EUI WORKING PAPERS IN ECONOMICS

EUI Working Paper ECO No. 94/5

Encompassing in Stationary Linear Dynamic Models

BERNADETTE GOVAERTS
DAVID F. HENDRY
JEAN-FRANÇOIS RICHARD

European University Institute, Florence

Please note
As from January 1990 the EUI Working Paper Series is divided into six sub-series, each sub-series is numbered individually (e.g. EUI Working Paper **ECO** No. 90/1).

EUROPEAN UNIVERSITY INSTITUTE, FLORENCE

ECONOMICS DEPARTMENT

EUI Working Paper **ECO** No. 94/5

Encompassing in Stationary Linear Dynamic Models

BERNADETTE GOVAERTS
DAVID F. HENDRY
JEAN-FRANÇOIS RICHARD

BADIA FIESOLANA, SAN DOMENICO (FI)

Printed in Italy in February 1994
European University Institute
Badia Fiesolana
I – 50016 San Domenico (FI)
Italy

Encompassing in Stationary Linear Dynamic Models

Bernadette Govaerts,
Solvay & Cie, Bruxelles, Belgium,

David F. Hendry
Nuffield College, Oxford, UK,

and

Jean-François Richard
University of Pittsburg, Pennsylvania, USA.

October 1993

Abstract

A model M1 encompasses a rival model M2 if M1 can explain M2's results. A Wald Encompassing Test (WET) checks if a statistic of interest to M2 coincides with an estimator of its predicted value under M1. We propose techniques for evaluating WETs in stationary, linear, dynamic, single equations with weakly exogenous regressors, extending results for strong exogeneity. Dynamics can constrain M1's predictions of M2's findings, so encompassing tests can differ from existing tests as examples illustrate. Their asymptotic power functions are compared with the outcomes in a small Monte Carlo. The results support the use of parsimonious encompassing tests.

Keywords: Encompassing; dynamics; weak exogeneity; parsimony; Monte Carlo.

This research was financed in part by grants B00220012 and R000231184 from the U.K. Economic and Social Research Council. Helpful comments from Jean-Pierre Florens, Grayham Mizon, Mark Steel and two anonymous referees are gratefully acknowledged. This paper was presented to an Econometrics Workshop at the EUI in September 1993.

Proofs and reprints should be sent to the second author at Nuffield College, Oxford, OX1 1NF UK.

1. Introduction

One model M_1 is said to encompass a rival model M_2 of the same variable y if the former can account for the results obtained by the latter. This notion is a natural component of a progressive research strategy and has been formalized in the econometric literature: see Hendry and Richard [1990] for an overview and bibliographic perspective.

We will first define encompassing at the level of estimated models, emphasizing dynamic equations, relate it to a more heuristic concept and to nesting, discuss testing for encompassing, then extend the definitions to conditional settings. Our discussion draws on Florens, Hendry and Richard [1991] to which the reader is referred for a formal analysis.

Let M_1 and M_2 denote two competing parametric dynamic models with no exogenous variables. Their (sequential) density functions for a common vector $\boldsymbol{y}_t$ are $f(\boldsymbol{y}_t | \boldsymbol{Y}_{t-1}, \boldsymbol{\alpha})$ for $\boldsymbol{\alpha} \in \mathcal{A}$ and $g(\boldsymbol{y}_t | \boldsymbol{Y}_{t-1}, \boldsymbol{\delta})$ for $\boldsymbol{\delta} \in \mathcal{D}$. The matrix $\boldsymbol{Y}_{t-1} = (\boldsymbol{y}_{t-1}, ..., \boldsymbol{y}_1)$, thereby grouping observations prior to period t. Observations start at $t = 1$, and initial conditions are assumed to be known.[1] The models are called estimated in that they are provided with estimators $\hat{\boldsymbol{\alpha}}_T$ and $\hat{\boldsymbol{\delta}}_T$ for any finite sample $\boldsymbol{Y}_T$. Let $\hat{M}_1 = (M_1, \hat{\boldsymbol{\alpha}}_T)$ and $\hat{M}_2 = (M_2, \hat{\boldsymbol{\delta}}_T)$.

Definition 1: $\hat{M}_1$ exactly encompasses $\hat{M}_2$ if and only if there exists a sequence of functions $\{\boldsymbol{\lambda}_T\}$ such that $\hat{\boldsymbol{\delta}}_T = \boldsymbol{\lambda}_T(\hat{\boldsymbol{\alpha}}_T)$, M_1-almost surely.

When attention is restricted to a subset $\boldsymbol{\delta}_1$ of $\boldsymbol{\delta}$, definition 1 applies subject to the qualification 'with respect to $\hat{\boldsymbol{\delta}}_{1T}$'.

In general, we would not expect encompassing to hold exactly for finite sample sizes, even if M_1 were the data generation process (DGP). A weaker requirement is that of asymptotic encompassing, which we refer to as encompassing when no ambiguity arises:

Definition 2: $\hat{M}_1$ asymptotically encompasses $\hat{M}_2$ (denoted by $\hat{M}_1 \mathcal{E} \hat{M}_2$) if and only if there exists a function $\boldsymbol{\lambda}$ such that $\hat{\boldsymbol{\delta}}_T = \boldsymbol{\lambda}(\hat{\boldsymbol{\alpha}}_T) + O_p(T^{-\frac{1}{2}})$, M_1-almost surely.[2]

The concept of asymptotic encompassing captures the heuristic notion that the DGP ought to encompass all competing models. If M_1 were the DGP, asymptotic encompassing would

---

[1] This assumption is largely for notational convenience: the specification of M_1 and M_2 can be extended to incorporate sampling distributions for $\boldsymbol{Y}_0$.

[2] The order of convergence can be adapted to the class of models under consideration. $1/\sqrt{T}$ is appropriate for the present paper where we consider stationary processes.

generally hold under appropriate technical conditions, with $\lambda(\alpha)$ being given by the *plim* of $\hat{\delta}_T$ under M_1. In particular, if $\hat{\delta}_T$ were a pseudo-maximum likelihood (ML) estimator, $\lambda(\alpha)$ would coincide with the pseudo-true value associated with the Kullback-Leibler information criterion (KLIC): see e.g. Sawa [1977], Kent [1986] or Gourieroux and Montfort [1991].

If we define nesting such that it implies the existence of a function $\delta = d(\alpha)$, and use an estimation procedure which is equivariant over such a function, then nesting implies encompassing. However, nesting is often defined by the weaker requirement that the KLIC of M_2 relative to M_1 be zero (see e.g. Pesaran [1987] or Gourieroux and Montfort, 1991). Then nesting also implies asymptotic encompassing for estimators that are asymptotically equivalent to (pseudo-) ML. Naturally, encompassing does not imply nesting as illustrated by the DGP encompassing all rival models, whether formally nested or not within it. The distinction between the two concepts of nesting is irrelevant to the analysis below and so $M_1 \subset M_2$ reads as 'M_1 is nested in M_2' without qualification.

From the perspective of econometric modelling, an important concept is parsimonious encompassing which determines whether a 'simple' model is capable of capturing the salient features of a more 'general' model within which it is nested.

Definition 3: $\hat{M}_1$ parsimoniously encompasses $\hat{M}_2$ ($\hat{M}_1 \mathcal{E}_p \hat{M}_2$) if and only if (*i*) $M_1 \subset M_2$ and (*ii*) $\hat{M}_1 \mathcal{E} \hat{M}_2$.

The concept of parametric encompassing in Mizon and Richard [1986] is closely related to definitions 1 and 2 (asymptotically), but restricts λ_T to be a pseudo-true value. That restriction creates conceptual problems (such as loss of transitivity) but delivers an operational concept that has important unifying features for testing nested and non-nested hypotheses.

Tests of whether or not $\hat{M}_1$ encompasses $\hat{M}_2$ are generically based on measures of divergence between $\hat{\delta}_T$ and $\lambda_T(\hat{\alpha}_T)$ for some suitable choice of $\{\lambda_T\}$ (see Florens *et al*, 1991). Typically one aims at selecting a sequence $\{\lambda_T\}$ which minimizes the divergence of $\hat{M}_2$ relative to $\hat{M}_1$, but this often results in intractable functional optimization problems, so that approximate solutions must be considered. Given the previous discussion, possible candidates are (finite sample or asymptotic) pseudo-true values. Wald encompassing test

(WET) statistics, introduced by Mizon and Richard [1986] and central to the objectives of the present paper, rely upon pseudo-true values to examine whether or not the encompassing difference $\sqrt{T}[\hat{\delta}_T - \lambda(\hat{\alpha}_T)]$ is significant on M_1 (see section 2 below).

Consider the case where the competing models include a vector of weakly exogenous variables. As discussed in Engle, Hendry and Richard [1983], weak exogeneity implies that there can be no efficiency gain in designing estimators which depend on the specification of the exogenous variables' process. Hence no such estimators will be considered in the present paper. Nevertheless, the sampling distribution of $\hat{\delta}_T$ on M_1 is bound to depend on the characteristics of the exogenous process since, in particular, M_2 is mis-specified from the viewpoint of M_1. Hence generalizations of definitions 1-3 for conditional models require explicit consideration of the exogenous process and an issue of robustness arises.

Let M_c denote a sequential model for the exogenous variables r_t, with density function $h(r_t | S_{t-1}, \tau)$ where S_{t-1} regroups past observations on $s_t' = (y_t : r_t')$ and $\tau \in \Upsilon$ is a nuisance parameter. Let $\hat{M}_i^c = (\hat{M}_i, \hat{M}_c)$ for i = 1, 2.

Definition 4: $\hat{M}_1$ encompasses $\hat{M}_2$ given $\hat{M}_c$ if and only if $\hat{M}_1^c \, \mathcal{E} \, \hat{M}_2^c$ relative to $\hat{\delta}_T$.

It is unreasonable in most applications to require that M_c be more than just an auxiliary model, which could be severely mis-specified for the DGP, so we extend the definitions to a class of competing models $\mathcal{M}_c$ (if only for consideration of robustness):

Definition 5: $\hat{M}_1$ encompasses $\hat{M}_2$ given $\hat{\mathcal{M}}_c$, if and only if there exists $\hat{M}_c$ in $\hat{\mathcal{M}}_c$ such that $\hat{M}_1^c \, \mathcal{E} \, \hat{M}_2^c$ relative to $\hat{\delta}_T$.

The analysis of the choice of regressor problem in Mizon and Richard [1986] and Florens, Hendry and Richard [1988] can be reinterpreted within the context of these definitions. However, since they only considered cases where y_t does not Granger-cause r_t (see Granger, 1969), so that r_t is strongly exogenous in the sense of Engle *et al.* [1983], the analysis simplifies considerably, and for practical purposes can be treated as a 'fixed regressor' analysis. The question naturally arises as to whether their results remain valid once Granger causality feedbacks from y_t to r_t are allowed.

To simplify notation for the rest of the analysis, we do not explicitly distinguish between models and their estimated counterparts, using M_i as a shorthand for both. Following

Hendry and Richard [1982], the limit of $\hat{\mathsf{M}}_i$ under the DGP, equivalent to the reduction of M_i from the DGP, is called an empirical model (see Hendry, 1993), and it is the characteristics of the empirical model which determine the outcome of any modelling exercise.

