T E C H N I C A L
R E P O R T

0620

BAYESIAN ENCOMPASSING SPECIFICATION TEST UNDER NOT COMPLETELY KNOW PARTIAL OBSERVABILITY

ALMEIDA C., and M. MOUCHART

I A P S T A T I S T I C S
N E T W O R K

INTERUNIVERSITY ATTRACTION POLE

http://www.stat.ucl.ac.be/IAP

Bayesian Encompassing specification test under not completely known partial observability *

Carlos ALMEIDA † and Michel MOUCHART ‡

Institut de statistique
Université catholique de Louvain
20, Voie du Roman Pays
1348 Louvain-la-Neuve
Belgium

May 30, 2006

Abstract

We present a Bayesian specification test in presence of partial observability and using the encompassing principle. The general theory is developed where the partial observability is known up to a Euclidean parameter, to be estimated from data. This general development is illustrated with an example where only a linear combination of a latent vector is observed; thus, in the example, the partial observability is known up the vector defining the observed linear combination. Some identifiability issues are treated and the example shows the operationality and the pitfall of the proposed test.

Keywords: Bayesian encompassing, partial observability, nonparametric specification test, Dirichlet priors.

AMS 2000 subject classification: 62G09, 62B05, 62F03, 62F15

*Financial support from the IAP research network Nr P5/24 of the Belgian State (Federal Office for Scientific, Technical and Cultural Affairs) is gratefully acknowledged. Comments by .....

†tel. + 32 10 47 91 74, fax: + 32 10 47 30 32, e-mail: almeida@stat.ucl.ac.be

‡tel. + 32 10 47 43 18, fax: + 32 10 47 30 32, e-mail: mouchart@stat.ucl.ac.be (Corresponding author)

1 Introduction

The motivational example for this work has been to develop a test for the normality hypothesis on latent variables supposedly generating, by discretization, manifest ordinal variables. This normality assumption is indeed a standard one when estimating a covariance structure model (or, LISREL type models) on the basis of manifest variables blending continuous and ordinal (*i.e.* ordered discrete) variables; for the use of the normality assumption in that class of models see Muthén (1983, 1984); Jöreskog *et al.* (2002) and for details on discretization model see Almeida and Mouchart (2003a,b).

In covariance structure models, the normality hypothesis of the latent variables has an important role. As matter of fact, this hypothesis permits to reduce the inference process to the analysis of the empirical means, variances and covariances as these empirical moments represent a sufficient reduction. In the discretization model, the normality assumption of the latent variables permits the same reduction at the level of latent variables. But given the unidentifiability of the marginal distributions of the latent variables, only the correlations matrix can be identified. Therefore the inference concentrates on the analysis and the estimation of this matrix; this matrix is called *polychoric correlations matrix*, and can be estimated by maximum likelihood, see in Olsson (1979) or with ad-hoc methods as in Jöreskog (1994).

In general, in models involving some kind of marginalization to the manifest variables, the natural parametrization of the complete model (*i.e.* involving both latent and manifest variables) is not identified by the observable variable. The models considered here correspond to models called *not completely known partial observability models*, where this means that the manifest variable is a deterministic function of the latent variables, known up to a Euclidean parameter only.

As another example of partial observability, one may also consider the observation of a linear combination of latent variables (with unknown coefficients). In Econometrics, this is indeed the case of the permanent income hypothesis, where permanent demand and income are observable with errors, only.

Formally, the models involving partial observability can be described by means of a *structural model*: $\xi \mid \theta \sim P_{\xi|\theta}$, along with an *observability process* defined by $X = g(\xi, \alpha)$ where g is known, $\alpha \in \mathbb{R}^q$ and ξ is a vector of latent variables. The case where g is not a function of α (or α is known), has been treated in Almeida and Mouchart (2005) under the heading of *completely known partial observability*; the case where $\alpha \in \mathbb{R}^q$ is unknown, but the functional form of g is known will be called *not completely known partial observabilitity* and is the object of this paper.

