The Pseudo-True Score Encompassing Test for Non-Nested Hypotheses

Yi-Ting Chen

Institute of Social Science and Philosophy

Academia Sinica

Chung-Ming Kuan

Institute of Economics

Academia Sinica

First version: October 5, 1998

This version: January 26, 2000

The author for correspondence: C.-M. Kuan, Institute of Economics, Academia Sinica, Taipei 115, Taiwan; e-mail: ckuan@ieas.econ.sinica.edu.tw.

We thank Songnian Chen, Yongmiao Hong, Cheng Hsiao, and Lung-Fei Lee for very useful comments. We are also greateful to two anonymous referees and the associate editor whose comments and suggestions lead to a much improved version of this paper. All remaining errors are ours.

Abstract

Well known encompassing tests are usually difficult to implement because it is difficult to compute the pseudo-true value of the quasi-maximum likelihood estimator. In this paper, we propose a more operational encompassing test that does not involve such pseudo-true value. Instead, the proposed test relies on the "pseudo-true score" which is relatively easier to evaluate. We show that this test is asymptotically equivalent to the Wald and score encompassing tests and has a wider applicability than the conditional mean encompassing test of Wooldridge (1990a). Our simulations confirm that the proposed test compares favorably with the J and JA tests.

JEL Classification: C22, C52

Keywords: conditional mean encompassing test, encompassing principle, non-nested hypotheses, pseudo-true score

1 Introduction

In many econometric applications, there often exist non-nested specifications that can characterize the same variable of interest. For example, linear regression models with distinct sets of regressors are non-nested, and so are the logit and probit models for a binary dependent variable. In the time series context, the autoregressive (AR) and moving average (MA) models, the logistic and exponential smooth transition AR models, the bilinear and autoregressive conditional heteroskedasticity (ARCH) models, and the generalized ARCH (GARCH) and exponential GARCH (EGARCH) models are also pairs of non-nested models. When non-nested models are available, it is practically important to test if a postulated model is correctly specified.

The encompassing principle of Hendry and Richard (1982), Mizon (1984), and Mizon and Richard (1986) is a leading approach to derive non-nested tests under the framework of quasi-maximum likelihood; see also Gouriéroux, Monfort, and Trognon (1983). This principle asserts that a correctly specified model should be able to predict or explain the statistical results of its competitors. Thus, for a statistic of the alternative model, its probability limit under the null hypothesis should depend on the null model. This limit is known as the "pseudo-true value" of this statistic. An encompassing test is then based on the difference between a properly chosen statistic of the alternative model and the sample counterpart of its pseudo-true value. For example, the celebrated test of Cox (1961, 1962) is an encompassing test. Smith (1992) also constructed encompassing tests for models estimated by the generalized method of moments.

As the quasi-maximum likelihood estimate (QMLE) of a nonlinear model does not have a closed form in general, it is difficult to derive its pseudo-true value and compute the corresponding sample counterpart. Unfortunately, the well known Wald and score encompassing tests rely on such pseudo-true values and hence cannot be easily applied to nonlinear models. In practice, researchers prefer convenient non-nested tests, such as the J test of Davidson and MacKinnon (1981) and its finite-sample correction, the JA test of Fisher and McAleer (1981); see McAleer (1995) for a comprehensive review. To circumvent this difficulty, Wooldridge (1990a) proposed the conditional mean encompassing (CME) test which does not require estimating the pseudo-true value of the QMLE. The CME test focuses on conditional mean specifications, but it is not applicable to test non-nested specifications of other conditional moments. On the other hand, Pesaran and Pesaran (1993) suggested to estimate pseudo-true values by simulations; their approach is still cumbersome for practitioners, however.

In this paper we propose a more operational encompassing test that is asymptotically equivalent to the Wald and score encompassing tests. The proposed test is based on the sample counterpart of the "pseudo-true score", i.e., the limit of the expected value of the score function from the alternative model, where the expectation is taken with respect to the null model. Similar to the CME test, the proposed test does not depend on the pseudo-true value of the QMLE from the alternative model. As the score function, unlike the QMLE, usually has an analytic expression, its expectation is relatively easy to evaluate. Our test can therefore be implemented quite easily. We also demonstrate that the proposed test is applicable to censored (truncated) regressions and conditional variance specifications but the CME test is not. Our simulations indicate that this test has good finite sample performance and compares favorably with the J and JA tests.

This paper proceeds as follows. In Section 2, we propose the new encompassing test based on the pseudo-true score. We then discuss various applications of the proposed test and its relationship with the CME test in Section 3. Some simulation results are reported in Section 4. Section 5 concludes the paper.

2 The Pseudo-True Score Encompassing Test

Let $\eta_t = (y_t, w_t')'$ denote a vector of observations at time t and $\mathfrak{F}_{t-1}$ denote the sigma algebra generated by $(w_t, \eta_{t-1}, \dots, \eta_1)$. Suppose that there are two competing specifications for the density of y_t conditional on $\mathfrak{F}_{t-1}$:

$$\mathrm{M}_f : f_t(y_t|\mathfrak{F}_{t-1}, \alpha), \quad \alpha \in \mathcal{A} \subseteq \mathbb{R}^p,$$

$$\mathrm{M}_g : g_t(y_t|\mathfrak{F}_{t-1}, \beta), \quad \beta \in \mathcal{B} \subseteq \mathbb{R}^q,$$

where f_t and g_t are postulated density functions. We are interested in testing the null hypothesis that M_f is correctly specified, in the sense that there is an $\alpha_o \in \mathcal{A}$ such that $f_t(y_t|\mathfrak{F}_{t-1}, \alpha_o)$ is the true conditional density of y_t for every t, against the alternative hypothesis of M_g model. If f_t and g_t are the same conditional density function and $\mathcal{A} \subseteq \mathcal{B}$, then the null hypothesis is said to be nested in the alternative hypothesis (or equivalently, M_f is nested in M_g). If these two specifications are not nested in each other, they are said to be non-nested.

2.1 Notations

Let $S_{fT}(\alpha)$ and $S_{gT}(\beta)$ denote, respectively, the score functions of M_f and M_g:

$$S_{fT}(\alpha) := \frac{1}{T}\sum_{t=1}^{T} \nabla_\alpha \log f_t(y_t|\mathfrak{F}_{t-1}, \alpha),$$

$$S_{gT}(\beta) := \frac{1}{T}\sum_{t=1}^{T} \nabla_\beta \log g_t(y_t|\mathfrak{F}_{t-1}, \beta),$$

where $\nabla_\alpha \log f_t$ is the $p \times 1$ gradient vector of $\log f_t$ with respect to α, and $\nabla_\beta \log g_t$ is the $q \times 1$ gradient vector of $\log g_t$ with respect to β. The resulting QMLEs are $\hat{\alpha}_T$ and $\hat{\beta}_T$, so that $S_{fT}(\hat{\alpha}_T) = 0$, and $S_{gT}(\hat{\beta}_T) = 0$. Also let $H_{fT}(\alpha) = \nabla_{\alpha'} S_{fT}(\alpha)$ and $H_{gT}(\beta) = \nabla_{\beta'} S_{gT}(\beta)$ denote the corresponding Hessian matrices.