The two objectives of our paper can now be formulated. First, we propose general techniques for the evaluation of WET statistics for single equations estimated from stationary, linear, dynamic systems where the regressors contain lagged dependent and weakly exogenous variables. We derive conditions under which results obtained for strong exogeneity extend to weak exogeneity. Because of feedbacks from the endogenous to the conditioning variables, the rival models do not fully characterize the joint data density, which needs to be completed by an auxiliary system linking the non-modelled variables. The formulation of the completing model is important for the validity and power of the encompassing tests and section 3 proposes a general approach which ensures at least a consistent test procedure.

The other objective is to explore the fact that dynamics often constrain the predictions which one model can make of another's findings, and encompassing tests that exploit such information can differ from existing tests. Consider the rival dynamic models:

M_1: $y_t \;=\; \beta y_{t-1} + \varepsilon_{1t}$;

M_2: $y_t \;=\; \gamma y_{t-2} + \varepsilon_{2t}$,

where $|\beta| < 1$ (to ensure stationarity), and each model assumes its error to be independent normal, mean zero, with variance σ_i^2, denoted $\mathsf{IN}(0,\sigma_i^2)$. Then, M_1 predicts γ to be β^2, but also predicts the presence of residual autocorrelation in M_2, since it views M_2 as:

M_2^*: $y_t \;=\; \beta^2 y_{t-2} + \varepsilon_{1t} + \beta\varepsilon_{1t-1}$.

These factors determine M_1's prediction of the estimate of γ in M_2. The WET statistic is equivalent to the usual F-test for deleting y_{t-2} from the linear nesting model:

M_n: $y_t \;=\; b y_{t-1} + c y_{t-2} + \upsilon_t$.

Both tests are approximations in finite samples due to the dynamics.

If we switch the roles of the competing models, then M_2 predicts β to be zero, and predicts the presence of an autoregressive error in M_1. The F-test has the same form, but now tests if $b = 0$, whereas the form of the WET statistic is noticeably different from the previous case due to the regressor in M_2 being at a longer lag than that in M_1. Such considerations extend to more interesting dynamic models which include regressors that are weakly exogenous for the parameters of interest under M_1.

The structure of the paper is as follows. Section 2 reviews results on encompassing tests in previously studied models, and section 3 extends those findings to stationary, linear dynamic models. The examples in section 4 illustrate implementation issues, including test power considerations. Section 5 concludes, and comments on extensions to integrated processes. The paper draws on Govaerts [1987] where additional results, proofs and examples are found (a copy is available from the first author upon request).

2. ENCOMPASSING TEST STATISTICS

This section summarizes the definitions and results needed for our later analysis. Proofs and additional details can be found in the literature noted in the introduction. We focus on parametric encompassing tests in the sense of Mizon and Richard [1986].

2.1 WET statistics

Ignoring for the present any complications arising from the treatment of exogenous variables, let $\tilde{\boldsymbol{\alpha}}$ denote a consistent estimator of $\boldsymbol{\alpha}$ under M_1 and $\tilde{\boldsymbol{\phi}}$ a statistic of interest in the context of M_2. The pseudo-true value $\boldsymbol{\phi}_{\alpha}$ of $\tilde{\boldsymbol{\phi}}$ under M_1 is:

$$\boldsymbol{\phi}_{\alpha} = plim_{M_1} \tilde{\boldsymbol{\phi}}. \tag{2.1}$$

Let $\tilde{\Delta}_{\phi}$ denote the encompassing difference relative to $\tilde{\boldsymbol{\phi}}$, namely the difference between $\tilde{\boldsymbol{\phi}}$ and an estimate of its pseudo-true value $\boldsymbol{\phi}_{\alpha}$:

$$\tilde{\Delta}_{\phi} = \tilde{\boldsymbol{\phi}} - \boldsymbol{\phi}_{\tilde{\alpha}} \tag{2.2}$$

The limiting distribution of $\sqrt{T}\,\tilde{\Delta}_{\phi}$ on M_1 is:

$\sqrt{T}\,\tilde{\Delta}_{\phi} \underset{M_1}{\overset{d}{\rightarrow}} N(0, V_{\alpha}[\sqrt{T}\,\tilde{\Delta}_{\phi}])$ where " $\underset{M_1}{\overset{d}{\rightarrow}}$ " means 'is asymptotically distributed on M_1 as'. Also, $V_{\alpha}[\sqrt{T}\,\tilde{\Delta}_{\phi}]$ is the asymptotic variance matrix of $\sqrt{T}\,\tilde{\Delta}_{\phi}$.

Using the estimated variance $V_{\tilde{\alpha}}[\sqrt{T}\,\tilde{\Delta}_{\phi}]$, a WET statistic with respect to $\tilde{\boldsymbol{\phi}}$ is given by:

$$\eta_W(\tilde{\phi}) = T\,\tilde{\Delta}_{\phi}' V^{+}_{\tilde{\alpha}}[\sqrt{T}\,\tilde{\Delta}_{\phi}]\tilde{\Delta}_{\phi} \underset{M_1}{\overset{d}{\rightarrow}} \chi^2(\rho) \tag{2.3}$$

where ρ is the rank of $V_{\tilde{\alpha}}[\sqrt{T}\,\tilde{\Delta}_{\phi}]$ and the superscript $^{+}$ denotes the Moore-Penrose inverse.

2.2 The choice of regressors problem

The choice of regressors problem usually takes the form:

$$\mathsf{M}_1: y = X\beta + \varepsilon_1, \ \ \varepsilon_1 \sim \mathsf{N}(0,\sigma^2 I_T), \ \ \alpha = (\beta,\sigma^2), \tag{2.4}$$

$$\mathsf{M}_2: y = Z\gamma + \varepsilon_2, \ \ \varepsilon_2 \sim \mathsf{N}(0,\tau^2 I_T), \ \ \delta = (\gamma,\tau^2), \tag{2.5}$$

where X and Z are full column rank matrices of dimensions $T\times k$ and $T\times l$ respectively. We assume for convenience that X and Z have no common regressors; the general case is easily treated at the cost of more cumbersome algebra. In the rest of the paper, the overscript $\tilde{}$ denotes (pseudo-) ML estimators, which coincide with ordinary least squares (OLS), except that the sums of squares in variance estimators are divided by T.

Another possible rival model for M_1 is M_n, the linear nesting model of M_1 and M_2:

$$\mathsf{M}_n: y = Xb + Zc + e, \ \ e \sim \mathsf{N}(0,\nu^2 I_T), \ \ d = (b,c,\nu^2). \tag{2.6}$$

We consider four classes of WET statistics, reflecting which parameters in M_2 or M_n are of interest to the builder of M_1. WET statistics against M_2 (non-nested encompassing) will be denoted by $\eta_\bullet$ and against M_n (parsimonious encompassing) by $\eta_\bullet^{\mathrm{p}}$.

(a) Complete WET statistic (CWET) $\eta_C = \eta_W(\tilde{\delta})$ $\eta_C^{\mathrm{p}} = \eta_W(\tilde{d})$;

(b) Simplification WET statistic (SWET) $\eta_S = \eta_W(\tilde{\gamma})$ $\eta_S^{\mathrm{p}} = \eta_W(\tilde{c})$;

(c) Hausman WET statistic (HWET) $\eta_H^{\mathrm{p}} = \eta_W(\tilde{b})$;

(d) Variance WET statistic (VWET) $\eta_V = \eta_W(\tilde{\tau}^2)$.

Their precise expressions are found e.g. in Florens *et al.* [1988]. The statistics η_H and η_V^{p} have been dropped because the Hausman WET statistic is not applicable in M_2 (since M_1 and M_2 have no common regressors), and η_V^{p} is not defined because the encompassing difference for the variance of the nesting model M_n is $\mathsf{O}(T^{-1})$.

A further important test in this context is the conventional F-test of the null hypothesis $c = 0$ in M_n. Two versions of it can be given: F_C, the conventional small sample F-statistic, and η_F, an asymptotically equivalent version:

$$\eta_F = \tilde{\sigma}^{-2} y' Q_X Z(Z' Q_X Z)^{-1} Z' Q_X y \ \underset{\mathsf{M}_1}{\overset{\mathrm{d}}{\rightarrow}} \ \chi^2(l), \tag{2.7}$$

where $Q_X = I - X(X'X)^{-1}X'$, so that:

$$F_C = \frac{T-k-l}{Tl}\, \eta_F\,(1 - T^{-1}\eta_F)^{-1}. \tag{2.8}$$

The WET statistics defined above and the F-statistic η_F are closely related as shown in proposition 2.1 below. In the rest of the paper, equations of the forms:

(a) $\eta_{\bullet} = \eta_F$; (b) $\eta_{\bullet} \underset{\mathsf{M}_1}{\approx} \eta_F$; and (c) $\eta_{\bullet} \underset{\mathsf{M}_1}{\simeq} \eta_F$,

respectively denote (a) finite sample equality; (b) large sample equality ($\eta_{\bullet} = \eta_F + O_p(T^{-1})$ on M_1); and (c) weak asymptotic equivalence ($\eta_{\bullet} = \eta_F + o_p(1)$ on M_1), where $O_p(T^{-\lambda})$ and $o_p(T^{-\lambda})$ are defined as in Mann and Wald [1943] and White [1984]. Using this notation, the main results available in the literature are:

Proposition 2.1: *Under the assumption that* **X** *and* **Z** *are strongly exogenous, we have*:

$$\tilde{\gamma} - \gamma_{\tilde{\alpha}} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'Q_X y\,; \tag{2.9}$$

$$\tilde{\tau}^2 - \tau^2_{\tilde{\alpha}} = -\frac{1}{T}(y + X\tilde{\beta})'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'Q_X y; \tag{2.10}$$

$$\eta_C \underset{\mathsf{M}_1}{\simeq} \eta_S = \eta_F = \eta_S^{\mathrm{p}} = \eta_C^{\mathrm{p}}\,. \tag{2.11}$$

See e.g. Florens *et al.* [1988] for proofs. In other words we find:

(*i*) The SWET statistic of M_1 against M_2 is equal to the *F*-statistic, and is asymptotically equivalent to the CWET statistic;

(*ii*) The SWET statistic of M_1 against M_n is equal to both the *F*-statistic and the CWET statistic;

(*iii*) $\mathsf{M}_1 \mathcal{E}\, \mathsf{M}_2$ if and only if $\mathsf{M}_1 \mathcal{E}_p\, \mathsf{M}_n$; and:

(*iv*) Simplification encompassing entails variance encompassing in static linear models; or equivalently, the implicit null hypothesis of η_S is included in the implicit null hypothesis of η_V (see Florens *et al.* [1988] for similar results for η_H and η_H^{p}).

A major objective of the present paper is to investigate under what conditions proposition 2.1 remains valid for weak exogeneity. For reasons of space, we restrict attention to the modified *F*-statistic η_F, together with the four WET statistics η_S, η_S^{p}, η_C and η_C^{p}.

