# THE ASYMPTOTIC LOCAL STRUCTURE OF THE COX MODIFIED LIKELIHOOD-RATIO STATISTIC FOR TESTING NON-NESTED HYPOTHESES

JERZY SZROETER
*University College London*

It is shown that the Cox modified likelihood-ratio statistic for testing partially non-nested hypotheses $H_0$ and $H_1$ is asymptotically equivalent to a bilinear form in nondegenerate asymptotically normal random vectors for sequences of data-generating processes converging to the intersection of $H_0$ and $H_1$ but not necessarily belonging to either $H_0$ or $H_1$. One of the asymptotically normal vectors is the complete parametric encompassing vector of Mizon and Richard, while the other is a close relative. The results are valid regardless of whether or not the data-generating process is exponential and imply that the Cox statistic is not generally asymptotically locally normal. This corrects an assumption made in recent literature.

## 1. INTRODUCTION

The purpose of this paper is to derive the asymptotic local structure of the Cox [1,2] modified likelihood-ratio statistic for testing a hypothesis $H_0$ versus a non-nested alternative $H_1$. By "asymptotic local" we mean "along a sequence, indexed by sample size, of true data distributions (data-generating processes) approaching some data distribution that lies in the intersection of $H_1$ and $H_0$." The special case where the sequence lies entirely within the originally defined $H_0$ or $H_1$ is of particular interest as these are the very hypotheses for which the Cox statistic is constructed. For such an analysis to be possible, we have to assume that the hypotheses $H_0$ and $H_1$ are partially non-nested in the sense of Pesaran [14]. The distribution of the Cox statistic for sequences of data-generating processes whose limit lies outside the intersection of the originally defined $H_0$ and $H_1$ is, in fact, less problematic than the distribution for limits within that intersection. It is for limits within the intersection that nonstandard issues of non-normality arise. These are the focus of the present paper.

In the light of the analyses of Cox [1, pp. 118–119] and Gourieroux, Monfort, and Trognon [7], it would appear that the Cox statistic will usually be asymptotically linear in the complete parametric encompassing (CPE) test vector of Mizon [12] and Mizon and Richard [13] under $H_0$ for a *fixed* data distribution of the exponential family. However, Kent [10, pp. 338–339] has shown that there exist certain degenerate cases in which the Cox statistic is asymptotically equivalent to a quadratic form in the CPE vector under $H_0$ for a fixed data distribution of the exponential family. Kent's demonstration made use of the special assumption that the distribution in $H_1$ nearest to any true distribution in $H_0$ always lies in the intersection of $H_0$ and $H_1$. Vuong [15] showed that the asymptotic distribution of the basic likelihood-ratio, as distinct from the Cox centered likelihood-ratio, is that of a quadratic form in normal variates whenever the distribution in $H_1$ nearest to a fixed true distribution (not necessarily belonging to $H_0$ or $H_1$) is observationally equivalent to the distribution in $H_0$ nearest that true distribution. In the present paper, we will prove that the asymptotic local structure of the Cox statistic itself is bilinear in asymptotically normal vectors, regardless of whether or not the data distribution itself is exponential. One of these vectors is the CPE vector, while the other is a close relative of it.

The analysis of the paper is of potential use as a first step toward a correct assessment of the asymptotic local power of the Cox test. In seeking to derive that power, Pesaran [14, p. 83, Theorem 3.1] appeared to show that the asymptotic local distribution of the Cox statistic under $H_1$ is normal. His proof was ingeniously arranged to take advantage of the very tempting assumption that the asymptotic distribution of the Cox statistic along a local sequence of data-generating processes under $H_0$ is normal just as it is for a fixed data-generating sequence under $H_0$ (but lying outside the intersection of $H_0$ and $H_1$). The results of the present paper imply that this assumption is false. The asymptotic local distribution of the Cox statistic is not normal either under $H_0$ or under $H_1$.

The plan of the paper is as follows. In Section 2, we set out the framework of analysis. In Section 3, we present and discuss the main result. This is proved in Section 5, following a careful listing of technical regularity conditions in Section 4.


## 2. THE FRAMEWORK OF ANALYSIS

Let $y$ denote an $np$-dimensional vector of $n$ observations (not necessarily i.i.d.) on $p$ variables. Let $f_n(y|\theta), g_n(y|\phi), k_n(y|\psi)$ denote density functions for $y$ with respect to some underlying measure $\nu_n$ on $\mathbb{R}^{np}$, where $\theta, \phi, \psi$ are parameter vectors belonging to the closed subsets $\Theta \subset \mathbb{R}^{\dim(\theta)}$, $\Phi \subset \mathbb{R}^{\dim(\phi)}$, $\Psi \subset \mathbb{R}^{\dim(\psi)}$, respectively. These density functions are assumed to have a common support on $\mathbb{R}^{np}$. They may be conditional on a common set

of exogenous variables, but for simplicity such variables are notationally suppressed.

Letting $d_n(y)$ denote a density with respect to $\nu_n$ for the "true" distribution of $y$, we consider the following hypotheses:

$$H_0: \qquad d_n(y) = f_n(y|\theta) \qquad \text{a.s. } (d_n) \text{ for some } \theta \in \Theta, \tag{2.1}$$

$$H_1: \qquad d_n(y) = g_n(y|\phi) \qquad \text{a.s. } (d_n) \text{ for some } \phi \in \Phi. \tag{2.2}$$

We assume that the hypotheses are partially non-nested in the sense that neither does $H_0$ imply $H_1$ nor does $H_1$ imply $H_0$, but $H_0$ and $H_1$ can both be true. Thus there exist parameter values $\theta^*$ and $\phi^*$ such that

$$f_n(y|\theta^*) = g_n(y|\phi^*) \qquad \text{a.s. } (f_n). \tag{2.3}$$

Let $\Theta^*$ and $\Phi^*$ denote the sets of all $\theta^*$-values and $\phi^*$-values, respectively, for which equation (2.3) holds. Furthermore, let $\Psi^*$ denote the set of all $\psi^*$-values such that

$$k_n(y|\psi^*) = f_n(y|\theta^*) = g_n(y|\phi^*) \qquad \text{a.s. } (k_n) \tag{2.4}$$

for some $\theta^* \in \Theta^*$ and some $\phi^* \in \Phi^*$. For further discussion of the notion of partial non-nesting, the reader may refer to Pesaran [14] and Vuong [15].