In what follows, we will always assume that regularity conditions ensuring the consistency and asymptotic normality of the QMLEs hold and that the information matrix equality holds under the null hypothesis; see e.g., White (1994) for more detailed discussion of those conditions. We will denote $\mathrm{plim}_{f(\alpha)}$ as the probability limit under M_f. Similarly, $\mathbb{E}_{f(\alpha)}$, $\mathrm{var}_{f(\alpha)}$, and $\mathrm{cov}_{f(\alpha)}$ are, respectively, the operators of expectation, variance and covariance, taken with respect to M_f. If an operator is given without subscript, it is taken with respect to the true data generating process. The limiting expected Hessian matrices under M_f are:

$$H_f(\alpha) := \lim_{T\to\infty} \mathbb{E}_{f(\alpha)}[H_{fT}(\alpha)],$$

$$H_g(\alpha, \beta) := \lim_{T\to\infty} \mathbb{E}_{f(\alpha)}[H_{gT}(\beta)].$$

The limiting information matrices under M_f are:

$$B_f(\alpha) := \lim_{T\to\infty} \mathrm{var}_{f(\alpha)}[\sqrt{T} S_{fT}(\alpha)],$$

$$B_g(\alpha, \beta) := \lim_{T\to\infty} \mathrm{var}_{f(\alpha)}[\sqrt{T} S_{gT}(\beta)].$$

Also denote

$$K_{fg}(\alpha, \beta) := \lim_{T\to\infty} \mathrm{cov}_{f(\alpha)}[\sqrt{T} S_{fT}(\alpha), \sqrt{T} S_{gT}(\beta)],$$

and $K_{gf}(\alpha, \beta) := K'_{fg}(\alpha, \beta)$.

For the QMLE $\hat{\beta}_T$ of the alternative model M_g, let

$$\beta(\alpha) = \mathrm{plim}_{f(\alpha)}\ \hat{\beta}_T.$$

Note that $\beta(\alpha)$ is the minimizer of the limit of the average Kullback-Leibler information criterion and therefore solves

$$\lim_{T\to\infty} \mathbb{E}_{f(\alpha)}[S_{gT}(\beta)] = 0; \tag{1}$$

see e.g., White (1994). In particular, $\beta(\alpha_o)$ is the probability limit of $\hat{\beta}_T$ under the null hypothesis of $f(\alpha_o)$ and solves (1) when the expectation is taken with respect to $f(\alpha_o)$. As such, $\beta(\alpha_o)$ is known as the pseudo-true value of $\hat{\beta}_T$ and will be referred to as the "pseudo-true parameter". When $\beta(\alpha_o)$ is approximated by $\beta(\hat{\alpha}_T)$, let $\hat{\beta}(\hat{\alpha}_T)$ denote its sample counterpart.

Under the null hypothesis, $\hat{H}_{fT} = H_{fT}(\hat{\alpha}_T)$ is consistent for $H_{fo} := H_f(\alpha_o)$, and $\hat{H}_{gT} = H_{gT}(\hat{\beta}_T)$ is consistent for

$$H_{go} := H_g(\alpha_o, \beta(\alpha_o)).$$

We also write $\hat{B}_{fT}$ and $\hat{B}_{gT}$ as estimators for B_f and B_g based on the QMLEs $\hat{\alpha}_T$ and $\hat{\beta}_T$, respectively. Then, under the null hypothesis, $\hat{B}_{fT}$ is consistent for $B_{fo} := B_f(\alpha_o)$, and $\hat{B}_{gT}$ is consistent for

$$B_{go} := B_g(\alpha_o, \beta(\alpha_o)).$$

Similarly, $\hat{K}_{fg}$ is an estimator based on $\hat{\alpha}_T$ and $\hat{\beta}_T$, and it is consistent for

$$K_{fgo} := K_{fg}(\alpha_o, \beta(\alpha_o)),$$

under the null hypothesis. Consistent estimation of those matrices has been studied by, e.g., White (1984), Newey and West (1987), and Andrews (1991), among others.

2.2 The Proposed Test

Both the Wald and score encompassing tests depend on $\hat{\beta}(\hat{\alpha}_T)$, the sample counterpart of the pseudo-true parameter $\beta(\alpha_o)$. In particular, the Wald encompassing test checks if

$$\hat{H}_{gT}\sqrt{T}(\hat{\beta}_T - \hat{\beta}(\hat{\alpha}_T)) \tag{2}$$

is sufficiently close to zero. The score encompassing test checks if $\sqrt{T}S_{gT}(\hat{\beta}(\hat{\alpha}_T))$ is sufficiently close to zero. When the alternative model is nonlinear such that $\hat{\beta}_T$ does not have a closed form, it is typically difficult to evaluate $\beta(\alpha_o)$ and therefore its sample counterpart. This difficulty prevents researchers from applying those encompassing tests to nonlinear models and motivates the CME test of Wooldridge (1990a) and the simulation method of Pesaran and Pesaran (1993).

In contrast with traditional encompassing tests, we now take the score function as the statistic of interest and consider its pseudo-true value. Let

$$\Psi_g(\alpha,\beta) := \lim_{T\to\infty} \mathbb{E}_{f(\alpha)}[S_{gT}(\beta)].$$

Then $\Psi_g(\alpha_o,\beta)$ may be interpreted as the "pseudo-true score" because it is obtained from the expectation with respect to the true density function $f(\alpha_o)$. In view of (1),

$$\Psi_g(\alpha_o,\beta(\alpha_o)) = 0,$$

i.e., the pseudo-true score evaluated at $\beta(\alpha_o)$ is zero. It is clear that under the null hypothesis, $\Psi_g(\hat{\alpha}_T,\hat{\beta}_T)$ should be close to $\Psi_g(\alpha_o,\beta(\alpha_o))$. In accordance with the encompassing principle, an encompassing test can be constructed by checking if $\Psi_g(\hat{\alpha}_T,\hat{\beta}_T)$ is sufficiently close to zero.

It can be shown that under the null hypothesis,

$$\sqrt{T}\Psi_g(\hat{\alpha}_T,\hat{\beta}_T) = \hat{H}_{gT}\sqrt{T}[\hat{\beta}_T - \beta(\hat{\alpha}_T)] + o_p(1); \tag{3}$$

a detailed derivation of this result is given in the Appendix. Note that Ψ_g is not observable because it is the limit of the expectation of S_{gT}. Replacing Ψ_g and $\beta(\hat{\alpha}_T)$ in (3) by their finite sample counterparts: $\hat{\Psi}_g$ and $\hat{\beta}(\hat{\alpha}_T)$, we have

$$\sqrt{T}\hat{\Psi}_g(\hat{\alpha}_T,\hat{\beta}_T) = \hat{H}_{gT}\sqrt{T}[\hat{\beta}_T - \hat{\beta}(\hat{\alpha}_T)] + o_p(1), \tag{4}$$

where the right-hand side is the basis of the Wald encompassing test (2). It is well known that

$$\hat{H}_{gT}\sqrt{T}[\hat{\beta}_T - \hat{\beta}(\hat{\alpha}_T)] \stackrel{A}{\sim} N(0,\Omega_o),$$

where $\Omega_o = B_{go} - K_{gfo}B_{fo}^{-1}K_{fgo}$; see e.g., Mizon and Richard (1986). It follows that $\sqrt{T}\hat{\Psi}_g(\hat{\alpha}_T,\hat{\beta}_T)$ is also asymptotically distributed as $N(0,\Omega_o)$.

