Reversed Score and Likelihood Ratio Tests*

Geert Dhaene
Ghent University

Olivier Scaillet
Université Catholique de Louvain

October 2000

Abstract

Two extensions of a parametric model are proposed, each one involving the score function of an alternative parametric model. We show that the encompassing hypothesis is equivalent to standard conditions on the score of each of the extended models. The condition on the first extension gives rise to the standard score encompassing test, while the condition on the second extension induces a so-called reversed score encompassing test. A similar logic is applied to the likelihood ratio, generating a likelihood ratio and a reversed likelihood ratio encompassing test. The ensued test statistics can be based on simulations if certain calculations are too difficult to carry out analytically. We study the first order asymptotic properties of the proposed test statistics under general conditions.

JEL classification: C5, C3.
Key-words: score test, likelihood ratio test, encompassing, simulation-based inference.

*We are grateful to Christian Gouriéroux for helpful comments. Address correspondence to: Geert Dhaene, Ghent University, Faculty of Economics and Business Administration, Hoveniersberg 24, B-9000 Gent, Belgium. Tel.: (+32) 9 2648978. Fax: (+32) 9 2648995. E-mail: geert.dhaene@rug.ac.be.

1 Introduction

Specification tests of parametric models are a central theme in the econometric literature. A standard approach is to confront a given parametric model with another, often non-nested, parametric model (see Gouriéroux and Monfort [1994] for a review), and therefore such tests are oriented towards this particular alternative model. The constraint underlying most of these tests is in fact the encompassing condition (see e.g. Mizon and Richard [1986], Hendry and Richard [1990], Smith [1994], Gouriéroux and Monfort [1995], Dhaene [1997], Dhaene, Gouriéroux and Scaillet [1998]), but not always (see Vuong [1989]). Another approach exploits moment conditions implied by the model under test without having a specific alternative model in mind. Information matrix tests (White (1982)) and unconditional and conditional moment tests (Newey [1985], Tauchen [1985], Bierens [1991]) are examples of the latter approach.

The approach taken in this paper falls into the former category. An arbitrary conditional parametric model is tested against another arbitrary, possibly non-nested, conditional parametric model. We expand on results reported in Gouriéroux and Monfort [1995] and Dhaene [1997], where score and likelihood ratio encompassing tests were proposed. These tests, and the new tests we propose, are generated by exponentially tilting the model under test in two alternative directions, each one involving the score function of the alternative model. Intuitively, the new tests are obtained from reversing the roles of the true distribution generating the data and the pseudo-true distribution of the model under test. This leads to what we call reversed score and likelihood ratio tests. The tests rely on simulations in order to avoid the need for analytic calculations of certain expectations in any particular application. In a recent paper, Chen and Kuan [2000] propose what they call the pseudo-true score encompassing test for non-nested hypotheses, which is based on essentially the same idea of reversing the roles of the two distributions just mentioned. The main differences with the present paper are as follows. We provide a heuristic argument, based on model extensions, which unifies the standard and the reversed score tests. Furthermore, we also apply the idea to the likelihood ratio test, we consider nested as well as non-nested hypotheses, we propose simulation-based versions of the tests, and provide robust asymptotic theory.

The framework is briefly presented in Section 2. Section 3 introduces two extensions of the model under test, obtained by exponential tilting. It also restates the encompassing condition in terms of these extensions and gives the intuition underlying the reversed score and likelihood ratio tests. The basic test statistics are presented in Section 4. Their first order asymptotic properties are studied in Section 5, in descending order of generality. Section 6 concludes.

2 Framework

We consider an arbitrary pair of conditional, possibly non-nested, possibly misspecified, parametric models for independent and identically distributed data.

Let X and Y be random vectors taking values x and y in $\mathbb{R}^k$ and $\mathbb{R}^l$, respectively, and let P_X be the true marginal distribution of X and $P_{Y|X}$ the true conditional distribution of Y, given X. Assume that the available data are T independent drawings (x_t, y_t), $t = 1, \ldots, T$, from P_X and $P_{Y|X}$. Let $\mathcal{G} = \{F_{\mathcal{G}}(\alpha) \,|\, \alpha \in \Omega_\alpha \subset \mathbb{R}^m\}$ and $\mathcal{H} = \{F_{\mathcal{H}}(\beta) \,|\, \beta \in \Omega_\beta \subset \mathbb{R}^n\}$ be parametric models of $P_{Y|X}$. It is assumed that the distributions $F_{\mathcal{G}}(\alpha)$, $F_{\mathcal{H}}(\beta)$ and $P_{Y|X}$ admit conditional density functions $f_{\mathcal{G}}(y|x;\alpha)$, $f_{\mathcal{H}}(y|x;\beta)$ and $p_0(y|x)$, respectively, relative to some measure μ not depending on x, α and β. It is also assumed that the expectations of the log density functions exist whenever they are taken.

Accounting for the possibility that $\mathcal{G}$ is misspecified, i.e. $P_{Y|X} \notin \mathcal{G}$, and likewise for $\mathcal{H}$, it is of interest to define the pseudo-true values of α and β with respect to P_X and $P_{Y|X}$ (see e.g. Sawa [1978]):

$$\alpha_0 = \arg\max_{\alpha\in\Omega_\alpha} E_X E_0 \log f_{\mathcal{G}}(Y|X;\alpha)$$
$$\beta_0 = \arg\max_{\beta\in\Omega_\beta} E_X E_0 \log f_{\mathcal{H}}(Y|X;\beta),$$

where the mathematical expectations E_X and E_0 are taken with respect to P_X and $P_{Y|X}$, respectively. We assume that α_0 and β_0 exist, are unique and interior to Ω_α and Ω_β, respectively.

We shall be interested in testing $\mathcal{G}$ against $\mathcal{H}$. Therefore, we also define the pseudo-true value of β with respect to P_X and $F_{\mathcal{G}}(\alpha)$,

$$\beta_\alpha = \arg\max_{\beta\in\Omega_\beta} E_X E_\alpha \log f_{\mathcal{H}}(Y|X;\beta),$$

where the mathematical expectation E_α is taken with respect to $F_{\mathcal{G}}(\alpha)$. We assume that β_α exists, is unique and interior to Ω_β and is continuously differentiable with respect to α. By definition, $\mathcal{G}$ encompasses $\mathcal{H}$, written $\mathcal{G}\,\mathcal{E}\,\mathcal{H}$, if $\beta_0 = \beta_{\alpha_0}$. It is well known that the implicit null hypothesis of many tests of $\mathcal{G}$ against $\mathcal{H}$ is characterized by the condition that $\mathcal{G}\,\mathcal{E}\,\mathcal{H}$. See e.g. Mizon and Richard [1986], Gouriéroux and Monfort [1995], and Dhaene [1997].