The Cox [1] modified likelihood-ratio statistic for testing $H_0$ versus $H_1$ is given by the expression

$$\log[f_n(y|\hat{\theta})] - \log[g_n(y|\hat{\phi})] - \hat{E}_0[\log[f_n(y|\hat{\theta})] - \log[g_n(y|\hat{\phi})]] \tag{2.5}$$

where $\hat{\theta}$ and $\hat{\phi}$ denote maximizers of $\log[f_n(y|\theta)]$ and $\log[g_n(y|\phi)]$, respectively. The term beginning with $\hat{E}_0$ in (2.5) represents the "best estimate" of the value we would expect the log-likelihood to take under $H_0$ (Cox [1, p. 114]). Of the various possible versions of such a best estimate, we focus here on the following form:

$$\hat{E}_0[\log[f_n(y|\hat{\theta})] - \log[g_n(y|\hat{\phi})]]$$

$$= \left[ \int \{\log[f_n(y|\theta)] - \log[g_n(y|\phi)]\} f_n(y|\theta) \, d\nu_n \right]_{\theta=\hat{\theta}, \phi=\hat{\phi}}. \tag{2.6}$$

This form, specialized to the case of i.i.d. observations, was used by White [17] and Pesaran [14].

It is important to know the asymptotic $(n \to \infty)$ distribution of expression (2.5), suitably scaled, under the following possible specifications for $d_n(y)$.

Specification 1. $d_n(y) = f_n(y|\theta)$ for some fixed value $\theta$ independent of $n$ and lying in the set $\Theta \setminus \Theta^*$ where $\Theta^*$ is defined after equation (2.3).

Specification 2. $d_n(y) = f_n(y|\theta_n)$ for some sequence of values $\{\theta_n\} \subset \Theta$ such that

$$\theta_n = \theta^* + n^{-1/2}\xi \tag{2.7}$$

where $\theta^*$ and $\xi$ are finite vectors having some fixed values independent of $n$, with $\theta^*$ satisfying equation (2.3).

Specification 3.  $d_n(y) = g_n(y \mid \phi)$ for some fixed value $\phi$ independent of $n$ and lying in the set $\Phi \setminus \Phi^*$ where $\Phi^*$ is defined after equation (2.3).

Specification 4.  $d_n(y) = g_n(y \mid \phi_n)$ for some sequence of values $\{\phi_n\} \subset \Phi$ such that

$$\phi_n = \phi^* + n^{-1/2}\zeta \tag{2.8}$$

where $\phi^*$ and $\zeta$ are finite vectors having some fixed values independent of $n$, with $\phi^*$ satisfying equation (2.3).

Specifications 1,3 and 2,4 are particular cases of the following more general Specifications 5 and 6, respectively, where the parametric density function $k_n(y \mid \psi)$ need not necessarily belong to either of the families $H_0$ or $H_1$.

Specification 5.  $d_n(y) = k_n(y \mid \psi)$ for some fixed value $\psi$ independent of $n$ and lying in the set $\Psi \setminus \Psi^*$ where $\Psi^*$ is defined between equations (2.3) and (2.4).

Specification 6.  $d_n(y) = k_n(y \mid \psi_n)$ for some sequence of values $\{\psi_n\} \subset \Psi$ such that

$$\psi_n = \psi^* + n^{-1/2}\lambda \tag{2.9}$$

where $\psi^*$ and $\lambda$ are finite vectors having some fixed values independent of $n$, with $\psi^*$ satisfying equation (2.4).

Under Specification 1 and suitable regularity conditions, the expression (2.5) appropriately scaled is distributed asymptotically as a unit normal variate (Cox [1], Mizon and Richard [13], and White [17]). Under Specification 3 and suitable regularity conditions, the probability limit of $n^{-1}$ times expression (2.5) is finite and nonzero, thus essentially making the Cox test consistent against $H_1$.

The asymptotic distribution of expression (2.5) under Specification 4 with $\zeta$ non-null is required as a preliminary to a correct analysis of the local power of the Cox test against distributions within the original class (2.2). As we will show, this asymptotic distribution is not normal. This finding appears to conflict with that of Pesaran [14]. The conflict is easily explained: Pesaran's derivation of the asymptotic distribution of the scaled Cox statistic under Specification 4 was so constructed as to exploit the premise that the asymptotic distribution of that statistic under Specification 2 is unit normal just as it is under Specification 1. The premise is given as Assumption 3 on page 81 of Pesaran [14]. The statement of that assumption makes no distinction between Specifications 1 and 2, yet the statement and proof of Theorem 3.1 on page 83 of Pesaran [14] implicitly takes Assumption 3 to be valid under

Specification 2. By writing the Cox statistic for $H_0$ versus $H_1$ as a function of the Cox statistic for $H_1$ versus $H_0$ and then invoking his Assumption 3 for the latter, Pesaran [14, pp. 83–84] appears to obtain the asymptotic local distribution of the Cox statistic for Specification 4. As we will show, however, Pesaran's Assumption 3, while correct for Specification 1, is invalid for Specification 2. The asymptotic distribution of the basic Cox statistic (2.5), hence that of the scaled Cox statistic, is not even normal under Specification 2, let alone unit normal. This result is in keeping with the practical finding that the unit normal distribution does not always appear to be an adequate representation of the empirical distribution of the Cox statistic (as, for example, in the Monte Carlo studies of Godfrey and Pesaran [6] and others reported in McAleer [11, pp. 179–183]).

For Specifications 2, 4, 6 with $\xi = 0$, $\zeta = 0$, $\lambda = 0$, the asymptotic distribution of just the likelihood-ratio itself (as distinct from Cox's centered version) can be obtained from Theorem 3.3 of Vuong [15, p. 313] as that of a weighted sum of central chi-square variates. Equation (A.7) of Vuong [15, Appendix, p. 327] shows that the simple likelihood-ratio is asymptotically equivalent to a quadratic form in $(\hat{\theta} - \theta^*)$ and $(\hat{\phi} - \phi^*)$, in the notation of the present paper. The additional contributions of the present paper could be viewed as: (1) attending to the complications introduced by Cox's centering of the basic likelihood-ratio, (2) allowing $\xi$, $\zeta$, and $\lambda$ to be nonzero, and (3) revealing the local connection with complete parametric encompassing. While our analysis holds good for a sequence of data densities not necessarily belonging to the intersection of $H_0$ and $H_1$, we do require that the sequence be local to that intersection. Thus, $\theta^*$ and $\phi^*$ must index densities lying within that intersection. In Vuong's analysis, the parameter values $\theta^*$ and $\phi^*$ are those associated with data densities of $H_0$ and $H_1$ closest (in an information-theoretical sense) to some "true" data density not necessarily belonging to the intersection of $H_0$ and $H_1$.

