計量經濟學報, 제16권 제2호(2005. 6. 30.), 27-45

A Modified Cox Test for Time Series Models

Donggeun Kim*

We propose a new approach based on conditional means and variances to avoid the computational difficulties of the traditional Cox test. This approach can be extended to more complicated time series models. Monte Carlo experiments are performed to investigate the potential applicability of the proposed test. Empirical applications to two different non-linear error equation models are also examined.

Keywords: Cox Test, Conditional Mean, Conditional Variance,
Modified Cox Test, Non-Linear Error Equation Models.
JEL Classifications: C1

I. Introduction

Since testing non-nested specification models was pioneered by Cox (1961, 1962) as a modification of the Neyman-Pearson maximum-likelihood ratio, it has been one of the main interests among econometricians. However, the application of the non-nested Cox test has been restricted to rather simple linear or non-linear regression models because of mainly its complicated and, in many cases, intractable derivation of the pseudo-true value in the second component of the Cox test. [see, for example, Pesaran and Deaton (1978), Gourieroux,

* Assistant Professor, Department of Economics, Ajou University,
E-mail: kimdongg@ajou.ac.kr

투고일: 2004. 11. 2 심사일: 2004. 11. 2 최종심사완료일: 2005. 6. 2

Monfort, and Trognon (1983), and Mizon and Richard (1986)]. The maximum likelihood estimate (MLE) of a non-linear model does not have a closed form and obtaining the analytical derivation of the pseudo-true value and its finite sample estimation may not be possible.

There have been alternative studies to avoid these computational difficulties. To examine these, see the artificial nesting model of Davidson and Mackinnon (1981) and the method of stochastic simulation by Pesaran and Pesaran (1993) among them. Especially, Bera and Higgins (1997) [BH, hereafter] apply the stochastic simulation method to the non-linear equation error models of the autoregressive conditional heteroscedasticity (ARCH) and the bilinear models. BH assume conditional normality in each model. However, applying the traditional Cox test in time series models and possibly in dynamic panel data models is very challenging because potentially very severe computational efforts arise from computing the unconditional expectation of the differenced log likelihood functions.

In this paper we reexamine the problems of applying Cox test to univariate time series models and propose a new approach based on the first two conditional moments given all past information. Like BH, we use a Gaussaian log-likelihood function. We think this approach improves on BH approach in several ways. First, our approach is computationally simpler than the BH approach because we apply the Cox logic not to the joint distribution of all T observations, but to the sequence of conditional distribution. Second, our approach can be more robust than the BH approach because we focus on the conditional density for each time period. Also, we can test other distributional features such as conditional variance because our approach uses the first two conditional moments.

The plan of this paper is as follows. In section 2 we briefly review of the traditional Cox test, describe how to modify and apply our approach to a sequence of conditional distributions and obtain tests under normality condition. In section 3 we provide Monte Carlo experiments of this modified Cox test. ln section 4 we apply our proposed tests to three time series data sets studies by BH and we draw conclusions in section 5.

II. A Modified Cox Test for Time Series

1. Motivation and General Concepts

Suppose that $y_t, t=1,2,\cdots,T$ are T individually, identically distributed random variables. Let $M_1: f(y_t, \alpha)$ and $M_2: g(y_t, \beta)$ denote two competing probability density functions, where α, β are their unknown parameters, respectively. Let the null hypothesis be that M_1 is correctly specified and the alternative hypothesis be that M_2 is correctly specified. If M_1 is not nested in M_2, and M_2 is not nested in M_1, then it is said that the two hypotheses are non-nested. Under the null hypothesis, the traditional Cox test is

$$T_f = \{L_f(\hat{\alpha}) - L_g(\hat{\beta})\} - E_{\hat{\alpha}}\{L_f(\hat{\alpha}) - L_g(\hat{\beta})\} \tag{1}$$

where $L_f(\hat{\alpha}), L_g(\hat{\beta})$ are the maximized log likelihood functions under M_1 and M_2, respectively and $\hat{\alpha}, \hat{\beta}$ are their maximum log likelihood estimators. The test statistic is based on the difference between the log likelihood ratio and its expected estimate under the null hypothesis. 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 non-zero under non-nested hypotheses. So the Cox test takes the deviation between the maximum log likelihood ratio and its expected value under the null hypothesis. If M_1 is correctly specified model, T_f should be close to zero while a large deviation from zero constitutes evidence against the null hypothesis. The standardized Cox test statistic, $\frac{\sqrt{T}\, T_f}{\hat{V}_f^{1/2}}$, where $\hat{V}_f$ is a consistent estimator of the asymptotic variance of $\sqrt{T} T_f$, is asymptotically distributed as N(0,1). [see, White (1982)]