Note that the underlying distributions P_X and $P_{Y|X}$ are crucial in determining whether or not $\mathcal{G}\,\mathcal{E}\,\mathcal{H}$. The score functions of $\mathcal{G}$ and $\mathcal{H}$ are defined as

$$s_{\mathcal{G}}(y|x;\alpha) = \frac{\partial}{\partial\alpha}\log f_{\mathcal{G}}(y|x;\alpha)$$

and

$$s_{\mathcal{H}}(y|x;\beta) = \frac{\partial}{\partial\beta}\log f_{\mathcal{H}}(y|x;\beta),$$

respectively. It is assumed that the score functions are continuously differentiable in the parameters, that their expectations exist whenever they are taken, that

$$\begin{aligned} E_X E_0 s_{\mathcal{G}}(Y|X;\alpha) = 0 \quad &\text{only if} \quad \alpha = \alpha_0, \\ E_X E_0 s_{\mathcal{H}}(Y|X;\beta) = 0 \quad &\text{only if} \quad \beta = \beta_0, \\ E_X E_\alpha s_{\mathcal{H}}(Y|X;\beta) = 0 \quad &\text{only if} \quad \beta = \beta_\alpha, \end{aligned}$$

and that the matrices $E_X E_0[s_{\mathcal{G}}(Y|X;\alpha_0)s'_{\mathcal{G}}(Y|X;\alpha_0)]$, $E_X E_0[s_{\mathcal{H}}(Y|X;\beta_0)s'_{\mathcal{H}}(Y|X;\beta_0])$ and $E_X E_\alpha[s_{\mathcal{H}}(Y|X;\beta_\alpha)s'_{\mathcal{H}}(Y|X;\beta_\alpha])$ exist and are positive definite. Then, defining the score quantity

$$s_1 = E_X E_0 s_{\mathcal{H}}(Y|X;\beta_{\alpha_0})$$

and the likelihood ratio (LR) quantity

$$l_1 = E_X E_0[\log f_{\mathcal{H}}(Y|X;\beta_{\alpha_0}) - \log f_{\mathcal{H}}(Y|X;\beta_0)],$$

it is obvious that $\mathcal{G}\,\mathcal{E}\,\mathcal{H}$ is equivalent to $s_1 = 0$ and also to $l_1 = 0$. This property has led to the development of score encompassing tests, based on estimates of s_1 (Gouriéroux and Monfort [1995]), and LR encompassing tests, based on estimates of l_1 (Smith [1994] and Dhaene [1997]). The purpose of this paper is to introduce tests that are based on quantities similar to s_1 and l_1, in particular the quantities obtained from s_1 and l_1 by reversing the roles of $P_{Y|X}$ and $F_{\mathcal{G}}(\alpha_0)$. A heuristic argument for doing so is presented in the next section.

3 Model extensions

Consider the following extension of $\mathcal{G}$:

$$\mathcal{G}_1 = \{F^1_{\mathcal{G}}(\alpha,\lambda_1) \,|\, (\alpha,\lambda_1) \in \Omega_\alpha \times \mathrm{I\!R}^n\},$$

where the distribution $F^1_{\mathcal{G}}(\alpha, \lambda_1)$ has the following density function relative to μ:

$$f^1_{\mathcal{G}}(y|x;\alpha,\lambda_1) = \frac{f_{\mathcal{G}}(y|x;\alpha)\exp(\lambda_1' s_{\mathcal{H}}(y|x;\beta_{\alpha_0}))}{E_\alpha \exp(\lambda_1' s_{\mathcal{H}}(y|x;\beta_{\alpha_0}))}.$$

The density $f^1_{\mathcal{G}}(y|x;\alpha,\lambda_1)$ is obtained from $f_{\mathcal{G}}(y|x;\alpha)$ by exponential tilting (Barndorff-Nielsen and Cox [1989]). Observe that $\mathcal{G} \subset \mathcal{G}_1$ and that the parameter vector (α, λ_1) need not be identified. Instead of putting $\beta = \beta_{\alpha_0}$ in the random vector $s_{\mathcal{H}}(Y|X;\beta)$, one may alternatively put $\beta = \beta_0$, leading to another extension of $\mathcal{G}$:

$$\mathcal{G}_2 = \{F^2_{\mathcal{G}}(\alpha,\lambda_2) \,|\, (\alpha,\lambda_2) \in \Omega_\alpha \times \mathrm{I\!R}^n\},$$

where the distribution $F^2_{\mathcal{G}}(\alpha, \lambda_2)$ has the following density function relative to μ:

$$f^2_{\mathcal{G}}(y|x;\alpha,\lambda_2) = \frac{f_{\mathcal{G}}(y|x;\alpha)\exp(\lambda_2' s_{\mathcal{H}}(y|x;\beta_0))}{E_{\alpha_0} \exp(\lambda_2' s_{\mathcal{H}}(y|x;\beta_0))}.$$

The density $f^2_{\mathcal{G}}(y|x;\alpha,\lambda_2)$ is also obtained from $f_{\mathcal{G}}(y|x;\alpha)$ by exponential tilting, but in a different direction. As before, $\mathcal{G} \subset \mathcal{G}_2$ and (α, λ_2) need not be identified. The motivation for considering the extended models $\mathcal{G}_1$ and $\mathcal{G}_2$ comes from the following proposition.

Proposition 1 *The following equivalences hold:*

$$\begin{aligned}
\mathcal{G}\,\mathcal{E}\,\mathcal{H} \quad &\Longleftrightarrow \quad E_X E_0 \log f^1_{\mathcal{G}}(Y|X;\alpha,\lambda_1) \textit{ has a local maximum at } (\alpha,\lambda_1) = (\alpha_0, 0)\\
&\Longleftrightarrow \quad E_X E_0 s_{\mathcal{H}}(Y|X;\beta_{\alpha_0}) = 0,\\
\mathcal{G}\,\mathcal{E}\,\mathcal{H} \quad &\Longleftrightarrow \quad E_X E_0 \log f^2_{\mathcal{G}}(Y|X;\alpha,\lambda_2) \textit{ has a local maximum at } (\alpha,\lambda_2) = (\alpha_0, 0)\\
&\Longleftrightarrow \quad E_X E_{\alpha_0} s_{\mathcal{H}}(Y|X;\beta_0) = 0.
\end{aligned}$$

Proof. The score functions associated with $\mathcal{G}_1$ and $\mathcal{G}_2$ are

$$s^1_{G}(y|x;\alpha,\lambda_1) = \begin{pmatrix} s_{\mathcal{G}}(y|x;\alpha) - \dfrac{E_\alpha[s_{\mathcal{G}}(Y|x;\alpha)\exp(\lambda_1' s_{\mathcal{H}}(Y|x;\beta_{\alpha_0}))]}{E_\alpha \exp(\lambda_1' s_{\mathcal{H}}(Y|x;\beta_{\alpha_0}))} \\[2ex] s_{\mathcal{H}}(y|x;\beta_{\alpha_0}) - \dfrac{E_\alpha[s_{\mathcal{H}}(Y|x;\beta_{\alpha_0})\exp(\lambda_1' s_{\mathcal{H}}(Y|x;\beta_{\alpha_0}))]}{E_\alpha \exp(\lambda_1' s_{\mathcal{H}}(Y|x;\beta_{\alpha_0}))} \end{pmatrix}$$

and

$$s^2_{\mathcal{G}}(y|x;\alpha,\lambda_2) = \begin{pmatrix} s_{\mathcal{G}}(y|x;\alpha) - \dfrac{E_\alpha[s_{\mathcal{G}}(Y|x;\alpha)\exp(\lambda_2' s_{\mathcal{H}}(Y|x;\beta_0))]}{E_\alpha \exp(\lambda_2' s_{\mathcal{H}}(Y|x;\beta_0))} \\[2ex] s_{\mathcal{H}}(y|x;\beta_0) - \dfrac{E_\alpha[s_{\mathcal{H}}(Y|x;\beta_0)\exp(\lambda_2' s_{\mathcal{H}}(Y|x;\beta_0))]}{E_\alpha \exp(\lambda_2' s_{\mathcal{H}}(Y|x;\beta_0))} \end{pmatrix},$$