A fully general analysis under Specification 5 and its nonparametric extensions is outside the scope of this paper. Vuong [15, p. 318, Theorem 5.1(i)] has noted that, under his version of our Specification 5, the basic *uncentered* (but scaled) likelihood-ratio would be asymptotically unit normal if $H_0$ and $H_1$ were not even partially non-nested but strictly disjoint and if the data densities of $H_0$ and $H_1$ closest to $k_n(y|\psi)$ were in fact *equally* close to $k_n(y|\psi)$. For some further ideas pertaining to general analysis under Specification 5, see the papers by Davidson and MacKinnon [3,4,5], Mizon and Richard [13], and Vuong [15].

## 3. THE MAIN RESULTS

For each fixed $\theta \in \Theta$, let us define the vector $a(\theta)$ as

$$a(\theta) \equiv \text{probability limit of } \hat{\phi} \text{ under } H_0 \text{ (equation (2.1))}. \tag{3.1}$$

For each fixed $\phi \in \Phi$, let us define the vector $b(\phi)$ as

$$b(\phi) = \text{probability limit of } \hat{\theta} \text{ under } H_1 \text{ (equation (2.2))}. \tag{3.2}$$

We now define vectors $\hat{S}$ and $\hat{h}$ as

$$\hat{S} = \hat{\phi} - a(\hat{\theta}), \tag{3.3}$$

$$\hat{h} = \hat{\phi} - a(b(\hat{\phi})). \tag{3.4}$$

$\hat{S}$ may be recognized as the complete parametric encompassing test vector of Mizon and Richard [13]. In practice, of course, the function $a(\cdot)$ might not be precisely known and some estimate of it would have to be used in order to implement the CPE test, but this issue is not our concern here. The core theorem of the present paper shows that the Cox statistic (2.5) is locally bilinear in the asymptotically normal vectors $n^{1/2}\hat{S}$ and $n^{1/2}\hat{h}$. To state the theorem, we need first to introduce some definitions. We define the matrices $J_0(\theta)$, $J_0$, $J_1(\phi)$, $J_1$, $M(\theta, \phi)$, $M$, $C_0(\theta, \psi)$, $C_0$, $C_1(\phi, \psi)$, $C_1$, $V_s$, $V_h$, $V$, $\Omega$, and the vectors $\alpha$ and $\mu$ as follows:

$$J_0(\theta) = \lim_{n \to \infty} \frac{1}{n} \int \left[ \frac{\partial \log[f_n(y \mid \theta)]}{\partial \theta} \right] \left[ \frac{\partial \log[f_n(y \mid \theta)]}{\partial \theta} \right]' f_n(y \mid \theta) \, d\nu_n \tag{3.5}$$

$$J_0 = J_0(\theta^*) \tag{3.6}$$

$$J_1(\phi) = \lim_{n \to \infty} \frac{1}{n} \int \left[ \frac{\partial \log[g_n(y \mid \phi)]}{\partial \phi} \right] \left[ \frac{\partial \log[g_n(y \mid \phi)]}{\partial \phi} \right]' g_n(y \mid \phi) \, d\nu_n \tag{3.7}$$

$$J_1 = J_1(\phi^*) \tag{3.8}$$

$$M(\theta, \phi) = \lim_{n \to \infty} \frac{1}{n} \int \left[ \frac{\partial \log[f_n(y \mid \theta)]}{\partial \theta} \right] \left[ \frac{\partial \log[g_n(y \mid \phi)]}{\partial \phi} \right]' f_n(y \mid \theta) \, d\nu_n \tag{3.9}$$

$$M = M(\theta^*, \phi^*) \tag{3.10}$$

$$C_0(\theta, \psi) = \lim_{n \to \infty} \frac{1}{n} \int \left[ \frac{\partial \log[f_n(y \mid \theta)]}{\partial \theta} \right] \left[ \frac{\partial \log[k_n(y \mid \psi)]}{\partial \psi} \right]' f_n(y \mid \theta) \, d\nu_n \tag{3.11}$$

$$C_0 = C_0(\theta^*, \psi^*) \tag{3.12}$$

$$C_1(\phi, \psi) = \lim_{n \to \infty} \frac{1}{n} \int \left[ \frac{\partial \log[g_n(y \mid \phi)]}{\partial \phi} \right] \left[ \frac{\partial \log[k_n(y \mid \psi)]}{\partial \psi} \right]' g_n(y \mid \phi) \, d\nu_n \tag{3.13}$$

$$C_1 = C_1(\phi^*, \psi^*) \tag{3.14}$$

$$V_s = J_1^{-1} - J_1^{-1} M' J_0^{-1} M J_1^{-1} \tag{3.15}$$

$$V_h = [I - J_1^{-1} M' J_0^{-1} M] J_1^{-1} [I - M' J_0^{-1} M J_1^{-1}] \tag{3.16}$$

$$V = \begin{bmatrix} V_s & \vdots & V_h \\ \cdots & \cdots & \cdots \\ V_h & \vdots & V_h \end{bmatrix} \tag{3.17}$$

$$\Omega = \begin{bmatrix} J_0^{-1} & \vdots & J_0^{-1}MJ_1^{-1} \\ \cdots & \cdots & \cdots \\ J_1^{-1}M'J_0^{-1} & \vdots & J_1^{-1} \end{bmatrix} \tag{3.18}$$

$$\alpha = [(J_0^{-1}C_0)', (J_1^{-1}C_1)']'\lambda \tag{3.19}$$

$$\mu = [(C_1 - M'J_0^{-1}C_0)'J_1^{-1}, C_1'V_s]'\lambda. \tag{3.20}$$

The core results of this paper are given by the following theorem.

