Parsimonious Encompassing: An Application to Non-Nested Hypotheses and Hausman Specification Tests

by

Jean-Pierre Florens*, David F. Hendry**, and Jean-François Richard***

DP#: 87-09

September 1987

* GREMAQ and Laboratoire de Statistique et de Probabilité. Université de Toulouse I et III.

** Nuffield College, Oxford.

*** CORE, Université Catholique de Louvain and ISDS, Duke University.

Abstract

A notion of parsimonious encompassing is analyzed within an information matrix framework. Hausman specification tests are reinterpreted within a class of Wald encompassing tests. Assumptions are discussed under which encompassing a non-nested alternative is equivalent to parsimoniously encompassing a minimal nesting model.
1. Introduction

Hausman specification tests and tests for non-nested hypotheses have received considerable attention in the recent econometric literature. The two references of direct relevance to the object of our paper are Holly (1982) and Mizon and Richard (1986) where additional references can be found. Mizon and Richard discuss a notion of parametric encompassing whereby a model is required to account for results derived under an alternative model. In particular, they argue that the specification tests can usefully be reinterpreted within this encompassing framework. (See Hendry and Richard (1987) for a survey).

Here we emphasize a notion of parsimonious encompassing, as outlined in Hendry and Richard (1982), whereby one analyzes whether a "simple" model can encompass a larger nesting model. For analytical convenience we only consider estimators that are asymptotically equivalent to Maximum Likelihood (ML) estimators so that our analysis is conducted in terms of information matrices. Hausman specification tests are then reinterpreted within this framework. Also we provide conditions under which encompassing a non-nested hypothesis is shown to be equivalent to parsimoniously encompassing a minimal nesting model.

The paper is organized as follows: the assumptions and notation are introduced in section 2; the notion of parsimonious encompassing is discussed in section 3, and it is applied to non-nested hypotheses testing in section 4 where Hausman specification tests are also considered. In section 5 we apply our analysis to the choice of regressions problem.

2. Parametric Encompassing

Let $Y_n$ denote an $nxp$ matrix of observations on a random vector $y \in \mathbb{R}^p$. Let $M$ denote a model characterized by the data densities $(p(Y_n | \alpha); n \rightarrow \infty)$ where $\alpha$
is a finite-dimensional identified parameter.

**Assumption 1:** $\hat{\alpha}_n$, the Maximum Likelihood Estimator (MLE) of $\alpha$, is a sufficient statistic.

Let us briefly comment upon the role of assumption 1 in our analysis. Below we introduce models that are nested within $\mathcal{M}$, and the analysis requires that the constrained MLEs associated with such models are (exact) functions of $\hat{\alpha}_n$. Assumption 1 is sufficient for that purpose but is clearly not necessary and hence could probably be weakened in specific cases. Note that assumption 1 holds within the exponential family of distributions. In section 4, we also discuss the situation where the constrained MLEs under consideration are only asymptotically functions of $\hat{\alpha}_n$.

**Assumption 2:** $\sqrt{n} \hat{\alpha}_n$ is asymptotically normally distributed as

\[
\sqrt{n} (\hat{\alpha}_n - \alpha) \overset{\mathcal{L}}{\rightarrow} N(0, J^{-1})
\]

where $J_{aa}$ is the information matrix associated with $\mathcal{M}$.

Let $\beta$ denote a parameter associated with a "rival" model $\mathcal{N}$ and let $\hat{\beta}_n$ denote the Pseudo-Maximum Likelihood Estimator (PMLE) of $\beta$. Assuming it exists\(^1\), the pseudo-true value of $\hat{\beta}_n$ is defined as

\[
\beta(\alpha) = \lim_{\mathcal{M}} \hat{\beta}_n
\]

The statistic

\[
\hat{\phi}_n = \hat{\beta}_n - \beta(\hat{\alpha}_n)
\]

---

\(^1\)See Gourieroux et al. (1984) for a discussion of the notion of pseudo-true value and of its relevance in the context of PMLE. Note that, under the usual technical conditions, assumptions 1 and 2 are sufficient to validate the existence of $\beta(\alpha)$ and the asymptotic normality of its estimator $\beta(\hat{\alpha}_n)$. 