The proposed Pseudo-true Score Encompassing (PSE) test statistic is then

$$\mathfrak{S}_T = T\hat{\Psi}_g(\hat{\alpha}_T,\hat{\beta}_T)'\hat{\Omega}^{-}\hat{\Psi}_g(\hat{\alpha}_T,\hat{\beta}_T), \tag{5}$$

where $\hat{\Omega}$ is a consistent estimator for Ω_o, and $\hat{\Omega}^-$ is its generalized inverse. Note that for a matrix A, its generalized inverse satisfies $AA^-A = A$; see e.g., Rao and Mitra (1971). The asymptotic normality result above implies

$$\mathfrak{S}_T \overset{A}{\sim} \chi^2(r),$$

where r is the rank of $\hat{\Omega}$. In view of (4) and (2), the PSE test is asymptotically equivalent to the Wald and score encompassing tests. If one is only interested in testing some partial specification (such as the conditional mean or conditional variance specification), the proposed test can be implemented using only the corresponding sub-vector of the pseudo-true score.

Comparing to the score encompassing test which is based on the score evaluated at the estimated pseudo-true parameter $\hat{\beta}(\hat{\alpha}_T)$, the proposed test relies on the pseudo-true score evaluated at $\hat{\alpha}_T$ and $\hat{\beta}_T$. The advantages of our approach are obvious. First, the score function usually has an analytic form so that its pseudo-true value is relatively easy to derive. Second, by considering the pseudo-true score function we are able to eliminate the pseudo-true parameter because $\Psi_g(\alpha_o, \beta(\alpha_o)) = 0$ in the linear expansion, as shown in the Appendix. Therefore, there is no need to estimate the pseudo-true parameter, and the proposed test depends only on the QMLEs from the null and alternative models but not on $\hat{\beta}(\hat{\alpha}_T)$. These advantages make the proposed test more operational than existing encompassing tests, as can be seen in Section 3.

Finally we note that a consistent estimator of Ω_o is

$$\hat{\Omega} = \hat{B}_{gT} - \hat{K}_{gf}\hat{B}_{fT}^{-1}\hat{K}_{fg}.$$

This is also the estimator for Ω_o in the Wald encompassing test and hence an "unconstrained" estimator. More directly, we may estimate Ω_o using the sample counterpart of $\mathrm{var}(\sqrt{T}\hat{\Psi}_g(\hat{\alpha}_T, \hat{\beta}_T))$. Such an estimator usually takes into account the specification of M_f and can be viewed as a "constrained" estimator for Ω_o.

3 Applications

In this section we discuss various applications of the proposed PSE test and its relationship with the CME test. In particular, our examples sequentially show the following: (1) the PSE test and the CME test are the same in standard nonlinear regressions; (2)

the PSE test is a natural test for binary choice models and asymptotically equivalent to a particular CME test; (3) the PSE test is valid in censored and truncated regressions, but the CME test is not; (4) the PSE test can be applied to test conditional variance specifications, but the CME test cannot. In what follows, let x_t and z_t be two distinct vectors containing elements of $(w_t, \eta_{t-1}, \ldots, \eta_1)$ such that they are not subvectors of each other.

3.1 Nonlinear Regressions

Suppose that there are two specifications of the conditional distribution of y_t:

$$\mathrm{M}_f : y_t|\mathfrak{F}_{t-1} \sim N(m_t(x_t, \gamma), \sigma_f^2),$$
$$\mathrm{M}_g : y_t|\mathfrak{F}_{t-1} \sim N(\mu_t(z_t, \delta), \sigma_g^2).$$

Apart from normality, they are virtually specifications of the conditional mean of y_t. Alternatively, we can write them as nonlinear regressions:

$$\begin{aligned} \mathrm{M}_f :&\ \ y_t = m_t(x_t, \gamma) + \varepsilon_t, \\ \mathrm{M}_g :&\ \ y_t = \mu_t(z_t, \delta) + \nu_t, \end{aligned} \tag{6}$$

where the conditional distributions of ε_t and ν_t are, respectively, $N(0, \sigma_f^2)$ and $N(0, \sigma_g^2)$. The parameter vector of M_f is $\alpha = (\gamma', \sigma_f^2)'$, and the parameter vector of M_g is $\beta = (\delta', \sigma_g^2)'$. For notational simplicity, we write $m_t(x_t, \gamma)$ as m_t, $m_t(x_t, \gamma_o)$ as m_{ot}, and $m(x_t, \hat{\gamma}_T)$ as $\hat{m}_t$; μ_t, μ_{ot}, and $\hat{\mu}_t$ are similarly defined.

The score function of M_g with respect to δ is

$$S_{g\delta}(\beta) = \frac{1}{T\sigma_g^2} \sum_{t=1}^{T} \nabla_\delta \mu_t [m_{ot} + \varepsilon_t - \mu_t].$$

When M_f is correctly specified, ε_t is orthogonal to all random variables that are $\mathfrak{F}_{t-1}$-measurable, so that the PSE test of conditional mean specifications is based on

$$\hat{\Psi}_{g\delta}(\hat{\alpha}_T, \hat{\beta}_T) = \frac{1}{T\hat{\sigma}_g^2} \sum_{t=1}^{T} \nabla_\delta \hat{\mu}_t [\hat{m}_t - \hat{\mu}_t] = -\frac{1}{T\hat{\sigma}_g^2} \sum_{t=1}^{T} \nabla_\delta \hat{\mu}_t \hat{\varepsilon}_t, \tag{7}$$

where $\hat{\varepsilon}_t = y_t - \hat{m}_t$.

On the other hand, the CME test relies on the nonlinear least squares (NLS) estimators $\tilde{\gamma}_T$ and $\tilde{\delta}_T$. Similar to previous notations, we write $\tilde{m}_t$ and $\tilde{\mu}_t$ for m_t and μ_t

evaluated at the NLS estimators. The key ingredient of the CME test is

$$-\frac{1}{T}\sum_{t=1}^{T}\nabla_\delta\tilde{\mu}_t\tilde{\varepsilon}_t, \tag{8}$$

where $\tilde{\varepsilon}_t = y_t - \tilde{m}_t$. As the NLS estimator $\tilde{\gamma}_T$ ($\tilde{\delta}_T$) is also the QMLE $\hat{\gamma}_T$ ($\hat{\delta}_T$), (8) differs from (7) only by the scaling factor $1/\hat{\sigma}_g^2$ which may be ignored in computing the test statistic. Thus, the PSE test obtained under conditional normality and homoskedasticity is the same as the CME test; they are different otherwise.