respectively. Putting $(\alpha,\lambda_1)=(\alpha,\lambda_2)=(\alpha_0,0)$ and taking expectations yields

$$\begin{aligned} E_X E_0 s^1_{\mathcal{G}}(Y|X;\alpha_0,0) &= \begin{pmatrix} E_X E_0 s_{\mathcal{G}}(Y|X;\alpha_0) - E_X E_{\alpha_0} s_{\mathcal{G}}(Y|X;\alpha_0) \\ E_X E_0 s_{\mathcal{H}}(Y|X;\beta_{\alpha_0}) - E_X E_{\alpha_0} s_{\mathcal{H}}(Y|X;\beta_{\alpha_0}) \end{pmatrix} \\ &= \begin{pmatrix} 0 \\ E_X E_0 s_{\mathcal{H}}(Y|X;\beta_{\alpha_0}) \end{pmatrix} \end{aligned}$$

and

$$\begin{aligned} E_X E_0 s^2_{\mathcal{G}}(Y|X;\alpha_0,0) &= \begin{pmatrix} E_X E_0 s_{\mathcal{G}}(Y|X;\alpha_0) - E_X E_{\alpha_0} s_{\mathcal{G}}(Y|X;\alpha_0) \\ E_X E_0 s_{\mathcal{H}}(Y|X;\beta_0) - E_X E_{\alpha_0} s_{\mathcal{H}}(Y|X;\beta_0) \end{pmatrix} \\ &= \begin{pmatrix} 0 \\ E_X E_{\alpha_0} s_{\mathcal{H}}(Y|X;\beta_0) \end{pmatrix}. \end{aligned}$$

Given the assumptions made earlier, it follows that $\mathcal{G}\,\mathcal{E}\,\mathcal{H}$ if and only if the functions $E_X E_0 \log f^1_{\mathcal{G}}(Y|X;\alpha,\lambda_1)$ and $E_X E_0 \log f^2_{\mathcal{G}}(Y|X;\alpha,\lambda_2)$ have a stationary point at $(\alpha,\lambda_1)=(\alpha_0,0)$ and $(\alpha,\lambda_2)=(\alpha_0,0)$, respectively. Now we need to show that, if $\beta_0=\beta_{\alpha_0}$, the stationary point $(\alpha_0,0)$ is indeed a local maximum of the functions involved. First, fixing $\lambda_1=0$, $E_X E_0 \log f^1_{\mathcal{G}}(Y|X;\alpha,0)$ attains a global maximum at $\alpha=\alpha_0$, by definition. Secondly, fixing $\alpha=\alpha_0$, we find, if $\beta_0=\beta_{\alpha_0}$,

$$\left[\frac{\partial^2}{\partial\lambda_1\partial\lambda_1'} E_X E_0 \log f^1_{\mathcal{G}}(Y|X;\alpha_0,\lambda_1)\right]_{\lambda_1=0} = -E_X E_0[s_{\mathcal{H}}(Y|X;\beta_0)s'_{\mathcal{H}}(Y|X;\beta_0)].$$

The latter matrix is negative definite by assumption, hence $E_X E_0 \log f^1_{\mathcal{G}}(Y|X;\alpha_0,\lambda_1)$ attains a local maximum at $\lambda_1=0$. The proof is complete by noting that the functions $f^1_{\mathcal{G}}$ and $f^2_{\mathcal{G}}$ are identical when $\beta_0=\beta_{\alpha_0}$. (*Q.E.D.*)

The proposition is in several respects similar to Theorem 1 in Chesher and Smith [1997], which restates moment conditions in terms of an extended parametric density. Here, an encompassing condition is restated in terms of extended parametric densities. The proposition shows that $\mathcal{G}\,\mathcal{E}\,\mathcal{H}$ if and only if the extensions of $\mathcal{G}$ carrying the score function of $\mathcal{H}$ do not alter the pseudo-true value associated with $\mathcal{G}$, at least not locally. In a sense, the extensions are thus ineffective in bringing $\mathcal{G}$ closer to $P_{Y|X}$, according to the Kullback-Leibler (1951) Information Criterion. Further, the condition $\mathcal{G}\,\mathcal{E}\,\mathcal{H}$ is restated in terms of properties of the score function $s_{\mathcal{H}}$ in relation to the distributions $P_{Y|X}$ and $F_{\mathcal{G}}(\alpha_0)$. Interestingly, the two properties mirror each other in the sense that each one, compared to the other, *reverses* the roles of $P_{Y|X}$ and $F_{\mathcal{G}}(\alpha_0)$. After all, this should not come as a surprise since, for *given* α_0, the distributions $P_{Y|X}$ and $F_{\mathcal{G}}(\alpha_0)$ play a symmetric role in the definition of encompassing. Thus, we are led to define the reversed score quantity

$$s_2 = E_X E_{\alpha_0} s_{\mathcal{H}}(Y|X;\beta_0),$$

and, applying the same logic, the reversed LR quantity

$$l_2 = E_X E_{\alpha_0}[\log f_{\mathcal{H}}(Y|X;\beta_0) - \log f_{\mathcal{H}}(Y|X;\beta_{\alpha_0})].$$

The quantities s_2 and l_2 share the property with s_1 and l_1 that $\mathcal{G}\,\mathcal{E}\,\mathcal{H}$ is equivalent to $s_2 = 0$ and also to $l_2 = 0$. This property enables us to develop reversed score encompassing tests, based on estimates of s_2, and reversed LR encompassing tests, based on estimates of l_2.

One may wonder whether the same reasoning of reversing the roles of $P_{Y|X}$ and $F_{\mathcal{G}}(\alpha_0)$ can also be applied to the Wald encompassing test to yield something interesting. The Wald encompassing test (Gouriéroux and Monfort [1995]) is based on estimates of the Wald quantity, defined as $w_1 = \beta_0 - \beta_{\alpha_0}$. The reversed Wald quantity would then be $w_2 = \beta_{\alpha_0} - \beta_0 = -w_1$, which obviously does not lead to an interesting new test. The reason for this finding is that $P_{Y|X}$ and $F_{\mathcal{G}}(\alpha_0)$ play similar roles in w_1, apart from the sign. Hence, reversing their roles doesn't lead to anything new. Looking back now at s_1 and l_1, we clearly see that $P_{Y|X}$ and $F_{\mathcal{G}}(\alpha_0)$ play essentially different roles. This is why reversing them happens to be fruitful.

4 Test statistics

Given the sample (x_t, y_t), $t = 1, \ldots, T$, of independent observations from P_X and $P_{Y|X}$, we seek to develop tests of the hypothesis that $\mathcal{G}\,\mathcal{E}\,\mathcal{H}$. It follows from the properties derived in the previous section that estimates of the quantities s_1, l_1, s_2 and l_2 and of their covariance matrices naturally lead to tests of $\mathcal{G}\,\mathcal{E}\,\mathcal{H}$. Note that this hypothesis is weaker than the hypothesis that $\mathcal{G}$ is correctly specified, i.e. $P_{Y|X} \in \mathcal{G}$. Hence estimates of the same quantities are also suited for testing the hypothesis that $\mathcal{G}$ is correctly specified. A distinguishing feature between tests of $\mathcal{G}\,\mathcal{E}\,\mathcal{H}$ and tests of $P_{Y|X} \in \mathcal{G}$ is that, for the latter tests the distribution theory is usually based on the assumption that $\mathcal{G}$ is correctly specified, whereas for the former tests the distribution theory can at most be based on the assumption that $\mathcal{G}\,\mathcal{E}\,\mathcal{H}$. The distribution theory presented in this paper considers the most general case, i.e. where $\mathcal{G}$ possibly does not encompass $\mathcal{H}$.