3
compares the observed value of \( \hat{\beta}_n \) with an estimate of its pseudo-true value on \( \mathbb{H} \).

**Assumption 3:** \( \sqrt{n} \hat{\phi}_n \) is asymptotically normally distributed as

\[
\sqrt{n} \hat{\phi}_n \overset{\text{L}}{\rightarrow} N(0, V_{\alpha}(\hat{\phi}))
\]  (2.4)

Note that under suitable regularity conditions such as those discussed by Huber (1961) or, more recently, by Kent (1982) and White (1982), the covariance matrix \( V_{\alpha}(\hat{\phi}) \) is related to that of \( \hat{\alpha}_n \) in (2.1) and of \( \hat{\beta}_n \), which is denoted by \( V_{\alpha}(\hat{\beta}) \), as follows:

\[
V_{\alpha}(\hat{\phi}) = V_{\alpha}(\hat{\beta}) - D\gamma^{-1}_{\alpha\alpha} D'
\]  (2.5)

where \( D = \partial \beta(\alpha)/\partial \alpha' \). The following definition is taken from Mizon and Richard (1986):

**Definition 1:** \( \mathbb{H} \) (parametrically) encompasses \( \mathcal{H} \) with respect to \( \hat{\beta}_n \) if \( \hat{\phi}_n \) does not differ significantly from zero (based on the sampling distribution of \( \hat{\beta}_n \) on \( \mathbb{H} \)).

Under assumption 3 a Wald Encompassing Test (WET) statistic with respect to \( \hat{\beta}_n \) is given by:

\[
\eta_{\mathbb{H}}(\hat{\beta}_n) = n \hat{\beta}_n^T V_{\hat{\alpha}_n}^{-1}(\hat{\phi}) \hat{\beta}_n \overset{\text{L}}{\rightarrow} \chi^2(r)
\]  (2.6)

where \( r = \text{rank} V_{\hat{\alpha}_n}^{+}(\hat{\phi}) \) and \( V_{\hat{\alpha}_n}^{+}(\hat{\phi}) \) denotes the Moore-Penrose inverse of \( V_{\alpha}(\hat{\phi}) \), evaluated at \( \alpha = \hat{\alpha}_n \), thus allowing for the fact that the latter can be singular.

Applications of the notion of parametric encompassing in the field of non-nested hypotheses testing are discussed in Mizon (1984) and Mizon and
Richard (1986) where it is shown that a large number of test statistics can be provided with encompassing interpretations.

3. Parsimonious Encompassing

We now consider the case where \( \mathcal{N} \) is nested within \( \mathbb{M}^2 \). More specifically, at the cost of considering implicit reparameterizations, it proves convenient to assume that \( \alpha \) can be partitioned as \( \alpha = (b, c) \) where \( b \in \mathbb{R}^k \) and \( c \in \mathbb{R}^l \) in such a way that \( \mathcal{N} \) is associated with the hypothesis \( c = 0 \), under which \( b = \beta \). Following assumption 1, the PMLE of \( \beta \) is then a function of \( \hat{n} = (\hat{b}_n, \hat{c}_n) \) only, say

\[
\hat{\beta}_n = h(\hat{b}_n, \hat{c}_n)
\]  

(3.1)

It is assumed that \( h \) is differentiable with respect to \( b \) and \( c \). The pseudo-true value of \( \hat{\beta}_n \) on \( \mathbb{M} \) is then given by \( h(b, c) \) so that \( \sqrt{n}\hat{\beta}_n \) is identically equal to zero. In this sense, \( \mathbb{M} \) "exactly" encompasses any model \( \mathcal{N} \) which is nested within it. It follows that, conceptually at least, \( \mathbb{M} \) could always be made to encompass a rival model by being extended to a more general nesting model (see Hendry and Richard, 1982). Such a possibility is clearly of little practical value in worlds where limited sample evidence enforces parsimony.