The loss of information due to the partial observability raises substantial problems of statistical inference. In general, the loss of information implies a loss identification and a loss of power; this fact can be illustrated, for example, in censored data, where the precision of estimators and the power of the tests depend on the expected proportion of censored data. As a pathological example, the case of the sign observation, where the loss of information does not permit any inference about the form of the distribution of the latent variable, has been presented in Almeida and Mouchart (2005).

Once the partial observability does not correspond to a sufficient reduction, a loss of identifiability is to be expected. Although the model is identified at the level of latent variable (*i.e.* θ is identified by ξ), the statistical model is typically not identified (*i.e.* (α, θ) is not identified by X) even if the partial observability process is identified (*i.e.* α identified by X); see Mouchart and Oulhaj (2003) for a study on the sufficiency and identification relations under partial observability. Hereafter the structural models at the level of latent variables will be considered as identified. This lost of identifiability due to partial observability requires a particular care for the actual meaning and for a correct interpretation of the hypothesis involved in a testing procedure.

Next section gives general results for testing by Bayesian encompassing the form of a distribution under partial observability; results are also obtained for making the proposed test interpretable from a structural modeling approach. In Section 3 the observation of a linear combination of latent variables is presented as an example and a simulation study is developed. Finally Section 5 presents some concluding remarks and conclusions.

2 General Results

The encompassing principle may viewed as a possible approach for replacing an "old" theory by another "new" theory and can be expressed as follows: The empirical findings explained in the framework of the old theory should also be explained in the framework of the new theory. This idea has been applied for testing non-nested hypotheses starting with the pioneering work of Cox (1961, 1962). Thus, the encompassing test leads to analyze, in the framework of the preferred model, the behavior of statistics of interest within the context of the non-preferred model. For a detailed study of the encompassing testing in classical sense and in parametric framework see Mizon and Richard (1986).

In Bayesian framework, encompassing testing has been sketched in Florens *et al.* (1990, section 3.5) and developed in a general setup and applied to parametric models by Florens and Mouchart (1993), later Florens *et al.* (2003) presents a Bayesian specification encompassing

test using as alternative model a nonparametric Bayesian specification. Here, we extend that idea for a case where only partial observation is available. For the sake of easier reference, a short summary on Bayesian encompassing is sketched in the Appendix A.

2.1 The two Bayesian models

In the general setup, the *Bayesian specification of the structural models* corresponding to the null, $\mathcal{E}^0$, and alternative hypotheses, $\mathcal{E}^1$, are given by:

$$\mathcal{E}^0 : \quad (\theta, \xi) \sim Q^0_{\theta,\xi} = M^0_\theta \otimes P^0_{\xi|\theta} \tag{1}$$

$$\mathcal{E}^1 : \quad (\psi, \xi) \sim Q^1_{\psi,\xi} = M^1_\psi \otimes P^1_{\xi|\psi} \tag{2}$$

where θ and ψ are parameters characterizing the respective models of latent variable ξ. We use $P^1_{\xi|\psi} = \psi$ for specifying a general non-parametric alternative.

Some words about notations may be in order. On the one hand we handle probability measures and transition probabilities (implicitly assuming the existence of regular versions of conditional probabilities) rather than densities for we deal with undominated families of distributions (because of some degeneracies and because of nonparametric alternative model). On the other hand, probability measures and transitions are denoted by capital letters with upper and lower indices. Upper indices mark different models or different measures (on a same space) whereas lower indices denote random variables (often under identification with the σ-field generated by these variables); when lower indices are not present, we refer to an implicitly defined complete joint distribution. Often we combine a probability measure and a transition of probability by a Markovian product denoted $\otimes$; more explicitly, when $Q^0_{\theta,\xi}$ is defined by $M^0_\theta \otimes P^0_{\xi|\theta}$, we mean that for any measurable set A on the θ-space and B on the ξ-space the probability of the rectangle $A \times B$ is defined as:

$$Q^0_{\theta,\xi}(A \times B) = \int_{\theta \in A} P^0_{\xi|\theta}(B)\, dM^0_\theta, \tag{3}$$

and the probability measure $Q^0_{\theta,\xi}$ is obtained as the unique extension of (3) to the σ-field generated by the rectangles, based on the θ-space and on the ξ-space.

