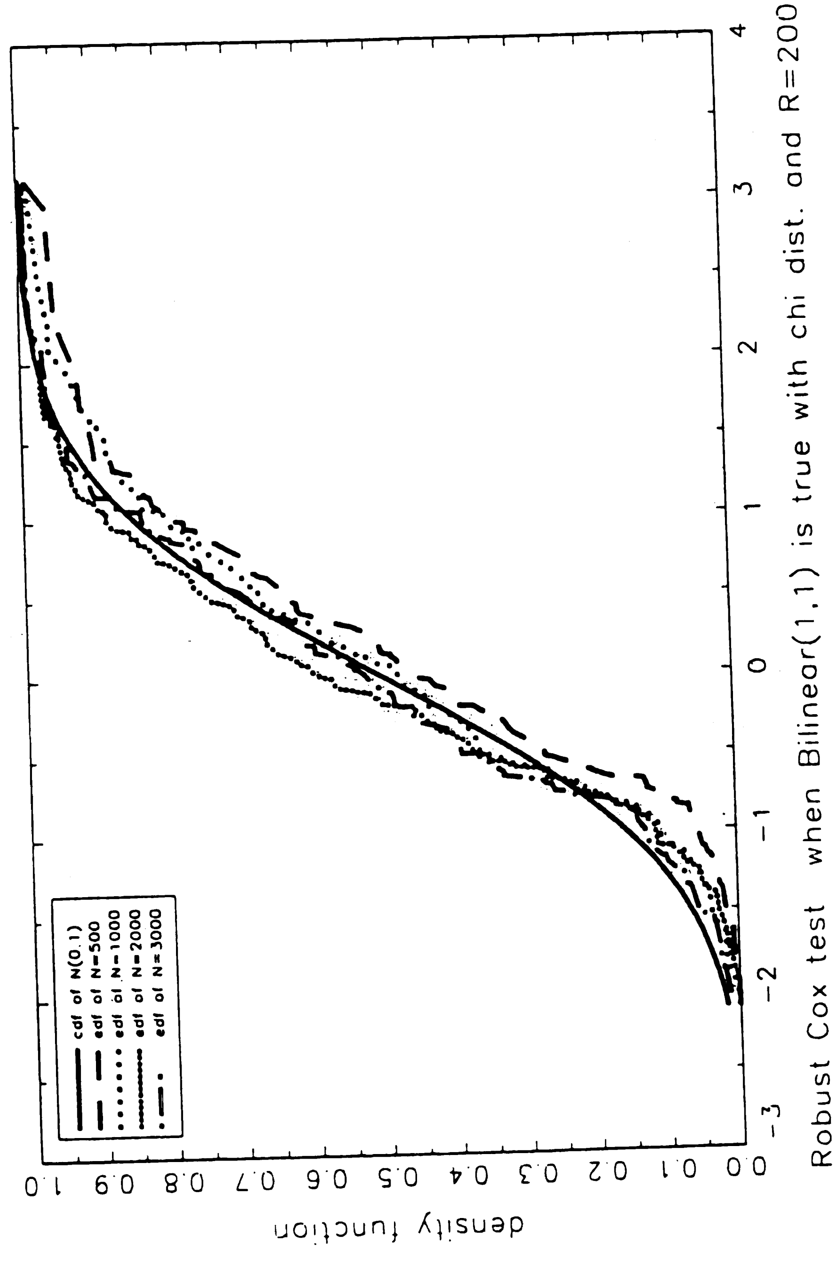


Figure2.3 The edfs of the modified Cox test. vs. the cdf of N(0,1)

Figure2.4 The edfs of the modified Cox test vs. the cdf of N(0,1)

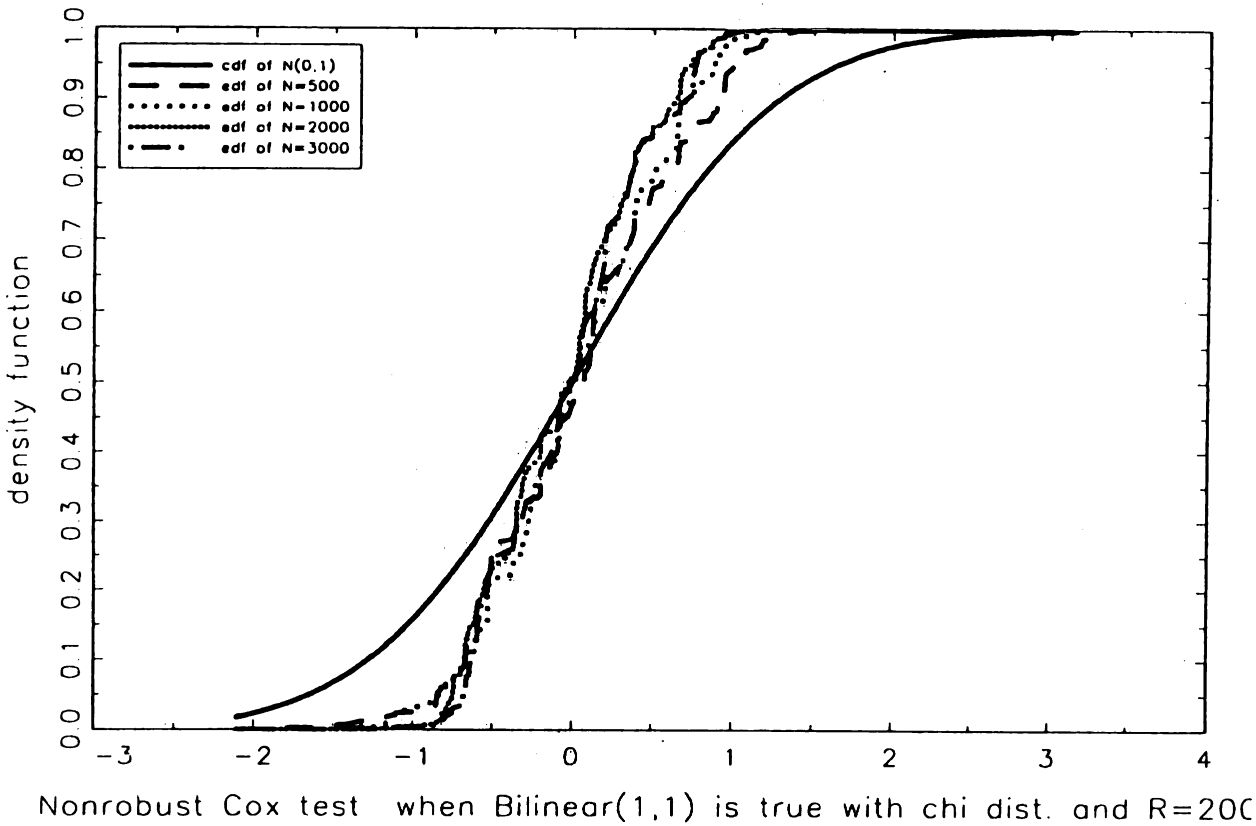


Figure2.5 The edfs of the modified Cox test vs. the cdf of N(0,1)

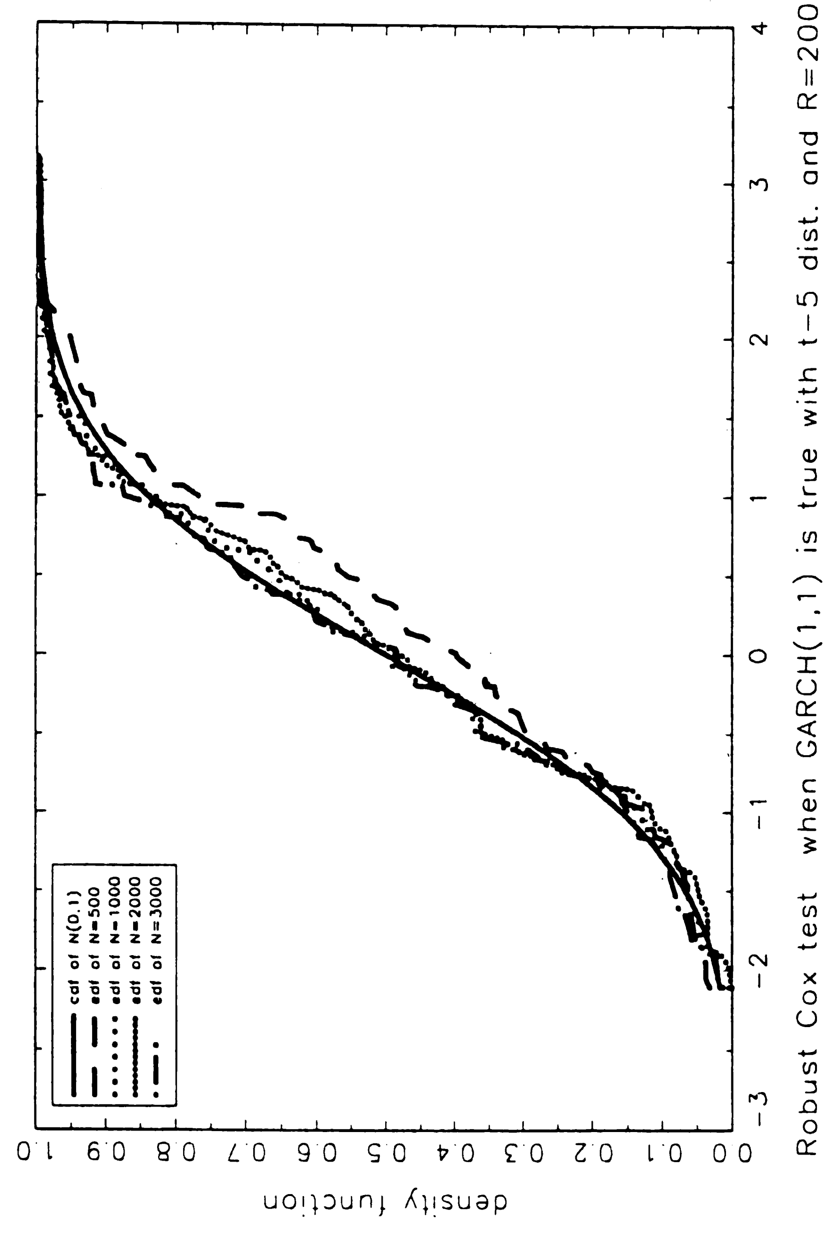
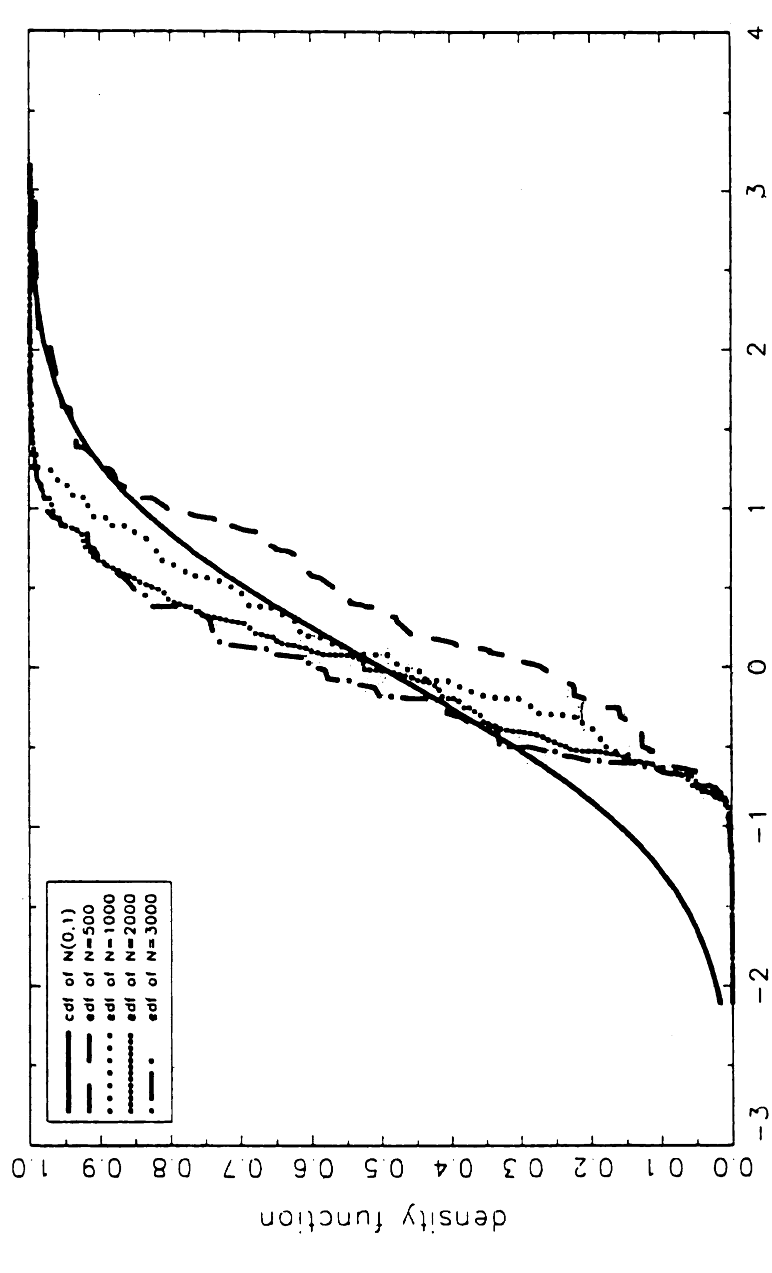
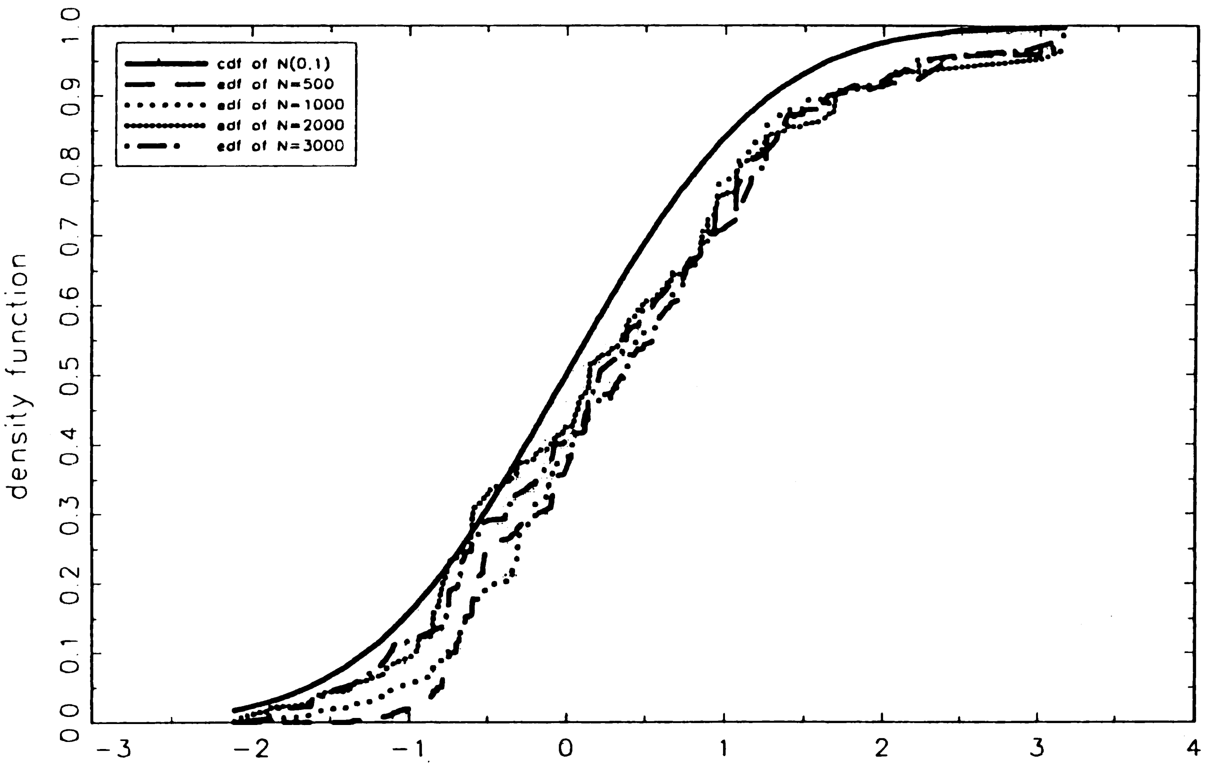