3. STATIONARY LINEAR DYNAMIC MODELS

3.1 The models

Let y_t denote the value of the dependent variable, $\boldsymbol{r}_t$ the set of all current variables believed weakly exogenous for the parameters of interest by either investigator, and p the maximum lag length. Define $\boldsymbol{s}_t' = (y_t, \boldsymbol{r}_t')$, $\boldsymbol{s}_{(t-1)}' = (\boldsymbol{s}_{t-1}, \dots \boldsymbol{s}_{t-p})$ and $\boldsymbol{f}_t' = (\boldsymbol{s}_t', \boldsymbol{s}_{(t-1)}')$. Initial conditions are assumed to be known. The rival models M_1 and M_2 and the nesting model M_n are formulated as:

$$M_1\colon y_t | r_t, s_{(t-1)} \sim IN(\beta' x_t, \sigma^2); \qquad (3.1)$$

$$M_2\colon y_t | r_t, s_{(t-1)} \sim IN(\gamma' z_t, \tau^2); \qquad (3.2)$$

$$M_n\colon y_t | r_t, s_{(t-1)} \sim IN(b' x_t + c' z_t, \nu^2), \qquad (3.3)$$

where x_t and z_t are selections of k and l variables in $(r_t, s_{(t-1)})$. To derive the WET statistics, an auxiliary process is needed for the non-modelled variables. The completing model is defined as:

$$M_c : r_t | s_{(t-1)} \sim IN(\Gamma w_t, \Sigma), \qquad (3.4)$$

where $w_t \subseteq s_{(t-1)}$, and Γ is a matrix of unrestricted parameters. This model may, but need not have, an economic interpretation as it is purely instrumental in the method used to derive the WET statistics.

From (3.1) and (3.4), the completed model $M_1^c = (M_1, M_c)$ is a vector autoregressive (VAR) process in s_t. Let α denote the complete set of parameters of M_1^c: $\alpha = (\beta, \sigma^2, \Gamma, \Sigma)$ and suppose that they ensure the stationarity of M_1^c. The notation matches that of section 2, in that α denotes the parameters of the augmented M_1 model needed to derive the predicted values of the statistics in M_2 when there is feedback from the dependent variable onto the conditioning variables of either model. Two reformulations can be given for M_1^c; first, the Markovian representation (i.e. the companion form):

$$M_1^c\colon f_t | f_{t-1} \sim IN(\Pi(\alpha) f_{t-1}, \Omega(\alpha)), \qquad (3.5)$$

where $\Pi(\alpha)$ and $\Omega(\alpha)$ are known functions of α. Secondly, the marginal distribution of f_t:

$$M_1^c\colon f_t \sim N(0, \Psi(\alpha)) \qquad (3.6)$$

where $\Psi(\alpha) = \Omega(\alpha) + \Pi(\alpha)\Psi(\alpha)\Pi(\alpha)'$ is the marginal variance-covariance matrix of the variables evaluated under M_1^c. The ML estimator of $\Psi(\alpha)$ is obtained by replacing α by $\tilde{\alpha}$ and is denoted below by $\tilde{\Psi} = \Psi(\tilde{\alpha})$. The unconditional variance-covariance matrix of any subvector of f_t, such as x_t, will be denoted by using corresponding subscripts such as Ψ_{XX}.

We complete the definitions of the models with some further notation: $\mathcal{M}$ denotes the class of stationary VAR processes of the form (3.5) where Π and Ω are unrestricted, and $\mathcal{M}_1$ is the sub-set of $\mathcal{M}$ including only models for which M_1 is the process generating the $\{y_t\}$.

3.2 Derivation of WET statistics

We now derive the WET statistics relative to $\tilde{\delta} = (\tilde{\gamma}, \tilde{\tau}^2)$ for (complete) encompassing of M_2. The derivation is conceptually straightforward and technicalities are omitted unless they are essential for the development of the argument (see Govaerts [1987] for detailed derivations of the relevant expressions).

Encompassing methodology in conditional dynamic models requires that statistics are explicitly derived under the joint model $M_1^c = (M_1, M_c)$ and not just under M_1. Because the completing model M_c is instrumental in the analysis, it influences the values of the WET statistics and hence the outcomes of the tests. Consequently, a careful choice of M_c is required. This is discussed in section 3.3.

The method is based on the following result in Hannan [1970]. Let $\boldsymbol{A}$ denote the unconstrained estimator of the second-order moment matrix of $\boldsymbol{f}_t$:

$$\boldsymbol{A} = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{f}_t \boldsymbol{f}'_t, \tag{3.7}$$

then the asymptotic distribution of vec$\boldsymbol{A}$, on M_1^c, is given by:

$$\sqrt{T}\text{vec}(\boldsymbol{A} - \boldsymbol{\Psi}(\boldsymbol{\alpha})) \underset{M_1^c}{\overset{d}{\rightarrow}} \mathsf{N}(\boldsymbol{0}, \boldsymbol{\Phi}(\boldsymbol{\alpha})), \tag{3.8}$$

where $\boldsymbol{\Phi}(\boldsymbol{\alpha})$ is an expression in $\boldsymbol{\Psi}(\boldsymbol{\alpha})$ which follows from Hannan's formulae. Govaerts [1988] proposes an algorithm for evaluating $\boldsymbol{\Phi}(\boldsymbol{\alpha})$ which is based on a Jordan canonical form representation of the system (3.5). Since $\tilde{\delta}$ is a known function of $\boldsymbol{A}$, the encompassing differences:

$$\tilde{\Delta}_\delta = \tilde{\delta} - \delta_{\tilde{\alpha}} = (\tilde{\gamma} - \gamma_{\tilde{\alpha}}, \tilde{\tau}^2 - \tau^2_{\tilde{\alpha}})$$

are as follows. First:

$$\tilde{\gamma} = (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{y} \quad \text{and} \quad \gamma_\alpha = \boldsymbol{\Psi}_{ZZ}^{-1}\boldsymbol{\Psi}_{ZY} = \boldsymbol{\Psi}_{ZZ}^{-1}\boldsymbol{\Psi}_{ZX}\boldsymbol{\beta},$$

so that:

$$\begin{aligned}\tilde{\gamma} - \gamma_{\tilde{\alpha}} &= (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{y} - \tilde{\boldsymbol{\Psi}}_{ZZ}^{-1}\tilde{\boldsymbol{\Psi}}_{ZX}\tilde{\boldsymbol{\beta}} \\ &= (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{Q}_X\boldsymbol{y} + ((\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{X} - \tilde{\boldsymbol{\Psi}}_{ZZ}^{-1}\tilde{\boldsymbol{\Psi}}_{ZX})\tilde{\boldsymbol{\beta}}.\end{aligned} \tag{3.9}$$

Next:

$$\tilde{\tau}^2 = \frac{1}{T}\boldsymbol{y}'\boldsymbol{Q}_Z\boldsymbol{y}$$

and:

$$\begin{aligned}\tau^2_\alpha &= \Psi_{YY} - \Psi_{YZ}\Psi^{-1}_{ZZ}\Psi_{ZY} \\ &= \sigma^2 + \beta'\Psi_{XX}\beta - \beta'\Psi_{XZ}\Psi^{-1}_{ZZ}\Psi_{ZX}\beta \\ &= \sigma^2 + \beta'\Psi_{XX\cdot Z}\beta,\end{aligned}$$

so that $\tau^2_\alpha \geq \sigma^2$, showing that variance dominance remains necessary. Further:

$$\tilde{\tau}^2 - \tau^2_{\tilde{\alpha}} = -\tfrac{1}{T}(y+X\tilde{\beta})'Z(Z'Z)^{-1}Z'Q_X y + \tilde{\beta}'(\tfrac{1}{T}X'Q_Z X - \tilde{\Psi}_{XX\cdot Z})\tilde{\beta}. \quad (3.10)$$

Compared to (2.9) and (2.10), these expressions show that the difference between the dynamic and static case is principally based on the values of the differences:

$$D_\gamma = (Z'Z)^{-1}Z'X - \tilde{\Psi}^{-1}_{ZZ}\tilde{\Psi}_{ZX}; \quad (3.11)$$

$$D_{\tau^2} = \tfrac{1}{T}X'Q_Z X - \tilde{\Psi}_{XX\cdot Z} \quad (3.12)$$

or, collecting these together in a sufficient (but not necessary) formulation:

$$D = \tfrac{1}{T}\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z \end{bmatrix} - \begin{bmatrix} \tilde{\Psi}_{XX} & \tilde{\Psi}_{XZ} \\ \tilde{\Psi}_{ZX} & \tilde{\Psi}_{ZZ} \end{bmatrix}. \quad (3.13)$$

This is the difference between an unrestricted estimator of the joint regressor second moment matrix, and an estimator thereof restricted by the implications from M_1 for the dynamic behaviour of the regressors. For testing parsimonious encompassing of M_1 against M_n, Z is replaced by the total set of regressors (X, Z) in the previous formulae. In that case, all the formulae can be simplified, since projecting X on $\bar{Z} = (X, Z)$ ensures that:

$$(\bar{Z}'\bar{Z})^{-1}\bar{Z}'X = \tilde{\Psi}^{-1}_{ZZ}\tilde{\Psi}_{ZX} = (I{:}0) \text{ and } \tfrac{1}{T}X'Q_Z X = \tilde{\Psi}_{XX\cdot Z} = 0. \quad (3.14)$$

Thus, in nested models with valid completing assumptions, encompassing differences are identical in static and dynamic cases, but that does not imply the equality of the corresponding WET statistics as shown below.

The formula to derive $\mathsf{V}_\alpha[\sqrt{T}(\tilde{\delta}-\delta_{\tilde{\alpha}})] = \mathsf{V}_\alpha[\sqrt{T}\,\tilde{\Delta}_\delta]$ follows from the fact that $\tilde{\Delta}_\delta$, as a function of $\tilde{\alpha}$ and $\tilde{\delta}$, is a known, continuous and differentiable function of A in (3.7):

$$\tilde{\Delta}_\delta = h(A). \quad (3.15)$$

Hence, the asymptotic distribution of $\tilde{\Delta}_\delta$ on M^c_1 is given by:

$$\sqrt{T}\,\tilde{\Delta}_\delta \xrightarrow[\mathsf{M}^c_1]{d} \mathsf{N}(0, H(\alpha)\Phi(\alpha)H(\alpha)'), \quad (3.16)$$

where $H(\alpha) = plim_\alpha(\partial h(A)/\partial \text{vec}A')$. Analytical evaluation of $H(\alpha)$ is impractical for most models (even simple ones) but its numerical evaluation is general and based on elementary matrix operations. Nevertheless, under restricted conditions (detailed below) the analytical expression of $\mathsf{V}_{\alpha}[\sqrt{T}\,\tilde{\Delta}_{\delta}]$ is known, or almost known, and a direct comparison with the static case can be made.