**THEOREM.** *Take as given the technical regularity conditions set out by Assumptions 1–11 in Section 4 of this paper. Then, under Specification 6 with equation (2.9) (which includes Specifications 2 and 4 with equations (2.7) and (2.8) as special cases), the Cox statistic (2.5) is asymptotically (as $n \to \infty$) equivalent in probability to the negative of the random quantity*

$$(n^{1/2}\hat{h})'V_s^+(n^{1/2}\hat{S}), \tag{3.21}$$

*which in turn converges in probability to the random variable*

$$n(\hat{\phi} - \phi^*)'[J_1(\hat{\phi} - \phi^*) - M'(\hat{\theta} - \theta^*)], \tag{3.22}$$

*where $V_s^+$ is any positive semidefinite matrix satisfying $V_s V_s^+ V_s = V_s$. Moreover,*

$$n^{1/2}(\hat{S}', \hat{h}')' \xrightarrow{d} N(\mu, V) \tag{3.23}$$

*and*

$$n^{1/2}[(\hat{\theta} - \theta^*)', (\hat{\phi} - \phi^*)']' \xrightarrow{d} N(\alpha, \Omega) \tag{3.24}$$

*where $\alpha, \mu$ and $V, \Omega$ are the partitioned vectors and matrices given by equations (3.17)–(3.20).* ■

The proof of the above theorem is presented in Section 5. Here we point out some of the important features of the theorem. First, we look at the special forms which emerge for the mean vectors $\mu$ and $\alpha$ appearing in the limit distributions of (3.23) and (3.24) under Specifications 2 and 4. When Specification 2 is in force as a particular case of Specification 6, we can replace the general density function $k_n(y|\psi)$ by $f_n(y|\theta)$, the density function characterizing $H_0$. We also replace the $\lambda$ in equation (2.9) of Specification 6 with the $\xi$ in equation (2.7) of Specification 2. With these replacements and given equation (2.3), we see in formulae (3.9)–(3.20) that $C_0 = J_0$, $C_1 = M'$, and that

$$\alpha = \begin{bmatrix} \xi \\ \cdots \\ J_1^{-1}M'\xi \end{bmatrix}, \qquad \mu = \begin{bmatrix} 0_{\dim(\phi)} \\ \cdots \\ V_s M'\xi \end{bmatrix}, \tag{3.25}$$

where $0_{\dim(\phi)}$ denotes a null column vector of dimension $\dim(\phi)$. On the other hand, when Specification 4 is in force as a particular case of Specification 6, we find by analogous replacements that

$$\alpha = \begin{bmatrix} J_0^{-1} M \zeta \\ \cdots \\ \zeta \end{bmatrix}, \qquad \mu = \begin{bmatrix} V_s J_1 \zeta \\ \cdots \\ V_s J_1 \zeta \end{bmatrix}. \tag{3.26}$$

It is important to distinguish the structure (3.25) of the vector $\mu$ under Specification 2 from its structure (3.26) under Specification 4. It is this difference which potentially gives the Cox test nontrivial local power against $H_1$. Heuristically, using (3.25) and (3.26) in (3.21) and (3.23), we see that, in its asymptotic local distribution, the Cox statistic has an expected value equal to $-\text{trace}[V_s^+ V_h]$ under Specification 2 but equal to $-\{\zeta' J_1 V_s J_1 \zeta + \text{trace}[V_s^+ V_h]\}$ under Specification 4. The latter expected value is strictly smaller than the former provided only that $V_s J_1 \zeta$ is non-null. Under the more general Specification 6, the expected value of the asymptotic local distribution of the Cox statistic is equal to $-\{\lambda'(C_1 - M'J_0^{-1} C_0)' J_1^{-1} V_s^+ V_s C_1 \lambda + \text{trace}[V_s^+ V_h]\}$. While the first term in parentheses may well be nonzero when Specification 2 does not hold, its sign is unknown. So a *one-sided* test based on the Cox statistic could have very poor power properties. Interestingly, however, when the models $H_0$ and $H_1$ are (locally) orthogonal in the sense that the matrix $M$ given by equation (3.10) is null, the first term in parentheses reduces to the *non-negative* quantity $\lambda' C_1 J_1^{-1} C_1 \lambda$. In this case, it can be seen from (3.15), (3.16), (3.17), (3.20), and (3.23) that the probability limit of $n^{1/2}(\hat{S} - \hat{h})$ degenerates to a zero vector, hence expression (3.21) becomes asymptotically equivalent to the CPE test quantity $n\hat{S}' V_s^+ \hat{S}$.

Turning now to a second important feature of the above theorem, we note that expressions (3.21) and (3.22) are bilinear forms in limiting normal vectors. Consequently, the asymptotic local distribution of the Cox statistic is not standard, let alone normal. Moreover, we have seen above that the expected value of this distribution equals $-\text{trace}[V_s^+ V_h]$ under Specification 2. Thus, not only is the asymptotic local distribution of the Cox statistic non-normal, but it is also noncentral even for data densities belonging to $H_0$ and local to the intersection of $H_0$ and $H_1$. These facts are not inconsistent with widely obtained results in Monte Carlo studies (see, for example, Godfrey and Pesaran [6] and McAleer [11]) showing a tendency for the empirical size of the Cox test to exceed its nominal size.

A third interesting feature of the theorem is the form (3.17) of the variance matrix $V$ in (3.23). Kent [10, p. 335] has observed a similar structure for the joint distribution of his versions of our $\hat{\phi}$ and $a(\hat{\theta})$. The form (3.17) implies that

$$V_s = V_h + V_{(S-h)} \tag{3.27}$$

where $V_{(S-h)}$ denotes the variance matrix of the asymptotic distribution of $n^{1/2}(\hat{S} - \hat{h})$. The rough intuition behind result (3.27) comes from equations (3.3) and (3.4), where it can be seen that the "estimator" $\hat{h}$ is in the nature of a "restricted" version of the "estimator" $\hat{S}$, obtained by applying the "constraint" $\theta = b(\phi)$ in a certain way. Thus one might expect the variance matrix of $\hat{S}$ to exceed that of $\hat{h}$. These observations suggest the possibility of a Hausman [9] test of $H_1$ (rather than $H_0$) based on checking the statistical significance of the difference $(\hat{S} - \hat{h})$. But, by (3.3) and (3.4), $(\hat{S} - \hat{h}) = [a(b(\hat{\phi})) - a(\hat{\theta})]$. Therefore a Hausman test using $(\hat{S} - \hat{h})$ could not be locally more powerful than the CPE test of $H_1$ based on $[b(\hat{\phi}) - \hat{\theta}]$.

We end this section by noting that, in practical applications, the Cox statistic (2.5) is scaled by some estimate of its asymptotic standard error. While different estimates may well be asymptotically equivalent under the fixed data distributions of Specifications 1 and 3, they will generally have different asymptotic structures along the sequences (2.7), (2.8), and (2.9) of Specifications 2, 4, and 6. These structures are beyond the scope of the present paper, but they would certainly need to be investigated before one could draw any final definitive conclusions about the asymptotic local size and power of the Cox test as implemented in practice.