Figure2.6 The edfs of the modified Cox test vs. the cdf of N(0,1)



Nonrobust Cox test when GARCH(1,1) is true with t-5 dist. and R=20C

Figure2.7 The edfs of the modified Cox test vs. the cdf of N(0,1)



Robust Cox test when Bilinear(1,1) is true with t-5 dist. and R=200

Figure2.8 The edfs of the modified Cox test vs. the cdf of N(0,1)

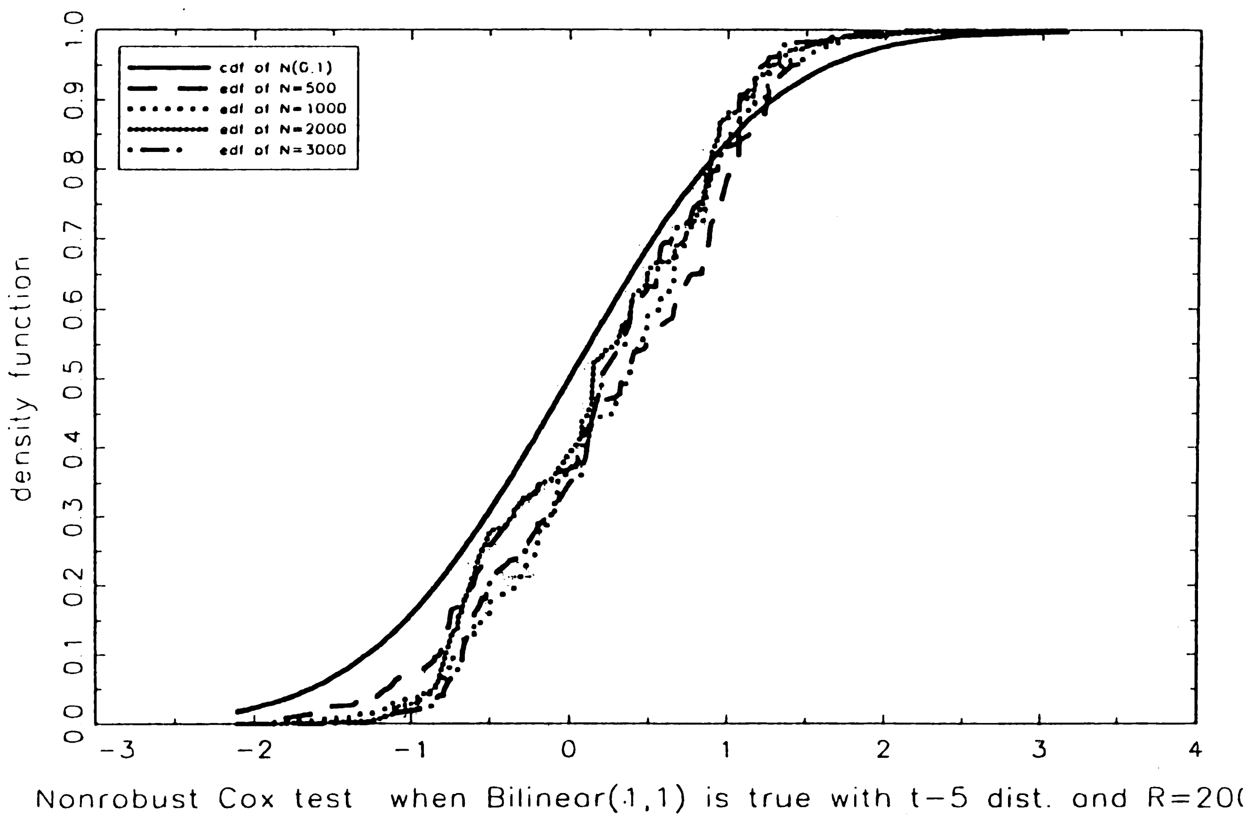


Figure2.9 The edfs of the modified Cox test vs. the cdf of N(0,1)

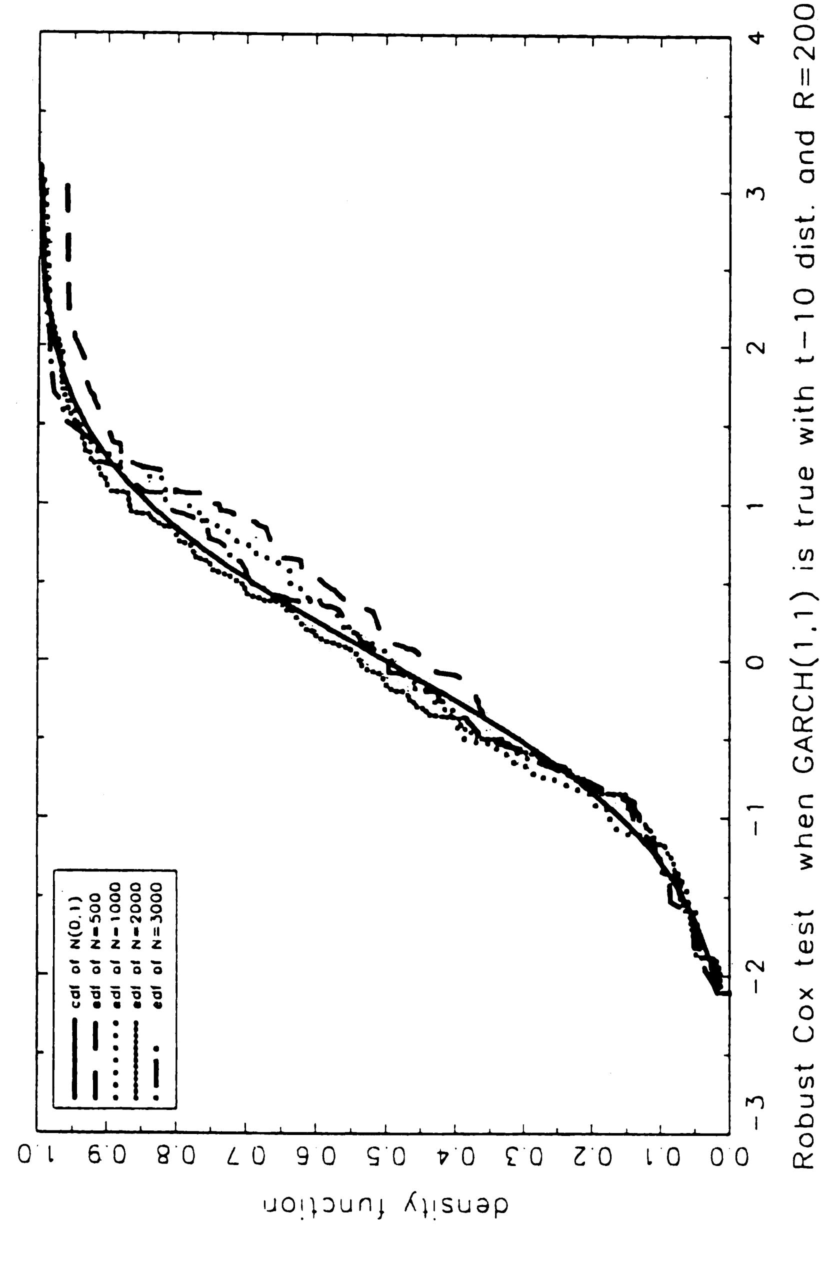


Figure2.10 The edfs of the modified Cox test vs. the cdf of N(0,1)