Instead of directly testing (8), Wooldridge (1990a) further suggests to check if $\tilde{\varepsilon}_t$ are uncorrelated to the residuals of the multivariate regression that projects $\nabla_\delta\tilde{\mu}_t$ onto $\nabla_\gamma\tilde{m}_t$. The resulting test is robust in the sense that its asymptotic null distribution does not depend on additional second-moment assumption. That is, the robustified CME test is valid whether y_t are conditionally homoskedastic or not.

In the time series context, for example, one may want to distinguish between the logistic and exponential smooth transition AR models because both of them are capable of describing similar nonlinear characteristics. For this purpose, it is straightforward to derive the PSE (CME) test which can serve as an alternative to the artificial nested test of Teräsvirta and Anderson (1992).

3.2 Binary Choice Models

Let y_t be the indicator variable of the unobservable, endogenous variable y_t^* such that $y_t = 1$ if $y_t^* > 0$ and $y_t = 0$ otherwise. Consider two specifications of the conditional mean of y_t^*:

$$\mathrm{M}_f: \; y_t^* = m_t(x_t,\gamma) + \varepsilon_t,$$

$$\mathrm{M}_g: \; y_t^* = \mu_t(z_t,\delta) + \nu_t.$$

where ε_t and ν_t have conditional cumulative distribution functions $F_t(\cdot;\vartheta)$ and $G_t(\cdot;\varrho)$, respectively. It is often postulated that F_t and G_t are symmetric about the origin and do not change with t so that $1 - F_t(-m_t;\vartheta) = F(m_t;\vartheta)$. Thus, the conditional mean of y_t under M_f is

$$\mathbb{E}_{f(\alpha)}[y_t|\mathfrak{F}_{t-1}] = 1 - F_t(-m_t;\vartheta) = F(m_t;\vartheta),$$

where $\alpha = (\gamma', \vartheta')'$. Similarly, the condition mean under M_g is $G(\mu_t; \varrho)$. We can also express M_f and M_g as two nonlinear regressions:

$$\begin{aligned} \mathrm{M}_f: \; & y_t = F(m_t; \vartheta) + u_t, \\ \mathrm{M}_g: \; & y_t = G(\mu_t; \varrho) + v_t. \end{aligned} \tag{9}$$

Note that u_t and v_t in (9) cannot be normally distributed and that they have conditional variances $F(1-F)$ and $G(1-G)$; cf. (6).

Given the quasi-log-likelihood function of M_g:

$$L_{gT}(\beta) = \frac{1}{T}\sum_{t=1}^{T} y_t \log[G(\mu_t; \varrho)] + \frac{1}{T}\sum_{t=1}^{T}(1-y_t)\log[1 - G(\mu_t; \varrho)],$$

where $\beta = (\delta', \varrho')'$, we have

$$\begin{aligned} \mathbb{E}_{f(\alpha_o)}[S_{g\delta}(\beta)] &= \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}_{f(\alpha_o)}\left[\nabla_\mu G(\mu_t; \varrho)\nabla_\delta \mu_t \frac{F(m_{ot}; \vartheta_o) - G(\mu_t; \varrho)}{G(\mu_t; \varrho)(1 - G(\mu_t; \varrho))}\right], \\ \mathbb{E}_{f(\alpha_o)}[S_{g\varrho}(\beta)] &= \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}_{f(\alpha_o)}\left[\nabla_\varrho G(\mu_t; \varrho) \frac{F(m_{ot}; \vartheta_o) - G(\mu_t; \varrho)}{G(\mu_t; \varrho)(1 - G(\mu_t; \varrho))}\right]. \end{aligned}$$

Similar to previous notations, we write $\hat{F}_t$, $\hat{G}_t$, $\nabla_\mu \hat{G}_t$, $\nabla_\delta \hat{\mu}_t$ and $\nabla_\varrho \hat{G}_t$ when the corresponding functions are evaluated at the QMLEs. The PSE test is then based on

$$\hat{\Psi}_g(\hat{\alpha}_T, \hat{\beta}_T) = \begin{bmatrix} \frac{1}{T}\sum_{t=1}^{T} \nabla_\mu \hat{G}_t \nabla_\delta \hat{\mu}_t [\hat{F}_t - \hat{G}_t]/[\hat{G}_t(1 - \hat{G}_t)] \\ \frac{1}{T}\sum_{t=1}^{T} \nabla_\varrho \hat{G}_t [\hat{F}_t - \hat{G}_t]/[\hat{G}_t(1 - \hat{G}_t)]. \end{bmatrix}. \tag{10}$$

Two leading binary choice models are the probit and logit models which correspond to the standard normal and logistic distribution functions, respectively. For these models, ϑ and ϱ are known; hence the PSE test involves only the first component of (10).

Let $\tilde{\gamma}_T$ and $\tilde{\delta}_T$ be the NLS estimators of (9). The key ingredient of the CME test is

$$\frac{1}{T}\sum_{t=1}^{T} \nabla_\mu \tilde{G}_t \nabla_\delta \tilde{\mu}_t [\tilde{F}_t - \tilde{G}_t], \tag{11}$$

where $\tilde{F}_t$, $\tilde{G}_t$, $\nabla_\mu \tilde{G}_t$, and $\nabla_\delta \tilde{\mu}_t$ are the functions evaluated at the NLS estimators. This differs from the first component of (10) in two respects. First, the NLS estimators are different from the corresponding QMLEs. Second, (11) does not involve the estimated conditional variance of y_t. As such, the PSE test and the NLS-based CME test need not be the same when the conditional distributions of (9) are non-normal or heteroskedastic. Although NLS estimation here ignores conditional heteroskedasticity of data, the robust procedure of the CME test still yields a limiting chi-squared null distribution (but is not asymptotically efficient).

As shown in Wooldridge (1990a), the CME test also admits weighted NLS (WNLS) estimators whose weight functions do not have to be correctly specified conditional variance functions. The WNLS-based CME test can also be robustified against potential misspecifications of the conditional variance function. For binary dependent variables, their conditional variances have specific forms. With properly estimated conditional variances as weight functions, the WNLS estimators are asymptotically equivalent to the QMLEs. Therefore, the CME test based on these particular WNLS estimators and the PSE test are asymptotically equivalent.

From the examples in these two subsections we can see that the CME test does not require auxiliary assumptions other than conditional mean specifications. Hence, it is particularly useful when conditional densities are not (or cannot be) completely specified. In some cases, data characteristics may have implications on conditional moments. For example, the specific form of the conditional variance in binary choice models is mainly an implication of binary data and hence is not an ad hoc assumption. This differs from, say, Poisson regressions because the specific form of the conditional variance in Poisson regressions is not necessarily a consequence of count data. In the former case, we can incorporate data characteristics into the likelihood function, and the resulting PSE test is a natural choice, even though an asymptotically equivalent CME test is also available. In the latter case, a WNLS-based CME test may be reasonable if one is reluctant to believe that the conditional mean and variance are the same.

3.3 Censored and Truncated Regressions

In this subsection we take the Tobit model as an example and derive the PSE test. It is interesting to note that while the PSE test is still a natural choice for censored and truncated regressions, the CME test is inappropriate because it ignores data characteristics.