The pseudo-maximum likelihood estimators $\hat{\alpha}$ and $\hat{\beta}$ solve

$$\max_{\alpha \in \Omega_\alpha} \frac{1}{T} \sum_{t=1}^{T} \log f_{\mathcal{G}}(y_t|x_t;\alpha)$$

and

$$\max_{\beta \in \Omega_\beta} \frac{1}{T} \sum_{t=1}^{T} \log f_{\mathcal{H}}(y_t | x_t; \beta),$$

respectively. Under regularity conditions such as given in WHITE [1982], $\hat{\alpha} \overset{\text{a.s.}}{\to} \alpha_0$ and $\hat{\beta} \overset{\text{a.s.}}{\to} \beta_0$. For any $\alpha \in \Omega_\alpha$, let $y_t^h(\alpha)$, $t = 1, \ldots, T$ and $h = 1, \ldots, H$, be independent drawings from $F_{\mathcal{G}}(\alpha)$, given x_t. For any $h = 1, \ldots, H$, the simulated pseudo-maximum likelihood estimator $\hat{\beta}_\alpha^h$ is defined to solve

$$\max_{\beta \in \Omega_\beta} \frac{1}{T} \sum_{t=1}^{T} \log f_{\mathcal{H}}(y_t^h(\alpha) | x_t; \beta).$$

Under similar regularity conditions, $\hat{\beta}_\alpha^h \overset{\text{a.s.}}{\to} \beta_\alpha$ and $\hat{\beta}_{\hat{\alpha}}^h \overset{\text{a.s.}}{\to} \beta_{\alpha_0}$. Here and in the sequel, stochastic limits are taken as $T \to \infty$, with H fixed, possibly at ∞. Then, define the simulated score and reversed score statistics as

$$\begin{aligned} \hat{s}_1 &= \frac{1}{TH} \sum_{h=1}^{H} \sum_{t=1}^{T} s_{\mathcal{H}}(y_t | x_t; \hat{\beta}_{\hat{\alpha}}^h), \\ \hat{s}_2 &= \frac{1}{TH} \sum_{h=1}^{H} \sum_{t=1}^{T} s_{\mathcal{H}}(y_t^h(\hat{\alpha}) | x_t; \hat{\beta}), \end{aligned}$$

respectively, and the simulated LR and reversed LR statistics as

$$\begin{aligned} \hat{l}_1 &= \frac{1}{TH} \sum_{h=1}^{H} \sum_{t=1}^{T} \left[\log f_{\mathcal{H}}(y_t | x_t; \hat{\beta}_{\hat{\alpha}}^h) - \log f_{\mathcal{H}}(y_t | x_t; \hat{\beta}) \right], \\ \hat{l}_2 &= \frac{1}{TH} \sum_{h=1}^{H} \sum_{t=1}^{T} \left[\log f_{\mathcal{H}}(y_t^h(\hat{\alpha}) | x_t; \hat{\beta}) - \log f_{\mathcal{H}}(y_t^h(\hat{\alpha}) | x_t; \hat{\beta}_{\hat{\alpha}}^h) \right], \end{aligned}$$

respectively. We have $\hat{s}_1 \overset{\text{a.s.}}{\to} s_1$, $\hat{s}_2 \overset{\text{a.s.}}{\to} s_2$, $\hat{l}_1 \overset{\text{a.s.}}{\to} l_1$ and $\hat{l}_2 \overset{\text{a.s.}}{\to} l_2$. The first order limit distributions of $\hat{s}_1$, $\hat{s}_2$, $\hat{l}_1$ and $\hat{l}_2$ are investigated in the next section.

5 Limit distributions

We need to introduce some additional notation. Let

$$l_{\mathcal{G}}(\alpha) = \frac{1}{T} \sum_{t=1}^{T} \log f_{\mathcal{G}}(y_t | x_t; \alpha)$$

and

$$l_{\mathcal{H}}(\beta) = \frac{1}{T} \sum_{t=1}^{T} \log f_{\mathcal{H}}(y_t | x_t; \beta)$$

be the normalized log likelihood functions of $\mathcal{G}$ and $\mathcal{H}$ based on the observed data (x_t, y_t), $t = 1, \ldots, T$, and let

$$l^h_{\mathcal{H}}(\beta; \alpha) = \frac{1}{T} \sum_{t=1}^{T} \log f_{\mathcal{H}}(y^h_t(\alpha) | x_t; \beta)$$

be the normalized log likelihood function of $\mathcal{H}$ based on the simulated data $(x_t, y^h_t(\alpha))$, $t = 1, \ldots, T$. Correspondingly, define the normalized score functions

$$\begin{aligned} s_{\mathcal{G}}(\alpha) &= \frac{\partial}{\partial \alpha} l_{\mathcal{G}}(\alpha), \\ s_{\mathcal{H}}(\beta) &= \frac{\partial}{\partial \beta} l_{\mathcal{H}}(\beta), \end{aligned}$$

and

$$s^h_{\mathcal{H}}(\beta; \alpha) = \frac{\partial}{\partial \beta} l^h_{\mathcal{H}}(\beta; \alpha).$$

5.1 Limit distributions under general conditions

For sufficiently large T, $\hat{\alpha}$ satisfies the first order condition $s_{\mathcal{G}}(\hat{\alpha}) = 0$. Expanding $s_{\mathcal{G}}(\hat{\alpha})$ in a Taylor series around $s_{\mathcal{G}}(\alpha_0)$, taking the probability limit of $[\partial s_{\mathcal{G}}(\alpha) / \partial \alpha']_{\alpha=\alpha_0}$ and rearranging yields the well known result (White [1982])

$$\sqrt{T}(\hat{\alpha} - \alpha_0) = \sqrt{T} K^{-1}_{\mathcal{G}} s_{\mathcal{G}}(\alpha_0) + o_p(1),$$

where

$$K_{\mathcal{G}} = -E_X E_0 \left[\frac{\partial}{\partial \alpha'} s_{\mathcal{G}}(\alpha) \right]_{\alpha=\alpha_0}.$$

Similarly,

$$\sqrt{T}(\hat{\beta} - \beta_0) = \sqrt{T} K^{-1}_{\mathcal{H}} s_{\mathcal{H}}(\beta_0) + o_p(1),$$

where

$$K_{\mathcal{H}} = -E_X E_0 \left[\frac{\partial}{\partial \beta'} s_{\mathcal{H}}(\beta) \right]_{\beta=\beta_0},$$

and

$$\sqrt{T}(\hat{\beta}^h_{\alpha_0} - \beta_{\alpha_0}) = \sqrt{T} \tilde{K}^{-1}_{\bar{\mathcal{H}}} s^h_{\mathcal{H}}(\beta_{\alpha_0}; \alpha_0) + o_p(1),$$

where

$$\tilde{K}_{\bar{\mathcal{H}}} = -E_X E_{\alpha_0} \left[\frac{\partial}{\partial \beta'} s^h_{\mathcal{H}}(\beta; \alpha) \right]_{\alpha=\alpha_0, \beta=\beta_{\alpha_0}}.$$