Therefore, we examine the more relevant issue of whether or not the "simpler" model \( \mathcal{N} \) (parsimoniously) encompasses the "larger" model \( \mathbb{M} \). Note that the pseudo-true value of \( (\hat{b}_n, \hat{c}_n) \) on \( \mathcal{N} \) is given by \( (\beta, 0) \). The information

\footnote{There are a number of ways in which one can define the concept of nested models. See, in particular, the discussion in Pesaran (1987). We adopt here a definition which is directly suited to the object of our paper and do not discuss explicitly its relationship to the definition in Pesaran.}
matrix $J_{\alpha \alpha}$ in (2.1) is then partitioned conformably with $\alpha = (b, c)$ as

$$
J_{\alpha \alpha} = \begin{pmatrix}
J_{bb} & J_{bc} \\
J_{cb} & J_{cc}
\end{pmatrix} \text{ on } \mathbb{R}, \quad
g_{\alpha \alpha} = \begin{pmatrix}
g_{\beta \beta} & g_{\beta \beta} \\
g_{\beta \alpha} & g_{\alpha \alpha}
\end{pmatrix} \text{ on } \mathbb{N}
$$

(3.2)

On $\mathbb{N}$, $\hat{\beta}_n$ coincides with the constrained MLE on $(b, c)$ in $\mathbb{R}$, subject to the valid restrictions $c = o$. Hence, as shown e.g. in Engle\(^3\) (1983, formula (95)):

$$
\sqrt{n} (\beta_n - \beta) = \sqrt{n} (\beta_n - \beta) + g_{\beta \beta}^{-1} g_{\beta \alpha} \sqrt{n} \hat{\varepsilon}_n + o_p(1)
$$

(3.3)

It follows that, under assumption 2,

$$
\sqrt{n} \begin{pmatrix}
\hat{\beta}_n - \beta \\
\hat{\varepsilon}_n
\end{pmatrix} \overset{d}{\rightarrow} N \begin{pmatrix}
o \\
o
\end{pmatrix}, \quad 
\begin{pmatrix}
g_{\beta \beta}^{-1} \\
g_{\beta \alpha}
\end{pmatrix} \begin{pmatrix}
o \\
o
\end{pmatrix}, \quad 
\begin{pmatrix}
g_{\beta \beta}^{-1} \\
g_{\beta \alpha}
\end{pmatrix}^{-1}
$$

(3.4)

and, hence, that

$$
\sqrt{n} \hat{\beta}_n = \sqrt{n} \begin{pmatrix}
\hat{\varepsilon}_n - \beta \\
\hat{\varepsilon}_n
\end{pmatrix} \overset{d}{\rightarrow} N \begin{pmatrix}
o \\
o
\end{pmatrix}, \quad 
\begin{pmatrix}
g_{\beta \beta}^{-1} \\
g_{\beta \alpha}
\end{pmatrix} \begin{pmatrix}
o \\
o
\end{pmatrix}, \quad 
\begin{pmatrix}
g_{\beta \beta}^{-1} \\
g_{\beta \alpha}
\end{pmatrix}^{-1}
$$

(3.5)

The asymptotic covariance matrix of $\sqrt{n} \hat{\beta}_n$ can be rewritten as

$$
V_{\beta}(\phi) = \Lambda_{\beta} \tilde{J}_{oo} \Lambda_{\beta}^{-1}
$$

(3.6)

with

---

\(^3\)This result also follows from the pioneering work by Aitchinson and Silvey (1958). If we let $q_b(b, c)$ denote the score vector relative to $b$, then $q_b(\beta_n, \hat{\varepsilon}_n) = q_b(\beta_n, o) = o$ by definition of the MLEs. Taylor series expansions of $q_b(\beta, o)$ around $(\beta_n, \hat{\varepsilon}_n)$ and $(\beta_n, o)$ respectively, yield the result.
\[ \mathcal{Y}_{\alpha\beta} = \mathcal{Y}_{\alpha\alpha} - \mathcal{Y}_{\alpha\beta} \mathcal{Y}_{\beta\beta}^{-1} \mathcal{Y}_{\beta\alpha} \] 

where \( \ell \) is the dimension of \( \mathcal{C} \) and \( I_\ell \) denotes the identity matrix of order \( \ell \).