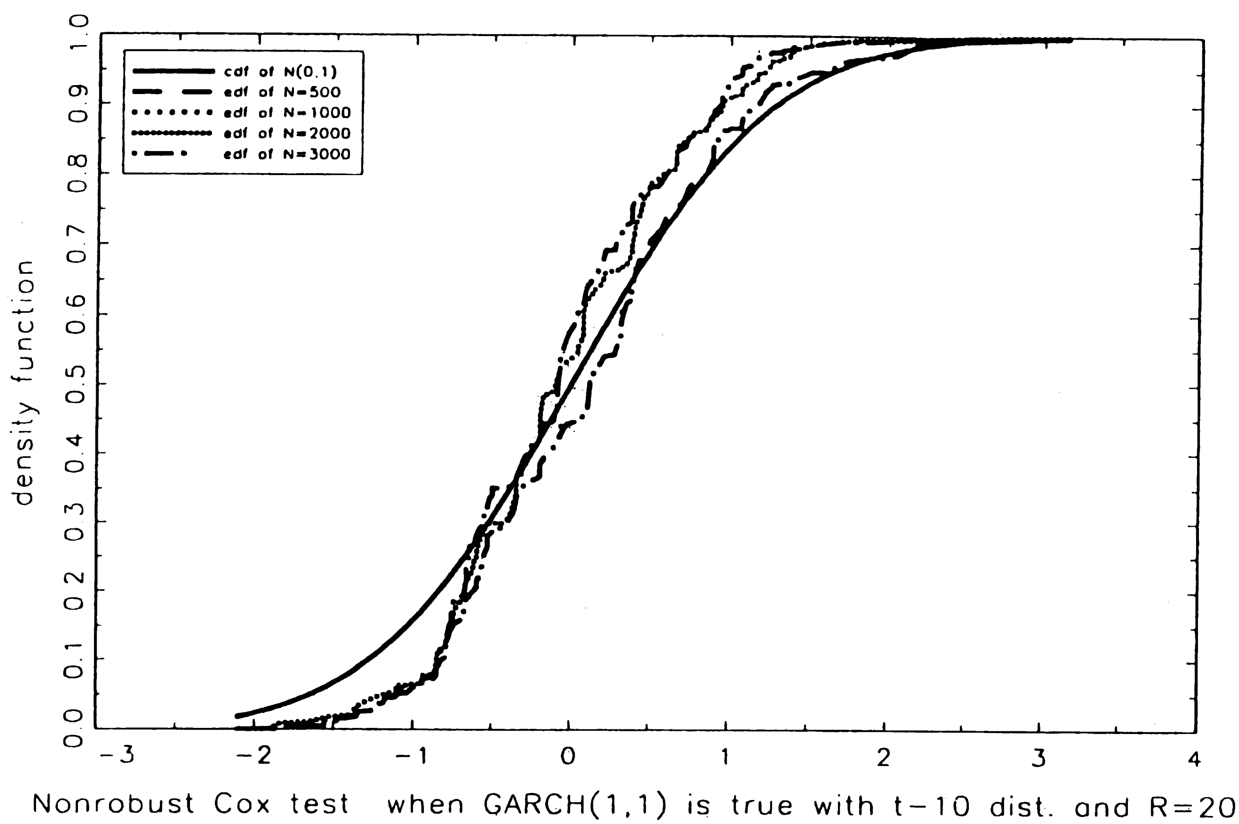


Figure2.11 The edfs of the modified Cox test vs. the cdf of N(0,1)

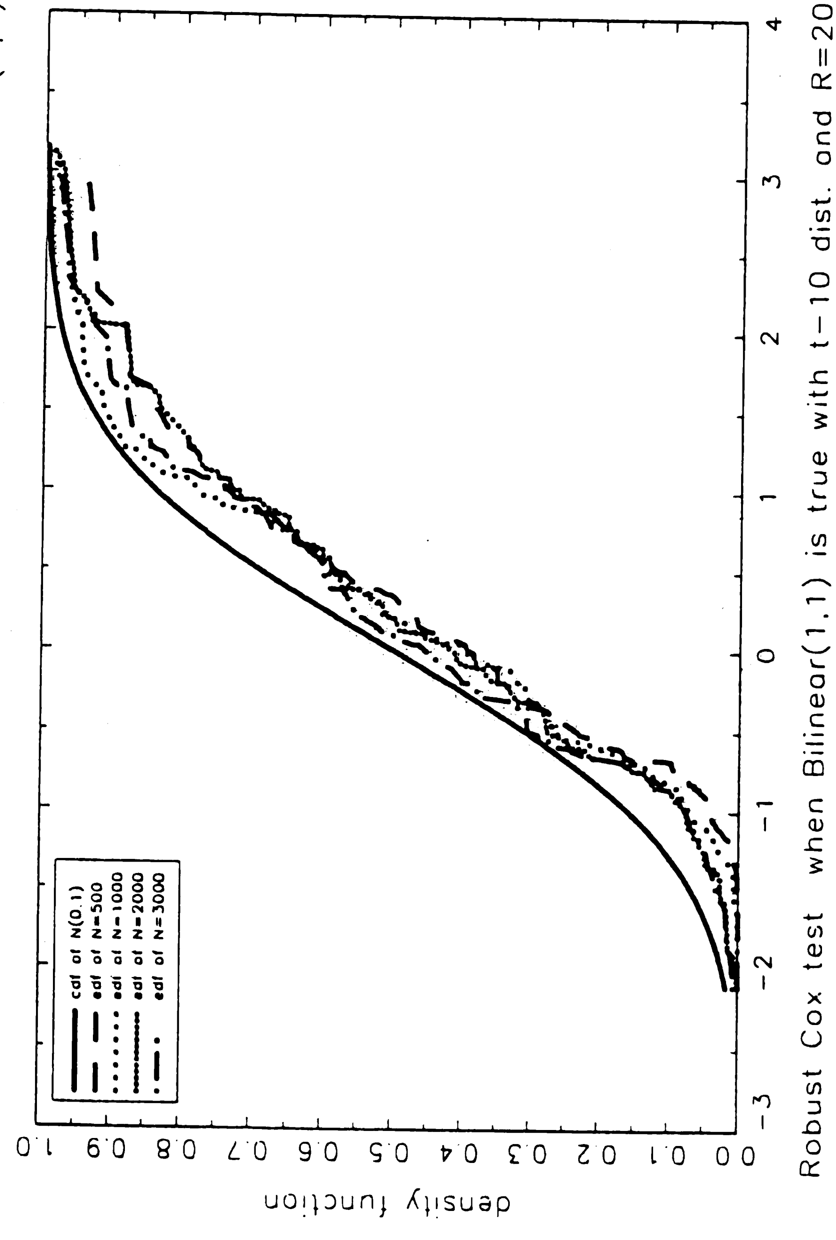
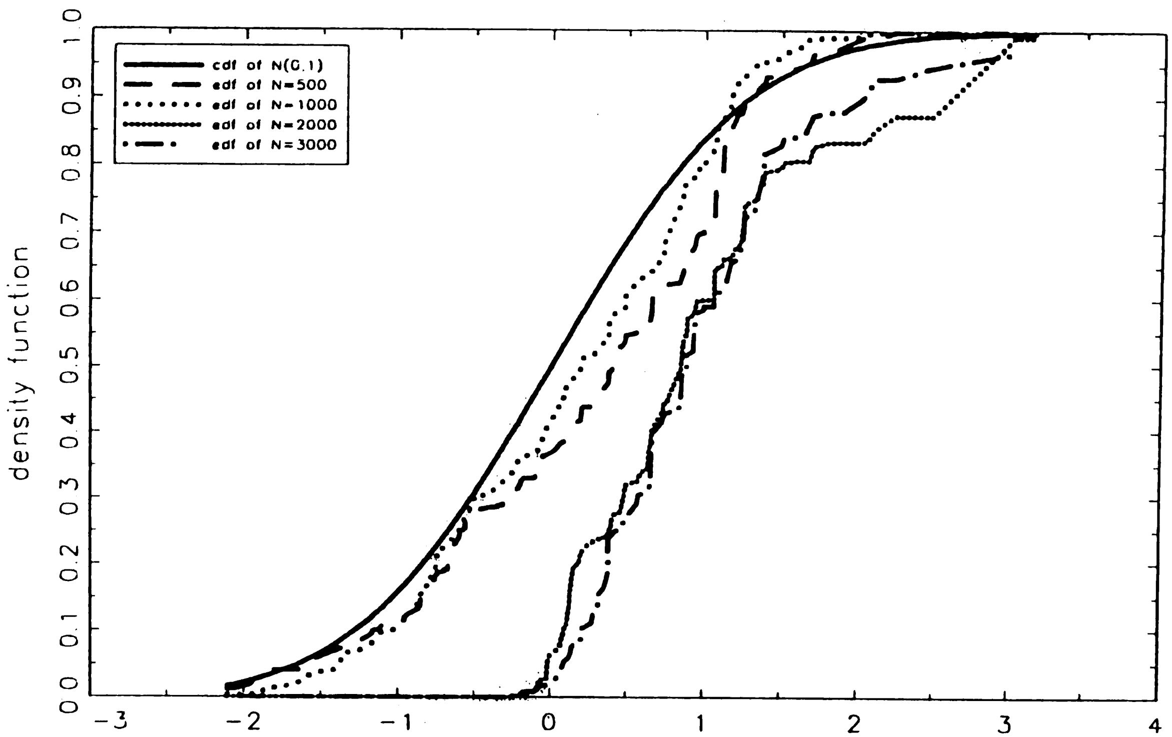
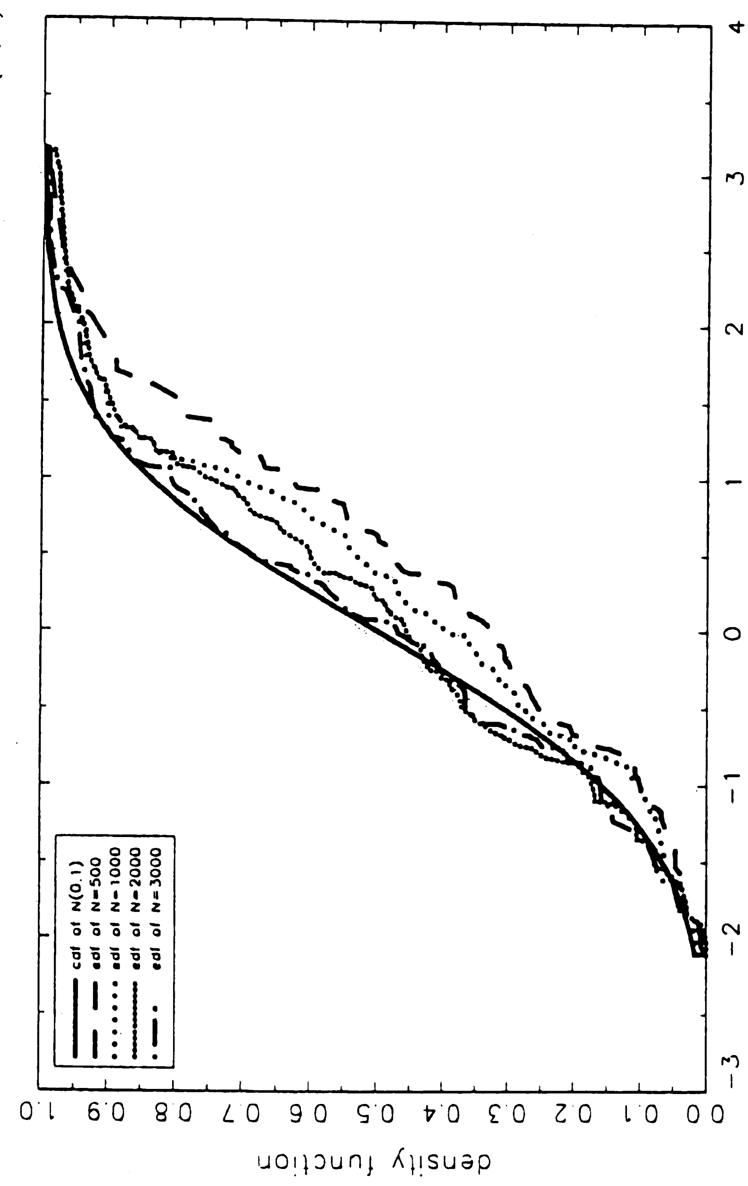