Further, expanding $\hat{\beta}^h_{\hat{\alpha}}$ around $\hat{\beta}^h_{\alpha_0}$ yields

$$\begin{aligned}\sqrt{T}(\hat{\beta}^h_{\hat{\alpha}} - \beta_{\alpha_0}) &= \sqrt{T}(\hat{\beta}^h_{a0} - \beta_{\alpha_0}) + \sqrt{T}B(\hat{\alpha} - \alpha_0) + o_p(1) \\ &= \sqrt{T}\tilde{K}^{-1}_{\bar{\mathcal{H}}} s^h_{\mathcal{H}}(\beta_{\alpha_0};\alpha_0) + BK^{-1}_{\mathcal{G}} s_{\mathcal{G}}(\alpha_0) + o_p(1),\end{aligned}$$

where (see Dhaene [1997])

$$B = \left[\frac{\partial \beta_\alpha}{\partial \alpha'}\right]_{\alpha=\alpha_0} = \tilde{K}^{-1}_{\bar{\mathcal{H}}} \tilde{J}_{\bar{\mathcal{H}}\mathcal{G}},$$

with

$$\tilde{J}_{\bar{\mathcal{H}}\mathcal{G}} = E_X E_{\alpha_0}\left[\frac{\partial}{\partial \beta} \log f_{\mathcal{H}}(Y|X;\beta) \frac{\partial}{\partial \alpha'} \log f_{\mathcal{G}}(Y|X;\alpha)\right]_{\alpha=\alpha_0,\beta=\beta_{\alpha_0}}.$$

Now, expanding $s_{\mathcal{H}}(\hat{\beta}^h_{\hat{\alpha}})$ around $s_{\mathcal{H}}(\beta_{\alpha_0})$ gives

$$\begin{aligned}\sqrt{T}(\hat{s}_1 - s_1) &= \sqrt{T}(s_{\mathcal{H}}(\beta_{\alpha_0}) - s_1) - \frac{\sqrt{T}}{H}\sum_{h=1}^{H} K_{\bar{\mathcal{H}}}(\hat{\beta}^h_{\hat{\alpha}} - \beta_{\alpha_0}) + o_p(1) \\ &= \sqrt{T}(s_{\mathcal{H}}(\beta_{\alpha_0}) - s_1) - \sqrt{T}K_{\bar{\mathcal{H}}}BK^{-1}_{\mathcal{G}} s_{\mathcal{G}}(\alpha_0) \\ &\quad -\sqrt{T}K_{\bar{\mathcal{H}}}\tilde{K}^{-1}_{\bar{\mathcal{H}}}\frac{1}{H}\sum_{h=1}^{H} s^h_{\mathcal{H}}(\beta_{\alpha_0};\alpha_0) + o_p(1),\end{aligned}$$

where

$$K_{\bar{\mathcal{H}}} = -E_X E_0\left[\frac{\partial}{\partial \beta'} s_{\mathcal{H}}(\beta)\right]_{\beta=\beta_{\alpha_0}}.$$

Expanding $s^h_{\mathcal{H}}(\hat{\beta};\hat{\alpha})$ around $s^h_{\mathcal{H}}(\beta_0;\alpha_0)$ gives

$$\begin{aligned}\sqrt{T}(\hat{s}_2 - s_2) &= \frac{\sqrt{T}}{H}\sum_{h=1}^{H}(s^h_{\mathcal{H}}(\beta_0;\alpha_0) - s_2) - \sqrt{T}\tilde{K}_{\mathcal{H}}(\hat{\beta} - \beta_0) \\ &\quad \sqrt{T}\tilde{J}_{\mathcal{H}\mathcal{G}}(\hat{\alpha} - \alpha_0) + o_p(1) \\ &= \frac{\sqrt{T}}{H}\sum_{h=1}^{H}(s^h_{\mathcal{H}}(\beta_0;\alpha_0) - s_2) - \sqrt{T}\tilde{K}_{\mathcal{H}}K^{-1}_{\mathcal{H}} s_{\mathcal{H}}(\beta_0) \\ &\quad +\sqrt{T}\tilde{J}_{\mathcal{H}\mathcal{G}}K^{-1}_{\mathcal{G}} s_{\mathcal{G}}(\alpha_0) + o_p(1),\end{aligned}$$

where

$$\begin{aligned}\tilde{J}_{\mathcal{H}\mathcal{G}} &= E_X E_{\alpha_0}\left[\frac{\partial}{\partial \beta}\log f_{\mathcal{H}}(Y|X;\beta)\frac{\partial}{\partial \alpha'}\log f_{\mathcal{G}}(Y|X;\alpha)\right]_{\alpha=\alpha_0,\beta=\beta_0} = \tilde{J}'_{\mathcal{G}\mathcal{H}} \\ \tilde{K}_{\mathcal{H}} &= -E_X E_{\alpha_0}\left[\frac{\partial}{\partial \beta'} s^h_{\mathcal{H}}(\beta;\alpha)\right]_{\alpha=\alpha_0,\beta=\beta_0}.\end{aligned}$$

This completes the asymptotic expansions for $\hat{s}_1$ and $\hat{s}_2$. Turning to $\hat{l}_1$, expanding $l_{\mathcal{H}}(\hat{\beta}^h_{\hat{\alpha}})$ around $l_{\mathcal{H}}(\beta_{\alpha_0})$ gives

$$\begin{aligned}
\sqrt{T}(\hat{l}_1 - l_1) &= \sqrt{T}(l_{\mathcal{H}}(\beta_{\alpha_0}) - l_{\mathcal{H}}(\beta_0) - l_1) + \frac{\sqrt{T}}{H}\sum_{h=1}^{H} s_{\mathcal{H}}(\beta_{\alpha_0})'(\hat{\beta}^h_{\hat{\alpha}} - \beta_{\alpha_0}) \\
&\quad -\sqrt{T} s_{\mathcal{H}}(\beta_0)'(\hat{\beta} - \beta_0) + o_p(1) \\
&= \sqrt{T}(l_{\mathcal{H}}(\beta_{\alpha_0}) - l_{\mathcal{H}}(\beta_0) - l_1) + \sqrt{T} s_1' B K_{\mathcal{G}}^{-1} s_{\mathcal{G}}(\alpha_0) \\
&\quad +\sqrt{T} s_1' \tilde{K}_{\bar{\mathcal{H}}}^{-1} \frac{1}{H}\sum_{h=1}^{H} s^h_{\mathcal{H}}(\beta_{\alpha_0};\alpha_0) + o_p(1),
\end{aligned}$$

where it was used that $s_{\mathcal{H}}(\beta_0) \overset{\text{a.s.}}{\to} 0$. Finally, for $\hat{l}_2$,