3.3 Choice of the completing model

The choice of different completing models in the comparison of M_1 to M_2 can lead to different expressions for $\tilde{\Psi}$, D_{γ}, D_{τ^2}, D, and hence can affect the encompassing statistics, occasionally inducing different conclusions for the corresponding encompassing tests. Consequently, the issue of robustness of the encompassing analysis against alternative choices of M_c arises. The orders in probability of the differences D_{γ}, D_{τ^2}, D are at the root of the discussion. If $D = 0$, the WET will be equivalent to the F-test; when $D \neq 0$, different cases must be studied. Six cases are retained, corresponding to the possible values of the encompassing differences, and the following concepts are defined:

(i) M_1^c is exactly moment efficient (EME) against M_2 iff $D = 0$;

(ii) M_1^c is exactly projection efficient (EPE) against M_2 iff $D_{\gamma} = 0$ and $D_{\tau^2} = 0$;

(iii) M_1^c is strongly moment efficient (SME) against M_2 iff $D \underset{\mathcal{M}}{\approx} 0$;

(iv) M_1^c is strongly projection efficient (SPE) against M_2 iff $D_{\gamma} \underset{\mathcal{M}}{\approx} 0$ and $D_{\tau^2} \underset{\mathcal{M}}{\approx} 0$;

(v) M_1^c is weakly moment efficient (WME) against M_2 iff $D \underset{\mathcal{M}_1}{\simeq} 0$;

(vi) M_1^c is weakly projection efficient (WPE) against M_2 iff $D_{\gamma} \underset{\mathcal{M}_1}{\simeq} 0$ and $D_{\tau^2} \underset{\mathcal{M}_1}{\simeq} 0$.

These concepts are related to each other as summarized in proposition 3.1:

Proposition 3.1:

$$\begin{array}{ccc} \text{EME} & \longrightarrow & \text{EPE} \\ \downarrow & & \downarrow \\ \text{SME} & \longrightarrow & \text{SPE} \\ \downarrow & & \downarrow \\ \text{WME} & \longrightarrow & \text{WPE} \end{array}$$

There are counter examples showing that no other relation exists between these concepts.

If none of the properties in (i)-(vi) holds, the robustness of the corresponding WET statistics is seriously compromised. In fact, if either of:

$$D_{\gamma} \underset{\mathcal{M}_1}{\neq} 0 \quad \text{or} \quad D_{\tau^2} \underset{\mathcal{M}_1}{\neq} 0,$$

then there exists at least one $(\mathsf{M}_1, \mathsf{M}_c^*) \in \mathcal{M}_1$ such that $plim_{\alpha}\tilde{\Delta}_{\gamma} \neq \mathbf{0}$ or $plim_{\alpha}\tilde{\Delta}_{\tau^2} \neq 0$ under $(\mathsf{M}_1, \mathsf{M}_c^*)$. As a consequence, when M_c is mis-specified, M_1^c can fail to encompass M_2 despite the fact that M_1 is the true model: example 4.7 illustrates this phenomenon. Fortunately, WPE can always be achieved as stated in proposition 3.2.

Proposition 3.2: *For any pair of rival models* M_1 *and* M_2 *of the form (3.1) and (3.2), there exists at least one completing model* M_c *such that* M_1^c *is WME and WPE against* M_2.

A sufficient condition to ensure WME (and so WPE) is that $plim\, \boldsymbol{D} = \mathbf{0}$, so M_1^c must generate the correct number and type of second moments of the joint set of regressors, which requires that M_1^c include a sufficiently rich dynamic specification. Parsimonious encompassing automatically ensures minimal robustness for WET statistics:

Proposition 3.3: *If* $\mathsf{M}_1 \subset \mathsf{M}_2$, *then* M_1^c *is EPE against* M_2 *for any choice of completing model* M_c.

More important are the properties of the WET statistics for the different possible values of $\boldsymbol{D}_{\gamma}$, $\boldsymbol{D}_{\tau^2}$ and $\boldsymbol{D}$ summarized in proposition 3.4.

Proposition 3.4 $\mathsf{M}_1^c \leftrightarrow \mathsf{M}_2$ $\mathsf{M}_1^c \leftrightarrow \mathsf{M}_n$

Test	CWET		SWET		F		SWET		CWET
EME	η_C	$\underset{\mathcal{M}_1}{\simeq}$	η_S	$=$	η_F	$=$	η_S^p	$=$	η_C^p
SME	η_C	$\underset{\mathcal{M}_1}{\simeq}$	η_S	$\underset{\mathcal{M}}{\approx}$	η_F	$\underset{\mathcal{M}}{\approx}$	η_S^p	$\underset{\mathcal{M}}{\approx}$	η_C^p
SPE	η_C	$\underset{\mathsf{M}_1^c}{\simeq}$	η_S	$\underset{\mathsf{M}_1^c}{\simeq}$	η_F	$\underset{\mathsf{M}_1^c}{\simeq}$	η_S^p	$\underset{\mathsf{M}_1^c}{\simeq}$	η_C^p
SPE+WME	η_C	$\underset{\mathcal{M}_1}{\simeq}$	η_S	$\underset{\mathcal{M}_1}{\simeq}$	η_F	$\underset{\mathcal{M}_1}{\simeq}$	η_S^p	$\underset{\mathcal{M}_1}{\simeq}$	η_C^p .

In every case, the WET statistics are analytically known and are easily computed. The first row (EME) directly generalizes (2.11) to dynamic models; SME ensures strong asymptotic equivalence under $\mathcal{M}$ across all the tests; SPE only delivers weak asymptotic equivalence under M_1^c; but SPE and WME together deliver weak asymptotic equivalence under $\mathcal{M}$.

The important question is whether or not, given a pair of models M_1 and M_2, there exists a completing model M_c that ensures SPE or a stronger property. Unfortunately, such a model does not always exist, and as stated in proposition 3.2, the best property which can always be achieved is WME. Following Govaerts [1987], an algorithm to build a completing model for a given pair of models M_1, M_2, based on a systematic comparison of the two lists of regressors in M_1 and M_2, is described in the appendix. The outcome is a list of regressors defining a minimal completing model, denoted here by M_c^m, with a number of features:

(i) If M_1 and M_2 are such that there exists a completing model M_c under which (M_1,M_c) is EME (or SME), then (M_1,M_c^m) is EME (or SME);

(ii) The model (M_1,M_c^m) is always WME and, hence, WPE;

(iii) M_c^m is minimal in that its list of regressors is included in the list associated with any other model M_c that satisfies EME, SME or WME.

(iv) If there exists a model M_c for which SPE is achieved, it will be nested in M_c^m but the latter will not necessarily be SPE.

(v) The minimal completing model is invariant to the choice of rival model M_n^* nesting $M_2^{\perp}$ and nested in M_n.

It follows that, unless the proprietor of M_1 has particular reasons for selecting a specific completing model M_c, selecting M_c^m as the completing model achieves robustness within the class of VAR completing models. Other choices for M_c need not achieve even this minimal property. Finally, based on the appendix algorithm, a sufficient condition for SME is available which does not need any additional calculations to check if M_1^c is SME:

Proposition 3.5 *If (i) the number of parameters of* (M_1,M_c) *is equal to the number of data second-order moments used in the evaluation of the estimators of these parameters (neglecting terms of* $O_p(\frac{1}{T})$*); and (ii) all the cross products included in* $\frac{1}{T}\boldsymbol{X}'\boldsymbol{X}$, $\frac{1}{T}\boldsymbol{X}'\mathbf{Z}$ *and* $\frac{1}{T}\mathbf{Z}'\mathbf{Z}$ *appear in the estimators of the parameters of* M_1^c,
then (M_1,M_c) *is SME against* M_2.

Intuitively, the stated conditions ensure that $\boldsymbol{D}$ is at most $O_p(\frac{1}{T})$ for all members of $\mathcal{M}$, which in turn was shown above to entail SME.

3.4 Power Comparisons

Although proposition 3.4 relates the various tests η_C, η_S, η_F, η_S^p and η_C^p under the null, it does not follow that they have the same power properties when M_1 is false. In particular, care is required in formulating the maintained hypothesis within which power is studied.

There are three distinct ways in which M_1 might be mis-specified relative to the assumed DGP. First, the alternative (denoted M_a) may differ from M_1 in the direction of M_2 but retain elements of M_1; second, M_a could be M_2 itself; finally, none of the elements in either M_1 or M_2 may be present in M_a. In each case, to correctly evaluate the power, M_1 must remain congruent with the data evidence despite being mis-specified, and in the context of dynamic models this aspect is not easy to achieve.

When M_a corresponds to M_n (e.g. the union of M_1 and M_2 less any redundant elements), then the *F*-test is bound to have the highest power in large samples since its error will be an innovation with minimum variance in the class, whereas M_1 will fail to have an innovation error if M_2 is dynamic. However, tests for innovation errors seem more appropriate for such a scenario than the encompassing tests discussed here. Example 4.2 illustrates this, and points up the advantages of parsimonious encompassing tests and general-to-simple modelling strategies.

When M_a does not correspond to M_n, power rankings are less clear-cut, and in small samples may depend on degrees of freedom. Example 4.3 illustrates this case.

4. EXAMPLES

The aim of this section is to illustrate the concepts and properties defined above using some simple examples. The choice of the completing model is discussed, explicit forms of the WET differences and statistics are given, most of the possible cases of moment and projection efficiency are illustrated, and the minimal completing model is given for each example. The *F*-test against M_n is asymptotically valid (i.e. it has the correct size asymptotically and is consistent against fixed alternatives), and is the parsimonious encompassing statistic.

Example 4.1: The static case

Static linear models were formulated in section 2 as:

M_1: $y_t = \beta' x_t + \varepsilon_{1t}$ with the claim that $\varepsilon_{1t} \sim IN(0, \sigma_1^2)$;

M_2: $y_t = \gamma' z_t + \varepsilon_{2t}$ with the claim that $\varepsilon_{2t} \sim IN(0, \sigma_2^2)$.

The minimal completing model derived from the algorithm is simply and minimally the unrestricted white-noise process:

M_c^m: $\boldsymbol{w}_t = \boldsymbol{u}_t$ where $\boldsymbol{u}_t \sim \mathsf{IN}(0, \Sigma)$ when $\boldsymbol{w}'_t = (\boldsymbol{x}'_t, \boldsymbol{z}'_t)$.

EME, EPE etc. are verified in this case since, for example, $\Psi_{WW} = plim_\alpha\, T^{-1}\boldsymbol{W}'\boldsymbol{W} = \Sigma$ and $\tilde{\boldsymbol{\Phi}}_{WW} = \hat{\Sigma} = T^{-1}\boldsymbol{W}'\boldsymbol{W}$, so that $\boldsymbol{D} = 0$. The results given in proposition 3.4 apply and coincide with those given in proposition 2.1. However, as discussed in section 1, invalidly restricting Σ can lead to rejection of M_1 even when it is the DGP.