For both models, a same *partial observability process* is defined by:

$$X = g(\xi, \alpha) = g_\alpha(\xi) \tag{4}$$

where the function g is known and α an unknown Euclidean parameter α. Clearly:

$$P^0_{X|\alpha,\theta,\xi} = P^1_{X|\alpha,\psi,\xi} = \delta_{\{X=g(\xi,\alpha)\}} \tag{5}$$

with $\delta_{\{\bullet\}}$ is the *unit mass measure* (or, Dirac measure). In general, (4) implies that for any γ and any probability R on (X, γ, α, ξ) we have:

$$X \perp\!\!\!\perp \gamma \mid \alpha, \xi; R. \tag{6}$$

We repeatedly use, in particular:

$$X \perp\!\!\!\perp \theta \mid \alpha, \xi; Q^0, \text{ and} \tag{7}$$

$$X \perp\!\!\!\perp \psi \mid \alpha, \xi; Q^1 \tag{8}$$

When completing these structural models in order to incorporate α and X, we shall also assume a separation between the structural and the observability model in the sense that the sampling distributions of ξ should not depend on α; more precisely, we always assume:

$$\alpha \perp\!\!\!\perp \xi \mid \theta; Q^0 \tag{9}$$

$$\alpha \perp\!\!\!\perp \xi \mid \psi; Q^1. \tag{10}$$

Thus, the *complete null model* $\mathcal{E}^0$ can be written as:

$$\mathcal{E}^0 : \quad (\theta, \alpha, \xi, X) \sim Q^0_{\theta,\alpha,\xi,X} = M^0_{\theta,\alpha} \otimes P^0_{\xi|\theta} \otimes P^0_{X|\xi,\alpha} \tag{11}$$

and the *complete alternative model* $\mathcal{E}^1$ can be written as:

$$\mathcal{E}^1 : \quad (\psi, \alpha, \xi, X) \sim Q^1_{\psi,\alpha,\xi,X} = M^1_{\psi,\alpha} \otimes P^1_{\xi|\psi} \otimes P^1_{X|\xi,\alpha} \tag{12}$$

For the Bayesian encompassing test , the null model needs to be extended by incorporating ψ, the parameter of the alternative one. The construction of this *extended model* $\mathcal{E}^*$ requires an *extended Bayesian Pseudo-True Value (BPTV) condition*, namely:

$$\psi \perp\!\!\!\perp \alpha, \xi \mid \theta; Q^{0,*}. \tag{13}$$

Condition (13) actually aggregates two conditions, namely:

$$\psi \perp\!\!\!\perp \alpha \mid \xi, \theta; Q^{0,*} \tag{14}$$

$$\psi \perp\!\!\!\perp \xi \mid \theta; Q^{0,*} \tag{15}$$

The first one gives a neutrality of the partial observability for interpreting ψ in $Q^{0,*}$ and the second one is the standard BPTV hypothesis within the structural model (see Florens and Mouchart, 1993, for details on BPTV condition). Note that (9) and (14) are jointly equivalent to

$$\alpha \perp\!\!\!\perp (\xi, \psi) \mid \theta; Q^{0,*}. \tag{16}$$

Under (9) and (13), the probability measure $Q^{0,*}$ defining the *extended complete model* can be written in two equivalent forms, namely:

$$Q^{0,*}_{\theta,\alpha,\psi,\xi,X} = M^0_{\theta,\alpha} \otimes P^0_{\xi|\theta} \otimes M_{\psi|\theta} \otimes P^0_{X|\xi,\alpha} \tag{17}$$

$$= M^0_{\theta} \otimes P^0_{\xi|\theta} \otimes M_{\psi|\theta} \otimes M^0_{\alpha|\theta} \otimes P^0_{X|\xi,\alpha} \tag{18}$$

where $M_{\psi|\theta}$ is a BPTV. The *extended statistical model*, obtained after integrating the latent variable ξ, can be written under (13) as:

$$Q^{0,*}_{\theta,\alpha,\psi,X} = M^0_{\theta,\alpha} \otimes P^0_{X|\theta,\alpha} \otimes M_{\psi|\theta}. \tag{19}$$