Suppose that the endogenous variable y_t^* is censored such that the observed variable $y_t = y_t^*$ when $y_t^* > 0$ and $y_t = 0$ otherwise. Consider two competing (conditional) mean specifications for y_t^*:

$$\mathrm{M}_f \ : \ y_t^* = m(x_t, \gamma) + \varepsilon_t,$$

$$\mathrm{M}_g \ : \ y_t^* = \mu(z_t, \delta) + \nu_t,$$

which can also be expressed as

$$\mathrm{M}_f \ : \ y_t = \begin{cases} m(x_t, \gamma) + \varepsilon_t, & \text{if } \varepsilon_t > -m(x_t, \gamma), \\ 0, & \text{if } \varepsilon_t \leq -m(x_t, \gamma), \end{cases}$$

$$\mathrm{M}_g \ : \ y_t = \begin{cases} \mu(z_t, \delta) + \nu_t, & \text{if } \nu_t > -\mu(z_t, \delta), \\ 0, & \text{if } \nu_t \leq -\mu(z_t, \delta). \end{cases}$$

These are two Tobit models when ε_t and ν_t are assumed to be i.i.d. $N(0, \sigma_f^2)$ and $N(0, \sigma_g^2)$, respectively.

Let ϕ and Φ be the density and cumulative distribution functions of the standard normal random variables, respectively. The quasi-likelihood function of M_g is

$$\prod_{t=1}^{T} \left[\frac{1}{\sigma_g} \phi \left(\frac{y_t - \mu(z_t, \delta)}{\sigma_g} \right) \right]^{\lambda_t} \left[1 - \Phi \left(\frac{\mu(z_t, \delta)}{\sigma_g} \right) \right]^{(1-\lambda_t)},$$

where $\lambda_t = 1$ when $y_t^* > 0$ and $\lambda_t = 0$ otherwise. The score function of the average quasi-log-likelihood function is

$$S_{g\delta}(\beta) = \frac{1}{T\sigma_g^2} \sum_{t=1}^{T} \nabla_\delta \mu_t \left[\lambda_t (y_t - \mu_t) - (1 - \lambda_t) \frac{\sigma_g \phi(\mu_t/\sigma_g)}{1 - \Phi(\mu_t/\sigma_g)} \right].$$

It is shown in the Appendix that the key ingredient of the PSE test is

$$\begin{aligned} \hat{\Psi}(\hat{\alpha}_T, \hat{\beta}_T) = \frac{1}{T\hat{\sigma}_{gT}^2} \sum_{t=1}^{T} \{ \nabla_\delta \hat{\mu}_t \, [(\hat{m}_t - \hat{\mu}_t) \Phi(\hat{m}_t/\hat{\sigma}_{fT}) + \hat{\sigma}_{fT} \phi(\hat{m}_t/\hat{\sigma}_{fT}) \\ - \ \hat{\sigma}_{gT} \phi(\hat{\mu}_t/\hat{\sigma}_{gT}) \frac{1 - \Phi(\hat{m}_t/\hat{\sigma}_{fT})}{1 - \Phi(\hat{\mu}_t/\hat{\sigma}_{gT})} \bigg] \bigg\}. \end{aligned} \tag{12}$$

By taking m_t (μ_t) as the specification of the conditional mean of y_t, we could also compute a CME test. It is clear that this cannot be the same as the PSE test. In fact, it is well known that least-squares estimation is not suitable for censored (truncated) regressions because it ignores data censoring (truncation). Thus, the CME test and its robustified version, whether it is NLS-based or WNLS-based, is not valid in these cases.

3.4 Conditional Variance Specifications

In addition to conditional mean functions, econometricians may also be interested in modelling other aspects of conditional distributions, such as conditional variance and conditional skewness. While the PSE test for such non-nested specifications can be easily obtained, the CME test, by construction, cannot be applied.

Consider two general specifications for both conditional mean and conditional variance:

$$\mathrm{M}_f: \; y_t = m_t(x_t, \gamma) + \varepsilon_t, \quad \mathrm{var}[\varepsilon_t|\mathfrak{F}_{t-1}] = h_t(x_t, \vartheta),$$

$$\mathrm{M}_g: \; y_t = \mu_t(z_t, \delta) + \nu_t, \quad \mathrm{var}[\nu_t|\mathfrak{F}_{t-1}] = k_t(z_t, \varrho),$$

where $h_t(x_t, \vartheta)$ and $k_t(z_t, \varrho)$ are the conditional variance functions of y_t under M_f and M_g, respectively. For example, h_t may be a GARCH specification, whereas k_t may be an EGARCH specification.

Suppose that $\varepsilon_t|\mathfrak{F}_t \sim N(0, h_t)$ and $\nu_t|\mathfrak{F}_t \sim N(0, k_t)$. Let $\alpha = (\gamma', \vartheta')'$ and $\beta = (\delta', \varrho')'$. For M_f, we can write

$$\varepsilon_t^2 = h_t(x_t, \vartheta) + u_t.$$

Note that when M_f with $\alpha = \alpha_o$ is correctly specified, $\mathbb{E}_{f(\alpha_o)}[\varepsilon_t|\mathfrak{F}_{t-1}] = 0$ and $\mathbb{E}_{f(\alpha_o)}[u_t|\mathfrak{F}_{t-1}] = 0$. Therefore, we have

$$\mathbb{E}_{f(\alpha_o)}[S_{g\delta}(\beta)] = \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}_{f(\alpha_o)}\left[\nabla_\delta \mu_t [m_{ot} - \mu_t]/k_t\right],$$

$$\mathbb{E}_{f(\alpha_o)}[S_{g\varrho}(\beta)] = \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}_{f(\alpha_o)}\left[\nabla_\varrho k_t \left[(m_{ot} - \mu_t)^2 + h_{ot} - k_t\right]/2k_t^2\right],$$

where h_{ot}, k_t, and $\nabla_\varrho k_t$ denote, respectively, $h_t(x_t, \vartheta_o)$, $k_t(z_t, \varrho)$, and $\nabla_\varrho k_t(z_t, \varrho)$. We also let $\hat{h}_t$, $\hat{k}_t$, and $\nabla_\varrho \hat{k}_t$ be $h_t(x_t, \hat{\vartheta}_T)$, $k_t(z_t, \hat{\varrho}_T)$, and $\nabla_\varrho k_t(z_t, \hat{\varrho}_T)$. The PSE test then depends on

$$\hat{\Psi}_g(\hat{\alpha}_T, \hat{\beta}_T) = \begin{bmatrix} \frac{1}{T}\sum_{t=1}^{T} \nabla_\delta \hat{\mu}_t \left[\hat{m}_t - \hat{\mu}_t\right]/\hat{k}_t \\ \frac{1}{2T}\sum_{t=1}^{T} \nabla_\varrho \hat{k}_t \left[(\hat{m}_t - \hat{\mu}_t)^2 + \hat{h}_t - \hat{k}_t\right]/\hat{k}_t^2 \end{bmatrix}. \tag{13}$$

When only the conditional variance specifications concern us, the conditional means are usually specified as the same function in M_f and M_g. Then, the first component of (13) becomes zero and is not needed, and the second component simplifies to

$$\hat{\Psi}_{g\varrho}(\hat{\alpha}_T, \hat{\beta}_T) = \frac{1}{2T}\sum_{t=1}^{T} \nabla_\varrho \hat{k}_t(\hat{h}_t - \hat{k}_t)/\hat{k}_t^2. \tag{14}$$

The PSE test based on (14) can be applied to distingish between competing ARCH and GARCH specifications and provides a useful alternative to the test of Engle and Ng (1993). Moreoever, the term $(\hat{h}_t - \hat{k}_t)/\hat{k}_t$ in (14) can be viewed as a generalized residual, so that the PSE test can also be robustified along the line in Wooldridge (1990b).