What makes difficult to apply the Cox test is that it requires computing the unconditional expectation of the differenced log likelihood functions under the null hypothesis that is not significantly analytical or tractable in many cases. White (1994) shows the difficulties of computing the expected value of the average log likelihood ratio under the null hypothesis.

$$E_{M_1}[\widehat{L}_{f_n} - \widehat{L}_{g_n}] \equiv \frac{1}{T}\int (\log f_T(y^T, \hat{\alpha}) - \log g^T(y^T, \hat{\beta}))f^T(y^T, \hat{\alpha})dv^T(y^T) \tag{2}$$

$$= \int \frac{1}{T}\sum_{t=1}^{T}(\log f_t(y_t, \hat{\alpha}) - \log g_t(y^t, \hat{\beta}))\prod_{t=1}^{T} f_t(y^t, \hat{\alpha})dv^T(y^T) \tag{3}$$

$$= \frac{1}{T}\sum_{t=1}^{T}[\int (\log f_t(y^T, \hat{\alpha}) - \log g_t(y^T, \hat{\beta}))f_t(y^t, \hat{\alpha})dv^t(y^t)] \tag{4}$$

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

The vt-fold integration in equation (4) causes severe computational difficulties of the unconditional expectation. Cox (1961, 1962) assumed that observations are independent. This reduces the computational difficulties in some degree and the vt-fold integral in equation (4) is reduced as v-fold integral

$$= \frac{1}{T}\sum_{t=1}^{T}[\int (\log f_t(y_t, \hat{\alpha}) - \log g_t(y_t, \hat{\beta}))f_t(y_t, \hat{\alpha})dv_t(y_t)] \tag{5}$$

However, the computational effort still remains, even though it is much more tractable than before. In addition, the analytical intractability outlasts when we apply the Cox test to the time series applications that include the lagged dependent variables as conditioning variables. To avoid these difficulties, we propose an alternative approach to computing the conditional expectation for each time t. By doing so, we can achieve some substantial simplifications of the Cox test in some important applications including ARCH and GARCH in time series.

2. A Modified Cox Test

Let $\{y_t, z_t;\ t = 1, 2, \cdots, T\}$ be a sequence of observable random variable with $y_t 1 \times J$, and $z_t 1 \times K$; y_t is the vector of dependent variables, and z_t is the vector of explanatory variables. Define $x_t = (z_t, y_{t-1}, z_{t-1}, \cdots, z_1, y_1)$ for each t. We assume that the conditional distribution, $D(y_t | x_t)$, follows a normal distribution. Suppose the two competing non-nested parametric models f_t and g_t representing $D(y_t | x_t)$ are given as

$$M_1: f_t(y_t | x_t; \Theta), \Theta \in \Theta \subseteq R^p, \ t = 1, 2, \cdots, T \quad (6)$$

$$M_2: g_t(y_t | x_t; \delta), \delta \in \Delta \subseteq R^q, \ t = 1, 2, \cdots, T \quad (7)$$

Let $\widehat{\Theta}$, $\widehat{\delta}$ be the maximum likelihood estimates that maximize $\frac{1}{T} \sum_{t=1}^{T} log[f_t(y_t | x_t; \Theta)]$ and $\frac{1}{T} \sum_{t=1}^{T} log[g_t(y_t | x_t; \delta)]$ respectively[1].