Hence \( \mathcal{V}_\beta(\hat{\phi}) \) is of rank \( \ell \) and its Moore-Penrose inverse is given by

\[ \mathcal{V}_\beta(\hat{\phi})^{+} = \mathcal{A}_\beta^{-} \mathcal{Y}_{\alpha\alpha} \mathcal{A}_\beta^{+} \]  

where \( \mathcal{A}_\beta^{+} = (\mathcal{A}_\beta \mathcal{A}_\beta)^{-1} \mathcal{A}_\beta^{+} \) so that \( \mathcal{A}_\beta^{+} \mathcal{A}_\beta = I_\ell \). Depending on which specific parameters in \( \mathcal{N} \) are of special interest to the proprietor of \( \mathcal{N} \) we can distinguish three basic classes of WET statistics (of \( \mathcal{N} \) relative to \( \mathcal{M} \)):

(i) The complete WET statistic:

\[ \eta_{\mathcal{N}}(\hat{\alpha}_n) = n \hat{\Phi}_n \mathcal{V}_\beta(\hat{\phi})^{+} \hat{\phi}_n \]  

where \( \hat{\phi}_n \) and \( \mathcal{V}_\beta(\hat{\phi}) \) are defined in \((3.5)\) and \((3.8)\) respectively;

(ii) The Hausman WET statistic which is relative to the coefficients that are not subject to the restrictions associated with \( \mathcal{X} \):

\[ \eta_{\mathcal{N}}(\hat{\beta}_n) = n \hat{\beta}_n \mathcal{V}_\beta^{-1}(\hat{\phi}_b) \hat{\beta}_b \]  

where \( \hat{\beta}_b = \hat{\beta}_n - \hat{\beta}_n \) and, following \((3.5)\), \( \mathcal{V}_\beta(\hat{\phi}_b) = \mathcal{Y}_{\beta\alpha}^{-1} \mathcal{Y}_{\beta\beta}; \)

(iii) The simplifications WET statistic which is relative to the coefficients that are restricted within \( \mathcal{X} \):

\[ \eta_{\mathcal{N}}(\hat{\alpha}_n) = n \hat{\varepsilon}_n \mathcal{V}_\beta^{-1}(\hat{\varepsilon}) \hat{\alpha}_n \]  

where, following \((3.5)\), \( \mathcal{V}_\beta(\hat{\varepsilon}) = \mathcal{Y}_{\alpha\alpha}^{-1} \).

The following equivalences hold on \( \mathcal{N} \):

\[ \mathcal{Y}_{\alpha\beta} = \mathcal{Y}_{\alpha\alpha} - \mathcal{Y}_{\alpha\beta} \mathcal{Y}_{\beta\beta}^{-1} \mathcal{Y}_{\beta\alpha} \]
(i) Following formulae (3.3) and (3.9), the complete WET statistics is asymptotically equivalent to the simplification WET statistics:

\[ \eta_n(\hat{\beta}_n) \xrightarrow{\mathcal{L}} \eta_n(\hat{\beta}_n) \xrightarrow{\mathcal{L}} \chi^2(\ell) \]

where \( \ell = \text{rank } V_\beta(\hat{\phi}) = \text{rank } T_{\infty, \beta} \). Hence \( \ell \) represents the maximum degrees of freedom for an encompassing test of \( \mathcal{X} \) relative to \( \mathcal{M} \);

(ii) The Hausman WET statistic, which reproduces the usual expression for a Hausman specification test when the estimators to be compared are respectively the constrained and unconstrained MLE's, is asymptotically equivalent to the WET statistic relative to \( T_{\beta_0, \beta_0} \). Hence, as noted by Holly (1982), it suffers a loss of degrees of freedom relative to the (direct) simplification test when \( \text{rank } T_{\beta_0} < \ell \). The two are asymptotically equivalent when \( k \geq \ell \) and \( \text{rank } T_{\beta_0} = \ell \).

4. Non-nested Hypotheses Testing

Let \( \mathcal{X}^\perp \) with parameter \( \gamma \) denote the submodel of \( \mathcal{M} \) associated with the hypothesis \( b=0 \), under which \( \gamma=\mathcal{C} \). The models \( \mathcal{X} \) and \( \mathcal{X}^\perp \) are mutually non-nested and \( \mathcal{M} \) is a minimal nesting model. As we discuss below we can equally consider any non-nested alternative \( \mathcal{X}^\star \) such that \( \mathcal{X}^\perp \subset \mathcal{X}^\star \subset \mathcal{M} \) (where "\( \subset \)" reads as "is nested within") preserving thereby the minimal nesting property of \( \mathcal{M} \).