As long as the structural process acts independently of the partial observability process, one might assume that α and θ are a priori independent, more specifically along with

$$\alpha \perp\!\!\!\perp \theta; Q^0 \tag{20}$$

$$\alpha \perp\!\!\!\perp \psi; Q^1 \tag{21}$$

(7) and (9) provides in the null and in the alternative models Bayesian cuts between ξ and X (in the distribution Q^0 of (X, ξ, α, θ) and in the distribution Q^1 of (X, ξ, α, ψ) respectively). Next theorem shows that (20) also implies, in the extended complete model, a same Bayesian cut between ξ and X (in the distribution $Q^{0,*}$ of (X, ξ, α, θ)). On Bayesian cuts, see *e.g.* Florens *et al.* (1990, section 3.4)

Theorem 2.1. *Using the extended BPTV condition (13), we have under (20), that:*

$$\alpha \perp\!\!\!\perp (\xi, \psi, \theta); Q^{0,*} \tag{22}$$

Proof. Clearly (22) is equivalent to the properties (a) $\alpha \perp\!\!\!\perp \xi, \theta$ and (b) $\alpha \perp\!\!\!\perp \psi \mid \xi, \theta$. Property (a) is equivalent to (20) and $\alpha \perp\!\!\!\perp \xi \mid \theta$, implied by (16), and property (b) is also implied by (16) ■

Two "natural" approaches might be considered for elaborating the encompassing test statistics. A first possibility starts by developing the encompassing test at the level of the latent variables, *i.e.* a distance (or discrepancy) d^* between posteriors at the level of latent variables, and a test statistics might be the expectation of this distance conditionally on available data:

$$d_1(X) = E\,[\,d^*(M^{0,*}_{\psi|\xi}, M^1_{\psi|\xi}) \mid X\,], \tag{23}$$

but this statistic is usually not operational. A second approach consists in evaluating

$$d_2(X) = d^*(M^{0,*}_{\psi|X}, M^1_{\psi|X}). \tag{24}$$

In both cases, these test statistics must be calibrated against P^0_X, the predictive distribution under the model $\mathcal{E}^0$.

Moreover, Florens *et al.* (2003) shows that a more operational version of the encompassing test statistic may be obtained by replacing $d^*(M^{0,*}_{\psi|X}, M^1_{\psi|X})$ by $d^*(M^{0,*}_{\lambda|X}, M^1_{\lambda|X})$ where $\lambda = h(\psi)$, a finite dimensional functional defined on ψ.

2.2 Identification and Encompassing testing

Once the partial observability is not completely known (*i.e.* α unknown), it should be noted that the partial observation is not likely to identify the complete parameters α and θ (resp.ψ) in the null (resp. alternative) model, even though we have assumed that θ (resp. ψ) is identified by ξ. This is so unless X were a sufficient statistic in both models, see Mouchart and Oulhaj (2003), a situation out of the interest of this paper.

Let us define γ_X (resp. ω_X) as a minimal sufficient parameter for the sampling distribution of X in the null (resp. alternative) model. Thus in the null model $\gamma_X \subset \theta \vee \alpha$, and in the alternative model the parameter identified by the statistical model $P^1_{\xi|\theta} \circ g_\alpha^{-1}$ is $\omega_X = \psi \circ g_\alpha^{-1} \subset \psi \vee \alpha$. These identified parameters are such that:

$$(a)\ X \perp\!\!\!\perp \theta, \alpha \mid \gamma_X; Q^0 \qquad (b)\ X \perp\!\!\!\perp \psi, \alpha \mid \omega_X; Q^1. \tag{25}$$

In the notation used here, random elements have been identified with the respective σ-field generated by itself, "$\vee$" denotes the minimal σ-field generated by the union of the operated σ-

fields and "$\subset$" is the usual inclusion relation. Thus $\gamma_X \subset \theta \vee \alpha$ can be heuristically interpreted as saying that there is a measurable function h such that $\gamma_X = h(\theta, \alpha)$ we shall also say that γ_X is a "subparameter" of (θ, α); similarly for ω_X as measurable function of ψ and α.