## 4. TECHNICAL REGULARITY CONDITIONS

In all that follows, $\theta^*$, $\phi^*$, and $\psi^*$ denote values interior to the sets $\Theta$, $\Phi$, and $\Psi$ such that equation (2.4) holds. The functions $a(\theta)$, $\theta \in \Theta$, and $b(\phi)$, $\phi \in \Phi$, are given by equations (3.1) and (3.2). In stating the definitions and assumptions of this section, we will take $k_n(y|\psi)$ to be the "true" density with respect to which probability statements are made, unless otherwise stated. This involves no loss of generality since $k_n(y|\psi)$ and $\psi$ can always be specialized to (read as) $f_n(y|\theta)$ and $\theta$, respectively, if $H_0$ is "true," or to $g_n(y|\phi)$ and $\phi$, respectively, if $H_1$ is "true." The definitions and assumptions of this section allow the "true" value of $\psi$ to follow the sequence (2.9) (which specializes to (2.7) or (2.8), respectively, if $H_0$ or $H_1$ is "true"), but amount to little more than restatements in convenient form of assumptions and basic ideas that have been thoroughly discussed by Cox [1,2], Gourieroux, Monfort, and Trognon [7,8], Mizon and Richard [13], and White [16,17]. We therefore state the definitions and assumptions without further comment.

DEFINITION 1. *For each fixed $\psi \in \Psi$, the vectors $q(\psi)$ and $r(\psi)$ are defined as*

$$q(\psi) \equiv \text{probability limit of } \hat{\theta} \text{ when } d_n(y) = k_n(y|\psi), \tag{4.1}$$

$$r(\psi) \equiv \text{probability limit of } \hat{\phi} \text{ when } d_n(y) = k_n(y|\psi). \tag{4.2}$$

*Note that the pair $q(\psi), r(\psi)$ specializes to the pair $\theta, a(\theta)$ or to the pair $b(\phi), \phi$ according as the pair $k_n(y|\psi)$, $\psi$ specializes to the pair $f_n(y|\theta), \theta$ or to the pair $g_n(y|\phi), \phi$.*   ∎

DEFINITION 2. *For each $\psi$, $q_n(\psi)$ and $r_n(\psi)$ will denote values of $\theta \in \Theta$ and $\phi \in \Phi$, respectively, which maximize the expectations*

$$\int \log[f_n(y|\theta)]\, k_n(y|\psi)\, d\nu_n \quad and \quad \int \log[g_n(y|\phi)]\, k_n(y|\psi)\, d\nu_n,$$

*respectively, and are such that the functions $q_n(\psi)$ and $r_n(\psi)$, $\psi \in \Psi$, satisfy the regularity conditions set out in Assumptions 1 and 2 below.*   ∎

Assumption 1. There exists an open neighborhood of $\psi^*$ in which the functions $q_n(\cdot)$ and $r_n(\cdot)$ are unique and continuous, and in which the values $q_n(\psi), r_n(\psi)$ converge uniformly to $q(\psi), r(\psi)$, respectively.

Assumption 2. There exists an open neighborhood of $\psi^*$ in which the functions $q_n(\cdot)$ and $r_n(\cdot)$ have continuous first-order partial derivatives $\partial q_n(\psi)/\partial \psi'$ and $\partial r_n(\psi)/\partial \psi'$ which converge uniformly to the continuous first-order partial derivatives $\partial q(\psi)/\partial \psi'$ and $\partial r(\psi)/\partial \psi'$, respectively, of the function $q(\cdot)$ and $r(\cdot)$.

Assumption 3. For $\theta$ and $\phi$ in some open neighborhoods of $\theta^*$ and $\phi^*$, respectively,

$$\int \{\partial \log[f_n(y|\theta)]/\partial \theta\} f_n(y|\theta)\, d\nu_n = 0,$$

$$\int \{\partial \log[g_n(y|\phi)]/\partial \phi\} g_n(y|\phi)\, d\nu_n = 0.$$

DEFINITION 3. *Let the vectors $D_n(\theta, \phi)$ and $D_n$ be defined as*

$$D_n(\theta, \phi) = n^{-1/2}\{\partial \log[f_n(y|\theta)]/\partial \theta', \partial \log[g_n(y|\phi)]/\partial \phi'\}',$$

$$D_n = D_n(q(\psi_n), r_n(\psi_n)),$$   ∎

*where $\psi_n$ is as given by equation (2.9).*

DEFINITION 4. *Let the matrix $V_D$ be defined as*

$$V_D = \lim_{n\to\infty} \int [D_n(\theta^*, \phi^*)][D_n(\theta^*, \phi^*)]' k_n(y|\psi^*)\, d\nu_n.$$   ∎

Assumption 4.

$$V_D = \lim_{n\to\infty} \int [D_n][D_n]' k_n(y|\psi_n)\, d\nu_n = \lim_{n\to\infty} \int [D_n][D_n]' k_n(y|\psi^*)\, d\nu_n.$$

Assumption 5. Along the sequence (2.9) under Specification 6, the asymptotic distribution of the vector $D_n$ is $N(0, V_D)$.

Assumption 6. The matrices $J_0$ and $J_1$ defined by equations (3.5)–(3.8) satisfy the information matrix equalities

$$J_0 = -\left\{ \lim_{n \to \infty} \frac{1}{n} \int \left[ \frac{\partial^2 \log[f_n(y|\theta)]}{\partial \theta \, \partial \theta'} \right] f_n(y|\theta) \, d\nu_n \right\}_{\theta = \theta^*}$$

and

$$J_1 = -\left\{ \lim_{n \to \infty} \frac{1}{n} \int \left[ \frac{\partial^2 \log[g_n(y|\phi)]}{\partial \phi \, \partial \phi'} \right] g_n(y|\phi) \, d\nu_n \right\}_{\phi = \phi^*}.$$

Assumption 7. The matrices $J_0$ and $J_1$ defined by equations (3.5)–(3.8) satisfy the conditions

$$J_0 = -\operatorname*{plim}_{n \to \infty} \frac{1}{n} \left[ \frac{\partial^2 \log[f_n(y|\theta)]}{\partial \theta \, \partial \theta'} \right]$$

and

$$J_1 = -\operatorname*{plim}_{n \to \infty} \frac{1}{n} \left[ \frac{\partial^2 \log[g_n(y|\phi)]}{\partial \phi \, \partial \phi'} \right],$$

where $\theta$ and $\phi$ may be evaluated along any sequence of points converging mathematically or in probability to the points $\theta^*$ and $\phi^*$, respectively, and where the probability limits are taken along the sequence (2.9) under Specification 6 (of which sequences (2.7) and (2.8) under Specifications 2 and 4 are special cases).