Figure2.12 The edfs of the modified Cox test vs. the cdf of N(0,1)



Nonrobust Cox test when Bilinear(1,1) is true with t-10 dist. and R=20

Figure2.13 The edfs of the modified Cox test vs. the cdf of N(0,1)



Robust Cox test when GARCH(1,1) is true with chi-dist. and R=200

Figure2.14 The edfs of the modified Cox test vs. the cdf of N(0,1)

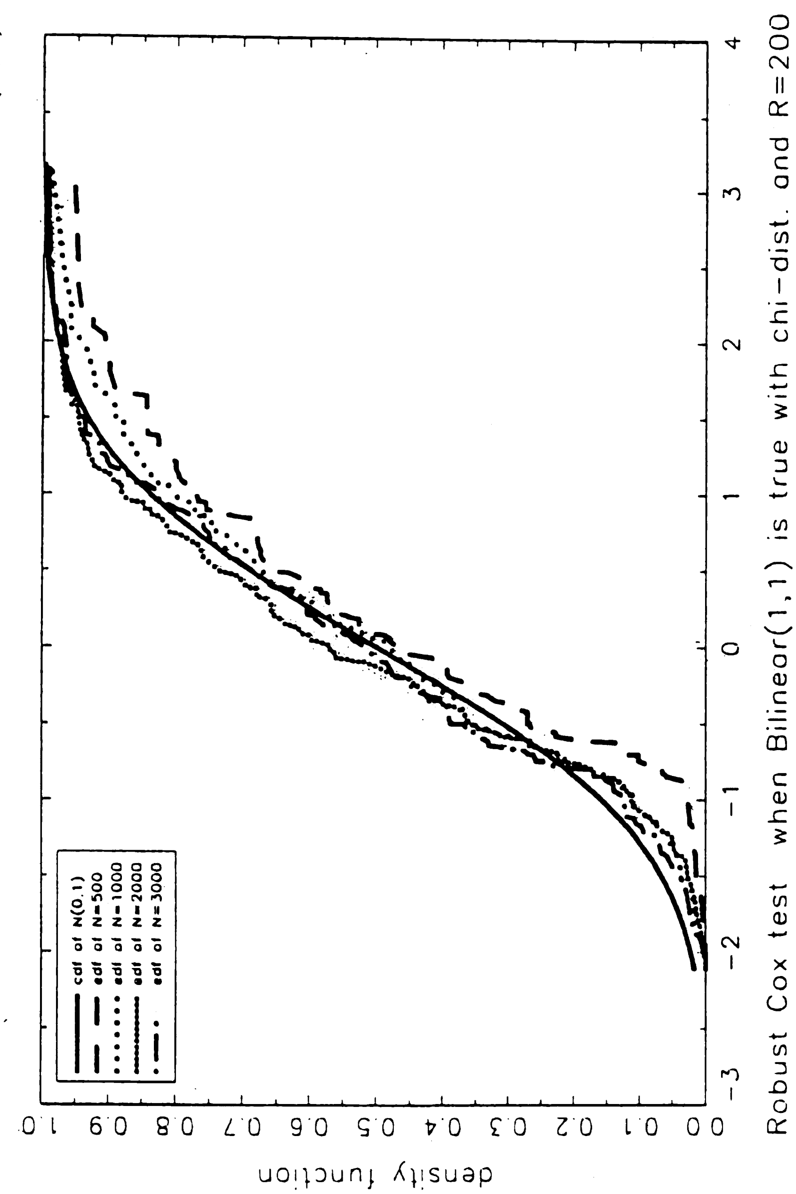
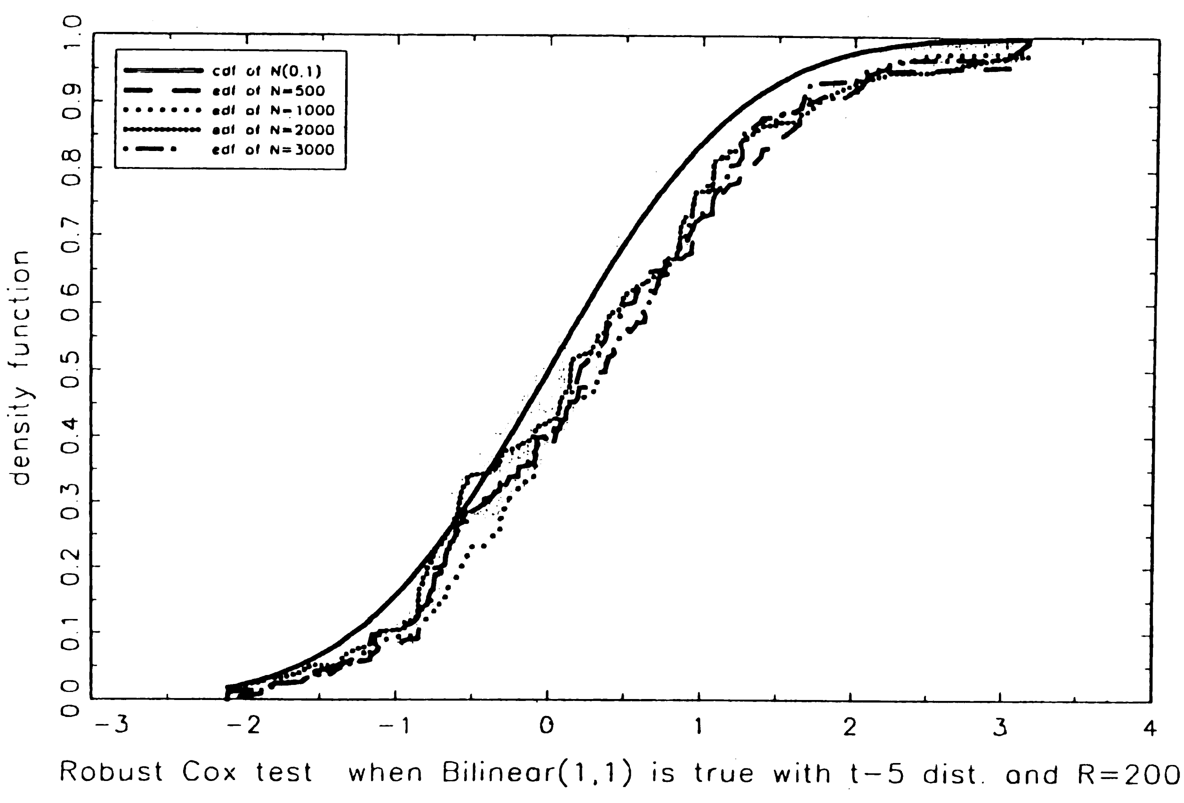


Figure2.15 The edfs of the modified Cox test vs. the cdf of N(0,1)



2.4 Empirical Application under Nonnormality

As noted earlier, high frequency financial time series ususally exhibit leptokurtosis and the distribution from these time series typically shows fatter tails than the normal distribution. Applying the modified Cox test, assuming normality to the situation under nonnormality, generally leads to invalid test inferences. As seen in Table 1.1 through 1.3 in section 1.3, all the empirical data sets reveal higher kurtosis than the normal distribution, so we suspect that the distributions of these time series follow the normal distribution. To investigate this, we perform the robust modified Cox test for these three time series data sets and compare the test results from the robust modified Cox test to the test results from the modified Cox test in the previous chapter.

Table 2.16 and 2.17 are the test results from the robust modified Cox test and Table 2.18 and 2.19 are the test results from an alternative way of the modified Cox test. First, when the null is the GARCH(1,1) model, the test values from the robust modified Cox test are smaller than those from the nonrobust test but both test results from the robust and the nonrobust modified Cox test could not reject the null hypothesis at any significance levels. Second, when the null is the bilinear model, the test values from two types of the robust modified Cox test are much smaller than those from Table 1.11, especially the test value of IP series but the test

Table 2.16: Test results:H_0:GARCH *vs.* H_1: Bilinear

	stochastically simulated Cox test	modified Cox test
S&P 500	.023	.114
British Pound	.196	-.011
Industrial Production	.533	.017

Table 2.17: Test results:H_0:Bilinear *vs.* H_1: GARCH

	stochasticly simulated Cox test	modified Cox test
S&P 500	-.910	-2.896
British Pound	-2.797	-6.289
Industrial Production	-1.643	-3.802

values between these two robust modified Cox tests are very similar: Test results in Table 2.16 vs. Table 2.18 and Table 2.17 vs. Table 2.19. And both test results reject the bilinear model as the null at any significance levels. The test results, in Table 2.16 through 2.19, are different from those in Table 1.10 and 1.11 and show evidence that these time series do not follow the normal distribution and the test statistics from the robust modified Cox test are more valid.

Table 2.18: Test results:H_0:GARCH *vs.* H_1: Bilinear

	stochasticly simulated Cox test	modified Cox test
S&P 500	.023	.179
British Pound	.196	-.051
Industrial Production	.533	.020

Table 2.19: Test results:H_0:Bilinear *vs.* H_1: GARCH

	stochasticly simulated Cox test	modified Cox test
S&P 500	-.910	-2.896
British Pound	-2.797	-7.058
Industrial Production	-1.643	-2.282

2.5 Conclusions

As noted earlier, most financial time series do not show evidence of the conditional normality. Thus, applying the modified Cox test, assuming normality to the non-normal situation, yields invalid test inferences. In this chapter, we have proposed the robust modified Cox test under nonnormality.

Monte Carlo simulation experiments suggest that the robust modified Cox test performs fairly well and can improve the validity of the test statistics and the actual size over the nonrobust modified Cox test under nonnormality. In comparison with the simulation results in the previous chapter, the means and standard deviations are slightly deviated from the standard normal distribution for some sample sizes such as N = 500. It is also shown that the robust modified Cox test oversizes but not significantly and the actual size approaches the nominal size as sample size increases. Evidence from Lumsdain(1995) suggested that the robust modified Cox test would perform relatively well under nonnomality. She compared the robust traditional test statistics to the nonrobust traditional test statistics under normality and showed the actual size and the test statistics are not close enough to the nominal levels even under normality. We infer that these results would be even worse under nonnormality. We also performed an alternative way of the robust modified Cox test with the conditional variance from equation (2.10). The simulation results are very similar to those from the robust modified Cox test in general but the actual size is usually understated while the robust modified Cox test usually slightly oversizes.

In some situations, as shown in Table 2.15, an alternative way of robust modified Cox test that we proposed here preforms very well and the simulation results are very close to normal. It is emphasized that this robust modified Cox test has computational advantage because it does not require computing every term in the conditional variance in equation (2.15).

We also have summarized the nonrobust Cox test with three time series data sets in the previous chapter. The robust test results are far different from the nonrobust test results, especially when the null is the bilinear model: the test values are much smaller than those from the nonrobust modified Cox test in absolute value for all three data sets.

Chapter 3

An Application of a Quasi-Modified Cox Test to Nonlinear Panel Data Models

3.1 Introduction

In many instances, the dependent variable takes on nonnegative integer values: for example, number of hospital visits in a given year, number of alpha particles emitted from a radioactive source during a given period of time or number of patents applied for and received by a firm during a year. When a variable takes on nonnegative integer values, it is referred to as a count variable. With the nonnegative property of count data, the most popular functional form for the conditional mean is the exponential function: $E(y \mid x) = \exp(x\beta)$, where y is a count variable and x is a vector of explanatory variables. When there are unobserved effects in count panel data models, we cannot simply apply the standard linear unobserved effects model if we want to impose nonnegativity of the conditional mean. Hausman, Hall, and Griliches (1984)(hereafter HHG) is a pioneering work that deals with the unobserved effects in count panel data analysis using the conditional maximum likelihood (CML) approach of Anderson (1970, 1972). HHG also presented an application to the patents and R&D expenditures relationship. Wooldridge (1999) showed the QCMLE is consistent and asymptotically normal just under the conditional mean assumption in the multiplicative models. He also showed that Poisson QMLE is robust if the conditional mean is correctly specified. But it will be inefficient, in general, unless the conditional varience is also correclty specified.