The second approach, that leads to (24), naturally raises the question whether $d_2(X)$ should be defined from the distance between the posterior distributions of the complete parameter ψ or of the identified parameter ω_X only. Let us therefore ask how far it is to legitimate to concentrate the encompassing test on the identified part of ψ, *i.e.* to evaluate

$$d_3(X) = d^*(M^{0,*}_{\omega_X|X}, M^1_{\omega_X|X}) \tag{26}$$

instead of $d_2(X)$. Intuitively, that would be legitimate if in the extended null model and in the alternative model the distributions conditional on the data and the identified parameters would no depend on the data. This is clearly the case for the alternative model in view of condition (25b). Next theorem and its corollary give conditions under which a similar property holds in the extended null model.

Theorem 2.2. *In $\mathcal{E}^{0,*}$, under (13),*

$$X \perp\!\!\!\perp \psi, \theta, \alpha \mid \gamma_X; Q^{0,*} \tag{27}$$

Proof. Indeed, in $Q^{0,*}$, we have that (13) implies $\psi \perp\!\!\!\perp X \mid \theta, \alpha$, which jointly with (25a) implies (27). ■

Corollary 2.3. *Under (13) the BPTV condition is fulfilled in the extended statistical model at the level of identified parameters, namely:*

$$X \perp\!\!\!\perp \omega_X \mid \gamma_X; Q^{0,*}. \tag{28}$$

Therefore, (13) permits us to write:

$$Q^{0,*}_{\theta,\alpha,\psi,X} = \left[M^0_{\gamma_X} \otimes M^{0,*}_{\omega_X|\gamma_X} \otimes P^0_{X|\gamma_X}\right] \otimes M^{0,*}_{\alpha,\theta,\psi|\gamma_X}. \tag{29}$$

Finally, next theorem requires a further condition on the BPTV in order to ensure that ω_X is sufficient, in the extended model, relatively to ψ.

Theorem 2.4. *Under the extended BPTV condition (13), if*

$$\psi \perp\!\!\!\perp \gamma_X \mid \omega_X; Q^{0,*} \tag{30}$$

then

$$\psi \perp\!\!\!\perp X \mid \omega_X; Q^{0,*} \tag{31}$$

Proof. Clearly (31) is implied by $\psi \perp\!\!\!\perp X, \gamma_X \mid \omega_X$ which is equivalent to (30) along with $\psi \perp\!\!\!\perp X \mid \gamma_X, \omega_X$ implied by (27). ■

Note that condition (30) means that the extension of the null model, into $\mathcal{E}^{0,*}$, is operated in such a way that the distribution of ψ, conditional on ω_X, the parameter identified by X in $\mathcal{E}^1$, should not depend on the parameter identified by X in the null model, *i.e.* γ_X. Condition (30) has a crucial role in the proof of Theorem 2.4, but is not likely to be testable. Therefore its acceptability is essentially a matter of contextual plausibility. Thus in any particular application, this assumption should be carefully scrutinised.

Theorem 2.4 permits us to write:

$$M^{0,*}_{\psi|X} = \int M^{0,*}_{\psi|\omega_X} dM^{0,*}_{\omega_X|X} \tag{32}$$

If, by specification we impose the coherence condition:

$$M^{0;*}_{\psi|\omega_X} = M^{1}_{\psi|\omega_X}, \tag{33}$$

a convenient distance for the posterior distributions of ψ can be defined through a distance between posterior distributions of ω_X, as in $d_3(X)$.

Summarising, under the condition (13), (30) and (33), d_3 is an adequate and more operational substitute of d_2, provided that the discrepancy function in use is such that when two probabilities share a common regular conditional probability, the discrepancy depends on a discrepancy between the two marginals only. From (29), conditions (30) and (33) are restrictions on the unspecified $M^{0,*}_{\alpha,\theta,\psi|\gamma_X}$.