4 Simulations

In this section we report some simulation results for linear regression models and binary choice models. We consider only the PSE test for conditional mean; the competing tests are the CME, J, and JA tests. In the experiments below, the sample sizes are $T = 50, 100$, the numbers of regressors are $p = 3, 5$, the number of replications is $R = 1000$, and the nominal size is 5%.

We first consider linear regression models with the data generating process (DGP):

$$y_t = w_{t,1} + \cdots + w_{t,p} + a_t, \qquad a_t \sim \text{i.i.d. } N(0,1),$$

where $w_{t,i} = (1-\lambda)x_{t,i} + \lambda z_{t,i}$, and $x_{t,i}$ and $z_{t,i}$ are all i.i.d. $N(0,1)$ random variables. The non-nested linear models are

$$\mathrm{M}_f: \; y_t = \sum_{i=1}^{p} \gamma_i x_{t,i} + \varepsilon_t, \quad \varepsilon_t|\mathfrak{F}_{t-1} \sim N(0, \sigma_f^2),$$

$$\mathrm{M}_g: \; y_t = \sum_{i=1}^{p} \delta_i z_{t,i} + \nu_t, \quad \nu_t|\mathfrak{F}_{t-1} \sim N(0, \sigma_g^2).$$

We then assess the test performance by setting $\lambda = -1, -0.9, \ldots, 0.9, 1$ (with increments 0.1). When $\lambda = 0$, M_f is correctly specified, so that the rejection frequency of a test is its empirical size. For other values of λ, the rejection frequencies of a test are empirical powers against the alternatives that deviate from the null to certain extent; in particular, $\lambda = 1$ implies M_g is correctly specified. As these models are conditional mean specifications, the proposed test is essentially the same as the CME test.

In Figures 1–4 we plot the rejection frequencies against different λ values. In Figures 1 and 2, the numbers of regressors are $p = 3$ and $p = 5$, respectively, and the sample size is $T = 50$. Figures 3 and 4 also correspond to $p = 3$ and $p = 5$, respectively, but are based on $T = 100$. We first observe that all but the J test have roughly correct sizes for $T = 50$ and 100. Hence, the power performance of the J test is exaggerated. Although the JA test indeed corrects the size problem of the J test, it also loses power as λ deviates from zero and performs much worse than the PSE test. Comparing to the CME test, the PSE test performs slightly better when $T = 50$ but has similar powers when $T = 100$.

We then consider the probit model (M_f) vs. the logit model (M_g), as discussed in Section 3.2. The DGP is

$$y_t^* = w_{t,1} + \ldots + w_{t,p} + a_t,$$

where $w_{t,i} = (1-\lambda)x_{t,i} + \lambda z_{t,i}$, $a_t = (1-\lambda)\varepsilon_t + \lambda\nu_t$, ε_t are i.i.d. $N(0,1)$ random variables and ν_t are i.i.d. logistic random variables. The models for y_t^* are linear:

$$\mathrm{M}_f: \; y_t^* = \sum_{i=1}^{p} \gamma_i x_{t,i} + \varepsilon_t,$$

$$\mathrm{M}_g: \; y_t^* = \sum_{i=1}^{p} \delta_i z_{t,i} + \nu_t.$$

We again consider $\lambda = -1, -0.9, \ldots, 0.9, 1$. In particular, M_f is correctly specified when $\lambda = 0$.

We also plot rejection frequencies against different λ values in Figures 5 ($p = 3, T = 50$), 6 ($p = 5, T = 50$), 7 ($p = 3, T = 100$), and 8 ($p = 5, T = 100$). It is easily seen that the J test again has very serious size distortions; hence its power performance is also exaggerated. In contrast with the simulations of the linear model, the JA test is still over-sized, especially for $p = 5$ and $T = 100$, but it also has very poor power performance. The PSE test maintains correct size in all cases, whereas the CME test may be under-sized when $p = 5$ and $T = 50$. In terms of powers, the PSE test performs similarly to the CME test for $p = 3$ but performs slightly better for $p = 5$. This suggests that the proposed test is relatively more stable for models of different complexity.

From these simulations, we find that the proposed test outperforms the J and JA tests but performs similarly to the CME test. This similarity is not surprising because the PSE test for conditional mean specifications and the CME test are quite close, as discussed in Sections 3.1 and 3.2. Note again that the CME test is not applicable to non-nested specifications of other distribution characteristics.

5 Conclusions

In this paper an operational encompassing test based on the pseudo-true score is proposed. An important feature of this test is that it does not involve the pseudo-true parameter of the alternative model; only the QMLEs of both the null and alternative models are needed to implement this test. We show that the proposed test can be conveniently applied to test various aspects of the postulated conditional distributions and therefore may be viewed as an extension of the CME test of Wooldridge (1990a). Our simulations indicate that the proposed test has correct sizes in all cases considered and compares favorably with other non-nested tests. Thus, the proposed test can serve as a useful complement to existing non-nested tests.

Finally, we note that, as the score encompassing test relies on the score evaluated at the estimated pseudo-true parameter, it is actually a parameter encompassing test based on score. By contrast, the proposed test checks if the pseudo-true score is close to zero and hence is truly a test of score encompassing.

Appendix

Proof of Equation (3): Let $f^T(y^T,\alpha) := \prod_{t=1}^T f_t(y_t|\mathfrak{F}_{t-1},\alpha)$ with $y^T = (y_1,\dots,y_T)$ and

$$\psi_{gT}(\alpha,\beta) := \mathbb{E}_{f(\alpha)}[S_{gT}(\beta)].$$

The limit of $\psi_{gT}(\alpha,\beta)$ is $\Psi_g(\alpha,\beta)$. Taking linear expansion of ψ_{gT} about $\alpha^\dagger$ and $\beta^\dagger$ yields

$$\begin{aligned}\psi_{gT}(\alpha,\beta) = \psi_{gT}(\alpha^\dagger,\beta^\dagger) &+ \nabla_{\alpha'}\psi_{gT}(\alpha^\dagger,\beta)(\alpha-\alpha^\dagger)\\ &+ \nabla_{\beta'}\psi_{gT}(\alpha,\beta^\dagger)(\beta-\beta^\dagger) + o(1),\end{aligned}$$

where

$$\begin{aligned}\nabla_{\alpha'}\psi_{gT}(\alpha,\beta) &= \int S_{gT}(\beta)[\nabla_{\alpha'}f^T(y^T,\alpha)/f^T(y^T;\alpha)]f^T(y^T;\alpha)\,dy^T\\ &= T\int S_{gT}(\beta)S_{fT}(\alpha)'f^T(y^T;\alpha)\,dy^T\\ &= \mathbb{E}_{f(\alpha)}[\sqrt{T}S_{gT}(\beta),\sqrt{T}S_{fT}(\alpha)'],\end{aligned}$$

because $S_{fT}(\alpha) = T^{-1}[\nabla_{\alpha'}f^T(y^T,\alpha)/f^T(y^T;\alpha)]$, and

$$\nabla_{\beta'}\psi_{gT}(\alpha,\beta) = \int \nabla_{\beta'}S_{gT}(\beta)f^T(y^T;\alpha)\,dy^T = \mathbb{E}_{f(\alpha)}[H_{gT}(\beta)].$$