The most popular distributional assumption for count data is the Poisson distribution. To remove the unobserved heterogeneity or fixed effects in the nonlinear count panel data analysis, the Fixed Effects Poisson (FEP) model was developed by HHG. But one of the shortcomings of the Poisson model is that the first two moments are the same. But in many applications the variance of a count variable is larger than the mean of it and we encounter overdispersion of the data. To solve this problem, the Fixed Effects Negative Binomial (FENB) model was developed as an alternative to the FEP. As shown in Wooldridge (1999), both the FEP and the FENB models have the same form of the conditional mean and are estimated by the multinomial QCML methodology. Like the GARCH and the Bilinear models in the previous chapters, the FEP and the FENB are two competing models in nonlinear count panel data analysis. It is worth while to note that the QCMLE of the FEP is consistent and asymptotically normal if the conditional first moment is correctly specified but, in general it is inefficient. On the contrary, the QCMLE of the FENB is not consistent unless the first two conditional moments are correctly specified because negative binomial is not in LEF but the FENB is usually more efficient than the FEP. Therefore, there is a robustness and efficiency trade-off between these two models. In principle, we could try the Cox test when we consider the specification testing between these two models because the Cox test applies to any two distributions and it is derived from the difference between log likelihood ratio and its expected value under the null. But using the original Cox test or even the modified Cox test from the previous chapter is very challenging task in this case. The log likelihood function of the FEP model is derived from Poisson distribution and the log likelihood function of the FENB model is derived from negative binomial distribution and these two log likelihood functions take very different forms from the normal log likelihood function. Therefore, to get the difference between these two different log likekihood ratio and its expected value is very complicated and computationally very difficult too. Instead, we want something computationally simpler and we derive a new Cox test (quasi-modified Cox test) using the property of normal quasi log-likelihood as shown, for example in Bollerslev and Wooldridge (1992). In this case the quasi-modifed Cox test is based only upon the implied conditional varience because the conditional means of these two models are the same. In section 2 we briefly explain these two models and the quasi-modified Cox test. In section 3 we present the applications of these models to the U.S. patents and R&D expenditures panel data, and then apply the quasi-modified Cox test to see if either model is rejected. Conclusions follow on section 4.

3.2 Two Competing Count Panel Data Models with the Unobserved Effects

Developed first by HHG, the FEP and the FENB models have been used as two competing counterparts in the nonlinear count panel data analysis. In this section we briefly discuss these two models.

We assume ramdom sampling from cross section and let $\{(y_{it}, x_{it}, \phi_i), i = 1, 2, \ldots, N, t = 1, 2, \ldots, T\}$ be a sequence of *i.i.d* random variables across i, but not t, where y_{it} denotes the discrete observable count variable, x_{it} is a vector of explanatory variables and ϕ_i is an unobserved random scalar. For the FEP model we assume that

$$y_{it} \mid x_i, \phi_i \sim Poisson(\phi_i \mu(x_{it}, \beta_0)), t = 1, 2, \ldots, T \tag{3.1}$$

$$y_{it}, y_{is} \ \ are \ \ independent \ \ conditional \ \ on \ \ x_{it}, \phi_i, \ \ t \neq s, \tag{3.2}$$

and the conditional mean of y_{it} is

$$\begin{aligned} E(y_{it} \mid x_i, \phi_i) &= E(y_{it} \mid x_{it}, \phi_i) \\ &= \phi_i \mu(x_{it}, \beta_0) \end{aligned} \tag{3.3}$$

HHG took the functional form of the conditional mean of y_{it} as an exponential function: $E(y_{it} \mid x_i, \phi_i) = \phi_i exp(x_{it}, \beta_0)$. Under assumptions (3.1) and (3.2), HHG used the CML techniques of Anderson(1970,1972) to estimate β, conditioning on the sum of the dependent variable across time, $\sum_{t=1}^{T} y_{it} = n_i$. HHG showed that

$$y_i \mid n_i, x_i, \phi_i \sim multinomial(n_i, p_{i1}(x_i, \beta_0), \ldots, p_{iT}(x_i, \beta_0)) \tag{3.4}$$

$$where \ \ p_{it} = exp(x_{it}, \beta_0) / \sum_{r=1}^{T} exp(x_{ir}, \beta_0) \ \ and \ \ \sum_{t=1}^{T} p_{it} = 1 \tag{3.5}$$

Eq(3.4) reveals that the distribution of $(y_{i1}, y_{i2}, \cdots, y_{iT})$ given (x_i, n_i) does not depend upon the unobserved effects ϕ_i. Therefore, the log likelihood function of the Poisson CML methodology by HHG can be written as

$$l_i(\beta)_{FEP} = \sum_{t=1}^{T}\Gamma(y_{it}+1) - \sum_{t=1}^{T} y_{it}\log\sum_{s=1}^{T}\exp(-(x_{it}-x_{is})\beta), \tag{3.6}$$

where $\Gamma(.)$ is gamma function.

Gourieroux, Monfort, and Trognon (1984) (hereafter GMT) showed that the multinomial QCMLE of the FEP is consistent even though the multinomial distribution is not correctly specified if

$$\begin{aligned} E(y_{it} \mid n_i, x_i) &= p_{it}(x_i, \beta_0)n_i \\ where \quad n_i &= \sum_{t=1}^{T} y_{it} \end{aligned} \tag{3.7}$$

However, Wooldridge (1999) argued that this is too restrictive and showed that, while the FEP estimator is derived under assumptions (3.1) and (3.2), it is consistent and asymptotically normal only under the conditional mean assumption (3.3).

On the other hand, $E(y_{it} \mid x_i, \phi_i) = Var(y_{it} \mid x_i, \phi_i) = \lambda_{it}$ where λ_{it} is the Poisson parameter from assumption (3.1). But it is not difficult to find, empirically, that the conditioanl variance of y_{it} is not the same as the conditional mean of y_{it}. More likely, the conditional variance of y_{it} is larger than the conditional mean of y_{it} or it is increasing with y_{it} in many cases. To solve this overdispersion problem, HHG used the negative binomial distribution and developed the Fixed Effects Negative Binomial (FENB) model as an alternative to the FEP. To derive the FENB by HHG, we assume that[1]

$$
\begin{aligned}
y_{it} \mid x_i, \phi_i &\sim NegativeBinomial(\mu(x_{it}, \beta_0), 1/\phi_i) && (3.8)\\
&\text{where } \phi_i \text{ is the unobserved effect and } \phi_i > 0 \\
& y_{it}, y_{ir} \text{ are independent conditional on } (x_{it}, \phi_i), \quad t \neq r && (3.9)\\
E(y_{it} \mid x_i, \phi_i) &= \phi_i \mu(x_{it}, \beta_0) && (3.10)
\end{aligned}
$$

Interestingly, under (3.8) to (3.10), the conditional mean is $E(y_{it} \mid n_i, x_i) = p_{it}(x_i, \beta_0) n_i$, which is the same as Eq. (3.7).

The conditional log likelihood function[2] for the FENB by HHG is

$$
\begin{aligned}
l_i(\beta)_{FENB} &= \sum_{t=1}^{T} (\log\Gamma(\mu_{it} + y_{it}) - \log\Gamma(\mu_{it}) - \log\Gamma(y_{it}+1)) \\
&\quad + \log\Gamma(\sum_{t=1}^{T} \mu_{it}) + \log\Gamma(n_i + 1) - \log\Gamma(\sum_{t=1}^{T} \mu_{it} + n_i) \quad (3.11)
\end{aligned}
$$

Under assumptions (3.8) to (3.10), the strict exogeneity of x_{it}, the CMLE of the FENB is consistent and asymptotically normal.

Now, we compare the Possion model and the Negative Binomial model analyzed in Wooldridge (1999). In Possion model,

$$
\begin{aligned}
E(y_{it} \mid x_{it}, \phi_i) &= \phi_i \mu(x_{it}, \beta_0) && (3.12)\\
&= Var(y_{it} \mid x_{it}, \phi_i) && (3.13)
\end{aligned}
$$

[1] We follow the notation from Wooldridge (1999)

[2] see HHG p.924 for more details

and the variance to mean ratio of the Possion model is unity. In Negative Binomial model,

$$E(y_{it} \mid x_{it}, \phi_i) = \phi_i \mu(x_{it}, \beta_0) \quad (3.14)$$

$$Var(y_{it} \mid x_{it}, \phi_i) = E(y_{it} \mid x_{it}, \phi_i)(1 + \phi_i) \quad (3.15)$$

and the variance to mean ratio of the NB model is $(1 + \phi_i) > 1$. The NB model shows the overdispersion and also allows the variance to mean ration to be different from each i.

The conditional mean of both Poisson and NB models conditional on the sum of dependant variable across time is $E(y_{it} \mid n_i, x_i) = p_{it}(x_i, \beta_0)n_i$. Next, we consider the conditional variance of both the FEP and the FENB models. Following HHG (1984), we first construct the conditional variance of the FEP model from the multinomial covariance matrix, $\Omega_i = diag(p_i) - p_i'p_i$ where $p_{it} = \mu(x_{it}, \beta_0)/\sum_{r=1}^{T} \mu(x_{ir}, \beta_0)$. From the fixed effects assumption, Ω_i is singular by construction. Therefore, we remove the first row and column to construct Ω_i, which is $(T-1) \times (T-1)$ matrix.[3] We derive the conditional variance of the FEP from the diagonal elements of Ω_i: $V(y_{it} \mid n_i, x_i) = (1 - p_{it})p_{it}$. Next, we derive the conditional variance of the FENB as we did that of the FEP but an extra term added from the NB assumption: $g_i = (\sum_{t=1}^{T} n_{it} + \sum_{t=1}^{T} \mu(x_{it}\beta_0))/(1 + \sum_{t=1}^{T} \mu(x_{it}\beta_0))$ and

[3] We can remove any time period from Ω_i. We take the first row and column for convenience.

$\Omega_{FENB} = g_i(diag(p_i) - p_i'p_i)$ and the conditional variance of the FENB is the diagonal terms of $\Omega : V(y_{it} \mid n_i, x_i) = g_i p_{it}(1 - p_{it})$.