Under standard regularity conditions, see, for example, Wooldridge (1994), the conditional MLE is consistent and $\sqrt{T}$-asymptotically normal. If M_1 is a correctly specified model, $\widehat{\Theta}$ is consistent for Θ_0, a true value of Θ, and asymptotically normal but $\widehat{\delta}$ generally converges not to its true value δ_0, but to some value δ^*, the "pseudo-true" value of δ that depends on Θ_0. Define

$$C_t(x_t; \Theta_0, \delta^*) \equiv E_{M_1}[\, log f_t(y_t | x_t; \Theta_0) - log g_t(y_t | x_t; \delta^*) | x_t] \quad (8)$$

Using equation (8) we can derive a modified Cox test

$$T_{M_1} = \frac{1}{T} \sum_{t=1}^{T} [\, (log f_t(y_t | x_t : \Theta_0) - log g_t(y_t | x_t; \delta^*)) - C_t(x_t; \Theta_0, \Theta^*)] \quad (9)$$

We can express equation (9) more formally when we parameterize the conditional means and variances of the two competing models, M_1 and M_2, by $\Theta_0 \in \Theta$ and $\delta_0 \in \Delta$. Assume that we have two competing models for conditional mean and variance:

$$M_1: E(y_t | x_t) = m_t(\Theta_0) \quad (10)$$

$$Var(y_t | x_t) = h_t(\Theta_0) \ where \ \ u_t(\Theta_0) \equiv y_t - m_t(\Theta_0) \quad (11)$$

and

$$M_2: E(y_t | x_t) = \mu_t(\delta_0) \quad (12)$$

1) We assume that the conditions for each process to be independent are satisfied.

$$Var(y_t|x_t) = \eta_t(\Theta_0) \quad where \quad \varepsilon_t(\delta_0) \equiv y_t - \mu_t(\delta_0) \tag{13}$$

Note that, for M_1 and M_2, the first two conditional moments in each model depend not on all elements of x_t, but on all of $(z_t, y_{t-1}, \cdots, y_1, z_1)$. Under normality, $logf_t(y_T|x_t;\Theta)$ and $logg_t(y_t|x_t;\delta)$ are

$$logf_t(y_t|x_t;\Theta) = -\frac{1}{2}log2\pi - \frac{1}{2}logh_t(\Theta) - \frac{1}{2}\frac{(y_t - m_t(\Theta))^2}{h_t(\Theta)} \tag{14}$$

$$logg_t(y_t|x_t;\delta) = -\frac{1}{2}log2\pi - \frac{1}{2}log\eta_t(\delta) - \frac{1}{2}\frac{(y_t - \mu_t(\delta))^2}{\eta_t(\delta)} \tag{15}$$

Plug equation (14) and (15) into equation (9) and replace the unknown parameters with the estimates, then the modified Cox test has a form of

$$\begin{aligned}\widehat{T}_{M_1} = \frac{1}{T}\sum_{t=1}^{T}[(y_t - m_t(\widehat{\Theta}))\frac{m_t(\widehat{\Theta}) - \mu_t(\widehat{\delta})}{\eta_t(\widehat{\delta})} \\ + \frac{u_t(\widehat{\Theta})^2 - h_t(\widehat{\Theta})}{2}(\frac{1}{\eta_t(\widehat{\delta})} - \frac{1}{h_t(\widehat{\Theta})})]\end{aligned} \tag{16}$$

A similar procedure holds for T_{M_2} when the null hypothesis is that M_2 is correctly specified but we do not cover it here. Furthermore, we prove the validity of all the inference procedures under the null hypothesis that M_1 is correctly specified.

Next, following Wooldridge (1990), Bolleslev and Wooldridge (1992), and White (1994), we can derive a test statistic and find the asymptotic variance of $\sqrt{T}\widehat{T}_{M_1}$ fairly easily. First, replace all estimates with the population parameters except $u_t(\widehat{\Theta})$ and $u_t(\widehat{\Theta})^2 - h_t(\widehat{\Theta})$ in equation (16). Second, apply mean-value expansion argument and replace the mean values with their probability limits, θ_0, and multiply $\sqrt{T}$ on both sides,[2] then