$$\begin{aligned}
\sqrt{T}(\hat{l}_2 - l_2) &= \frac{\sqrt{T}}{H}\sum_{h=1}^{H}(l^h_{\mathcal{H}}(\beta_0;\alpha_0) - l^h_{\mathcal{H}}(\beta_{\alpha_0};\alpha_0) - l_2) + \frac{\sqrt{T}}{H}\sum_{h=1}^{H} s^h_{\mathcal{H}}(\beta_0;\alpha_0)'(\hat{\beta} - \beta_0) \\
&\quad -\frac{\sqrt{T}}{H}\sum_{h=1}^{H} s^h_{\mathcal{H}}(\beta_{\alpha_0};\alpha_0)'(\hat{\beta}^h_{\hat{\alpha}} - \beta_{\alpha_0}) + \sqrt{T}(\tilde{\omega}_{\mathcal{G}\mathcal{H}} - \tilde{\omega}_{\mathcal{G}\bar{\mathcal{H}}})'(\hat{\alpha} - \alpha_0) + o_p(1) \\
&= \frac{\sqrt{T}}{H}\sum_{h=1}^{H}(l^h_{\mathcal{H}}(\beta_0;\alpha_0) - l^h_{\mathcal{H}}(\beta_{\alpha_0};\alpha_0) - l_2) + \sqrt{T} s_2' K_{\mathcal{H}}^{-1} s_{\mathcal{H}}(\beta_0) \\
&\quad +\sqrt{T}(\tilde{\omega}_{\mathcal{G}\mathcal{H}} - \tilde{\omega}_{\mathcal{G}\bar{\mathcal{H}}})' K_{\mathcal{G}}^{-1} s_{\mathcal{G}}(\alpha_0) + o_p(1),
\end{aligned}$$

using $s^h_{\mathcal{H}}(\beta_{\alpha_0};\alpha_0) \overset{\text{a.s.}}{\to} 0$, with

$$\begin{aligned}
\tilde{\omega}_{\mathcal{G}\mathcal{H}} &= E_X E_{\alpha_0}\left[\frac{\partial}{\partial\alpha}\log f_{\mathcal{G}}(Y|X;\alpha)\log f_{\mathcal{H}}(Y|X;\beta_0)\right]_{\alpha=\alpha_0}, \\
\tilde{\omega}_{\mathcal{G}\bar{\mathcal{H}}} &= E_X E_{\alpha_0}\left[\frac{\partial}{\partial\alpha}\log f_{\mathcal{G}}(Y|X;\alpha)\log f_{\mathcal{H}}(Y|X;\beta_{\alpha_0})\right]_{\alpha=\alpha_0}.
\end{aligned}$$

To summarize the expansions, let

$$\hat{d} = \begin{pmatrix} \hat{s}_1 \\ \hat{s}_2 \\ \hat{l}_1 \\ \hat{l}_2 \end{pmatrix}, \qquad d = \begin{pmatrix} s_1 \\ s_2 \\ l_1 \\ l_2 \end{pmatrix},$$

$$w_t = \begin{pmatrix} s_{\mathcal{G}}(y_t|x_t;\alpha_0) \\ s_{\mathcal{H}}(y_t|x_t;\beta_0) \\ s_{\mathcal{H}}(y_t|x_t;\beta_{\alpha_0}) - s_1 \\ \frac{1}{H}\sum_{h=1}^{H} s_{\mathcal{H}}(y_t^h(\alpha_0)|x_t;\beta_0) - s_2 \\ \frac{1}{H}\sum_{h=1}^{H} s_{\mathcal{H}}(y_t^h(\alpha_0)|x_t;\beta_{\alpha_0}) \\ \log f_{\mathcal{H}}(y_t|x_t;\beta_{\alpha_0}) - \log f_{\mathcal{H}}(y_t|x_t;\beta_0) - l_1 \\ \frac{1}{H}\sum_{h=1}^{H} \log f_{\mathcal{H}}(y_t^h(\alpha_0)|x_t;\beta_0) - \frac{1}{H}\sum_{h=1}^{H} \log f_{\mathcal{H}}(y_t^h(\alpha_0)|x_t;\beta_{\alpha_0}) - l_2 \end{pmatrix},$$

and

$$A = \begin{pmatrix} -K_{\bar{\mathcal{H}}} B K_{\mathcal{G}}^{-1} & 0 & I & 0 & -K_{\bar{\mathcal{H}}}\tilde{K}_{\bar{\mathcal{H}}}^{-1} & 0 & 0 \\ \tilde{J}_{\mathcal{H}\mathcal{G}} K_{\mathcal{G}}^{-1} & -\tilde{K}_{\mathcal{H}} K_{\mathcal{H}}^{-1} & 0 & I & 0 & 0 & 0 \\ s_1' B K_{\mathcal{G}}^{-1} & 0 & 0 & 0 & s_1' \tilde{K}_{\bar{\mathcal{H}}}^{-1} & 1 & 0 \\ (\tilde{\omega}_{\mathcal{G}\mathcal{H}} - \tilde{\omega}_{\mathcal{G}\bar{\mathcal{H}}})' K_{\mathcal{G}}^{-1} & s_2' K_{\mathcal{H}}^{-1} & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Then,

$$\sqrt{T}(\hat{d} - d) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} A w_t + o_p(1).$$

Observe that $E_X E_0 w_t = 0$. Assuming the existence of $V = E_X E_0(w_t w_t')$,

$$\sqrt{T}(\hat{d} - d) \xrightarrow{\text{d}} N(0, AVA'),$$

by the central limit theorem. Note that all the submatrices in A can be consistently estimated, and hence A itself, by replacing $E_X E_0$ by $\frac{1}{T}\sum_{t=1}^{T}$, E_X by $\frac{1}{T}\sum_{t=1}^{T}$, E_{α_0} by $E_{\hat{\alpha}}$ or by $\frac{1}{H}\sum_{h=1}^{H}$ and using $y_t^h(\hat{\alpha})$ in place of y_t, α_0 by $\hat{\alpha}$, β_0 by $\hat{\beta}$, β_{α_0} by $\frac{1}{H}\sum_{h=1}^{H}\hat{\beta}_{\hat{\alpha}}^h$, and $(\hat{s}_1, \hat{s}_2, \hat{l}_1, \hat{l}_2)$ by (s_1, s_2, l_1, l_2), successively. Similar replacements in w_t yield $\hat{w}_t$ and $\hat{V} = \frac{1}{T}\sum_{t=1}^{T}\hat{w}_t\hat{w}_t'$ as a consistent estimator of V. A consistent estimator of AVA' follows.

Inspection of Aw_t reveals that no general asymptotic equivalences hold between subvectors of $\hat{d}$. More precisely, there does not exist in general a fixed non-zero matrix C such that $\sqrt{T}C(\hat{d} - d) = o_p(1)$, because V is not of reduced rank in general and A has not reduced row rank in general. This implies, in particular, that no general asymptotic equivalences exist between $\hat{s}_1$, $\hat{s}_2$, $\hat{l}_1$ and $\hat{l}_2$. This finding, and the full characterization of the joint first order limit distribution of $\hat{s}_1$, $\hat{s}_2$, $\hat{l}_1$ and $\hat{l}_2$ opens perspectives for jointly exploiting the evidence contained in these statistics against any of the hypotheses $\mathcal{G}\,\mathcal{E}\,\mathcal{H}$ and $P_{Y|X} \in \mathcal{G}$, thereby gaining in power compared to the standard score or LR test. The unresolved problem for doing this is to control the (asymptotic) size of the joint test. A fully joint test would typically take a quadratic form in $\sqrt{T}\hat{d}$, weighted by a consistent estimate of $(AVA')^{+}$, and refer to the χ^2 distribution with appropriate degrees of freedom. As we show below, asymptotic equivalences do appear when $\mathcal{G}\,\mathcal{E}\,\mathcal{H}$ (a fortiori when $P_{Y|X} \in \mathcal{G}$), making AVA' a singular matrix. In many cases of interest, consistent estimates of AVA' have an asymptotic rank that exceeds the rank of AVA', which makes consistent estimation of $(AVA')^{+}$ a difficult task (see also Andrews [1989]). In other words, the main difficulty for building a test on the full vector $\hat{d}$ is that the rank of his covariance matrix depends on whether or not $\mathcal{G}\,\mathcal{E}\,\mathcal{H}$, which is precisely the hypothesis being tested.