The original Cox test is $T_f = \{L_f(\hat{\alpha}) - L_g(\hat{\beta})\} - E_{\hat{\alpha}}\{L_f(\hat{\alpha}) - L_g(\hat{\beta})\}$. The test statistic of the Cox test is based upon the difference between the log likelihood ratio and its expected estimate under the null. In principle, we can try the original Cox test but this may cause very severe computational difficulties. Instead, we use the first two conditional moments from the QCML methodology and construct a quasi-modified Cox test using the normal quasi-log likelihood framework. Bollerslev and Wooldridge (1992) showed that the normal log-likelihood and its expected values are maximized when the correct conditional mean and variance are used, even though the normality assumption is violated. Using this property, we now construct a quasi-modified Cox test. Let M_1 denote the model defined by Eqs. (3.1) and (3.2). Under M_1,

$$E(y_{it} \mid n_i, x_i) = p_{it}(x_i, \beta_0)n_i \tag{3.16}$$

$$Var(y_{it} \mid n_i, x_i) = (1 - p_{it}(x_i, \beta_0))p_{it}(x_i, \beta_0) \tag{3.17}$$

$$where \quad n_{it} = \sum_{t=1}^{T} y_{it},$$

$$and \quad p_{it} = \frac{\mu(x_{it}, \beta_0)}{\sum_{r=1}^{T} \mu(x_{ir}, \beta_0)},$$

Let M_2 denote the model by Eqs. (3.8) and (3.9), so that

$$E(y_{it} \mid n_i, x_i) = p_{it}(x_i, \theta_0)n_i \tag{3.18}$$

$$
\begin{aligned}
Var(y_{it} \mid n_i, x_i) &= g_i(1 - p_{it}(x_i, \theta_0))p_{it}(x_i, \theta_0) \qquad (3.19)\\
where \quad n_{it} &= \sum_{t=1}^{T} y_{it},\\
p_{it} &= \frac{\mu(x_{it}, \theta_0)}{\sum_{r=1}^{T} \mu(x_{ir}, \theta_0)},\\
and \quad g_i &= (\sum_{t=1}^{T} n_{it} + \sum_{t=1}^{T} \mu(x_{it}, \theta_0))/(1 + \sum_{t=1}^{T} \mu(x_{it}, \theta_0))
\end{aligned}
$$

We use these two conditional mean and variance from QCML methodolody and put them into the modified Cox test that we derived from previous chapter to get the test statistic. Now, the quasi-modified Cox test has a form of

$$
\begin{aligned}
\hat{T}_{M_1} &= \frac{1}{N(T-1)} \sum_{i=1}^{N} \sum_{t=1}^{T-1} \left[\left(y_{it} - p_{it}(x_i, \hat{\beta}) n_i \right) \left\{ \frac{p_{it}(x_i, \hat{\beta}) n_i - p_{it}(x_i, \hat{\theta}) n_i}{(1 - p_{it}(x_i, \hat{\theta})) p_{it}(x_i, \hat{\theta})} \right\} \right.\\
&+ \frac{u_{it}(\hat{\beta})^2 - (1 - p_{it}(x_i, \hat{\beta})) p_{it}(x_i, \hat{\beta})}{2}\\
&\left. \times \left(\frac{1}{g_i(1 - p_{it}(x_i, \hat{\theta})) p_{it}(x_i, \hat{\theta})} - \frac{1}{(1 - p_{it}(x_i, \hat{\beta})) p_{it}(x_i, \hat{\beta})} \right) \right] \qquad (3.20)
\end{aligned}
$$

Following the previous chapter[4], we can derive the asymptotic distribution of the quasi-modified Cox test:

$$
\begin{aligned}
\sqrt{N} \hat{T}_{M_1} &= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left[\left(\bar{y}_i - \bar{p}_i(x_i, \hat{\beta}) n_i \right) \left\{ \frac{\bar{p}_i(x_i, \hat{\beta}) n_i - \bar{p}_i(x_i, \hat{\theta}) n_i}{(1 - \bar{p}_i(x_i, \hat{\theta})) \bar{p}_i(x_i, \hat{\theta})} \right\} \right.\\
&+ \frac{\bar{u}_i(\hat{\beta})^2 - (1 - \bar{p}_i(x_i, \hat{\beta})) \bar{p}_i(x_i, \hat{\beta})}{2}\\
&\times \left(\frac{1}{g_i(1 - \bar{p}_i(x_i, \hat{\theta})) \bar{p}_i(x_i, \hat{\theta})} - \frac{1}{(1 - \bar{p}_i(x_i, \hat{\beta})) \bar{p}_i(x_i, \hat{\beta})} \right)\\
&\left. - \left(p \lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} \bar{\psi}(\hat{\beta}, \hat{\theta}) \right) \left(p \lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} \bar{A}_n(\hat{\beta}) \right)^{-1} \nabla_{\beta} \bar{l}_i(\beta)_{FEP} \right]\\
&+ o_p(1) \qquad (3.21)
\end{aligned}
$$

[4] see section 2.2 for more detail

And the variance of the robust quasi-modified Cox test, $\hat{V}_{T_1}$, is

$$\begin{aligned}
\hat{V}_{T_1} &= \frac{1}{N}\sum_{n=1}^{N}\Bigg[\hat{D}_{n1}^2(1-\bar{p}_i(x_i,\hat{\beta}))\bar{p}_i(x_i,\hat{\beta}) + \frac{\hat{D}_{n1}\hat{D}_{n2}}{2}\bar{u}_i(\hat{\beta})^3 \\
&+\frac{\hat{D}_{n2}^2}{4}(\bar{u}_i(\hat{\beta})^4 - ((1-\bar{p}_i(x_i,\hat{\beta}))\bar{p}_i(x_i,\hat{\beta}))^2) \\
&+\left(\frac{1}{N}\sum_{i=1}^{N}\bar{\psi}(\hat{\beta},\hat{\theta})\right)\left(\frac{1}{N}\sum_{i=1}^{N}\bar{A}_i(\hat{\beta})\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\bar{\psi}(\hat{\beta},\hat{\theta})\right)' \\
&-2\left(\frac{1}{N}\sum_{i=1}^{N}\bar{\psi}(\hat{\beta},\hat{\theta})\right)\left(\frac{1}{N}\sum_{i=1}^{N}\bar{A}_i(\hat{\beta})\right)^{-1} \\
&\times\Bigg(\hat{D}_{n1}\nabla_\beta\bar{p}_i(x_i,\hat{\beta})n_i + \frac{\hat{D}_{n2}}{4(1-\bar{p}_i(x_i,\hat{\beta}))\bar{p}_i(x_i,\hat{\beta})} \\
&\times(1-\frac{1}{(1-\bar{p}_i(x_i,\hat{\beta}))\bar{p}_i(x_i,\hat{\beta})}) \\
&\times\nabla_\beta\left((1-\bar{p}_i(x_i,\hat{\beta}))\bar{p}_i(x_i,\hat{\beta})\right)\Bigg)\Bigg]
\end{aligned} \tag{3.22}$$

$$\begin{aligned}
\textit{where}\quad \hat{D}_{n1} &\equiv \frac{1}{T-1}\sum_{t=1}^{T-1}\left(\frac{(p_{it}(x_{it},\hat{\beta})-p_{it}(x_{it},\hat{\theta}))n_i}{g_i p_{it}(x_{it},\hat{\theta})(1-p_{it}(x_{it},\hat{\theta})}\right), \\
\hat{D}_{n2} &\equiv \frac{1}{T-1}\sum_{t=1}^{T-1}\left(\frac{1}{g_i p_{it}(x_{it},\hat{\theta})(1-p_{it}(x_{it},\hat{\theta})}\right. \\
&\left.-\frac{1}{p_{it}(x_{it},\hat{\beta})(1-p_{it}(x_{it},\hat{\beta})}\right), \\
\bar{A}_i(\hat{\beta}) &\equiv \frac{1}{T-1}\left[\frac{\partial^2 l_i(\hat{\beta})_{FEP}}{\partial\beta\partial\beta'}\right], \\
\textit{and}\quad \bar{\psi}_i(\hat{\beta},\hat{\theta}) &\equiv \frac{1}{T-1}\sum_{t=1}^{T-1}\Bigg[\nabla_\beta\Bigg(\hat{u}_{it}(\hat{\beta})\hat{D}_{n1} \\
&-\frac{\hat{u}_{it}(\hat{\beta})^2 - p_{it}(x_{it},\hat{\beta})(1-p_{it}(x_{it},\hat{\beta}))}{2}\hat{D}_{n2}\Bigg)\Bigg]
\end{aligned}$$

And the robust quasi-modified Cox test statictics, $\frac{\sqrt{N}\hat{T}_1}{\hat{V}_{T_1}^{1/2}}$, follows aymptotically unit normal under the null hypothesis.

Table 3.1: Summary Statistics: the Patents and lnR&D Data

	mean	s.d	median	Min	Max	proportion of zeros
Patents	37.133	72.642	6.000	0.000	515	0.000
lnR&D	1.415	1.947	1.196	-3.849	7.034	.

3.3 An Empirical Application

3.3.1 An Application to U.S. Patents and R&D Data

In this section, we estimate the FEP and the FENB models under the CML framework using the U.S. patents and R&D expenditures data and apply the quasi-modified Cox test between these two competing models.

We examined the dynamic specification properties from the data on U.S. patents and R&D expenditures from 1970 to 1979. We obtained this U.S. patents and R&D spending data set, patrhgh.txt, from the data directory in the NBER website. This data set is a subset of the patents and R&D data used in HHG (1986), "Patents and R&D: Is there a Lag?", IER 27: 265-283. There are a total of 346 firms and 22 firms (about 6.4% of all the firms) have zero patents during all time periods and we deleted these firms from our data because these observations do not contribute to the estimation.

Table 3.1 presents the summary statistics of the dependent variable, patents, and the explanatory variable, lnR&D.

Table 3.2: Estimation Results for the Patents Model: Linear Time Trend

Parameter	the Fixed Effects Poisson	the Fixed Effects Neg Bin
lnR&D	0.428 (0.038)	0.261 (0.090)
lnR&D_1	-0.159 (0.048)	-0.112 (0.115)
lnR&D_2	0.021 (0.044)	0.042 (0.103)
lnR&D_3	0.174 (0.041)	0.114 (0.098)
lnR&D_4	0.090 (0.039)	0.178 (0.092)
lnR&D_5	0.259 (0.030)	0.224 (0.068)
time	-0.083 (0.003)	-0.080 (0.010)
Sum of lnR&D	0.813	0.707
log likelihood	-6069.156	-3935.991
Skewness of residuals	0.141	0.207
Kurtosis of residuals	7.430	7.559
Probability of Normality	0.000	0.000

∗ The standard errors are in the parentheses.

Table 3.2 presents the estimation results for a patents model with linear time trend using the FEP and the FENB estimators. In this table, both estimation results indicate that the contemporaneous effect of lnR&D is significant. The sums of lnR&D are similar but the sum of lnR&D of the FEP is slightly bigger than that of the FENB. In the FEP estimator, the sum of lnR&D is 0.813 but it is 0.707 in the FENB estimator. And the time coefficient is -8.3 per cent per year in the FEP estimator and -8 per cent per year in the FENB estimator. Table 3.3 presents the estimation results for patents model with a full set of year dummies by the FEP and the FENB estimators.

Table 3.3: Estimation Results for the Patents Model: Full Set of Year Dummies

Parameter	the Fixed Effects Poisson	the Fixed Effects Neg Bin
lnR&D	0.407 (0.039)	0.245 (0.090)
lnR&D_1	-0.115 (0.058)	-0.087 (0.120)
lnR&D_2	0.061 (0.045)	0.063 (0.108)
lnR&D_3	0.111 (0.042)	0.084 (0.099)
lnR&D_4	0.073 (0.040)	0.165 (0.094)
lnR&D_5	0.279 (0.030)	0.238 (0.069)
year76	-0.044 (0.014)	-0.052 (0.038)
year77	-0.077 (0.014)	-0.105 (0.040)
year78	-0.238 (0.015)	-0.233 (0.041)
year79	-0.320 (0.015)	-0.309 (0.042)
Sum of lnR&D	0.816	0.708
log likelihood	-6042.707	-3934.719
Skewness of residuals	0.229	0.259
Kurtosis of residuals	7.290	7.437
Probability of Normality	0.000	0.000

* The standard errors are in the parentheses.

Table 3.4: Estimation Results for the Patents Model: Linear Time Trend

Parameter	the Fixed Effects Poisson	the Fixed Effects Neg Bin
lnR&D	0.826 (0.009)	0.694 (0.020)
Time	-0.065 (0.009)	-0.079 (0.020)
Time*lnR&D	-0.008 (0.002)	-0.005 (0.005)
Sum of lnR&D	0.826	0.694
log likelihood	-6273.494	-3981.590
Skewness of residuals	-0.245	-0.295
Kurtosis of residuals	36.014	35.173
Probability of Normality	0.000	0.000

* The standard errors are in the parentheses.

Table 3.5: Estimation Results for the Patents Model: Linear Time Trend Only

Parameter	the Fixed Effects Poisson	the Fixed Effects Neg Bin
lnR&D	0.800 (0.006)	0.679 (0.014)
Time	-0.097 (0.003)	-0.096 (0.010)
Sum of lnR&D	0.800	0.679
log likelihood	-6280.996	-3982.094
Skewness of residuals	-0.392	-0.394
Kurtosis of residuals	38.446	37.159
Probability of Normality	0.000	0.000

* The standard errors are in the parentheses.