Passing to the limit we have

$$\Psi_g(\alpha,\beta) = \Psi_g(\alpha^\dagger,\beta^\dagger) + K_{gf}(\alpha^\dagger,\beta)(\alpha-\alpha^\dagger) + H_g(\alpha,\beta^\dagger)(\beta-\beta^\dagger).$$

Given $\alpha = \hat{\alpha}_T$, $\beta = \hat{\beta}_T$, $\alpha^\dagger = \alpha_o$, and $\beta^\dagger = \beta(\alpha_o)$,

$$\begin{aligned}\sqrt{T}\Psi_g(\hat{\alpha}_T,\hat{\beta}_T) &= \sqrt{T}\Psi_g(\alpha_o,\beta(\alpha_o)) + K_{gf}(\alpha_o,\hat{\beta}_T)\sqrt{T}(\hat{\alpha}_T-\alpha_o)\\ &\qquad + H_g(\hat{\alpha}_T,\beta(\alpha_o))\sqrt{T}[\hat{\beta}_T-\beta(\alpha_o)]\\ &= K_{gfo}\sqrt{T}(\hat{\alpha}_T-\alpha_o) + H_{go}\sqrt{T}[\hat{\beta}_T-\beta(\alpha_o)] + o_p(1).\end{aligned}$$

To prove (3), note that

$$\begin{aligned}H_{go}\sqrt{T}[\hat{\beta}_T-\beta(\hat{\alpha}_T)] &= H_{go}\sqrt{T}[\hat{\beta}_T-\beta(\alpha_o)] - H_{go}\sqrt{T}[\beta(\hat{\alpha}_T)-\beta(\alpha_o)]\\ &= H_{go}\sqrt{T}[\hat{\beta}_T-\beta(\alpha_o)] - H_{go}\nabla_{\alpha'}\beta(\alpha_o)\sqrt{T}(\hat{\alpha}_T-\alpha_o) + o_p(1).\end{aligned}$$

The proof is complete if we can show

$$K_{gfo} = -H_{go}\nabla_{\alpha'}\beta(\alpha_o) + o(1).$$

Consider now the following expansion about $\alpha^\dagger$:

$$\psi_{gT}(\alpha,\beta(\alpha)) = \psi_{gT}(\alpha^\dagger,\beta(\alpha^\dagger)) + \nabla_1\psi_{gT}(\alpha^\dagger,\beta(\alpha^\dagger))(\alpha-\alpha^\dagger) + \nabla_{\beta'}\psi_{gT}(\alpha^\dagger,\beta(\alpha^\dagger))\nabla_{\alpha'}\beta(\alpha^\dagger)(\alpha-\alpha^\dagger) + o(1),$$

where $\nabla_1\psi_{gT}$ is the gradient of ψ_{gT} with respect to the first argument so that, similar to previous derivations,

$$\nabla_1\psi_{gT}(\alpha^\dagger,\beta(\alpha^\dagger)) = \mathbb{E}_{f(\alpha^\dagger)}[\sqrt{T}S_{gT}(\beta(\alpha^\dagger)),\sqrt{T}S_{fT}(\alpha^\dagger)'],$$

and

$$\nabla_{\beta'}\psi_{gT}(\alpha^\dagger,\beta(\alpha^\dagger)) = \mathbb{E}_{f(\alpha^\dagger)}[H_{gT}(\beta(\alpha^\dagger))].$$

Again passing to the limit and setting $\alpha^\dagger = \alpha_o$, we have

$$\Psi_g(\alpha,\beta(\alpha)) = \Psi_g(\alpha_o,\beta(\alpha_o)) + [K_{gfo} + H_{go}\nabla_{\alpha'}\beta(\alpha_o)](\alpha-\alpha_o).$$

By (1), $\Psi_g(\alpha,\beta(\alpha)) = 0$ and $\Psi_g(\alpha_o,\beta(\alpha_o)) = 0$. It follows that

$$K_{gfo} + H_{go}\nabla_{\alpha'}\beta(\alpha_o) = 0,$$

a zero matrix. □

Proof of Equation (12): Let λ_{ot} denote the indicator variable from M_f with $\alpha = \alpha_o$. The pseudo-true score function of M_g is

$$\Psi_g(\alpha_o,\beta) = \lim_{T\to\infty}\frac{1}{T\sigma_g^2}\left[\sum_{t=1}^{T}\mathbb{E}_{f(\alpha_o)}[\lambda_{ot}\nabla_\delta\mu_t(m_{ot}-\mu_t)] - \mathbb{E}_{f(\alpha_o)}\left((1-\lambda_{ot})\nabla_\delta\mu_t\frac{\sigma_g\phi(\mu_t/\sigma_g)}{1-\Phi(\mu_t/\sigma_g)}\right) + \mathbb{E}_{f(\alpha_o)}[\nabla_\delta\mu_t\lambda_{ot}\varepsilon_t]\right].$$

By the law of iterated expectations,

$$\mathbb{E}_{f(\alpha_o)}[\nabla_\delta\mu_t\lambda_{ot}\varepsilon_t] = \mathbb{E}_{f(\alpha_o)}[\nabla_\delta\mu_t\mathbb{E}_{f(\alpha_o)}[\lambda_{ot}\varepsilon_t|\mathfrak{F}_{t-1}]],$$

and

$$\begin{aligned}\mathbb{E}_{f(\alpha_o)}[\lambda_{ot}\varepsilon_t|\mathfrak{F}_{t-1}] &= \mathbb{E}_{f(\alpha_o)}[\lambda_{ot}\mathbb{E}_{f(\alpha_o)}[\varepsilon_t|\mathfrak{F}_{t-1},\lambda_{ot}]],\\ &= \mathrm{P}_{f(\alpha_o)}(\lambda_{ot}=1)\mathbb{E}_{f(\alpha_o)}[\varepsilon_t|\mathfrak{F}_{t-1},\lambda_{ot}=1].\end{aligned}$$