M_c^m can be written in an equivalent projection form for the conditional distribution of $\boldsymbol{x}_t$ given z_t, jointly with the marginal process for z_t:

M_c^{m*}: $\boldsymbol{x}_t = \boldsymbol{\pi} z_t + \boldsymbol{u}^*_{1t}$ where $\boldsymbol{u}^*_{1t} \sim \mathsf{IN}(0, \Sigma_{XX\cdot Z})$;

$z_t = \boldsymbol{u}_{2t}$ with $\boldsymbol{u}_{2t} \sim \mathsf{IN}(0, \Sigma_{ZZ})$,

where $\boldsymbol{\pi} = \Sigma_{XZ}\Sigma_{ZZ}^{-1}$ and $\boldsymbol{u}^*_{1t} \perp \boldsymbol{u}_{2t}$. The first part of M_c^{m*} is the completing model used by Hendry and Richard [1983] to introduce the encompassing principle in the static case. They show that defining the projection of $\boldsymbol{x}_t$ on z_t in this way is sufficient to ensure EPE to M_1^c. We have:

$$\Psi_{ZZ}^{-1}\Psi_{ZX} = \Sigma_{ZZ}^{-1}\Sigma_{ZX} = \boldsymbol{\pi}' \quad \text{and} \quad \tilde{\boldsymbol{\pi}}' = (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{X}; \tag{4.1}$$
$$\Psi_{XX\cdot Z} = \Sigma_{XX\cdot Z} \quad \text{and} \quad \hat{\Sigma}_{XX\cdot Z} = \tfrac{1}{T}\boldsymbol{X}'\boldsymbol{Q}_Z\boldsymbol{X},$$

so that only strong exogeneity needs to be postulated for the marginal process of z_t. This claim remains true even if in fact z_t is Granger-caused by y, so complete robustness is achieved.

Example 4.2: Autoregressive models

Reconsider the simple dynamic model from section 1:

M_1: $y_t = \beta y_{t-1} + \varepsilon_{1t}$ where $\varepsilon_{1t} \sim \mathsf{IN}(0, \sigma^2)$;

M_2: $y_t = \gamma y_{t-2} + \varepsilon_{2t}$ with $\varepsilon_{2t} \sim \mathsf{IN}(0, \tau^2)$,

such that $|\beta| < 1$. Then, M_1 is itself the completing model, and we have:

$$\Psi_{ZX} = plim_\alpha \frac{1}{T}\boldsymbol{y}'_{-2}\boldsymbol{y}_{-1} = \beta\sigma^2/(1-\beta^2), \tag{4.2}$$

where (e.g.) $\boldsymbol{y}'_{-1} = (y_2 \ldots y_{T-1})$, and:

$$\Psi_{ZZ} = plim_\alpha \frac{1}{T}\boldsymbol{y}'_{-2}\boldsymbol{y}_{-2} = [\sigma^2/(1-\beta^2)] = \Psi_{XX}; \tag{4.3}$$

so that $\tilde{\Psi}_{ZZ}^{-1}\tilde{\Psi}_{ZX} = \tilde{\beta}$, and hence the restricted second moments are determined by M_1 alone. Thus:

$$(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X} = (\mathbf{y}'_{-2}\mathbf{y}_{-2})^{-1}\mathbf{y}'_{-2}\mathbf{y}_{-1} = \tilde{\beta} + \mathsf{O}_p(\tfrac{1}{T})$$
$$= (\mathbf{y}'_{-1}\mathbf{y}_{-1})^{-1}\mathbf{y}'_{-1}\mathbf{y} + \mathsf{O}_p(\tfrac{1}{T}), \qquad (4.5)$$

where the terms of $\mathsf{O}_p(\frac{1}{T})$ are due to lagging. Similarly:

$$\tfrac{1}{T}\mathbf{X}'\mathbf{Q}_Z\mathbf{X} = \tilde{\Psi}_{XX.Z} + \mathsf{O}_p(\tfrac{1}{T}), \qquad (4.6)$$

and hence M_1 is SPE against M_2. One can show that M_1 is also SME on replacing β and σ^2 by their estimators in the above formulae. This result guarantees that, in this simple dynamic case, the WET statistics are equivalent to the *F*-statistic as stated in proposition 3.4.

If we switch the two competing models so the rival models become:

$$\mathsf{M}_1^{\dagger}: y_t = \beta y_{t-2} + \varepsilon_{1t}$$
$$\mathsf{M}_2^{\dagger}: y_t = \gamma y_{t-1} + \varepsilon_{2t},$$

then the forms of the statistics change noticeably and $\mathsf{M}_1^{\dagger}$ is no longer SME but only WME. This example illustrates the differences that can occur between η_F, η_S and η_S^{p} when including lagged dependent variables in the list of regressors. The WET statistics can be evaluated (unconditionally on the regressors) either directly or by taking advantage of the general technique discussed in section 3.2 above. The completing model is now $\mathsf{M}_2^{\dagger}$ and the complete encompassing differences relative to the parameters of $\mathsf{M}_2^{\dagger}$ and M_n (defined in section 1) are:

$$\tilde{\Delta}_{\delta} = (1:-\tfrac{1}{T}\mathbf{y}'\mathbf{y}_{-1})'(\mathbf{y}'\mathbf{y})^{-1}\mathbf{y}'\mathbf{y}_{-1}; \qquad (4.7)$$

$$\tilde{\Delta}_{d} = (1:-\tfrac{1}{T}(\mathbf{y}'\mathbf{y})^{-1}\mathbf{y}'\mathbf{y}_{-1}:-\tfrac{1}{T}\mathbf{y}'\mathbf{Q}_{\mathbf{y}_{-2}}\mathbf{y}_{-1})'(\mathbf{y}'\mathbf{Q}_{\mathbf{y}_{-1}}\mathbf{y})^{-1}\mathbf{y}'\mathbf{Q}_{\mathbf{y}_{-2}}\mathbf{y}_{-1} \qquad (4.8)$$

where $\mathbf{Q}_{\mathbf{y}_{-i}} = \mathbf{I}_T\text{-}\mathbf{y}_{-i}(\mathbf{y}'_{-i}\mathbf{y}_{-i})^{-1}\mathbf{y}'_{-i}$.

The variance-covariance matrices of $\tilde{\Delta}_{\delta}$ and $\tilde{\Delta}_{d}$ on M_1 are both singular of rank 1 and so:

$$\eta_C = \tilde{\sigma}^{-2}(\mathbf{y}'\mathbf{Q}_{\mathbf{y}_{-2}}\mathbf{y}_{-1})^2(\mathbf{y}'\mathbf{y})^{-1} \underset{\mathsf{M}_1}{\overset{d}{\to}} \chi^2(1) \qquad (4.9)$$

$$\eta_C^{p} = \tilde{\sigma}^{-2}(\mathbf{y}'\mathbf{Q}_{\mathbf{y}_{-2}}\mathbf{y}_{-1})^2(\mathbf{y}'\mathbf{Q}_{\mathbf{y}_{-1}}\mathbf{y})^{-2}\,\mathbf{y}'\mathbf{y} \underset{\mathsf{M}_1}{\overset{d}{\to}} \chi^2(1) \qquad (4.10)$$

$$\eta_F = \tilde{\sigma}^{-2}(\mathbf{y}'\mathbf{Q}_{\mathbf{y}_{-2}}\mathbf{y}_{-1})^2(\mathbf{y}'\mathbf{Q}_{\mathbf{y}_{-1}}\mathbf{y})^{-1} \underset{\mathsf{M}_1}{\overset{d}{\to}} \chi^2(1). \qquad (4.11)$$

In this example, the three statistics are ordered as $\eta_C \leq \eta_F \leq \eta_C^{p}$. They are asymptotically equivalent on any model for which $T^{-1}\mathbf{y}'y_{-1} \xrightarrow{p} 0$ as $T \to \infty$ (i.e., in particular on M_1 or on local alternatives to M_1 which do not introduce first-order autocorrelation). An interesting difference between the two orderings of the hypotheses here is that the implicit projection of the F-test switches from the realizable dynamic model $y_t = \phi y_{t-1} + u_t$ to the forward projection of y_{t-2} on y_{t-1}, although M_c^m remains the same.

Govaerts [1987] examined the power of the three statistics in (4.9)-(4.11) by Monte Carlo simulations under local alternatives to M_1 of the form:

$$y_t = T^{-\frac{1}{2}}\nu_1 y_{t-1} + \nu_2 y_{t-2} + \zeta_t = \theta y_{t-1} + \nu_2 y_{t-2} + \zeta_t \quad \text{where} \quad \zeta_t \sim \mathsf{IN}(0, \sigma_\zeta^2). \tag{4.12}$$

After correction for size, η_C turns out to have lower finite sample power than either η_F or η_C^p. The asymptotic power functions for η_C and η_F under the local alternative in (4.12) can be derived as follows. First, the *plims* of the data second moments under (4.12) are:

$$\Sigma_{yy} = (1-\nu_2)\sigma_\zeta^2/[(1+\nu_2)\{(1-\nu_2)^2-\theta^2\}]; \quad \Sigma_{yy_{-1}} = \theta\Sigma_{yy}/(1-\nu_2); \quad \text{and:}$$

$$\Sigma_{yy_{-2}} = [\nu_2(1-\nu_2)+\theta^2]\Sigma_{yy}/(1-\nu_2).$$

Estimation of M_1 leads to population parameter values $\beta_p = plim\ \tilde{\beta}$, and $\sigma_p^2 = plim\ \tilde{\sigma}^2$:

$$\beta_p = \Sigma_{yy}^{-1}\Sigma_{yy_{-2}} = [\nu_2(1-\nu_2)+\theta^2]/(1-\nu_2), \tag{4.13}$$

and:

$$\sigma_p^2 = \sigma_\zeta^2(1+[(1-\nu_2)\theta^2/(1+\nu_2)\{(1-\nu_2)^2-\theta^2\}]). \tag{4.14}$$

On local alternatives like (4.12), the $\eta_\bullet$ tests become non-central $\chi^2(1,\mu^2)$ where (see Mizon and Hendry, 1980):

$$\mu^2(\eta_F) = \nu_1^2/(1-\nu_2^2) = T\theta^2/(1-\nu_2^2); \tag{4.15}$$

$$\mu^2(\eta_C) = \sigma_\zeta^2\nu_1^2\{(1-\nu_2)^2-\theta^2\}/\sigma_p^2(1-\nu_2^2)(1-\nu_2)^2$$

$$= \mu^2(\eta_F)[1-2\theta^2/(1-\nu_2^2)(1-\nu_2)] + o(T^{-1}) \leq \mu^2(\eta_F). \tag{4.16}$$

Two factors contribute to the power loss arising from the smaller non-centrality parameter: on (4.12), $\sigma_p^2 \geq \sigma_\zeta^2$, which η_F avoids; and $\beta_p \neq \beta$, so the residuals in M_1 are not white noise, whereas $\{\zeta_t\}$ is, so M_1 ceases to be congruent and is not a good basis for testing. Even though θ is $O(T^{-\frac{1}{2}})$, the asymptotic power loss is quite large as figure 1 shows for the parameter values $\theta = 0.1$, and $\nu_2 = 0.8$ over $T = 70, ..., 300$. The small sample power of the F-test is also shown, based on recursive Monte Carlo using PC-NAIVE (see Hendry, Neale and Ericsson, 1991), to illustrate the applicability of the asymptotic formulae.

Example 4.3: Strong moment efficiency (SME)

Let the two rival models be:

M_1: $y_t \;=\; \beta x_t + \varepsilon_{1t}$

M_2: $y_t \;=\; \gamma y_{t-1} + \varepsilon_{2t}$,

with the usual claims that $\varepsilon_{it} \sim \mathsf{IN}(0,\, \sigma_i^2)$. It seems natural to consider as a completing model, the realizable projection model:

M_c^m: $x_t \;=\; \phi y_{t-1} + u_t$ where $u_t \sim \mathsf{IN}(0,\, \tau^2)$.