2) $\sqrt{T}\widehat{T}_{M_1}$ depends on both $\widehat{\Theta}$ and $\widehat{\delta}$ but only the asymptotic distribution of $\sqrt{T}(\widehat{\Theta} - \Theta_0)$ affects the limiting distribution of $\sqrt{T}\widehat{T}_{M_1}$.

$$\sqrt{T}\,\widehat{T}_{M_1} = \sqrt{T}\,T_{M_1}(\Theta_0, \delta^*) - \frac{1}{T}\sum\left\{E\left[\left\{\frac{m_t(\Theta_0)-\mu_t(\delta^*)}{\eta_t(\delta^*)}\right\}\nabla_\Theta m_t(\Theta_0)\right.\right.$$
$$+\frac{1}{2}E\left[\left(\frac{1}{\eta_t(\delta^*)}-\frac{1}{h_t(\Theta_0)}\right)\nabla_\Theta h_t(\Theta_0)\right]\sqrt{T}(\widehat{\Theta}-\Theta_0)+o_p(1) \qquad (17)$$

where ∇_Θ is the gradient operator.

Since $\widehat{\Theta}$ is the MLE using Gaussian log-likelihood function, with suitable regularity condition, $\sqrt{T}(\widehat{\Theta}-\Theta_0) = -A_T^0(\Theta_0)^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T} s_t(\Theta_0)+op(1)$,

where $A_T^0(\Theta_0)\equiv \lim_{t\to\infty}\frac{1}{T}E[h_t(\Theta_0)]$, the limit of the expected value of the hessian of $logf_t(y_t|x_t;\Theta_0)$, and $s_t(\Theta_0)\equiv\nabla_\Theta logf_t(y_t|x_t;a_0)$. [see Bollerslev and Wooldridge (1992) and White (1994)]

Then, we can establish the first-order representation

$$\sqrt{T}\,\widehat{T}_{M_1} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[u_t(\Theta_0)\left\{\frac{m_t(\Theta_0)-\mu_t(\delta^*)}{\eta_t(\delta^*)}\right\}+\frac{v_t(\Theta_0)}{2}\left(\frac{1}{\eta_t(\delta^*)}-\frac{1}{h_t(\Theta_0)}\right)\right.$$
$$\left.+\Psi_T^0(\Theta_0,\delta^*)A_T^0(\Theta_0)^{-1}s_t(\Theta_0)\right]+o_p(1) \qquad (18)$$

where

$$v_t(\Theta_0)=\frac{u_t(\Theta_0)^2-h_t(\Theta_0)}{2} \quad \text{and}$$

$$\Psi_T^0(\Theta_0,\delta^*)= \frac{1}{T}\sum_{t=1}^{T}\left\{E\left[\left\{\frac{m_t(\Theta_0)-\mu_t(\delta^*)}{\eta_t(\delta^*)}\right\}\nabla_\Theta m_t(\Theta_0)\right]\right.$$
$$\left.+\frac{1}{2}E\left[\left(\frac{1}{\eta_t(\delta^*)}-\frac{1}{h_t(\Theta_0)}\right)\nabla_\Theta h_t(\Theta_0)\right]\right\}$$

Note that the elements in the summand in equation (18) are martingale difference sequence random variables. Thus, $\sqrt{T}\,\widehat{T}_{M_1}$ is asymptotically normally distributed with mean zero and its asymptotic variance,

$$V(\Theta_0,\delta^*) = \frac{1}{T}\sum_{t=1}^{T}E[D_t+\Psi_T^0(\Theta_0,\delta^*)A_T^0(\Theta_0)^{-1}s_t(\Theta_0)]^2 \qquad (19)$$

$$where\ D_t \equiv [\,u_t(\Theta_0)(\frac{m_t(\Theta_0)-\mu_t(\delta^*)}{\eta_t(\delta^*)}) + \frac{v_t(\Theta_0)}{2}(\frac{1}{\eta_t(\delta^*)} - \frac{1}{h_t(\Theta_0)})]$$