Assumption 8. The quantities

$$n^{-1} \int \log[f_n(y|\theta)] f_n(y|\theta) \, d\nu_n, \qquad n^{-1} \int \log[g_n(y|\phi)] f_n(y|\theta) \, d\nu_n$$

and, for each $y$, the quantities

$$n^{-1} \log[f_n(y|\theta)], \qquad n^{-1} \log[g_n(y|\phi)]$$

have continuous partial derivatives up to the third order with respect to $(\theta, \phi)$ in some open neighborhood of $(\theta^*, \phi^*)$. The third-order partial derivatives are of probability order unity as $n \to \infty$ with $(\theta, \phi)$ evaluated along any sequence of points converging mathematically or in probability to $(\theta^*, \phi^*)$.

Assumption 9. Differentiation up to the second order, with respect to $(\theta, \phi)$ in some open neighborhood of $(\theta^*, \phi^*)$, of the quantities

$$\int \log[f_n(y|\theta)] f_n(y|\theta) \, d\nu_n, \qquad \int \log[g_n(y|\phi)] f_n(y|\theta) \, d\nu_n$$

may be carried out under the integral.

Assumption 10. Given (3.5)–(3.14), the continuous partial derivative matrices $\partial q(\psi)/\partial \psi'$ and $\partial r(\psi)/\partial \psi'$ satisfy

$$\left[ \frac{\partial q(\psi)}{\partial \psi'} \right]_{\psi = \psi^*} = J_0^{-1} C_0, \qquad \left[ \frac{\partial r(\psi)}{\partial \psi'} \right]_{\psi = \psi^*} = J_1^{-1} C_1.$$

In particular, by the note after (4.1) and (4.2) in Definition 1 and the remarks in the first paragraph of Section 4, the continuous partial derivative matrices $\partial b(\phi)/\partial \phi'$ and $\partial a(\theta)/\partial \theta'$ satisfy

$$\left[ \frac{\partial b(\phi)}{\partial \phi'} \right]_{\phi = \phi^*} = J_0^{-1} M, \qquad \left[ \frac{\partial a(\theta)}{\partial \theta'} \right]_{\theta = \theta^*} = J_1^{-1} M'.$$

Assumption 11. The maximum-likelihood estimators $\hat{\theta}$ and $\hat{\phi}$ converge in probability to $\theta^*$ and $\phi^*$, respectively, and satisfy the first-order conditions for maxima of their respective log-likelihood functions with probability tending to unity as $n \to \infty$ under Specification 6 (which includes Specifications 2 and 4 as special cases).

## 5. PROOF OF THE THEOREM OF SECTION 3

It will be convenient first to prove statement (3.24) of the theorem. Accordingly, noting Assumptions 8 and 11, we may apply a mean-value theorem to the first-order conditions for maximizing $\log[f_n(y|\theta)]$ and $\log[g_n(y|\phi)]$ to obtain

$$n^{-1/2} \left[ \frac{\partial \log[f_n(y|q_n(\psi_n))]}{\partial \theta} \right] + n^{-1} \left[ \frac{\partial^2 \log[f_n(y|\theta)]}{\partial \theta \, \partial \theta'} \right]^{\nabla}$$

$$\times\, n^{1/2}(\hat{\theta} - q_n(\psi_n)) = 0 \qquad\qquad (5.1)$$

and

$$n^{-1/2} \left[ \frac{\partial \log[g_n(y|r_n(\psi_n))]}{\partial \phi} \right] + n^{-1} \left[ \frac{\partial^2 \log[g_n(y|\phi)]}{\partial \phi \, \partial \phi'} \right]^{\nabla}$$

$$\times\, n^{1/2}(\hat{\phi} - r_n(\psi_n)) = 0 \qquad\qquad (5.2)$$

where the superscript $\nabla$ attached to the second derivative matrices denotes that, in each element of those matrices, the parameters $\theta$ and $\phi$ are evaluated at values lying between $\hat{\theta}$ and $q_n(\psi_n)$ and between $\hat{\phi}$ and $r_n(\psi_n)$, respectively. Given Definition 3, Assumptions 1, 5, 7, 11, and the sequence (2.9), the above equations imply that

$$n^{1/2}[(\hat{\theta} - q_n(\psi_n))', (\hat{\phi} - r_n(\psi_n))']'$$

is asymptotically equivalent in probability to

$$\begin{bmatrix} J_0^{-1} & \vdots & 0 \\ \cdots & \cdots & \cdots \\ 0 & \vdots & J_1^{-1} \end{bmatrix} D_n.$$

Hence, by Assumption 5, Definition 4, and the definitions (3.5)–(3.18),

$$n^{1/2}[(\hat{\theta} - q_n(\psi_n))', \ (\hat{\phi} - r_n(\psi_n))']' \xrightarrow{d} N(0,\Omega). \tag{5.3}$$

Note that, by Definition 2, Assumption 1, and equation (2.4), we may write $q_n(\psi^*) = \theta^*$ and $r_n(\psi^*) = \phi^*$. Therefore

$$n^{1/2}[q_n(\psi_n) - \theta^*] = n^{1/2}[q_n(\psi_n) - q_n(\psi^*)], \tag{5.4}$$

$$n^{1/2}[r_n(\psi_n) - \phi^*] = n^{1/2}[r_n(\psi_n) - r_n(\psi^*)]. \tag{5.5}$$

Applying Assumptions 2 and 10 along sequence (2.9), we easily obtain that

$$n^{1/2}[q_n(\psi_n) - q_n(\psi^*)] \to J_0^{-1} C_0 \lambda, \tag{5.6}$$

$$n^{1/2}[r_n(\psi_n) - r_n(\psi^*)] \to J_1^{-1} C_1 \lambda. \tag{5.7}$$

Using (3.19) and (5.4)–(5.7), we see that

$$n^{1/2}[(q_n(\psi_n) - \theta^*)', \ (r_n(\psi_n) - \phi^*)']' \to \alpha. \tag{5.8}$$

Results (5.3) and (5.8) together imply that

$$n^{1/2}[(\hat{\theta} - \theta^*)', \ (\hat{\phi} - \phi^*)']' \xrightarrow{d} N(\alpha,\Omega), \tag{5.9}$$

which is precisely statement (3.24) of the theorem.