There now exists a wide range of non-nested test statistics—see e.g. MacKinnon (1983) for a survey, Davidson and MacKinnon (1987) for a discussion of the properties of some of these tests or Mizon and Richard (1986) where non-nested statistics are reinterpreted within the encompassing framework. Most non-nested hypotheses test statistics are deliberately constructed in such a way that they do not require the specification of a nesting model even when
there exists an obvious one, as with the choice of regressors problem.
Moreover, the more "conventional" procedures which explicitly rely upon the
existence of a nesting model have occasionally been dismissed on that ground.
The analysis which follows indicates, on the contrary, that the two approaches
are essentially equivalent.

The properties of the WET statistics with respect to \( \gamma_n \) on \( \mathcal{N} \) follow from
the analysis in section 3 since \( \mathcal{X} \) is nested within \( \mathcal{N} \). Hence, following
assumption 1, \( \gamma_n \) is an exact function of \( \hat{\gamma}_n = (\hat{b}_n, \hat{c}_n) \), say

\[
\gamma_n = g(\hat{b}_n, \hat{c}_n)
\]  \hspace{1cm} (4.1)

The pseudo-true value of \( \gamma_n \) on \( \mathcal{N} \) is then given by

\[
\gamma(\beta) = g(\beta, 0)
\]  \hspace{1cm} (4.2)

and the corresponding encompassing difference is

\[
\hat{\gamma}_n = g(\hat{b}_n, \hat{c}_n) - g(\hat{\beta}_n, 0)
\]  \hspace{1cm} (4.3)

Therefore, the following asymptotic equivalence holds under the usual
differentiability conditions

\[
\sqrt{n} \hat{\gamma}_n \overset{\mathcal{N}}{\sim} \mathcal{N}(\mathcal{X}, \mathcal{N})
\]

\[
\sqrt{n} \hat{\gamma}_n \overset{\mathcal{L}}{\rightarrow} N(0, V_{\beta}(\hat{\gamma}))
\]

(4.4)

where \( G(\beta, 0) \) denotes the matrix of partial derivatives of \( g \) evaluated at
\((b, c) = (\beta, 0)\), and

\[
V_{\beta}(\hat{\gamma}) = G(\beta, 0) \cdot V_{\beta}(\hat{\beta}) \cdot G'(\beta, 0)
\]  \hspace{1cm} (4.5)

Thus, under assumptions 1 and 2, the condition

9
rank \( G(\beta,0) = k \) \hspace{1cm} (4.6)

is necessary and sufficient for the following asymptotic equivalence

\[
\eta_w(\hat{\alpha}) \sim \eta_w(\hat{\gamma}) \rightarrow \chi^2(k) \hspace{1cm} (4.7)
\]

In words, the fact that \( \mathcal{N} \) parsimoniously encompasses \( \mathbb{M} \) (w.r.t. \( \hat{\alpha} \)) is asymptotically equivalent to the fact that it encompasses \( \mathcal{N}^\perp \) (w.r.t. \( \hat{\gamma} \)). This equivalence extends to the case of a non-nested alternative model \( \mathcal{N}_\star \) which nests \( \mathcal{N}^\perp \) and is nested within \( \mathcal{N} \):

\[
\mathcal{N}^\perp \subset \mathcal{N}_\star \subset \mathcal{N} \hspace{1cm} (4.8)
\]

Adopting a shorthand notation whereby "\( \mathcal{N} \mathcal{E}_p \mathbb{M} \)" and "\( \mathcal{N} \mathcal{E} \mathcal{N}_\star \)" read respectively as "\( \mathcal{N} \) parsimoniously encompasses \( \mathbb{M} \)" and "\( \mathcal{N} \) encompasses \( \mathcal{N}_\star \)" and restricting our attention to complete encompassing, we can establish the equivalence relative to \( \mathcal{N}_\star \) as follows:

(i) \( \mathcal{N} \mathcal{E} \mathcal{N}_\star \Rightarrow \mathcal{N} \mathcal{E} \mathcal{N}^\perp \Leftrightarrow \mathcal{N} \mathcal{E}_p \mathbb{M} \),

where the first implication follow from the fact that \( \mathcal{N}^\perp \subset \mathcal{N}_\star \);

(ii) \( \mathcal{N} \mathcal{E}_p \mathbb{M} \Rightarrow \mathcal{N} \mathcal{E} \mathcal{N}_\star \), since \( \mathcal{N}_\star \subset \mathbb{M} \).