We can estimate this asymptotic variance by removing the expected value and replacing Θ_0 and δ^* with their MLEs:

$$\widehat{V}_T = \frac{1}{T}\sum_{t=1}^{T}[\,\widehat{h}_t\widehat{D}_{t1}^2 + \frac{\widehat{h}_t^2}{4}\widehat{D}_{t2}^2 + \widehat{\Psi}_T\widehat{A}_T^{-1}\widehat{\Psi}'_T] \tag{20}$$

$$where\ \widehat{D}_{t1} = \frac{m_t(\widehat{\Theta}) - \mu_t(\widehat{\delta})}{\eta_t(\widehat{\delta})} \text{ and } \widehat{D}_{t2} = \frac{1}{\eta_t(\widehat{\delta})} - \frac{1}{h_t(\widehat{\Theta})}$$

Note that only the conditional third and fourth moments under normality are used to obtain this result.

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

- Proposition

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

1. Regularity conditions in Wooldridge (1994) hold.
2. $y_t|x_t$ is normally distributed.
3. Conditional mean and conditional variance exist and are estimated by the Gaussian MLE.

Then,

$$\widehat{T}_{M_1} = \frac{1}{T}\sum_{t=1}^{T}[\,(y_t - m_t(\widehat{\Theta}))\frac{m_t(\widehat{\Theta}) - \mu_t(\widehat{\delta})}{\eta_t} + \frac{u_t(\widehat{\Theta})^2 - h_t(\widehat{\Theta})}{2}(\frac{1}{\eta_t(\widehat{\delta})} - \frac{1}{h_t(\widehat{\Theta})})] \tag{21}$$

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

Note that the equation (21) is a function of $\widehat{\Theta}$ and $\widehat{\delta}$, the MLEs of Θ_0 and δ^*, 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 on conditional mean and variance specifications, is a more convenient method for a computational purpose. Now we apply this proposition as follows;

- Procedure

1. Obtain $\widehat{\Theta}$ and $\widehat{\delta}$, the ML estimates of Θ_0 and δ^*, save residuals, $\widehat{u}_t$, and the conditional variance, $\widehat{h}_t$ from the log likelihood function $\log f(y_t | x_t; \widehat{\Theta})$ and $\widehat{e}_t$, and $\widehat{\eta}_t$ from the log likelihood function $\log g_t(y_t | x_t; \widehat{\delta})$.
2. Compute $\widehat{D}_{t1}$, $\widehat{D}_{t2}$, $\widehat{v}_t$, $\widehat{s}_t$, $\widehat{\Psi}_t$, and $\widehat{A}_T$.
3. Compute

$$\sqrt{T}\widehat{T}_{M_1} = T^{-1/2}\sum_{t=1}^{T}\left[\widehat{u}_t\widehat{D}_{t1} + \frac{\widehat{v}_t}{2}\widehat{D}_{t2} + \widehat{\Psi}_T\widehat{A}_T^{-1}\widehat{s}_t\right] \text{ and}$$

$$\widehat{V}_{M_1} = T^{-1}\sum_{t=1}^{T}\left[\widehat{h}_t\widehat{D}_{t1}^2 + \frac{\widehat{h}_t^2}{4}\widehat{D}_{t2} + \widehat{\Psi}_T\widehat{A}_T^{-1}\widehat{\Psi}'_T\right]$$

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

Ⅲ. Simulation Experiments

We perform some simulation experiments to investigate the potential applicability of the modified Cox test. We take the generalized autoregressive conditional heteroskedasticity (GARCH) by Bollerslev (1986) and bilinearity by Granger and Anderson (1978) as two competing models for non-linear dependence in time series. Because GARCH and bilinear processes are both non-linear and have very similar unconditional distributions, although the conditional distributions are quite different, it is difficult to detect the true specification between these two 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.