Note that $\mathrm{P}_{f(\alpha_o)}(\lambda_{ot} = 1) = \Phi(m_{ot}/\sigma_{fo})$ and $\mathrm{P}_{f(\alpha_o)}(\lambda_{ot} = 0) = 1 - \Phi(m_{ot}/\sigma_{fo})$. Given that the conditional density function of ε_t is

$$f(\varepsilon_t = a|\mathfrak{F}_{t-1}, \lambda_{ot} = 1) = \frac{\phi(a/\sigma_{fo})/\sigma_{fo}}{\Phi(m_{ot}/\sigma_{fo})},$$

we thus have

$$\begin{aligned}
\mathbb{E}_{f(\alpha_o)}[\lambda_{ot}\varepsilon_t|\mathfrak{F}_{t-1}] &= \int_{-m_{ot}}^{\infty} (a/\sigma_{fo})\phi(a/\sigma_{fo})\,\mathrm{d}\,a \\
&= \frac{1}{2\sqrt{2\pi\sigma_{fo}^2}} \int_{m_{ot}^2}^{\infty} \exp(-v/2\sigma_{fo}^2)\,\mathrm{d}\,v \qquad (v := a^2) \\
&= \sigma_{fo}\phi(m_{ot}/\sigma_{fo}).
\end{aligned}$$

It follows that

$$\mathbb{E}_{f(\alpha_o)}[\nabla_\delta \mu_t \lambda_{ot}\varepsilon_t] = \mathbb{E}_{f(\alpha_o)}[\nabla_\delta \mu_t \sigma_{fo}\phi(m_{ot}/\sigma_{fo})].$$

Similarly, we also have

$$\begin{aligned}
&\mathbb{E}_{f(\alpha_o)}[\lambda_{ot}\nabla_\delta\mu_t(m_{ot} - \mu_t)] = \mathbb{E}_{f(\alpha_o)}[\nabla_\delta\mu_t(m_{ot} - \mu_t)\Phi(m_{ot}/\sigma_{fo})], \\
&\mathbb{E}_{f(\alpha_o)}\left[(1 - \lambda_{ot})\nabla_\delta\mu_t \frac{\sigma_g\phi(\mu_t/\sigma_g)}{1 - \Phi(\mu_t/\sigma_g)}\right] = \mathbb{E}_{f(\alpha_o)}\left[\nabla_\delta\mu_t\sigma_g\phi(\mu_t/\sigma_g)\frac{1 - \Phi(m_{ot}/\sigma_{fo})}{1 - \Phi(\mu_t/\sigma_g)}\right].
\end{aligned}$$

Thus, after taking into account the censoring scheme under M_f with $\alpha = \alpha_o$, the pseudo-true score function of M_g is

$$\begin{aligned}
\Psi_g(\alpha_o, \beta) = \lim_{T\to\infty} \frac{1}{T\sigma_g^2} \sum_{t=1}^{T} \mathbb{E}_{f(\alpha_o)} \Big\{\nabla_\delta\mu_t \Big[(m_{ot} - \mu_t)\Phi(m_{ot}/\sigma_{fo}) + \sigma_{fo}\phi(m_{ot}/\sigma_{fo}) \\
-\ \sigma_g\phi(\mu_t/\sigma_g)\frac{1 - \Phi(m_{ot}/\sigma_{fo})}{1 - \Phi(\mu_t/\sigma_{go})}\Big]\Big\},
\end{aligned}$$

from which we can see that (12) is the key ingredient of the PSE test. $\square$

References

[1] Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation, *Econometrica*, **59**, 817–858.

[2] Cox, D. R. (1961). Tests of separate families of hypotheses, in *Proceedings of the Fourth Berkeley Symposium*, **1**, 105–123, Berkeley: University of California Press.

[3] Cox, D. R. (1962). Further results on tests of separate families of hypotheses, *Journal of the Royal Statistical Society*, **B, 24**, 404–426.

[4] Davidson, R. and J. G. MacKinnon (1981). Several tests for model specification in the presence of alternative hypotheses, *Econometrica*, **49**, 781–793.

[5] Engle, R. F. and V. Ng (1993). Measuring and testing the impact of news on volatility, *Journal of Finance*, **48**, 1749–1778.

[6] Fisher, G. R. and M. McAleer (1981). Alternative procedures and associated tests of significance for non-nested hypotheses, *Journal of Econometrics*, **16**, 103–119.

[7] Gourieroux, C., A. Monfort, and A. Trognon (1983), Testing nested or non-nested hypotheses, *Journal of Econometrics*, **21**, 83–115.

[8] Gourieroux, C., A. Monfort, and A. Trognon (1984). Pseudo-maximum likelihood methods: Theory, *Econometrica*, **52**, 681–700.

[9] Hendry, D. F. and J. F. Richard (1982). On the formulation of empirical models in dynamic econometrics, *Journal of Econometrics*, **20**, 3–34.

[10] McAleer, M. (1995). The significance of testing empirical non-nested models, *Journal of Econometrics* , **67**, 149–171.

[11] Mizon, G. E. (1984). The encompassing approach in econometrics, in D. Hendry and K. F. Wallis eds., *Econometrics and Quantitative Econometrics*, 135–172, Oxford: Basil Blackwell.

[12] Mizon, G. E. and J. F. Richard (1986). The encompassing principle and its application to testing non-nested hypotheses, *Econometrica*, **54**, 657–678.

[13] Newey, W. K. and K. D. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica*, **55**, 703–708.

[14] Pesaran, M. H. and B. Pesaran (1993). A simulation approach to the problem of computing Cox's statistic for testing non-nested models, *Journal of Econometrics*, **57**, 377–392.

[15] Rao, C. R. and S. K. Mitra (1971). *Generalized Inverse of Matrices and Its Applications*, New York: Wiley.

[16] Smith, R. J. (1992). Non-nested tests for competing models estimated by generalized method of moments, *Econometrica*, **60**, 973–980.

[17] Teräsvirta, T. and H. M. Anderson (1992). Characterizing nonlinearities in business cycles using smooth transition autoregressive models, *Journal of Applied Econometrics*, **7**, S119–S136.

[18] White, H. (1984). *Asymptotic Theory for Econometricians*, Orlando: Academic Press.

[19] White, H. (1994). *Estimation, Inference and Specification Analysis*, New York: Cambridge University Press.

[20] Wooldridge, J. M. (1990a). An encompassing approach to conditional mean tests with applications to testing non-nested hypotheses, *Journal of Econometrics*, **45**, 331–350.

[21] Wooldridge, J. M. (1990b). A unified approach to robust, regression-based specification tests, *Econometric Theory*, **6**, 17–43.

Figure 1: Non-nested regressions: $T = 50$, $p = 3$.

Figure 2: Non-nested regressions: $T = 50$, $p = 5$.

Figure 3: Non-nested regressions: $T = 100$, $p = 3$.

Figure 4: Non-nested regressions: $T = 100$, $p = 5$.

Figure 5: Probit vs. Logit: $T = 50$, $p = 3$.

Figure 6: Probit vs. Logit: $T = 50$, $p = 5$.

Figure 7: Probit vs. Logit: $T = 100$, $p = 3$.

Figure 8: Probit vs. Logit: $T = 100$, $p = 5$.