The appendix algorithm for the minimal completing model also yields M_c^m.

Verifying that M_1^c is SME in this case is tedious if one wants to explicitly calculate the order of the difference $\boldsymbol{D}$ in section 3.3(i)-(vi) above, but fortunately the sufficient condition for SME in proposition (3.5) is applicable. The number of parameters in M_1^c is four: $\boldsymbol{\beta}, \sigma^2, \phi$, τ^2. The data second-order moments used in their estimation are $\frac{1}{T}\boldsymbol{x}'\boldsymbol{x}$, $\frac{1}{T}\boldsymbol{x}'\boldsymbol{y}_{-1}$, $\frac{1}{T}\boldsymbol{y}'\boldsymbol{y}$ and $\frac{1}{T}\boldsymbol{x}'\boldsymbol{y}$ and hence, condition (*i*) is verified. Second, in present notation, the cross products appearing in $\frac{1}{T}\boldsymbol{X}'\boldsymbol{X}$, $\frac{1}{T}\boldsymbol{X}'\boldsymbol{Z}$, and $\frac{1}{T}\boldsymbol{Z}'\boldsymbol{Z}$ are $\frac{1}{T}\boldsymbol{x}'\boldsymbol{x}$, $\frac{1}{T}\boldsymbol{x}'\boldsymbol{y}_{-1}$ and $\frac{1}{T}\boldsymbol{y}'\boldsymbol{y}$ which ensures that (*ii*) is satisfied. M_1^c is then SPE and the equivalences between WET and *F*-statistics given in proposition 3.4 hold.

If we switch the rival models:

$\mathsf{M}_1^\dagger$: $y_t \;=\; \beta y_{t-1} + \varepsilon_{1t}$ where $\varepsilon_{1t} \sim \mathsf{IN}(0,\, \sigma_1^2)$

$\mathsf{M}_2^\dagger$: $y_t \;=\; \gamma x_t + \varepsilon_{2t}$ where $\varepsilon_{2t} \sim \mathsf{IN}(0,\, \sigma_2^2)$,

the completing model remains the projection model M_c^m above. Consider the local alternative:

M_a: $y_t \;=\; \lambda_1 y_{t-1} + \lambda_2 z_t + \zeta_t$ where $\zeta_t \sim \mathsf{IN}(0,\, \sigma_\zeta^2)$,

when:

$$z_t \;=\; \rho x_t + \upsilon_t \quad \text{with} \quad \upsilon_t \sim \mathsf{IN}(0,\, \sigma_\upsilon^2). \tag{4.17}$$

In this example, direct substitution of (4.17) into M_a shows that β and ε_{1t} in $\mathsf{M}_1^\dagger$ are:

$$\beta \;=\; (\lambda_1 + \lambda_2\rho\phi) \quad \text{and} \quad \varepsilon_{1t} \;=\; (\zeta_t + \lambda_2\upsilon_t + \lambda_2\rho u_t), \tag{4.18}$$

whereas e_t in M_n is $(\zeta_t + \lambda_2\upsilon_t)$, so that the asymptotic non-centrality of the *F*-test is:

$$\mu^2(\eta_F) \;=\; (\lambda_2^2\rho^2\sigma_u^2)/(\lambda_2^2\sigma_\upsilon^2+\sigma_\zeta^2). \tag{4.19}$$

Further, the asymptotic encompassing difference is:

$$plim\ \tilde{\Delta}_{\phi} = plim\ (\tilde{\gamma}-\gamma_{\tilde{\alpha}}) = \lambda_2\rho\sigma_u^2/\sigma_x^2, \tag{4.20}$$

with a limiting variance of $\sigma_u^2(\lambda_2^2\sigma_\upsilon^2+\sigma_\zeta^2)/\sigma_x^4$, so that:

$$\mu^2(\eta_C) = (\lambda_2^2\rho^2\sigma_u^2)/(\lambda_2^2\sigma_\upsilon^2+\sigma_\zeta^2), \tag{4.21}$$

and so the two tests are asymptotically equivalent under a local alternative that keeps $M_1^{\dagger}$ congruent.

The use of other completing models may generate WET statistics that satisfy neither strong projection nor moment efficiency conditions, and even worse choices can lose weak efficiency. For example, consider:

M_c: $x_t = u_t$.

Under M_1^c, $\Psi_{XZ} = \Psi_{xy_{-1}} = 0$, and the encompassing difference on γ becomes:

$$\tilde{\gamma} - \gamma_{\tilde{\alpha}} = (\mathbf{y}'_{-1}\mathbf{y}_{-1})^{-1}\mathbf{y}'_{-1}\mathbf{y},$$

from which η_S can be derived. Since M_1^c is neither WME nor WPE, there exists at least one $M_1^* = (M_1, M_c^*) \in \mathcal{M}_1$ such that $D_{\gamma M_1^*} \neq 0$. If the DGP is (M_1, M_c^*) where:

M_c^*: $x_t = \phi x_{t-1} + u_t$,

then $\eta_S \to \infty$ under the DGP when T increases, and hence the WET will, in most cases, reject even though the model M_1 is correct. Such an outcome demonstrates the dangers of using inappropriate completing models and encourages the design of completing models with the aim of ensuring a minimum of robustness for the resulting WET statistics.

Example 4.4: Exact projection efficiency (EPE)

Let:

M_1: $y_t = \beta x_t + \varepsilon_{1t}$

M_2: $y_t = \gamma_1 y_{t-1} + \gamma_2 y_{t-2} + \varepsilon_{2t}$.

The minimal completing model is given by the projection of x_t on y_{t-1} and y_{t-2}:

M_c^m: $x_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + u_t$.

Formulae (4.1) in example 1 can be applied to check that EPE is satisfied here. This does not imply the equality of the WET statistics with the F-statistic because $V_{\alpha}[\sqrt{T}\,\tilde{\Delta}]$ differs from the corresponding matrix obtained in the static case. Nevertheless, from proposition 3.4, the encompassing statistics are $\underset{\mathcal{M}_1}{\simeq}$ to η_F because the choice of the minimal completing model ensures that M_1^c is also WME.

Example 4.5: Strong projection efficiency (SPE)

This example differs from the last in lagging the regressors of the two models by one period:

M_1: $y_t = \beta x_{t-1} + \varepsilon_{1t}$

M_2: $y_t = \gamma_1 y_{t-2} + \gamma_2 y_{t-3} + \varepsilon_{2t}$.

The projection model remains the same:

M_c: $x_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + u_t$,

and is included in the minimal completing model given by:

M_c^m: $x_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_3 y_{t-3} + \phi_4 x_{t-1} + u_t$.

Then (M_1, M_c) is SPE for the same reasons as before. The difference between EPE and SPE is due to lagging; for example, $\frac{1}{T}\boldsymbol{x}'_{-1}\boldsymbol{y}_{-2}$ in $(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{X}$ is estimated by $\frac{1}{T}\boldsymbol{x}'\boldsymbol{y}_{-1}$ in $\tilde{\boldsymbol{\Psi}}_{ZZ}^{-1}\tilde{\boldsymbol{\Psi}}_{ZX}$.

It can also be shown that WME is not achieved for M_c, which implies that η_S and η_F are not equivalent under any model of $\mathcal{M}_1$: for example, consider $(M_1, M_c^m) \in \mathcal{M}_1$. Despite that difficulty, M_c probably remains a good choice for the completing model since M_c^m ensures no more than WME.

Example 4.6: Weak moment efficiency (WME)

Consider:

M_1: $y_t = \beta x_{t-1} + \varepsilon_{1t}$

M_2: $y_t = \gamma y_{t-1} + \varepsilon_{2t}$,

with:

M_c^m: $x_t = \phi x_{t-1} + u_t$.

For these M_1 and M_2 models, no completing model exists such that SME or SPE is satisfied. M_c^m ensures WME. The resulting simplification WET statistic η_S is completely different from the *F*-statistic. For example, the encompassing difference on γ has the form:

$$\tilde{\gamma} - \gamma_{\tilde{\alpha}} = (\boldsymbol{y}'\boldsymbol{y})^{-1}(\boldsymbol{y}'\boldsymbol{Q}_{x_{-1}}\boldsymbol{y}_{-1} + \boldsymbol{y}'\boldsymbol{Q}_{x_{-1}}\boldsymbol{x}\tilde{\boldsymbol{\beta}}) \quad (4.22)$$

which tests if $\text{cov}(y_t, y_{t-1}|x_{t-1}) + \beta\text{cov}(y_t, x_t|x_{t-1}) = 0$. The *F*-statistic tests if $\text{cov}(y_t, y_{t-1}|x_{t-1})$ is zero. The different possible WET statistics are given in Govaerts [1987] jointly with an analysis of their asymptotic power against local alternatives and Monte Carlo simulations to compare their small sample properties. For example, when M_1 relates y_t to y_{t-2} and M_2 relates y_t to y_{t-3}, the encompassing statistic can be more powerful than the F-test.

5. Conclusion

The general analysis in section 3 and the examples in section 4 reveal that encompassing in linear stationary dynamic processes raises new issues. The need to take account of feedbacks from lagged dependent variables forces the explicit introduction of a completing model, and the choice of its formulation is important if the encompassing tests are to be robust to how the completing model is specified. Six levels of efficiency of the completing model were distinguished, and illustrated by examples that highlighted which features induced which consequences.

The resulting analysis reproduces that previously established for strong exogeneity when the models are static. However, in dynamic systems, a poor choice of the completing model can lead to rejection of the correct hypothesis as shown in the example 4.3.

Stationarity is an important assumption in the approach adopted here because of the central role played by (3.7) and (3.8). In integrated systems with cointegrated relationships, Hendry and Mizon [1993] develop asymptotically valid encompassing tests for linear equations or sub-systems against each other, or the VAR when the latter is the DGP. Similar generalizations should hold for the class of equations of interest above. In particular, parsimonious encompassing is the final check on a reduction sequence, by which stage, mapping to $I(0)$ variables will usually have occurred. However, weak exogeneity only holds in cointegrated systems if the error corrections in the equation of interest do not enter other equations of the system, so that enforces a necessary condition for the present analysis to apply. Further, when weak exogeneity is violated due to the presence of common cointegrating vectors, the limiting distributions are no longer linear mixtures of normals and inference can be distorted (see e.g. Phillips and Loretan [1991] and Hendry, 1993). Nevertheless, both the present analysis and cointegration theory emphasize the primary role of the system in sustaining inference even when interest is in individual equations.

Overall, our analysis favours the use of parsimonious encompassing tests or *F*-tests as these are more robust to the specification of the completing model, are not restricted to paired comparisons between models and complement general-to-specific modelling strategies. Further, they are invariant to extensions in the specification of the rival model up to the union of all the non-redundant regressors in both models. Finally, when the alternative hypothesis makes the model under test non-congruent, there is a potentially serious power loss in encompassing tests (including non-nested tests) relative to parsimonious encompassing tests.