$$\begin{aligned} M_1:\ & y_{1,t}=\alpha_0+\alpha_1 y_{1,t-1}+u_t, \quad u_t|\,x_t\sim N(0,1) \\ & E(y_{1,t}\,|I_{t-1})=\alpha_0+\alpha_1 y_{t-1} \\ & Var(y_t|I_{t-1})=h_t=\kappa+\gamma\, u_{t-1}^2+\delta\, h_{t-1} \\ & \text{and } u_t=\sqrt{h_t v_t},\ where\ v_t\sim N(0,1) \end{aligned} \tag{22}$$

$$\begin{aligned} M_2:\ & y_{2,t}=\beta_0+\beta_1 y_{2,t-1}+\varepsilon_t,\ \ \varepsilon_t=b_{11}\varepsilon_{t-1}\xi_{t-1}+\xi_t,\ \ where\ \ \xi_t\sim N(0,1) \\ & E(y_{1,t}\,|I_{t-1})=\beta_0+\beta_1 y_{2,t-1}+b_{11}\varepsilon_{t-1}\xi_{t-1} \\ & Var(y_t|I_{t-1})=\sigma_\xi^2 \end{aligned} \tag{23}$$

We generate the artificial data in 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 $y_{1,t}=0.15+0.85\,y_{t-1}+u_t$ with a 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 $y_{2,t}=0.19+0.80y_{2,t-1}+\varepsilon_t\ where\ \varepsilon_t=0.385\varepsilon_{t-1}\xi_{t-1}+\xi_t$. The parameter values chosen for both models correspond to the empirical estimates of 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 newly generated data, we perform the testing experiments by setting different values of λ; $\lambda=0$ and $\lambda=1$.

If $\lambda=1$, then $y_t=y_{1,t}$, so M_1 becomes the correctly specified one, while M_2 is correctly specified if $\lambda=0$. The MLEs of these two specifications are calculated based on BHHH algorithm and the simulation results are calculated from 200 replications and with large sample sizes of 1000, 2000, 3000, and 5000 and 250, 500, and 750 for relatively small sample size properties. Let T1 denote the modified Cox test when M_1 is correctly specified and T2 denote 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> Simulation Results when GARCH (1,1) is true

sample size	N=1000		N=2000		N=3000		N=4000	
	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	-0.030	5.692	-0.369	6.129	0.449	-0.276
kurt	2.866	4.423	2.386	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
too high	0.005	0.000	0.030	0.000	0.010	0.000	0.040	0.000
too low	0.015	0.960	0.010	0.995	0.035	1.000	0.015	1.000

Note: 1) two-tailed test with $\alpha=.05$ and $\lambda=1$.
2) R.F. is rejection frequency.

In <Table 1>, we report the simulation results under the null hypothesis that the GARCH(1,1) model is correctly specified 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 also very close to the nominal size for all sample sizes except for N=1000, in which the actual size is slightly understated.

Table 2 reports the simulation results when the bilinear model is correctly specified for N=1000, 2000, 3000, and 5000. The unconditional 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.

<Table 2> Simulation Results when Bilinear (1,1) is true

sample size	N=1000		N=2000		N=3000		N=4000	
	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.(α =.05)	0.945	0.050	0.945	0.045	0.995	0.050	0.980	0.030
too high	0.000	0.020	0.000	0.005	0.000	0.025	0.000	0.000
too low	0.945	0.030	0.945	0.040	0.995	0.025	0.980	0.030

Note) 1) two-tailed test with $\alpha = .05$ and $\lambda = 0$.
2) R.F. is rejection frequency.

<Table 3> 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
too high	0.000	0.000	0.005	0.000	0.020	0.000
too low	0.030	0.725	0.015	0.880	0.020	0.935

Note: 1) two-tailed test with $\alpha = .05$ and $\lambda = 1$.
2) R.F. is rejection frequency.

<Table 3> reports the simulation results with relatively small sample sizes of 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 4>, the means are slightly bigger than zero in absolute value for all three sample sizes but this deviation decreases as sample size increases. The actual size and the rejection frequency are approximately equivalent to the nominal levels.

<Table 4> 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
too high	0.005	0.010	0.000	0.000	0.000	0.005
too low	0.930	0.060	0.940	0.070	0.945	0.050

Note: 1) two-tailed test with $\alpha = .05$ and $\lambda = 0$.
2) R.F. is rejection frequency.