In this table, only the contemporaneous effect of lnR&D is significant in both models except the last lag of lnR&D. In the FEP estimator, the sum of lnR&D is 0.816 but it is 0.708 in the FENB estimator. Table 3.4 presents the estimation results including only current lnR&D, time trend and the multiplication of these two variables. The coeffiecients of current lnR&D in both models are much higher than those in Table 3.2 and 3.3 but the sum of lnR&D are very similar. Table 3.5 presents the estimation results including current lnR&D and time trend only. The time trend coefficient for the FEP is -9.7 per cent and -9.6 per cent for the FENB. These coefficients are bigger in absolute value than those in Table 3.4. Not surprisingly, the standard errors in the FENB are much larger than those in the FEP and it is expected from the increased noise in the Negative Binomial specification.

Table 3.6: The quasi-modified Cox Test Results

	H_0:FEP vs. H_1:FENB	H_0:FENB vs. H_1:FEP
Nonrobust Cox test	-8.740	-6.318
Robust Cox test	-7.570	-6.531

* Test results from the patents model with Linear Time Trend

Table 3.7: The quasi-modified Cox Test Results

	H_0:FEP vs. H_1:FENB	H_0:FENB vs. H_1:FEP
Nonrobust Cox test	-0.725	-2.170
Robust Cox test	-0.706	-2.181

* Test results from the patents model with a full set of year dummies

3.3.2 The Quasi-Modified Cox Test Results

Our quasi-modified Cox test has been used to compare the correct specification between the FEP model and the FENB model.

Table 3.6 presents the quasi-modified Cox test results for the patents model with liner time trend. In this table, the nonrobust test results indicate that both models are rejected against the correctly specified model at any significance level. The Jarque-Bera test (probability of Normality) reveals that the residuals of both models are not distributed as normal in Table 3.2. And the robust quasi-modified Cox test results also reject both models to be correctly specified.

Table 3.7 presents the quasi-modified Cox test results for the patents model with a full set of year dummies. In this table, the nonrobust quasi-modified Cox test results show that the FENB model is rejected against the correct specification at the

Table 3.8: The quasi-modified Cox Test Results

	H_0:FEP vs. H_1:FENB	H_0:FENB vs. H_1:FEP
Nonrobust Cox test	-28.417	-6.678
Robust Cox test	-10.621	-7.043

* Test results from the patents model with linear time trend

5 per cen significance level but we fail to reject both models at 1 per cent significance level. The probability of normality in Table 3.3 indicates that the residuals from both models are not normally distributed.

The robust quasi-modified Cox test results are very close to the nonrobust test results and the FENB is rejected at 5 per cen significance level but both models failed to reject the null at a 1 per cent significance level. Table 3.8 presents the quasi-modified Cox test results for the patents model including only current lnR&D, time trend and the multiplication of these two variables. Both nonrobust and robust test results reveal that both models are rejected against the correctly specified model at any significance level. Table 3.9 presents the quasi-modified Cox test results for the patents model including current lnR&D and time trend only and shows that both models are aslo rejected at any significance level. Interestingly, including a full set of year dummies seems to play an important role to correct the model specification. We can suggest the role of time dummy with an example below. Suppose

$$
\begin{aligned}
Var(y_{it} \mid x_i, \phi_i, \gamma_t) &= \gamma_t \phi_i exp(x_{it}\beta) \\
&= \gamma_t \mu(x_{it}, \phi_i)
\end{aligned}
$$

Table 3.9: The quasi-modified Cox Test Results

	H_0:FEP vs. H_1:FENB	H_0:FENB vs. H_1:FEP
Nonrobust Cox test	-13.345	-10.088
Robust Cox test	-9.773	-11.437

* Test results from the patents model with linear time trend only

$$
\begin{aligned}
&= \phi_i exp(x_{it}\beta + \alpha_t) \\
where \quad \gamma_t &= exp(\alpha_t)
\end{aligned}
$$

If α_t is time dummy, then there is no more overdispersion problem when we include this time dummy in the model. Further, suppose γ_t is independent of (ϕ_i, x_i), then

$$
\begin{aligned}
Var(y_{it} \mid x_i, \phi_i) &= E\left[Var(y_{it} \mid x_i, \phi_i, \gamma_t) \mid x_i, \phi_i\right] \\
&\quad + Var\left[E(y_{it} \mid x_i, \phi_i, \gamma_t) \mid x_i, \phi_i\right] \\
&= \phi_i exp(x_{it}\beta) + \sigma_\gamma^2 \left[\phi_i exp(x_{it}\beta)\right]^2 \\
&> \phi_i exp(x_{it}\beta) = Var(y_{it} \mid x_i, \phi_i, \gamma_t)
\end{aligned}
$$

What we can infer from this example and possibly from the test results is that the overdispersion problem may be caused not by the distributional misspecification but by the parametric misspecification. And including this time dummy can correct the overdispersion problem and lead the Poisson model to be the correctly specified model.

3.4 Conclusion

In the count panel data models with unobserved effects, the FEP and the FENB models are frequently compared as two competing counterparts. The QMLE of the FEP is consistent if only the conditional mean is correctly specified but it is generally inefficient. On the contary, the QMLE of the FENB is not consistent unless the first two moments are correctly specified but it is more efficient than that of the FEP. Therefore, there is robustness and efficiency trade-off between these two models.

We applied the FEP and the FENB models to the U.S. patents and lnR&D expenditures relationship. The quasi-modified Cox test results indicate that including a full set of year dummies plays a major role for the correct model specification. When we include a full set of year dummies, the quasi-modified Cox test results become different from those without a full set of time dummies and both the FEP and the FENB models fail to be rejected against the correctly specified model at 1 per cent significance level while both models are rejected at any significance level if we include the linear time trend instead. We can conjecture from these test results that the overdispersion problem may not be a matter of the distributional misspecification but a matter of parametric misspecification. The further study is needed to find more correctly speicified model for the nonlinear count or continuous panel data model with unobserved effects.

Appendix A

Modified Cox test

Here we derive the modified Cox test. The original Cox test statistic can be written in terms of the information set available at time t as

$$\begin{aligned}\hat{T}_{M_1} &= \frac{1}{T}\sum_{t=1}^{T}\{\log f_t(y_t \mid I_{t-1};\theta_0) - \log g_t(y_t \mid I_{t-1};\delta^*)\} \\ &\quad -E_{M_1}[\frac{1}{T}\sum_{t=1}^{T}\{\log f_t(y_t \mid I_{t-1};\theta_0) - \log g_t(y_t \mid I_{t-1};\delta^*)\} \mid I_{t-1}] \quad \text{(A.1)}\end{aligned}$$

where θ_0 is the MLE under M_1 and δ^* is the MLE under M_2 when M_1 is correctly specified. Now we decompose the equation (A.1) by two terms and rewrite these two terms as

i) The first term;

$$\sum_{t=1}^{T}\{\log f_t(y_t \mid I_{t-1};\theta_0) - \log g_t(y_t \mid I_{t-1};\delta^*)\} \quad \text{(A.2)}$$

$$\log f_t(y_t \mid I_{t-1};\theta_0) = -\frac{1}{2}\log 2\pi - \frac{1}{2}\log h_t(\theta_0) - \frac{1}{2}\frac{(y_t - m_t(\theta_0))^2}{h_t(\theta_0)} \quad \text{(A.3)}$$

$$\log g_t(y_t \mid I_{t-1};\delta^*) = -\frac{1}{2}\log 2\pi - \frac{1}{2}\log \eta_t(\delta^*) - \frac{1}{2}\frac{(y_t - \mu_t(\delta^*))^2}{\eta_t(\delta*)} \quad \text{(A.4)}$$

From now on we define $\log f_t(y_t \mid I_{t-1};\theta_0) = \log f_t$, $\log g_t(y_t \mid I_{t-1};\delta^*) = \log g_t$, $h_t(\theta_0) = h_t$, $\eta_t(\delta^*) = \eta_t$, $m_t(\theta_0) = m_t$, and $\mu_t(\delta^*) = \mu_t$ for notational simplicity.

$$\log f_t - \log g_t = \frac{1}{2}\log \eta_t - \frac{1}{2}\log h_t - \frac{1}{2}\left[\frac{(y_t - m_t)^2}{h_t} - \frac{(y_t - \mu_t)^2}{\eta_t}\right] \quad \text{(A.5)}$$

$$so, \sum_{t=1}^{T}\{\log f_t - \log g_t\} = -\frac{1}{2}\sum_{t=1}^{T}\left[\frac{(y_t - m_t)^2}{h_t} - \frac{(y_t - \mu_t)^2}{\eta_t}\right] - \frac{T}{2}\{\log h_t - \log \eta_t\} \quad \text{(A.6)}$$

ii) The second term;

$$E_{M_1}\left[\sum_{t=1}^{T}\{\log f_t - \log g_t\} \mid I_{t-1}\right] = E_{M_1}\left[\frac{T}{2}\log \eta_t - \frac{T}{2}\log h_t + \frac{1}{2}\sum_{t=1}^{T}\{\frac{(y_t - \mu_t)^2}{\eta_t} - \frac{(y_t - m_t)^2}{h_t} \mid I_{t-1}\}\right] \quad \text{(A.7)}$$

$$= \frac{T}{2}\log \eta_t - \frac{T}{2}\log h_t + \frac{1}{2}\sum_{t=1}^{T} E_{M_1}\left\{\frac{(y_t - \mu_t)^2}{\eta_t} - \frac{(y_t - m_t)^2}{h_t} \mid I_{t-1}\right\} \quad \text{(A.8)}$$

$$= \frac{T}{2}\log \eta_t - \frac{T}{2}\log h_t - \frac{T}{2} + \frac{1}{2}\sum_{t=1}^{T} E_{M_1}\left\{\frac{(y_t - \mu_t)^2}{\eta_t} \mid I_{t-1}\right\} \quad \text{(A.9)}$$

Because

$$\left(\begin{array}{rcl} E_{M_1}\{(y_t - m_t)^2 \mid I_{t-1}\} & = & E_{M_1}\{u_t^2 \mid I_{t-1}\} \\ & = & h_t \end{array}\right.$$

$$= \frac{T}{2}\log \eta_t - \frac{T}{2}\log h_t - \frac{T}{2}$$

$$+\frac{1}{2}\sum_{t=1}^{T} E_{M_1}\left\{\frac{(y_t-m_t)^2+(m_t-\mu_t)^2}{\eta_t} \mid I_{t-1}\right\} \tag{A.10}$$

Because

$$\left(\begin{array}{ll}
(y_t-\mu_t)^2 & =(y_t-m_t+m_t-\mu_t)^2 \\
 & =(y_t-m_t)^2+(m_t-\mu_t)^2 \\
 & +2(y_t-m_t)(m_t-\mu_t) \\
 & =(y_t-m_t)^2+(m_t-\mu_t)^2+2u_t(m_t-\mu_t) \\
E_{M_1}\{(y_t-\mu_t)^2 \mid I_{t-1}\} & =E_{M_1}\{(y_t-m_t)^2 \mid I_{t-1}\} \\
 & +E_{M_1}\{(m_t-\mu_t)^2 \mid I_{t-1}\} \\
 & +2E_{M_1}\{(y_t-m_t)(m_t-\mu_t) \mid I_{t-1}\}, \quad and \\
E_{M_1}\{(y_t-m_t)(m_t-\mu_t) \mid I_{t-1}\} & =E_{M_1}\{u_t(m_t-\mu_t) \mid I_{t-1}\} \\
 & =E_{M_1}\{u_t \mid I_{t-1}\}E_{M_1}\{m_t-\mu_t \mid I_{t-1}\} \\
 & =0 \\
so, \quad E_{M_1}\{(y_t-\mu_t)^2 \mid I_{t-1}\} & =E_{M_1}\{(y_t-m_t)^2 \mid I_{t-1}\} \\
 & +E_{M_1}\{(m_t-\mu_t)^2 \mid I_{t-1}\} \\
 & =h_t+(m_t-\mu_t)^2
\end{array}\right.$$

$$= \frac{T}{2}\log\eta_t-\frac{T}{2}\log h_t-\frac{T}{2}+\frac{1}{2}\sum_{t=1}^{T}\frac{h_t}{\eta_t}+\frac{1}{2}\sum_{t=1}^{T}\frac{(m_t-\mu_t)^2}{\eta_t} \tag{A.11}$$