5.2 Limit distributions under the condition $\mathcal{G}\,\mathcal{E}\,\mathcal{H}$

The first order limit distribution of $\hat{d}$ when $\mathcal{G}\,\mathcal{E}\,\mathcal{H}$ is easily obtained using the results of the previous subsection. We then have $d = 0$ and

$$w_t = \begin{pmatrix} s_{\mathcal{G}}(y_t|x_t;\alpha_0) \\ s_{\mathcal{H}}(y_t|x_t;\beta_0) \\ s_{\mathcal{H}}(y_t|x_t;\beta_0) \\ \frac{1}{H}\sum_{h=1}^{H} s_{\mathcal{H}}(y_t^h(\alpha_0)|x_t;\beta_0) \\ \frac{1}{H}\sum_{h=1}^{H} s_{\mathcal{H}}(y_t^h(\alpha_0)|x_t;\beta_0) \\ 0 \\ 0 \end{pmatrix}.$$

Further, $K_{\mathcal{H}} = K_{\bar{\mathcal{H}}}$, $\tilde{K}_{\mathcal{H}} = \tilde{K}_{\bar{\mathcal{H}}}$, $\tilde{\omega}_{\mathcal{G}\mathcal{H}} = \tilde{\omega}_{\mathcal{G}\bar{\mathcal{H}}}$, $B = \tilde{K}_{\mathcal{H}}^{-1}\tilde{J}_{\mathcal{H}\mathcal{G}}$ and

$$A = \begin{pmatrix} -K_{\mathcal{H}}\tilde{K}_{\mathcal{H}}^{-1}\tilde{J}_{\mathcal{H}\mathcal{G}}K_{\mathcal{G}}^{-1} & 0 & I & 0 & -K_{\mathcal{H}}\tilde{K}_{\bar{\mathcal{H}}}^{-1} & 0 & 0 \\ \tilde{J}_{\mathcal{H}\mathcal{G}}K_{\mathcal{G}}^{-1} & -\tilde{K}_{\mathcal{H}}K_{\mathcal{H}}^{-1} & 0 & I & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix},$$

from which we obtain $\sqrt{T}\hat{l}_1 = o_p(1) = \sqrt{T}\hat{l}_2$ and the asymptotic equivalence

$$\sqrt{T}\hat{s}_1 = -K_{\mathcal{H}}\tilde{K}_{\mathcal{H}}^{-1}\sqrt{T}\hat{s}_2 + o_p(1).$$

We can be more precise about the limiting behaviour of $\hat{l}_1$ and $\hat{l}_2$ by considering the expansions

$$\begin{aligned} Tl_{\mathcal{H}}(\hat{\beta}^h_{\hat{\alpha}}) &= Tl_{\mathcal{H}}(\hat{\beta}) - \frac{T}{2}(\hat{\beta}^h_{\hat{\alpha}} - \hat{\beta})' K_{\mathcal{H}} (\hat{\beta}^h_{\hat{\alpha}} - \hat{\beta}) + o_p(1), \\ Tl^h_{\mathcal{H}}(\hat{\beta};\hat{\alpha}) &= Tl^h_{\mathcal{H}}(\hat{\beta}^h_{\hat{\alpha}};\hat{\alpha}) - \frac{T}{2}(\hat{\beta}^h_{\hat{\alpha}} - \hat{\beta})' \tilde{K}_{\mathcal{H}} (\hat{\beta}^h_{\hat{\alpha}} - \hat{\beta}) + o_p(1), \end{aligned}$$

wherefrom

$$\begin{aligned} -2T\hat{l}_1 &= \frac{T}{H}\sum_{h=1}^{H}(\hat{\beta}^h_{\hat{\alpha}} - \hat{\beta})' K_{\mathcal{H}} (\hat{\beta}^h_{\hat{\alpha}} - \hat{\beta}) + o_p(1), \\ -2T\hat{l}_2 &= \frac{T}{H}\sum_{h=1}^{H}(\hat{\beta}^h_{\hat{\alpha}} - \hat{\beta})' \tilde{K}_{\mathcal{H}} (\hat{\beta}^h_{\hat{\alpha}} - \hat{\beta}) + o_p(1). \end{aligned}$$

Upon gathering previous results,

$$\begin{aligned} \sqrt{T}(\hat{\beta}^h_{\hat{\alpha}} - \hat{\beta}) &= \sqrt{T}\tilde{K}^{-1}_{\mathcal{H}} s^h_{\mathcal{H}}(\beta_0;\alpha_0) + \sqrt{T}\tilde{K}^{-1}_{\mathcal{H}}\tilde{J}_{\mathcal{H}\mathcal{G}}K^{-1}_{\mathcal{G}}s_{\mathcal{G}}(\alpha_0) - \sqrt{T}K^{-1}_{\mathcal{H}}s_{\mathcal{H}}(\beta_0) \\ &= -\sqrt{T}K^{-1}_{\mathcal{H}}\hat{s}_1 + o_p(1) \\ &= -\sqrt{T}\tilde{K}^{-1}_{\mathcal{H}}\hat{s}_2 + o_p(1), \end{aligned}$$

yielding the asymptotic equivalences

$$\begin{aligned} -2T\hat{l}_1 &= T\hat{s}_1' K^{-1}_{\mathcal{H}}\hat{s}_1 + o_p(1) \\ &= T\hat{s}_2'\tilde{K}^{-1}_{\mathcal{H}}K_{\mathcal{H}}\tilde{K}^{-1}_{\mathcal{H}}\hat{s}_2 + o_p(1), \\ -2T\hat{l}_2 &= T\hat{s}_2'\tilde{K}^{-1}_{\mathcal{H}}\hat{s}_2 + o_p(1) \\ &= T\hat{s}_1'K^{-1}_{\mathcal{H}}\tilde{K}_{\mathcal{H}}K^{-1}_{\mathcal{H}}\hat{s}_1 + o_p(1). \end{aligned}$$

Note that $-2T\hat{l}_1$ and $-2T\hat{l}_2$ are not in general asymptotically equivalent. The limit distributions can be summarized as follows. Let

$$v_t = \begin{pmatrix} s_{\mathcal{G}}(y_t|x_t;\alpha_0) \\ s_{\mathcal{H}}(y_t|x_t;\beta_0) \\ \frac{1}{H}\sum_{h=1}^{H} s_{\mathcal{H}}(y^h_t(\alpha_0)|x_t;\beta_0) \end{pmatrix}$$

and

$$D = \begin{pmatrix} \tilde{J}_{\mathcal{H}\mathcal{G}}K^{-1}_{\mathcal{G}} & -\tilde{K}_{\mathcal{H}}K^{-1}_{\mathcal{H}} & I \end{pmatrix}.$$