IV. Empirical Application

We apply the proposed test to two competing non-linear error equation models with three time series data sets studied by BH: 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. The stochastic error equation follows non-linearity and we specify the GARCH model as the null hypothesis and the bilinear model as the alternative hypothesis or vise versa. The exogeneous variables are considered as autoregressive (AR) models. The model specifications are nested withen AR(2)-GARCH(1,1) and AR(2)-bilinear (2,1). We obtained these data sets from JBES data archives and follow the model specifications and the orders of autoregression that minimize the Schwarz criterion from BH. The two competing models are

$$M_1: \ y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + u_t$$
$$u_t | x_t \sim i.i.d\ (0,\ h_t) \ \ where \ \ h_t = \kappa + \gamma u_{t-1}^2 + \delta h_{t-1} \qquad (24)$$

$$M_2: \; y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \varepsilon_t$$
$$where \;\; u_t = b_{11} u_{t-1} \varepsilon_{t-1} + b_{12} u_{t-2} \varepsilon_{t-1} + e_t \text{ and } \varepsilon_t \sim i.i.d(0, \sigma_\varepsilon^2) \qquad (25)$$

<Table 5> Summary Statistics

	Mean	s.d.	Skew	Kurt	Max	Min	sample size
S&P 500	0.060	0.820	-0.651	8.767	3.468	-5.877	1138
BP	-0.023	0.477	0.032	4.762	1.960	-2.252	1210
IP	3.358	10.604	-0.646	5.669	37.699	-51.732	359

First, we take the daily S&P 500 stock index from January 4, 1978 to May 28, 1993. Next, we take the daily log exchange rate of the British pound (BP) to the U.S. dollar in a sample period from December 12, 1985 to February 28, 1991. As BH 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. These data sets are summarized in <Table 5>.

<Table 6> Estimated GARCH Models

S&P 500	$y_t = .053 + .066 y_{t-1} + u_t$ (.026) (.031) $h_t = .012 + .013 u_{t-1}^2 + .968 h_{t-1}$ (.011) (.006) (.019)	$l(\hat{\Theta}) = -1361.63$
BP	$y_t = .023 + . u_t$ (.013) $h_t = .010 + .059 u_{t-1}^2 + .897 h_{t-1}$ (.004) (.015) (.025)	$l(\hat{\Theta}) = -784.82$
IP	$y_t = 2.688 + .278 y_{t-1} + .104 y_{t-2} + u_t$ (.510) (.064) (.055) $h_t = 60.111 + .236 u_{t-1}^2 + .104 h_{t-1}$ (12.867) (.063) (.132)	$l(\hat{\Theta}) = -1301.32$

<Table 6> and <Table 7> report the estimation results of both models; the estimation results of GARCH models in <Table 6> and the bilinear models in <Table 7>.[3)]

<Table 7> Estimated Bilinear Models

S&P 500	$y_t = .017 + .101 y_{t-1} + u_t$ (.026) (.031) $u_t = .053 u_{t-1} \varepsilon_{t-1} + \varepsilon t$ (.012)	$l(\hat{\Theta}) = -1369.66$ $\hat{\sigma}_{\varepsilon}^2 = .650$
BP	$y_t = -.023 + u_t$ (.016) $u_t = .0007 u_{t-1} \varepsilon_{t-1} - .079 u_{t-2} \varepsilon_{t-1} + \varepsilon_t$ (.039) (.056)	$l(\hat{\Theta}) = -818.74$ $\hat{\sigma}_{\varepsilon}^2 = .227$
IP	$y_t = 2.340 + .321 y_{t-1} + .125 y_{t-2} + u_t$ (.661) (.073) (.055) $u_t = -.006 u_{t-1} \varepsilon_{t-1} + \varepsilon_t$ (.004)	$l(\hat{\Theta}) = -1311.15$ $\hat{\sigma}_{\varepsilon}^2 = 90.352$

<Table 8> Test Results ($H_0: GARCH$ vs. $H_1: Bilinear$)

	Bera & Higgins	Modified Cox Test
S&P 500	.023	.204
BP	.196	-.002
IP	.533	-.022

<Table 9> Test Results ($H_0: Bilinear$ vs. $H_1: GARCH$)

	Bera & Higgins	Modified Cox Test
S&P 500	-.910	-6.195
BP	-2.797	-5.413
IP	-1.643	-23.094

<Table 8> and <Table 9> present the modified Cox test results. <Table 8> reports the test results when the GARCH models is the null hypothesis

3) We estimated these models with the same data sets that BH used. As expected, our estimation results are very close to those of BH. Followed by the comment from an anonymous referee, those regression results are not reported here but can be provided by request.

and <Table 9> reports the test results when the bilinear model is the null hypothesis. In <Table 8>, 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 9, the absolute values of our test results are far much bigger than those of BH when the bilinear model is the null hypothesis and, although BH 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, all three series are rejected in our test results. In the test results by BH, S&P 500 cannot be rejected in both specifications but the GARCH models are preferred in all three data sets in our test results.