Note that expansions (5.1) and (5.2) continue to hold with $q_n(\psi_n)$ and $r_n(\psi_n)$ replaced by $\theta^*$ and $\phi^*$, respectively, and with the matrices there superscripted by the symbol $\nabla$ accordingly reevaluated. Given result (5.9) and Assumptions 7 and 11, such expansions imply

$$\plim_{n\to\infty}\left\{n^{-1/2}\left[\frac{\partial \log[f_n(y|\theta^*)]}{\partial \theta}\right] - J_0 n^{1/2}(\hat{\theta} - \theta^*)\right\} = 0, \tag{5.10}$$

$$\plim_{n\to\infty}\left\{n^{-1/2}\left[\frac{\partial \log[g_n(y|\phi^*)]}{\partial \phi}\right] - J_1 n^{1/2}(\hat{\phi} - \phi^*)\right\} = 0. \tag{5.11}$$

We are now in a position to proceed to a proof of those parts of the theorem centered on (3.21), (3.22), and (3.23).

Using Assumptions 7, 8, 11, and results (5.9)–(5.11), we obtain the following second-order expansions about $\theta^*$ and $\phi^*$:

$$\log[f_n(y|\hat{\theta})] = \log[f_n(y|\theta^*)] + \tfrac{1}{2}n(\hat{\theta} - \theta^*)'J_0(\hat{\theta} - \theta^*) + o_p(1), \tag{5.12}$$

$$\log[g_n(y|\hat{\phi})] = \log[g_n(y|\phi^*)] + \tfrac{1}{2}n(\hat{\phi} - \phi^*)'J_1(\hat{\phi} - \phi^*) + o_p(1). \tag{5.13}$$

Using Assumptions 8, 9, and the result (5.9), we obtain also the following expansion:

$$\left[\int \log[f_n(y|\theta)]f_n(y|\theta)\,d\nu_n\right]_{\theta=\hat\theta}$$

$$=\int \log[f_n(y|\theta^*)]f_n(y|\theta^*)\,d\nu_n$$

$$+\left\{\int\left[\frac{\partial\log[f_n(y|\theta^*)]}{\partial\theta'}\right]f_n(y|\theta^*)\,d\nu_n\right.$$

$$+\int \log[f_n(y|\theta^*)]\left[\frac{\partial f_n(y|\theta^*)}{\partial\theta'}\right]d\nu_n\bigg\}(\hat\theta-\theta^*)$$

$$+\frac{1}{2}(\hat\theta-\theta^*)'\left\{\int\left[\frac{\partial^2\log[f_n(y|\theta^*)]}{\partial\theta\,\partial\theta'}\right]f_n(y|\theta^*)\,d\nu_n\right.$$

$$+\int\left[\frac{\partial\log[f_n(y|\theta^*)]}{\partial\theta}\right]\left[\frac{\partial f_n(y|\theta^*)}{\partial\theta'}\right]d\nu_n$$

$$+\int\left[\frac{\partial f_n(y|\theta^*)}{\partial\theta}\right]\left[\frac{\partial\log[f_n(y|\theta^*)]}{\partial\theta'}\right]d\nu_n$$

$$+\int \log[f_n(y|\theta^*)]\left[\frac{\partial^2 f_n(y|\theta^*)}{\partial\theta\,\partial\theta'}\right]d\nu_n\bigg\}$$

$$\times(\hat\theta-\theta^*)+O_p(n^{-1/2}).$$

By virtue of Assumptions 3 and 6, the result (5.9), and the definition of the matrix $J_0$ by equations (3.5)–(3.6), the preceding expansion can be shortened to the following:

$$\left[\int \log[f_n(y|\theta)]f_n(y|\theta)\,d\nu_n\right]_{\theta=\hat\theta}$$

$$=\int \log[f_n(y|\theta^*)]f_n(y|\theta^*)\,d\nu_n$$

$$+\left\{\int \log[f_n(y|\theta^*)]\left[\frac{\partial f_n(y|\theta^*)}{\partial\theta'}\right]d\nu_n\right\}(\hat\theta-\theta^*)$$

$$+\frac{1}{2}n(\hat\theta-\theta^*)'\left\{J_0+n^{-1}\int \log[f_n(y|\theta^*)]\left[\frac{\partial^2 f_n(y|\theta^*)}{\partial\theta\,\partial\theta'}\right]d\nu_n\right\}$$

$$\times(\hat\theta-\theta^*)+o_p(1). \tag{5.14}$$

In a similar way to the expansion preceding equation (5.14), we obtain

$$
\left[ \int \log[g_n(y|\phi)] f_n(y|\theta) \, d\nu_n \right]_{\phi=\hat{\phi}, \, \theta=\hat{\theta}}
$$

$$
= \int \log[g_n(y|\phi^*)] f_n(y|\theta^*) \, d\nu_n
$$

$$
+ (\hat{\phi} - \phi^*)' \int \left[ \frac{\partial \log[g_n(y|\phi^*)]}{\partial \phi} \right] f_n(y|\theta^*) \, d\nu_n
$$

$$
+ (\hat{\theta} - \theta^*)' \int \log[g_n(y|\phi^*)] \left[ \frac{\partial f_n(y|\theta^*)}{\partial \theta} \right] d\nu_n
$$

$$
+ \frac{1}{2} (\hat{\phi} - \phi^*)' \left\{ \int \left[ \frac{\partial^2 \log[g_n(y|\phi^*)]}{\partial \phi \, \partial \phi'} \right] f_n(y|\theta^*) \, d\nu_n \right\} (\hat{\phi} - \phi^*)
$$

$$
+ (\hat{\phi} - \phi^*)' \left\{ \int \left[ \frac{\partial \log[g_n(y|\phi^*)]}{\partial \phi} \right] \left[ \frac{\partial f_n(y|\theta^*)}{\partial \theta'} \right] d\nu_n \right\} (\hat{\theta} - \theta^*)
$$

$$
+ \frac{1}{2} (\hat{\theta} - \theta^*)' \left\{ \int \log[g_n(y|\phi^*)] \left[ \frac{\partial^2 f_n(y|\theta^*)}{\partial \theta \, \partial \theta'} \right] d\nu_n \right\} (\hat{\theta} - \theta^*)
$$

$$
+ O_p(n^{-1/2}).
$$

Using Assumptions 3 and 6, the result (5.9), the definition of the matrix $M$ by equations (3.9)–(3.10), and noting equation (2.4), we can rewrite the above expansion as follows:

$$
\left[ \int \log[g_n(y|\phi)] f_n(y|\theta) \, d\nu_n \right]_{\phi=\hat{\phi}, \, \theta=\hat{\theta}}
$$

$$
= \int \log[g_n(y|\phi^*)] f_n(y|\theta^*) \, d\nu_n
$$

$$
+ (\hat{\theta} - \theta^*)' \int \log[g_n(y|\phi^*)] \left[ \frac{\partial f_n(y|\theta^*)}{\partial \theta} \right] d\nu_n
$$

$$
- \frac{1}{2} n(\hat{\phi} - \phi^*)' J_1 (\hat{\phi} - \phi^*) + n(\hat{\phi} - \phi^*)' M'(\hat{\theta} - \theta^*)
$$

$$
+ \frac{1}{2} (\hat{\theta} - \theta^*)' \left\{ \int \log[g_n(y|\phi^*)] \left[ \frac{\partial^2 f_n(y|\theta^*)}{\partial \theta \, \partial \theta'} \right] d\nu_n \right\} (\hat{\theta} - \theta^*) + o_p(1).
$$

$$
(5.15)
$$

Putting together the results (5.12)–(5.15), and using equation (2.4), we find that

$$\log[f_n(y|\hat{\theta})] - \log[g_n(y|\hat{\phi})]$$

$$- \left[ \int \{\log[f_n(y|\theta)] - \log[g_n(y|\phi)]\} f_n(y|\theta)\, d\nu_n \right]_{\theta=\hat{\theta}, \phi=\hat{\phi}}$$

$$= n(\hat{\phi} - \phi^*)'M'(\hat{\theta} - \theta^*) - n(\hat{\phi} - \phi^*)'J_1(\hat{\phi} - \phi^*) + o_p(1). \qquad (5.16)$$

Now, by Assumptions 10, 11, the result (5.9), and equation (3.15),

$$n^{1/2}\hat{h} \equiv n^{1/2}[\hat{\phi} - a(b(\hat{\phi}))] = [I - J_1^{-1}M'J_0^{-1}M]n^{1/2}(\hat{\phi} - \phi^*) + o_p(1)$$

$$= V_s J_1 n^{1/2}(\hat{\phi} - \phi^*) + o_p(1) \qquad (5.17)$$

while

$$n^{1/2}\hat{S} \equiv n^{1/2}[\hat{\phi} - a(\hat{\theta})] = n^{1/2}(\hat{\phi} - \phi^*) - J_1^{-1}M'n^{1/2}(\hat{\theta} - \theta^*) + o_p(1).$$

$$(5.18)$$

Using the results (5.9), (5.17), and (5.18), we find that

$$n^{1/2}(\hat{S}', \hat{h}')' \overset{d}{\to} N(\mu, V), \qquad (5.19)$$

where $\mu$ and $V$ are as given by equations (3.20) and (3.17) with (3.6)–(3.16).

Now, since $n^{1/2}\hat{S}$ is asymptotically normal with variance matrix $V_s$, we may write

$$n^{1/2}\hat{S} = n^{1/2}V_s V_s^+ \hat{S} + o_p(1), \qquad (5.20)$$

where $V_s^+$ is any positive semidefinite matrix such that $V_s V_s^+ V_s = V_s$.

Results (5.17), (5.19), and (5.20) imply that

$$(n^{1/2}\hat{h})'V_s^+(n^{1/2}\hat{S}) = n^{1/2}(\hat{\phi} - \phi^*)'J_1(n^{1/2}\hat{S}) + o_p(1). \qquad (5.21)$$

From results (5.18) and (5.21), it is readily seen that

$$-n\hat{h}'V_s^+\hat{S} = n(\hat{\phi} - \phi^*)'M'(\hat{\theta} - \theta^*) - n(\hat{\phi} - \phi^*)'J_1(\hat{\phi} - \phi^*) + o_p(1).$$

This finding, in conjunction with the result (5.16), completes the proof of the Theorem of Section 3. ∎

*REFERENCES*

1. Cox, D.R. Tests of separate families of hypotheses. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability* 1 (1961): 105–123.
2. Cox, D.R. Further results on tests of separate families of hypotheses. *Journal of the Royal Statistical Society, Series B,* 24 (1962): 406–424.
3. Davidson, R. & J.G. MacKinnon. Some non-nested hypothesis tests and the relations among them. *Review of Economic Studies* 49 (1982): 551–565.
4. Davidson, R. & J.G. MacKinnon. The interpretation of test statistics. *Canadian Journal of Economics* 18 (1985): 38–57.
5. Davidson, R. & J.G. MacKinnon. Implicit alternatives and the local power of test statistics. *Econometrica* 55 (1987): 1305–1329.

6. Godfrey, L.G. & M.H. Pesaran. Tests of non-nested regression models: Small-sample adjustments and Monte Carlo evidence. *Journal of Econometrics* 21 (1983): 133–154.
7. Gourieroux, C., A. Monfort & A. Trognon. Testing nested or non-nested hypotheses. *Journal of Econometrics* 21 (1983): 83–116.
8. Gourieroux, C., A. Monfort & A. Trognon. Pseudo maximum likelihood methods: Theory. *Econometrica* 52 (1984): 681–700.
9. Hausman, J. Specification tests in econometrics. *Econometrica* 46 (1978): 1251–1271.
10. Kent, J.T. The underlying structure of non-nested hypothesis tests. *Biometrika* 73 (1988): 333–343.
11. McAleer, M. Specification tests for separate models: A survey. In M.L. King & D.E.A. Giles (eds.), *Specification Analysis in the Linear Model*, Chapter 9 and pp. 146–196. London: Routledge and Kegan Paul, 1987.
12. Mizon, G.E. The encompassing approach in econometrics. In D.F. Hendry & K.F. Wallis (eds.), *Econometrics and Quantitative Economics*, Chapter 6 and pp. 135–172. Oxford: Basil Blackwell, 1984.
13. Mizon, G.E., & J.-F. Richard. The encompassing principle and its application to non-nested hypotheses. *Econometrica* 54 (1986): 657–678.
14. Pesaran, M.H. Global and partial non-nested hypotheses and asymptotic local power. *Econometric Theory* 3 (1987): 67–97.
15. Vuong, Q.H. Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica* 57 (1989): 307–333.
16. White, H. Maximum likelihood estimation of misspecified models. *Econometrica* 50 (1982): 1–25.
17. White, H. Regularity conditions for Cox's test of non-nested hypotheses. *Journal of Econometrics* 19 (1982): 301–318.