In some cases, such as the discretization model, the parameter α may be retrieved from the knowledge of ω_X and ψ. These situations permit us to strengthen the Theorem 2.4 as follows:

Corollary 2.5. *Under the conditions of Theorem 2.4, if furthermore:*

$$\alpha \perp\!\!\!\perp X \mid \omega_X, \psi; Q^{0,*}, \tag{34}$$

then ω_X is sufficient relatively to ψ and α, i. e.

$$(\alpha, \psi) \perp\!\!\!\perp X \mid \omega_X; Q^{0,*}. \tag{35}$$

Note that the condition (34) is trivially implied by $\alpha \subset \omega_X \vee \psi$. This corollary ensures in $\mathcal{E}^{0,*}$ a complete sufficiency of ω_X (w.r.t. α and ψ) rather than a partial sufficiency of ω_X (w.r.t. ψ only).

3 An Example

Let us illustrate the computations implied by the test developed so far by examining a simple example. Thus, we first sketch a parametric (normal) null model with a nonparametric alternative model along with a partial observability equation. Next we sketch the computations required to obtain the Bayesian encompassing test statistic, and its calibration. More specifically we sketch the steps necessaries for evaluating the posterior distributions of ω_X, the parameter identified by the data in the alternative model, namely: $M^{0,*}_{\omega_X|X}$ and $M^{1}_{\omega_X|X}$, entering the test statistic $d_3(X)$ in (26), and for simulating the distribution of this statistic under the null model in order to calibrate it. In both models, the simulation of the posterior distributions of ω_X requires, as a preliminary step the simulation of the posterior distributions of α. Under a continuous specification of the prior distribution of α, the posterior distributions of $(\alpha \mid X_1^n)$ can be simulated by acceptance rejection methods.

Finally, we control the operationality of the proposed procedure through a simulation exercise.

The Null and the Alternative Models Consider a linear model with errors of measurement:

$$X_i = \eta_i + \sigma_x \varepsilon_{x_i} \qquad X_i \in \mathbb{R} \tag{36}$$

$$Z_i = \zeta_i + \Sigma_{zz}^{\frac{1}{2}} \varepsilon_{z_i} \qquad Z_i \in \mathbb{R}^k \tag{37}$$

$$\eta_i = \beta' \zeta_i \tag{38}$$

$$\zeta_i \sim ind\ N_{(k)}(\mu_\zeta, \Sigma_{\zeta\zeta}) \tag{39}$$

where Var$(\varepsilon_{x_i}) = 1$ and Var$(\varepsilon_{z_i}) = I_k$. Let us denote $W_i = (X_i, Z_i')'$ and $\xi_i^* = (\zeta_i', \varepsilon_{x_i}, \varepsilon_{z_i}')'$. Defining a $(k+1) \times (2k+1)$-matrix A, we obtain a partial observability model:

$$W_i = A\xi_i^* \quad \text{where} \quad A = \begin{pmatrix} \beta' & 1 & 0' \\ I_k & 0 & I_k \end{pmatrix} \tag{40}$$

The test developed so far for testing the normality of the $(2k+1)$-dimensional vector ξ_i^* requires considerable computations. Let us illustrate the main difficulties to be faced in the

simplest case where $k = 1$ and only X_i is actually observable. Thus, let us consider a bivariate latent vector ξ_i a linear combination of which is observable, namely:

$$X_i = g_\alpha(\xi_i) = \alpha' \xi_i. \tag{41}$$

where $\xi_i = (\zeta_i', \varepsilon_{x_i})'$ and $\alpha = (\beta, \sigma_x)'$.