Now we combine these two terms back together, and we produce

$$\begin{aligned}
= & -\frac{T}{2}\{\log h_t-\log\eta_t\}-\frac{1}{2}\sum_{t=1}^{T}\frac{(y_t-m_t)^2}{h_t}+\frac{1}{2}\sum_{t=1}^{T}\frac{(y_t-\mu_t)^2}{\eta_t} \\
& -\left\{\frac{T}{2}\log\eta_t-\frac{T}{2}\log h_t\right\}+\frac{T}{2}-\frac{1}{2}\sum_{t=1}^{T}\frac{h_t}{\eta_t}-\frac{1}{2}\sum_{t=1}^{T}\frac{(y_t-\mu_t)^2}{\eta_t} & \text{(A.12)} \\
= & \frac{T}{2}-\frac{1}{2}\left\{\sum_{t=1}^{T}\frac{h_t}{\eta_t}+\sum_{t=1}^{T}\frac{(m_t-\mu_t)^2}{\eta_t}+\sum_{t=1}^{T}\frac{(y_t-m_t)^2}{h_t}-\sum_{t=1}^{T}\frac{(y_t-\mu_t)^2}{\eta_t}\right\} & \text{(A.13)} \\
= & \frac{T}{2}-\frac{1}{2}\sum_{t=1}^{T}\frac{u_t^2}{\eta_t}+\sum_{t=1}^{T}\left\{\frac{h_t+(m_t-\mu_t)^2-(y_t-\mu_t)^2}{\eta_t}\right\} & \text{(A.14)} \\
= & \frac{T}{2}-\frac{1}{2}\sum_{t=1}^{T}\frac{u_t^2}{h_t}+\sum_{t=1}^{T}\left\{\frac{(y_t-\mu_t)^2-h_t-(m_t-\mu_t)^2}{2\eta_t}\right\} & \text{(A.15)}
\end{aligned}$$

On the other hand, we can rewrite the numerator of the third term as

$$\begin{cases} (y_t - \mu_t)^2 & = (m_t + u_t - \mu_t)^2 \\ & = (m_t - \mu_t + u_t)^2 \\ & = (m_t - \mu_t)^2 + u_t^2 + 2u_t(m_t - \mu_t) \\ so, (y_t - \mu_t)^2 - h_t - (m_t - \mu_t)^2 & = (m_t - \mu_t)^2 + u_t^2 - 2u_t(m_t - \mu_t) \\ & -h_y - (m_t - \mu_t)^2 \\ & = u_t^2 - h_t - 2u_t(m_t - \mu_t) \end{cases}$$

$$= \frac{T}{2} - \frac{1}{2}\sum_{t=1}^{T}\frac{u_t^2}{h_t} + \sum_{t=1}^{T}\frac{u_t(m_t - \mu_t)}{\eta_t} + \sum_{t=1}^{T}\frac{u_t^2 - h_t}{2\eta_t} \tag{A.16}$$

$$= \frac{T}{2} - \sum_{t=1}^{T}\frac{u_t^2 - h_t + h_t}{2\eta_t} + \sum_{t=1}^{T}\frac{u_t(m_t - \mu_t)}{\eta_t} + \sum_{t=1}^{T}\frac{u_t^2 - h_t}{2\eta_t} \tag{A.17}$$

$$= \frac{T}{2} - \sum_{t=1}^{T}\frac{u_t^2 - h_t}{2h_t} - \sum_{t=1}^{T}\frac{h_t}{2h_t} + \sum_{t=1}^{T}\frac{u_t(m_t - \mu_t)}{\eta_t} + \sum_{t=1}^{T}\frac{u_t^2 - h_t}{2\eta_t} \tag{A.18}$$

$$= \frac{T}{2} - \frac{T}{2} + \frac{1}{2}\sum_{t=1}^{T}(u_t^2 - h_t)\left(\frac{1}{\eta_t} - \frac{1}{h_t}\right) + \sum_{t=1}^{T}(y_t - m_t)\frac{m_t - \mu_t}{\eta_t} \tag{A.19}$$

Therefore, the modified Cox test is derived as

$$\begin{aligned} \hat{T}_{M_1} &= \frac{1}{T}\sum_{t=1}^{T}\{\log f_t(y_t \mid I_{t-1};\theta_0) - \log g_t(y_t \mid I_{t-1};\delta^*)\} \\ &\quad -E_{M_1}[\frac{1}{T}\sum_{t=1}^{T}\{\log f_t(y_t \mid I_{t-1};\theta_0) - \log g_t(y_t \mid I_{t-1};\delta^*)\} \mid I_{t-1}] \\ &= \frac{1}{T}\sum_{t=1}^{T}\left\{(y_t - m_t)\frac{(m_t - \mu_t)}{\eta_t} + \frac{u_t^2 - h_t}{2}\left(\frac{1}{\eta_t} - \frac{1}{h_t}\right)\right\} \end{aligned} \tag{A.20}$$

Appendix B

Regularity Conditions

Suppose $y_t, t = 1, \cdots, T$ is a vector of *i.i.d* observations and we wish to compare when y_t has the density function $f(y_t, \theta)$ for some θ in Θ under the null hypothesis, H_f, and when y_t has the density function $g(y_t, \delta)$ for some δ in Δ under the alternative hypothesis, H_g. Then let θ_0 denote the true value of θ under H_f, let $\hat{\theta}$ be the MLE of θ_0 and let δ^* denote the value that $\hat{\delta}$, the QMLE of δ, converges to. For notational brevity, we state the regularity conditions in terms of $f(y, \theta)$ but these conditions are also applicable to $g(y, \delta)$ as well. Below are the regularity conditions for the existence and the consistency of QMLE(White, 1982).

1. The sequence of *i.i.d* observations $y_t, t = 1, \cdots, T$ have common joint distribution function G on Ω with measurable *Radon-Nikodým* density $g = dG/dv$.

2. *Radon-Nikodým* density $f(y, \theta) = dF(y, \theta)/dv$ where $F(y, \theta)$ is the family of distribution function is measurable in y for every θ in Θ, a compact subset of a $p-$dimensional *Euclidean* space, and continuous in θ for every y in Ω.

3. a) $| \log f(y, \theta) | \leq m(g)$ for all θ in Θ, where m is integrable with respect to G.

   b) $E(\log f(y_t, \theta))$ has a unique maximum at θ in Θ.

4. $\partial \log f(y, \theta)/\partial\theta_i, i = 1, \cdots, p$, are a measurable function of y for each θ in Θ and a continuously differentiable function of θ for each y in Ω.

5. $| \; \partial^2 \log f(y, \theta)/\partial\theta_i \cdot \partial\theta_j \; |$ and $| \; \partial \log f(y, \theta)/\partial\theta_i \cdot \partial \log f(y, \theta)/\partial\theta_j \; |, i, j = 1, \cdots, p$, are dominated by functions integrable with respect to G for all y in Ω and θ in Θ.

6. Define $A(\theta) \equiv \{E(\partial^2 \log f(y, \theta)/\partial\theta_i \cdot \partial\theta_j)\}$,

   and $B(\theta) \equiv \{E(\partial \log f(y, \theta)/\partial\theta_i \cdot \partial \log f(y\theta)/\partial\theta_j)\}$,

   a) θ is interior to Θ,

   b) $A(\theta)$ and $B(\theta)$ are nonsingular.

Under these conditions, $\sqrt{T}(\hat{\theta} - \theta_0)$ is asymptotically normally distributed.

Bibliography

[1] Bera, A. and Higgins, M., 1997, *ARCH and Bilinearity as Competing Models for Nonlinear Dependence*, Journal of Business & Economic Statistics, 15, 43-50

[2] Bollerslev,T.,1986,*Generalized Autoregressive Conditional Heteroscedasticity*,Journal of Econometrics, 31, 307-327

[3] Bollerslev, T. and Wooldridge, J. M., 1992, *Quasi-Maximum Likelihood Estimation and Inference in Dynamic Models With Time-Varying Covariances*, Econometric Reviews, 11, 143-172

[4] Kuan, C. and Chen, Y., 2002, *The pseudo-True Score Encompassing Test for Non-Nested Hypotheses*, Journal of Econometrics, vol. 106, 271-195

[5] Cox, D.R.,1961,*Tests of Separate Families of Hypotheses*, in Proceedings of the 4^{th} Berkeley Symposium (1), Berkeley: University of California Press, 105-123

[6] Cox, D.R.,1962, *Further Results on Tests of Separate Families of Hypothesis*, Journal of the Royal Statistical Society, Series B, 24, 406-424

[7] Davidson, R. and Mackinnon, J.G.,1981,*Several Tests for Model Specification in the Presence of Alternative Hypotheses*, Econometrica, 49, 781-793

[8] Davidson, R. and Mackinnon, J.G.,1984, *Model Specification Tests Based on Artificial Linear Regressions*, International Economic Review, 25, 485-502

[9] Engle, R.,1982, *Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation*, Econometrica, 50, 987-1008

[10] Fisher, G.R. and McAleer, M., 1979, *On the Interpretation of the Cox test in Econometrics*, Economics Letter, 4, 145-150

[11] Gourieroux, C., Monfort, A., and Trognon, A., 1983, *Testing Nested or Non-Nested Hypotheses*, Journal of Econometrics, 21, 83-115

[12] Gourieroux, C. and Monfort, A., 1994, *Testing Non-Nested Hypotheses*,in: R.F.Engle and D.L.McFadden, Eds., Elsevier Science, Handbook of Econometrics, vol.4, 2583-2633

[13] Granger, C.W. and Anderson, A.P.,1978, *An Introduction to Bilinear Time Series Models*, Göttingen:Vandenhoeck & Ruprecht

[14] Hausman, J., Hall, B.H., and Griliches, Z., 1984, *Econometric Models for Count Data with an Application to the Patents-R & D Relationship*, Econometrica, vol. 52, 909-938

[15] Lumsdaine, R.L., 1991, *Essays on Time Series Econometrics*, unpublished Ph.D. dissertation, Harvard University, Dept. of Economics

[16] Lumsdaine, R.L., 1995, *Finite-Sample Properties of the Maximum Likelihood Estimator in GARCH(1,1) and IGARCH(1,1) Models: A Monte Carlo Investigation*, Journal of Business & Economic Statistics, vol. 13, 1-10

[17] Mizon, G.E. and Richard, J., 1986*The Encompassing Principle and its Applications to Testing Non-Nested Hypothesis*, Econometrica, 54, 657-678

[18] Pesaran, M.H. and Deaton, A.S., 1978, *Testing Non-Nested Nonlinear Regression Models*, Econometrica, 46, 677-694

[19] Pesaran, M.H. and Pesaran, B., 1993, *A Simulation Approach to the Problem of Computing Cox's Statistic for Testing Non-Nested Models*, Journal of Econometrics, 57, 377-392

[20] Quandt, R.E., 1974, *A Comparison of Methods for Testing Non-Nested Hypotheses*, Review of Economics and Statistics, 56, 92-99

[21] White, H., 1982, *Regularity Conditions for Cox's Test of Non-Nested Hypotheses*, Journal of Econometrics, 19, 307-327

[22] White, H.,1994, *Estimation, Inference and Specification Analysis*, Econometric Society Monographs: Cambridge University Press

[23] Wooldridge, J.M., 1990, *An Encompassing Approach to Conditional Mean Tests with Applications to Testing Non-Nested Hypotheses*, Journal of Econometrics, 45, 331-350

[24] Wooldridge, J.M., 1990, *A Unified Approach to Robust, Regression-Based Specification Tests*, Economics Theory, 6, 17-43

[25] Wooldridge, J.M., 1999, *Distribution-free Estimation of Some Nonlinear Panel Data Models*, Journal of Econometrics, 90, 77-97