Appendix: Minimal Completing Models

Sample moments are denoted by either $\boldsymbol{A}$ (matrices) or $\boldsymbol{a}$ (scalars and vectors) appropriately subscripted. In particular, let:

$$\boldsymbol{A} = \frac{1}{T}\begin{bmatrix} \boldsymbol{X}'\boldsymbol{X} & \boldsymbol{X}'\boldsymbol{Z} \\ \boldsymbol{Z}'\boldsymbol{X} & \boldsymbol{Z}'\boldsymbol{Z} \end{bmatrix} = \begin{bmatrix} A_{XX} & A_{XZ} \\ A_{ZX} & A_{ZZ} \end{bmatrix} \tag{A.1}$$

The ML estimator of $\psi(\alpha)$ is given by $\tilde{\psi} = \psi(\hat{\alpha})$ where $\hat{\alpha} = (\hat{\beta}, \hat{\sigma}^2, \hat{\Gamma}, \hat{\Sigma})$ regroups the OLS estimators of the parameters in M_1 and M_c as given by (3.1) and (3.4) respectively. Hence EME requires that $\boldsymbol{A} = \psi(\hat{\alpha})$, while SME (WME) requires that $\boldsymbol{A} \approx \psi(\hat{\alpha})$ on $\mathcal{M}$ ($\mathcal{M}_1$). The following relationships hold:

$$\begin{aligned} \boldsymbol{A}_{Xy} &= \boldsymbol{A}_{XX}\hat{\beta} & \boldsymbol{A}_{WR} &= \boldsymbol{A}_{WW}\hat{\Gamma} \\ \boldsymbol{A}_{yy} &= \hat{\alpha}^2 + \hat{\beta}'\boldsymbol{A}_{XX}\hat{\beta} & \boldsymbol{A}_{RR} &= \hat{\Sigma} + \hat{\Gamma}'\boldsymbol{A}_{WW}\hat{\Gamma} \end{aligned} \tag{A.2}$$

Let L_B denote a list of all the distinct elements in $(\boldsymbol{A}_{yy}, \boldsymbol{A}_{Xy}, \boldsymbol{A}_{RR}, \boldsymbol{A}_{WR})$. Conditionally on $\boldsymbol{A}_{XX}$ and $\boldsymbol{A}_{WW}$, (A.2) defines a 1-1 mapping between L_B and $\hat{\alpha}$, subject to the usual restrictions for moment matrices (symmetry and positivity). However, in order for $\boldsymbol{A}$ itself to be a function of $\hat{\alpha}$, it has to be the case that $\boldsymbol{A}_{XX}$ and $\boldsymbol{A}_{WW}$, which are included in $\boldsymbol{A}$, can be retrieved from L_B. For ease of discussion, we restrict attention to the case where *all* the restrictions between $\boldsymbol{A}$, $\boldsymbol{A}_{WW}$ and L_B originate from the exclusion of regressors in M_1 and M_c.[3] Let L_A denote a list of all the distinct elements in $\boldsymbol{A}$ and $\boldsymbol{A}_{WW}$, then:

$$\mathsf{L}_A \subseteq \mathsf{L}_B \tag{A.3}$$

is - for all practical purposes - necessary and sufficient for $\boldsymbol{A}$ to be a function of $\hat{\alpha}$. A formal proof of that assertion requires explicitly stating a number of technical conditions and goes beyond the objectives of the current paper. At a more heuristic level, if condition (A.3) holds, then (A.2) can be solved (recursively) for $\boldsymbol{A}$: at step j of the recursion, $\boldsymbol{A}_{XX}$ and $\boldsymbol{A}_{WW}$ are set at the values obtained on step (j-1) and (A.2) is solved for L_B; $\boldsymbol{A}_{XX}$ and $\boldsymbol{A}_{WW}$ are then updated and the procedure is repeated until convergence (which follows from a fixed point theorem). If, on the other hand, $\boldsymbol{A}$ is a function of $\hat{\alpha}$, then (A.3) holds. If, indeed, an element of L_A were not included in L_B, then we could assign an arbitrary value to that element and proceed as just described. That element would never be revised, contradicting

[3] Accounting for more general linear restrictions among the variables in M_1^c raises no conceptual problems but necessitates additional algebraic manipulations of the relevant moment matrices.

the assertion that $\boldsymbol{A}$ is a function of $\hat{\alpha}$ alone.

The algorithm we propose aims to select a (minimal) set of regressors for w_t in such a way that condition (A.3) holds. Two additional issues deserve attention before we can describe the algorithm:

(1) The only difference between EME and SME lies in the treatment of initial and terminal observations in the sample. For example, under EME, moments such as $T^{-1}\boldsymbol{x}'\boldsymbol{y}_{-1}$ and $T^{-1}\boldsymbol{x}'_{-1}\boldsymbol{y}_{-2}$ are treated as different entities while, under SME, they are conflated with each other. Our proposed algorithm trivially accommodates that distinction;

(2) $\boldsymbol{w}_t$ consists of *lagged* variables only, while $\boldsymbol{r}_t$ regroups all *current* exogenous variables. It follows that $\boldsymbol{A}_{RW}$ cannot include cross-moments between *x*s and leading *y*s. If any such moments are included in L_A, then EME (SME) cannot be obtained. Let $\bar{x}_t$ consist of all regressors that are excluded from M_1. The following asymptotic equivalence holds on M_1:

$$\boldsymbol{A}_{\bar{x}y} \underset{\mathcal{M}_1}{\approx} \boldsymbol{A}_{XX}\boldsymbol{A}_{XX}^{-1}\boldsymbol{a}_{xy} \tag{A.4}$$

Both $\boldsymbol{A}_{XX}$ and $\boldsymbol{a}_{xy}$ are already covered by an analysis of condition (A.3). Hence, for the purpose of achieving WME, we can replace components of $\boldsymbol{A}_{\bar{x}y}$ in L_A by the corresponding elements of $\boldsymbol{A}_{XX}$ and proceed.

The proposed algorithm follows from the above discussion. We initially set $w_t = \emptyset$ and accordingly define L_A and L_B. We then examine whether condition (A.3) holds. Each time an element of L_A is found to be missing in L_B, w_t is 'augmented' according to one of the following two (mutually exclusive) schemes:

Type-A augmentation: direct augmentation of w_t such that the missing element is included in the augmented L_B;

Type-B augmentation: the missing element belongs to $\boldsymbol{a}_{\bar{x}y}$; it is replaced by the corresponding elements in $\boldsymbol{A}_{XX}$, and w_t is then augmented in such a way that the latter are included in the augmented L_B.

The algorithm is finite and generates a 'minimal' completing model under which WME holds. Further, if it necessitates only type-A augmentations, then EME (SME) obtains. All the minimal completing models M_c^m, to which we refer in section 4, have been obtained by application of this algorithm.

REFERENCES

Cox, David R., 1961, Tests of Separate Families of Hypotheses, Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Vol. I, 105-123, University of California Press, Berkeley.

Cox, David R., 1962, Further Results on Tests of Separate Families of Hypotheses, Journal of the Royal Statistical Society, B, 24, 406-424.

Engle, Robert F., Hendry, David F. and Richard, Jean-François, 1983, Exogeneity, Econometrica, 51, 277-304.

Florens, Jean-Piérre, Hendry, David F. and Richard, Jean-François, 1988, Parsimonious Encompassing: An Application to Non-Nested Hypotheses and Hausman Specification Tests, ISDS discussion paper, Duke University, N.C., USA.

Florens, Jean-Piérre, Hendry, David F. and Richard, Jean-François, 1991, Encompassing and Specificity, Cahier 91c, GREMAQ, University of Toulouse.

Gourieroux, Christian, Monfort, Alain and Trognon, Alain, 1984, Pseudo-Maximum Likelihood Methods: Theory, Econometrica, 52, 681-700.

Gourieroux, Christian and Monfort, Alain, 1991, Testing Non-nested Hypotheses, mimeo, INSEE.

Gourieroux, Christian and Monfort, Alain, 1992, Testing, Encompassing and Simulating Dynamic Econometric Models, mimeo, INSEE.

Govaerts, Bernadette, 1987, Application of the Encompassing Principle to Linear Dynamic Models, Ph.D. Dissertation, Universite Catholique de Louvain, Belgium.

Govaerts, Bernadette, 1988, A Note on a Method to Compute the Asymptotic Distribution of the Sample Second-Order Moments of Dynamic Linear Normal Variables, Communications in Statistics.

Granger, Clive W.J., 1969, Investigating Causal Relations by Econometric Models and Cross-Spectral Methods, Econometrica, 37, 424-438.

Hannan, Edward G., 1970, Multiple Time Series, John Wiley: New York.

Hendry, David F., 1993, Lectures on Econometric Methodology. Forthcoming, Oxford University Press.

Hendry, David F. and Mizon, Grayham E., 1993, Evaluating Dynamic Econometric Models by Encompassing the VAR, p272-300 in Peter C.B. Phillips (ed.), Models, Methods and Applications of Econometrics. Oxford: Basil Blackwell.

Hendry, David F., Neale, Adrian J. and Ericsson, Neil R., 1991, PC-NAIVE: An Interactive Program for Monte Carlo Experimentation in Econometrics. Oxford: Oxford Institute of Economics and Statistics.

Hendry, David F. and Richard, Jean-François, 1982, On the Formulation of Empirical Models in Dynamic Econometrics, Journal of Econometrics, 20, 3-33.

Hendry, David F. and Richard, Jean-François, 1983, The Econometric Analysis of Economic Time Series, International Statistical Review, 51, 111-163.

Hendry, David F. and Richard, Jean-François, 1990, Recent Developments in the Theory of Encompassing, p393-440 in Bernard Cornet and Henry Tulkens (eds.), Contributions to Operations Research and Econometrics. The XXth Anniversary of CORE, Cambridge, Mass.: M.I.T. Press.

Kent, John T. (1986), "The Underlying Nature of Nonnested Hypothesis Tests", Biometrika, 73, 333-43.

Mann, H.B. and Wald, Abraham, 1943, On the Statistical Treatment of Linear Stochastic Difference Equations, Econometrica, 11, 173-220.

Mizon, Grayham E., 1984, The Encompassing Approach in Econometrics, p135-72 in D.F. Hendry and K.F. Wallis (eds.), Econometrics and Quantitative Economics, Oxford: Basil Blackwell.

Mizon, Grayham E. and Hendry, David F., 1980, An Empirical Application and Monte Carlo Analysis of Tests of Dynamic Specification, Review of Economic Studies, 47, 21-46.

Mizon, Grayham E. and Richard, Jean-François, 1986, The Encompassing Principle and its Application to Non-Nested Hypotheses, Econometrica, 54, 657-78.

Pesaran, Hashem M., 1987, Global and Partial Non-nested Hypotheses and Asymptotic Local Power, Econometric Theory, 3, 69-90.

Phillips, Peter C.B. and Loretan, Mico, 1991, Estimating Long-run Economic Equilibria, Review of Economic Studies, 53, 473-95.

Sawa, Takeshi, 1978, Information Criteria for Discriminating Among Alternative Regression Models, Econometrica, 46, 1273-92.

White, Halbert (ed.), 1983, Non-nested Models, Special Issue of Journal of Econometrics, 21.

White, Halbert, 1984, Asymptotic Theory for Econometricians, New York: Academic Press Inc.