We want to compare a completely normally distributed (with known variances) null model with a non parametric alternative model both satisfying (9)-(10) and (20)-(21), namely:

$$\alpha \perp\!\!\!\perp (\xi, \theta); Q^0 \tag{42}$$

$$\alpha \perp\!\!\!\perp (\xi, \psi); Q^1 \tag{43}$$

More explicitly:

$$\mathcal{E}^0 : \quad \begin{cases} \xi \mid \theta, \alpha \ \sim \ N_{(2)}(\theta, A_0) \\ \theta \mid \alpha \ \sim \ N_{(2)}(0, B_0) \\ \alpha \ \sim \ M_\alpha^0 \end{cases} \qquad \mathcal{E}^1 : \quad \begin{cases} \xi \mid \psi, \alpha \ \sim \ \psi \\ \psi \mid \alpha \ \sim \ \mathcal{D}i(n_0^1 N_{(2)}(0, C_0)) \\ \alpha \ \sim \ M_\alpha^1. \end{cases} \tag{44}$$

The corresponding statistical models, reduced by integration of the latent variables, are:

$$\mathcal{E}^0 : \quad \begin{cases} X \mid \theta, \alpha \ \sim \ N(\alpha'\theta, \alpha' A_0 \alpha) \\ \theta \mid \alpha \ \sim \ N_{(2)}(0, B_0) \\ \alpha \ \sim \ M_\alpha^0 \end{cases} \qquad \mathcal{E}^1 : \quad \begin{cases} X \mid \psi, \alpha \ \sim \ \omega_X \\ \omega_X \mid \alpha \ \sim \ \mathcal{D}i(n_0^1 N(0, \alpha' C_0 \alpha)) \\ \alpha \ \sim \ M_\alpha^1. \end{cases} \tag{45}$$

where, as in Section 2.2, $\omega_X = \psi \circ g_\alpha^{-1}$ is the parameter of $\mathcal{E}^1$ identified by X.

Posterior distribution of α under the null model We are using an algorithm where the simulation of the posterior distribution of the parameter of interest is based on the simulation of a posterior predictive distribution requiring the posterior distribution of $(\alpha \mid X_1^n)$. This one may be obtained as follows. The null model $\mathcal{E}^0$ conditionally to α is:

$$\begin{pmatrix} X \\ \theta \end{pmatrix} \mid \alpha; \mathcal{E}^0 \ \sim \ N_{(3)} \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \alpha'(A_0 + B_0)\alpha & \alpha' B_0 \\ B_0 \alpha & B_0 \end{pmatrix} \right]. \tag{46}$$

For an n-size sample, the joint distribution of $X_1^n = (X_1, \ldots, X_n)$ and θ conditionally to α

is:

$$\begin{pmatrix} X_1^n \\ \theta \end{pmatrix} \mid \alpha; \mathcal{E}^0 \ \sim \ N_{(n+2)} \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_n & \Sigma_{X\theta} \\ \Sigma_{\theta X} & B_0 \end{pmatrix} \right] \tag{47}$$

with:

$$\Sigma_n = \alpha' A_0 \alpha \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix} + \alpha' B_0 \alpha \begin{pmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 1 \end{pmatrix} \tag{48}$$

$$\Sigma_{\theta X} = \Sigma'_{X\theta} = (B_0\alpha, \dots, B_0\alpha)\,. \tag{49}$$

Using a continuous prior distribution, $m^0(\alpha)$, the posterior will be also continuous, moreover:

$$m^0(\alpha \mid X_1^n) \propto m^0(\alpha)\, p(X_1^n \mid \alpha), \tag{50}$$

where $p(X_1^n \mid \alpha)$ is the density function of a normal distribution with the mean and the variance given above.

Posterior distribution of α under the alternative model In the alternative model $\mathcal{E}^1$, Dirichlet process properties imply that

$$X \mid \alpha; \mathcal{E}^1 \ \sim \ N(0, \alpha' C_0 \alpha) \tag{51}$$

Conditionally on α, the posterior distribution of the identified parameter ω_X is a Dirichlet process:

$$\omega_X \mid \alpha, X_1^n; \mathcal{E}^1 \ \sim \ \mathcal{D}i(n_0^1 N(0, \alpha' C_0 \alpha) + n\hat{F}_n) \tag{52}$$

where $\hat{F}_n$ denotes the empirical distribution of the sample X_1^n. After integrating α out, the posterior distribution of the identified parameter ω_X is accordingly a mixture of Dirichlet processes, more specifically:

$$M^1_{\omega_X|X_1^n} = \int M^1_{\omega_X|\alpha, X_1^n}\, dM^1_{\alpha|X_1^n}. \tag{53}$$