Our analysis unambiguously suggests that, within the class of models for which assumptions 1 and 2 apply, it is misleading to discuss non-nested hypotheses testing without reference to a (possibly implicit) nesting model, especially when there exists an obvious one as for the choice of regressors problem we discuss in section 5 below. Though it may occasionally prove convenient to avoid specifying a nesting model \( \mathbb{M} = \mathcal{N} \mathcal{N}^\perp \) (since there is a limited class of models for which explicit nesting is operational), it is nevertheless the case that denying the relevance of an (implicit) nesting
model leads to misunderstanding the very notion of non-nested hypotheses testing. The recurrent dismissal of the conventional F-test statistic for the choice of regressors problem provides an interesting illustration of such a pervasive misunderstanding.

The question naturally arises as to whether or not assumption 1---and, more specifically, the implied identity (4.1)---can be released. We note, in particular, that in the context of linear dynamic models, as discussed e.g. in Govaerts (1987) and Govaerts et al. (1987), the identity (4.1) is occasionally replaced by a weaker asymptotic equivalence of the form

$$\hat{\gamma}_n = g(\hat{\theta}_n, \hat{\delta}_n) + O(n^{-1/2})$$

(4.9)

Though under (4.9) the pseudo-true value of $\hat{\gamma}_n$ on $\mathcal{I}$ is still given by (4.2), the factors of order $n^{-1/2}$ cannot be neglected in the evaluation of the asymptotic covariance matrix of the encompassing difference $\sqrt{n}\hat{\delta}_n$ and the asymptotic equivalence (4.4) no longer holds. It may then be the case that the (complete) encompassing test against a non-nested rival model has fewer degrees of freedom than the (complete) parsimonious encompassing test against the (minimal) nesting model. An example is given in Govaerts (1987, example 4.1).

5. The Choice of Regressors Problem

The choice of regressors problem has been widely used for illustrative purposes in the literature on non-nested hypotheses testing---see e.g. Pesaran (1974). In this section we adopt to our context some of the results derived in Mizon and Richard (1986) for the non-nested choice of regressors problem and extend them to the case of parsimonious encompassing. Proofs are generally straightforward and, hence, are kept to a minimum.
We limit our attention to the case where the regressors are held "fixed", under the implicit assumption that they are strongly exogenous in the terminology of Engle et al. (1983). Analyses of the stochastic regressors case are found e.g. in Covaerts (1987) or in Hendry and Richard (1987).

Let the competing models be

\[ X: y = X\beta_1 + X_2\beta_2 + u, \quad u \sim N(0, \sigma^2 I_n) \quad (5.1) \]

\[ X_2: y = Z\gamma + v = Z_1\gamma_1 + Z_2\gamma_2 + v, \quad v \sim N(0, \tau^2 I_n) \quad (5.2) \]

where \( y \in \mathbb{R}^n \), \( X \in \mathbb{R}^{nxm} \), \( Z_1 \in \mathbb{R}^{nxk_1} \), \( Z_2 \in \mathbb{R}^{nxk_2} \), \( \beta_1 \in \mathbb{R}^{k_1} \), \( \gamma_1 \in \mathbb{R}^{k_2} \) (\( i = 1, 2 \)), \( k_1 + k_2 = k \) and \( k_1 + k_2 = \ell \). Let

\[ M_X = I_n - X(X'X)^{-1}X', \quad M_Z = I_n - Z(Z'Z)^{-1}Z'. \quad (5.3) \]