Now $Ev_t = 0$, and letting $\Sigma = E(v_t v_t')$ we have

$$\begin{aligned} \sqrt{T}\hat{s}_1 &\xrightarrow{\mathrm{d}} N(0, K_{\mathcal{H}}\tilde{K}^{-1}_{\mathcal{H}}D\Sigma D'\tilde{K}^{-1}_{\mathcal{H}}K_{\mathcal{H}}), \\ \sqrt{T}\hat{s}_2 &\xrightarrow{\mathrm{d}} N(0, D\Sigma D'), \\ -2T\hat{l}_1 &\xrightarrow{\mathrm{d}} M(\lambda(\tilde{K}^{-1}_{\mathcal{H}}K_{\mathcal{H}}\tilde{K}^{-1}_{\mathcal{H}}D\Sigma D')), \\ -2T\hat{l}_2 &\xrightarrow{\mathrm{d}} M(\lambda(\tilde{K}^{-1}_{\mathcal{H}}D\Sigma D')), \end{aligned}$$

where $M(\lambda(W))$ is the distribution of a weighted sum of independent χ^2 variates with weights equal to the eigenvalues of W. The matrices D and Σ and the necessary eigenvalues can be consistently estimated by the procedure outlined in the previous subsection. If we can determine the rank of the asymptotic covariance matrices of $\sqrt{T}\hat{s}_1$ and $\sqrt{T}\hat{s}_2$, asymptotic score and reversed score encompassing tests follow readily. Asymptotic LR and reversed LR encompassing tests follow also from the limit distribution given above. They require the calculation of critical values of weighted sum of chi-squares distributions, which can easily be obtained by simulation. Note that LR and reversed LR tests do not require the determination of the rank of a matrix.

5.3 Limit distributions under the condition $P_{Y|X} \in \mathcal{G}$

Further simplifications occur when $P_{Y|X} \in \mathcal{G}$. We have $F_{\mathcal{G}}(\alpha_0) = P_{Y|X}$, wherefrom $\tilde{K}_{\mathcal{H}} = K_{\mathcal{H}}$, yielding

$$\begin{aligned}
\sqrt{T}\hat{s}_1 &= -\sqrt{T}\tilde{J}_{\mathcal{H}\mathcal{G}}K_{\mathcal{G}}^{-1}s_{\mathcal{G}}(\alpha_0) + \sqrt{T}\left(s_{\mathcal{H}}(\beta_0) - \frac{1}{H}\sum_{h=1}^{H} s_{\mathcal{H}}^h(\beta_0;\alpha_0)\right) + o_p(1) \\
\sqrt{T}\hat{s}_2 &= \sqrt{T}\tilde{J}_{\mathcal{H}\mathcal{G}}K_{\mathcal{G}}^{-1}s_{\mathcal{G}}(\alpha_0) - \sqrt{T}\left(s_{\mathcal{H}}(\beta_0) - \frac{1}{H}\sum_{h=1}^{H} s_{\mathcal{H}}^h(\beta_0;\alpha_0)\right) + o_p(1)
\end{aligned}$$

and the asymptotic equivalences

$$\sqrt{T}\hat{s}_1 = -\sqrt{T}\hat{s}_2 + o_p(1)$$

and

$$\begin{aligned}
-2T\hat{l}_1 &= T\hat{s}_1' K_{\mathcal{H}}^{-1}\hat{s}_1 + o_p(1) \\
&= T\hat{s}_2' K_{\mathcal{H}}^{-1}\hat{s}_2 + o_p(1), \\
&= -2T\hat{l}_2 + o_p(1).
\end{aligned}$$

Note also that $s_{\mathcal{H}}(\beta_0)$ and $s_{\mathcal{H}}^h(\beta_0;\alpha_0)$, $h = 1,\ldots,H$, are conditionally independent and identically distributed, given x_t, $t = 1,\ldots,T$. Asymptotic score and reversed score tests and asymptotic LR and reversed LR tests of $P_{Y|X} \in \mathcal{G}$ can be constructed along the same lines as given in the previous subsection, taking advantage of the simplifications just mentioned.

6 Conclusion

We have outlined alternative procedures to the standard score and LR encompassing tests. They follow from restating the encompassing condition in terms of exponentially tilted models. Intuitively, the alternative procedures are obtained from reversing the roles of the true distribution generating the data and the pseudo-true distribution of the model under test. Application requires the models to be estimable by the method of maximum likelihood. No analytic calculations are needed beyond the analytic first and second derivatives of the log likelihood functions. The need to calculate mathematical expectations analytically is avoided by the use of any finite number of simulations from the model under test.

References

[1] Andrews, D.W.K. (1987), "Asymptotic Results for Generalized Wald Tests", *Econometric Theory*, **3**, 348–358.

[2] Barndorff-Nielsen, O.E. and D.R. Cox (1989), *Asymptotic Techniques for Use in Statistics*, London, Chapman and Hall.

[3] Bierens, H. (1994), *Estimation, Testing, and Specification of Cross-Section and Time Series Models*, Cambridge University Press, Cambridge.

[4] Bierens, H. (1991), "A Consistent Conditional Moment Test of Functional Form", *Econometrica*, **58**, 1443–1458.

[5] Chen, Y.-T. and C.-M. Kuan (2000), "The Pseudo-True Score Encompassing Test for Non-Nested Hypotheses", presented at the 8th World Congress of the Econometric Society, Seattle.

[6] Chesher, A. and R.J. Smith (1997), "Likelihood Ratio Specification Tests", *Econometrica*, **65**, 627–646.

[7] Dhaene, G. (1997), *Encompassing: Formulation, Properties and Testing*, Lecture Notes in Economic and Mathematical Systems, n^o 446, Springer Verlag, Berlin.

[8] Dhaene, G., C. Gouriéroux and O. Scaillet (1998), "Indirect Encompassing and Instrumental Models", *Econometrica*, **66**, 673–689.

[9] Gouriéroux, C. and A. Monfort (1994), "Testing Non-Nested Hypotheses", in *Handbook of Econometrics*, **IV**, ed. by R. Engle and D. McFadden, Elsevier Science, North-Holland, Amsterdam, 2583–2637.

[10] Gouriéroux, C. and A. Monfort (1995), "Testing, Encompassing and Simulating Dynamic Econometric Models", *Econometric Theory*, **11**, 195–228.

[11] Hendry, D.F. and J.-F. Richard (1990), "Recent Developments in the Theory of Encompassing", in *Contributions to Operations Research and Econometrics, The Twentieth Anniversary of CORE*, ed. by B. Cornet and H. Tulkens, MIT Press, Cambridge.

[12] Mizon, G.E. and J.-F. Richard (1986), "The Encompassing Principle and its Applications to Testing Non-nested Hypotheses", *Econometrica*, **54**, 657–678.

[13] Newey, W. (1985), "Maximum Likelihood Specification Testing and Conditional Moment Tests", *Econometrica*, **53**, 1047–1070.

[14] Kullback, S. and R.A. Leibler (1951), "On Information and Sufficiency," *Annals of Mathematical Statistics*, **22**, 79–86.

[15] Sawa, T. (1978), "Information Criteria for Discriminating among Alternative Regression Models", *Econometrica*, **46**, 1273–1292.

[16] Smith, R. (1994), "Consistent Tests of the Encompassing Hypothesis", CREST-DP 9403.

[17] Tauchen, G. (1985), "Diagnostic Testing and Evaluation of Maximum Likelihood Models", *Journal of Econometrics*, **30**, 415–443.

[18] Vuong, Q. (1989), "Lihelihood Ratio Tests for Model Selection and Non Nested Hypotheses", *Econometrica*, **57**, 303–333.

[19] White, H. (1982), "Maximum Likelihood Estimation of Misspecified Models", *Econometrica*, **50**, 1–26.