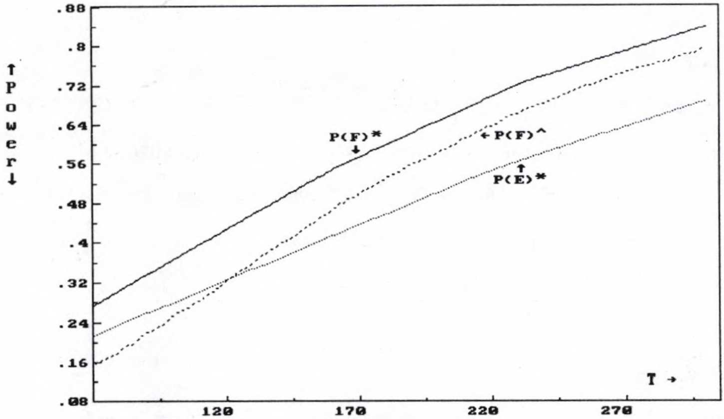

Figure 1

Asymptotic and finite sample power functions of the *F*-test and the encompassing test: P(*F*)* and P(*E*)* are asymptotic, P(*F*)^ is based on recursive Monte Carlo: example 4.2

Appendix A: Asymptotic Power Functions of η_F and η_C.

The rival models are:

M_1: $y_t = \beta y_{t-2} + \varepsilon_{1t}$

M_2: $y_t = \gamma y_{t-1} + \varepsilon_{2t}$.

The completing model is M_2. The η_F and η_C tests are given by equations (4.11) and (4.9) with $\eta_C \leq \eta_F$, but they are asymptotically equivalent on M_1. The local alternative to M_1 is:

(A1) $\quad y_t = T^{-\frac{1}{2}}\nu_1 y_{t-1} + \nu_2 y_{t-2} + \zeta_t = \theta y_{t-1} + \nu_2 y_{t-2} + \zeta_t$ where $\zeta_t \sim \mathsf{IN}(0, \sigma_\zeta^2)$.

The *plims* of the data second moments under (A1) are as follows:

(A2) $\quad \Sigma_{yy} = (1-\nu_2)\sigma_\zeta^2/[(1+\nu_2)\{(1-\nu_2)^2-\theta^2\}];$

(A3) $\quad \Sigma_{yy_{-1}} = \theta\Sigma_{yy}/(1-\nu_2);$

(A4) $\quad \Sigma_{yy_{-2}} = [\nu_2(1-\nu_2)+\theta^2]\Sigma_{yy}/(1-\nu_2).$

Thus, estimation of M_1 leads to population parameter values $\beta_p = plim\ \tilde{\beta}$, and $\sigma_p^2 = plim\ \tilde{\sigma}^2$:

(A5) $\quad \beta_p = \Sigma_{yy}^{-1}\Sigma_{yy_{-2}} = [\nu_2(1-\nu_2)+\theta^2]/(1-\nu_2),$

and:

(A6) $\quad \sigma_p^2 = \sigma_\zeta^2(1+[(1-\nu_2)\theta^2/(1+\nu_2)\{(1-\nu_2)^2-\theta^2\}]).$

On local alternatives like (A1), the $\eta_\bullet$ tests become non-central $\chi^2(1,\mu^2)$ where (see Mizon and Hendry, 1980):

(A7) $\quad \mu^2(\eta_F) = \nu_1^2/(1-\nu_2^2) = \mathrm{T}\theta^2/(1-\nu_2^2);$

(A8) $\quad \mu^2(\eta_C) = \sigma_\zeta^2\nu_1^2\{(1-\nu_2)^2-\theta^2\}/\sigma_p^2(1-\nu_2^2)(1-\nu_2)^2$

$\quad = \mu^2(\eta_F)[1-2\theta^2/(1-\nu_2^2)(1-\nu_2)] + o_p(\mathrm{T}^{-1})$

$\quad \leq \mu^2(\eta_F).$

Two factors contribute to the power loss arising from the uniformly smaller non-centrality parameter: on (A1), $\sigma_p^2 \geq \sigma_\zeta^2$, which η_F avoids; and $\beta_p \neq \beta$, so the residuals are not white noise, whereas ζ_t is. Even though θ is $\mathsf{O}(\mathrm{T}^{-\frac{1}{2}})$, the asymptotic power loss is quite large as figure 1 shows when $\theta = 0.1$, and $\nu_2 = 0.8$ for $\mathrm{T} = 10, ..., 300$.

EUI WORKING PAPERS

EUI Working Papers are published and distributed by the
European University Institute, Florence

Copies can be obtained free of charge
– depending on the availability of stocks – from:

The Publications Officer
European University Institute
Badia Fiesolana
I-50016 San Domenico di Fiesole (FI)
Italy

Please use order form overleaf

Publications of the European University Institute

Economics Department Working Paper Series

To Department of Economics **WP**
European University Institute
Badia Fiesolana
I-50016 San Domenico di Fiesole (FI)
Italy

From Name . . . .

Address. . . . .

. . . .

. . . .

. . . .

. . . .

(Please print)

❒ Please enter/confirm my name on EUI Economics Dept. Mailing List
❒ Please send me a complete list of EUI Working Papers
❒ Please send me a complete list of EUI book publications
❒ Please send me the EUI brochure Academic Year 1994/95

Please send me the following EUI **ECO** Working Paper(s):

No, Author . . . .
Title: . . . .
No, Author . . . .
Title: . . . .
No, Author . . . .
Title: . . . .
No, Author . . . .
Title: . . . .

Date . . . . Signature . . . .

Working Papers of the Department of Economics Published since 1993

ECO No. 93/1
Carlo GRILLENZONI
Forecasting Unstable and Non-Stationary Time Series

ECO No. 93/2
Carlo GRILLENZONI
Multilinear Models for Nonlinear Time Series

ECO No. 93/3
Ronald M. HARSTAD/Louis PHLIPS
Futures Market Contracting When You Don't Know Who the Optimists Are

ECO No. 93/4
Alan KIRMAN/Louis PHLIPS
Empirical Studies of Product Markets

ECO No. 93/5
Grayham E. MIZON
Empirical Analysis of Time Series: Illustrations with Simulated Data

ECO No. 93/6
Tilman EHRBECK
Optimally Combining Individual Forecasts From Panel Data

ECO NO. 93/7
Víctor GÓMEZ/Agustín MARAVALL
Initializing the Kalman Filter with Incompletely Specified Initial Conditions

ECO No. 93/8
Frederic PALOMINO
Informed Speculation: Small Markets Against Large Markets

ECO NO. 93/9
Stephen MARTIN
Beyond Prices Versus Quantities

ECO No. 93/10
José María LABEAGA/Angel LÓPEZ
A Flexible Demand System and VAT Simulations from Spanish Microdata

ECO No. 93/11
Maozu LU/Grayham E. MIZON
The Encompassing Principle and Specification Tests

ECO No. 93/12
Louis PHLIPS/Peter MØLLGAARD
Oil Stocks as a Squeeze Preventing Mechanism: Is Self-Regulation Possible?

ECO No. 93/13
Pieter HASEKAMP
Disinflation Policy and Credibility: The Role of Conventions

ECO No. 93/14
Louis PHLIPS
Price Leadership and Conscious Parallelism: A Survey

ECO No. 93/15
Agustín MARAVALL
Short-Term Analysis of Macroeconomic Time Series

ECO No. 93/16
Philip Hans FRANSES/Niels HALDRUP
The Effects of Additive Outliers on Tests for Unit Roots and Cointegration

ECO No. 93/17
Fabio CANOVA/Jane MARRINAN
Predicting Excess Returns in Financial Markets

ECO No. 93/18
Iñigo HERGUERA
Exchange Rate Fluctuations, Market Structure and the Pass-through Relationship

ECO No. 93/19
Agustín MARAVALL
Use and Misuse of Unobserved Components in Economic Forecasting

ECO No. 93/20
Torben HOLVAD/Jens Leth HOUGAARD
Measuring Technical Input Efficiency for Similar Production Units: A Survey of the Non-Parametric Approach

ECO No. 93/21
Stephen MARTIN/Louis PHLIPS
Product Differentiation, Market Structure and Exchange Rate Passthrough

ECO No 93/22
F. CANOVA/M. FINN/A. R. PAGAN
Evaluating a Real Business Cycle Model

ECO No 93/23
Fabio CANOVA
Statistical Inference in Calibrated Models

ECO No 93/24
Gilles TEYSSIÈRE
Matching Processes in the Labour Market in Marseilles. An Econometric Study

ECO No 93/25
Fabio CANOVA
Sources and Propagation of International Business Cycles: Common Shocks or Transmission?

ECO No. 93/26
Marco BECHT/Carlos RAMÍREZ
Financial Capitalism in Pre-World War I Germany: The Role of the Universal Banks in the Financing of German Mining Companies 1906-1912

ECO No. 93/27
Isabelle MARET
Two Parametric Models of Demand, Structure of Market Demand from Heterogeneity

ECO No. 93/28
Stephen MARTIN
Vertical Product Differentiation, Intra-industry Trade, and Infant Industry Protection

ECO No. 93/29
J. Humberto LOPEZ
Testing for Unit Roots with the k-th Autocorrelation Coefficient

ECO No. 93/30
Paola VALBONESI
Modelling Interactions Between State and Private Sector in a "Previously" Centrally Planned Economy

ECO No. 93/31
Enrique ALBEROLA ILA/J. Humberto LOPEZ/Vicente ORTS RIOS
An Application of the Kalman Filter to the Spanish Experience in a Target Zone (1989-92)

ECO No. 93/32
Fabio CANOVA/Morten O. RAVN
International Consumption Risk Sharing

ECO No. 93/33
Morten Overgaard RAVN
International Business Cycles: How much can Standard Theory Account for?

ECO No. 93/34
Agustín MARAVALL
Unobserved Components in Economic Time Series

ECO No. 93/35
Sheila MARNIE/John MICKLEWRIGHT
"Poverty in Pre-Reform Uzbekistan: What do Official Data Really Reveal?"

ECO No. 93/36
Torben HOLVAD/Jens Leth HOUGAARD
Measuring Technical Input Efficiency for Similar Production Units: 80 Danish Hospitals

ECO No. 93/37
Grayham E. MIZON
A Simple Message for Autocorrelation Correctors: DON'T

ECO No. 93/38
Barbara BOEHNLEIN
The Impact of Product Differentiation on Collusive Equilibria and Multimarket Contact

ECO No. 93/39
H. Peter MØLLGAARD
Bargaining and Efficiency in a Speculative Forward Market

***

ECO No. 94/1
Robert WALDMANN
Cooperatives With Privately Optimal Price Indexed Debt Increase Membership When Demand Increases

ECO No. 94/2
Tilman EHRBECK/Robert WALDMANN
Can Forecasters' Motives Explain Rejection of the Rational Expectations Hypothesis?

ECO No. 94/3
Alessandra PELLONI
Public Policy in a Two Sector Model of Endogenous Growth

ECO No. 94/4
David F. HENDRY
On the Interactions of Unit Roots and Exogeneity

ECO No. 94/5
Bernadette GOVAERTS/David F. HENDRY/Jean-François RICHARD
Encompassing in Stationary Linear Dynamic Models