The partitionings in (5.1) and (5.2) are such that

\[ M_X Z_1 = 0, \quad M_Z Z_1 = 0 \quad (5.4) \]

while \( M_X Z_2 \) and \( M_Z X_2 \) have full column ranks. It is assumed that \( \ell > 0 \), otherwise \( X_2 \in \mathbb{N} \) and encompassing would be automatic. The minimal nesting model \( M \) can be reparameterized in different ways depending on which regressors are chosen to represent the \((k+\ell_2)\)-dimensional column space of \((X, Z)\). The parameterizations of interest for our purpose are

\[ M: y = Xb + Z_2c_2 + \epsilon = X_2b_2 + Zc + \epsilon, \quad \epsilon \sim N(0, \nu^2 I_n) \quad (5.5) \]

\( M \) naturally satisfies assumptions 1 and 2. In fact, so far as the regression coefficients are concerned, assumption 2 applies also in finite samples and the corresponding part of the information matrix coincides with the regressors
sample moment matrix. Also \( \lambda^1 \) is given by the regression of \( y \) on \( Z_2 \) only, say

\[
\lambda^1: y = X_2 \tilde{\gamma}_2 + w, \quad w \sim N(0, \sigma^2 I_n)
\]

(5.6)

The MLE's and PMLE's are the usual least squares estimators and need not be reproduced here. \( \lambda \) and \( \lambda_\ast \) are nested within \( \lambda^1 \), according to the notion of nesting which was used in section 3, and the functions \( h \) and \( g \) in formulae (3.1) and (4.1) are characterized by the following identities

\[
\dot{\beta} = \tilde{\beta} + (X'X)^{-1}X'Z_2 \delta_2, \quad \dot{\delta}^2 = \delta^2 + \frac{1}{n} \tilde{e}_2'Z_2M_{\lambda}Z_2 \delta_2
\]

(5.7)

\[
\dot{\gamma} = \tilde{\gamma} + (Z'Z)^{-1}Z'X_2 \tilde{\delta}_2, \quad \dot{\delta}^2 = \delta^2 + \frac{1}{n} \tilde{e}_2'X_2M_{\lambda}X_2 \tilde{\delta}_2
\]

(5.8)

where the subscript \( n \), which is indicative of the sample size, has been deleted for notational convenience. We first discuss the case where the variances are known.

5.1 Known Variances

Straightforward algebraic manipulations lead to the following expression for the encompassing difference w.r.t. \( \dot{\gamma} \)

\[
\dot{\gamma}_{\lambda} - \dot{\gamma}_{\lambda_\ast} = A(Z_2'Z_1Z_2)^{-1}Z_2'X_2Y
\]

(5.9)

where \( \gamma_{\lambda} = (Z'Z)^{-1}Z'X\delta \), \( M_1 = I_n - Z_1(Z_1'Z_1)^{-1}Z_1' \), and

\[
A' = (-Z_2'Z_1(Z_1'Z_1)^{-1} : I_2)
\]

(5.10)

The (finite sample) covariance matrix of that difference is.
\[ v_2^\beta(\phi_2) = \sigma^2 AV_{22} A^\dagger \]  
(5.11)

where \( V_{22} = (Z'_2 M_1 Z_2)^{-1} Z'_2 M_2 Z_2 (Z'_2 M_1 Z_2)^{-1} \). Its Moore-Penrose inverse is

\[ v_2^\beta(\phi_2) = \frac{1}{\sigma^2} A^+ v_{22}^{-1} A^+ \]  
(5.12)

where \( A^+ = (A^\dagger A)^{-1} A^\dagger \) so that \( A^+ A = I_2 \). Hence, the WET statistics w.r.t. \( \hat{\gamma} \) is given by

\[ \eta_W(\hat{\gamma}) = \frac{1}{\sigma^2} \hat{\gamma} Z'_2 M_2 Z_2 \hat{\gamma} \rightarrow X^2(\ell_2) \]  
(5.13)

and coincides with the "conventional" statistic for testing the hypothesis that \( c_2 = 0 \) within \( \mathcal{M} \) when the variance is known.