Ⅴ. Conclusion

A new approach based on the conditional mean and variance specifications has been proposed in this paper. This modified Cox test has some attractive features. The major attraction among them is its computational simplification because it does not require computing the pseudo-true values. As this proposed test is based on the specification of the first two conditional moments, we can also test other distributional features such as conditional variance. Furthermore, it can be easily extended to the more complicated non-linear models in time series and possibly dynamic panel data 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. Empirical applications have been also considered in this paper. The test results from three different time series data sets indicate that the bilinear models are misspecified and the GARCH models are preferred in those data sets.

Acknowledgements

This paper was partially supported by Ajou University. I am grateful to the two anonymous referees for their valuable comments. Any remaining errors are my responsibility.

[References]

Bera, A. and Higgins, M.(1997), "ARCH and Bilinearity as Competing Models for Nonlinear Dependence," *Journal of Business and Economic Statistics,* 15, 43-50.

Bollerselv, T.(1986), "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 31, 307-327.

Bollerslev, T. and Wooldridge, J. M.(1992), "Quasi-Maximum Likelihood Estimation and Inference in Dynamic Models With Time-Varying Covariance," *Econometric Reviews*, 11, 143-172.

Cox, D. R.(1961), *Tests of Separate Families of Hypothesis*, in Proceedings of the 4th Berkeley Symposium (1), Berkeley: University of California Press, 105-123.

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

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

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

Engle, R.(1982), "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica,* 50, 987-1008.

Fisher, G. R. and McAleer, M.(1979), "On the Interpretation of the Cox Test in Econometrics," *Economics Letter*, 4, 145-150.

Gourieroux, C., Monfort, A. and Trognon, A.(1983), *Testing Non-Nested Hypothesis*, Elservier Science, Handbook of Econometrics, 4, 2583-2633.

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

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

Pesaran, M. H. and Deaton, A. S.(1978), "Testing Non-Nested Nonlinear Regression Models," *Econometrica*, 46, 677-694.

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.

White, H.(1982), "Regularity Conditions for Cox's Test of Non-Nested Hypotheses," *Journal of Econometrics*, 19, 307-327.

White, H.(1994), *Estimation, Inference and Specification Analysis*, Econometric Society Monographs: Cambridge University Press.

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

Wooldridge, J. M.(1990), "A Unified Approach to Robust, Regression-Based Specification Tests," *Economics Theory,* 6, 17-43.

Wooldridge, J. M.(1994), *Estimation and Inference for Dependent Processes,* Handbook of Econometrics, Vol. IV, Amsterdam: North-Holland, 2659-2738.

[Abstract]

시계열 모형 분석을 위한 보완된 콕스 검정법

김 동 근

본 논문은 전통적인 콕스 검정법이 가지고 있는 계산상의 어려움을 완화시키기 위해 조건부 평균과 조건부 분산을 이용한 보완된 콕스 검정법을 제시한다. 처음 두 조건부 적률을 이용한 이러한 접근방법은 주어진 관찰치 전체의 결합밀도함수를 구하지 않아도 되는 계산상의 이점과 시계열모형에 쉽게 적용될 수 있는 장점이 있다. 제안된 검정법에 대한 만족 할만 한 모의실험 결과를 얻었으며 비선형 오차 방정식 모형에 대한 실증분석결과 GARCH모형이 bilinear모형보다 더 올바른 모형으로 나타났다.

핵심용어 : 조건부 평균, 조건부 분산, 보완된 콕스 검정법, 비선형 오차 방정식 모형