Next consider the issue of whether or not \(\mathcal{W}_p \mathcal{M} \) w.r.t. \( \alpha = (b, c_2) \). The corresponding encompassing difference is given by

\[ \hat{\phi}_\alpha = \begin{pmatrix} \hat{b} - \beta \\ \hat{c}_2 \end{pmatrix} = B \hat{c}_2 \]  
(5.14)

with \( B = (-Z'_2 X (X'X))^{-1} : I_{\ell_2} \) so that

\[ v_2^\beta(\hat{\phi}_\alpha) = \sigma^2 B (Z'_2 M_2 Z_2)^{-1} B^\dagger \]  
(5.15)

\[ v_2^\beta(\hat{\phi}_\alpha) = \frac{1}{\sigma^2} B^+ Z'_2 M_2 Z_2 B^+ \]  
(5.16)

where \( B^+ = (B^\dagger B)^{-1} B^\dagger \) and \( B^+ B = I_{\ell_2} \). It follows that

\[ \eta_2(\hat{\alpha}) = \eta_W(\hat{\gamma}) = \eta_W(\hat{c}_2) = \frac{1}{\sigma^2} \hat{c}_2 Z'_2 M_2 Z_2 \hat{c}_2 \rightarrow X^2(\ell_2) \]  
(5.17)
confirming our findings in formulae (3.12) and (4.4) when asymptotic equivalences are replaced by finite sample identities.

5.2 Unknown Variances

The finite sample distributions of variance estimators being non Normal, the analysis under unknown variances partially reverts to investigating asymptotic equivalences. The encompassing difference w.r.t. $\hat{\sigma}^2$ can be written as

$$n(\hat{\sigma}^2 - \sigma^2) = -(y + X\beta)'Z'(\hat{\gamma} - \gamma)$$  \hspace{1cm} (5.18)

or, equivalently, taking advantage of formula (5.9)

$$\sqrt{n}(\hat{\sigma}^2 - \sigma^2) = -(\hat{\gamma}_2 + \gamma_{2\beta}) \cdot \frac{Z^ memoir_{2}XZ_2}{n} \cdot \sqrt{n} \hat{e}_2$$  \hspace{1cm} (5.19)

The following asymptotic equivalence holds

$$\sqrt{n}(\hat{\sigma}^2 - \sigma^2) \xrightarrow{\text{as.}} 2\gamma_{2\beta} \cdot \frac{Z^ memoir_{2}XZ_2}{n} \cdot \sqrt{n} \hat{e}_2$$  \hspace{1cm} (5.20)

where $\gamma_2$ is the subvector of $\gamma$ associated with the regressors in $Z_2$. Hence, the addition of $\hat{\sigma}^2$ to the list of unknown coefficients in $\mathcal{N}$ does not affect the form of the (asymptotic) WET statistic for complete encompassing and, in particular, adds no degrees of freedom to the test.

An interesting situation arises from the addition of $\nu^2$ to the list of unknown coefficients in $\mathcal{M}$. The pseudo-true value of $\nu^2$ on $\mathcal{N}$ is $\sigma^2$. It turns out that the joint limiting distribution of $\nu^2$ and $\sigma^2$ on $\mathcal{N}$ is singular:

$$\sqrt{n}\begin{pmatrix} \nu^2 - \sigma^2 \\ \sigma^2 - \sigma^2 \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N} \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, 2\sigma^4 \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \right)$$  \hspace{1cm} (5.21)
Hence the limiting distribution of the encompassing difference $\sqrt{n}(\hat{\nu}^2 - \hat{\delta}^2)$ is degenerate. Here again, by properties of the Moore-Penrose inverse, adding the variance to the unknown coefficients (in $\mathcal{M}$) does not affect the form of the WET statistic for complete (parsimonious) encompassing.

It is worth mentioning that, as shown e.g. in Pesaran (1974), the case where $\mathcal{K} \subset \mathcal{M}$ is precisely that for which the Cox-test statistic (as well as other one degrees of freedom non-nested test statistics) breaks down. Within our encompassing framework this problem is taken care of by using a non-degenerate asymptotic distribution for the appropriately scaled encompassing difference. The latter is given by

$$n(\hat{\nu}^2 - \hat{\delta}^2) = - \hat{\delta}_2^2 Z_2' \mathcal{K}_2 Z_2 \hat{\delta}_2$$

(5.22)

so that a valid test for variance encompassing in the nested case coincides with the conventional test of the null $c_2 = 0$ in $\mathcal{M}$. Hence that unknown variances contribute no additional degrees of freedom to test statistics for complete encompassing remains true for parsimonious encompassing.
REFERENCES