In order to obtain the distribution $M^1_{\alpha|X_1^n}$, we first make use of the following property of the Dirichlet processes:

$$P^1_{X_1^n|\alpha} \ = \bigotimes_{\{1 \le i \le n\}} P^1_{X_i|X_1^{i-1},\alpha} = \bigotimes_{\{1 \le i \le n\}} \frac{n_0^1 N(0, \alpha' C_0 \alpha) + (i-1)\hat{F}_{i-1}}{n_0^1 + (i-1)}, \tag{54}$$

where $\hat{F}_j$ is the empirical distribution using the first j observations (X_1^j). A Radon-Nikodym derivative of $P^1_{X_1^n|\alpha}$ is easily obtained relative to a measure $L_*^{(n)}$ to be defined. Let firstly μ_n be the measure defined as follows:

$$\mu_n(A) = \text{card}(A \cap \{X_1, \ldots, X_n\}) \text{ for any Borelian } A \subset \mathbb{R}, \tag{55}$$

let also L be the Lebesgue measure on the real numbers and denote the following Radon-Nikodym derivatives as follows:

$$f_0^{1,\alpha} = \frac{dN(0, \alpha C_0 \alpha)}{dL} \tag{56}$$

$$\hat{f}_i = \frac{d\hat{F}_i}{d\mu_n}. \tag{57}$$

As the measures L and μ_n are mutually singular, we may write:

$$\frac{dP^1_{X_i|X_1^{i-1},\alpha}}{d(L+\mu_n)}(x) = \frac{n_0^1 f_0^{1,\alpha}(x) + (i-1)\hat{f}_{i-1}(x)}{n_0^1 + (i-1)} \tag{58}$$

$$\frac{dP^1_{X_1^n|\alpha}}{dL_*^{(n)}}(x_1^n) = \prod_{\{1 \le i \le n\}} \frac{n_0^1 f_0^{1,\alpha}(x_i) + (i-1)\hat{f}_{i-1}(x_i)}{n_0^1 + (i-1)} \tag{59}$$

where $L_*^{(n)}$ denotes the n-times product measure of $(L + \mu_n)$. Therefore, if there exists a prior density $m^1(\alpha)$, the posterior distribution of α given X_1^n can be simulated using:

$$m^1(\alpha \mid X_1^n) \;\propto\; m^1(\alpha) \frac{dP^1_{X_1^n|\alpha}}{dL_*^{(n)}}(X_1^n). \tag{60}$$

Note that in a sample of size 1, the specifications of $C_0 = A_0 + B_0$ and of the same prior for α in both models $(M_\alpha^0 = M_\alpha^1)$ imply the same posterior distribution of α $(M^0_{\alpha|X} = M^1_{\alpha|X})$.

Encompassing Let us now build an encompassing test under assumptions (30) and (33) with a BPTV with the following structure:

$$M_{\omega_X|\gamma_X} = \int M^1_{\omega_X|\tilde{X}_1^n} dP^0_{\tilde{X}_1^n|\gamma_X} \;=\; E^0[\, M^1_{\omega_X|\tilde{X}_1^n} \mid \gamma_X \,] \tag{61}$$

where $\tilde{X}_1^n$ is a virtual sample generated from $P^0_{X_1^n|\gamma_X}$ - for a comment on this specification see Florens and Mouchart (1993). From Corollary 2.3, this specification ensures that

$$\omega_X \perp\!\!\!\perp X_1^n \mid \gamma_X; Q^{0,*}. \tag{62}$$

Let us now stipulate that the virtual sample $\tilde{X}_1^n$ is generated independently of the actual sample X_1^n:

$$\tilde{X}_1^n \perp\!\!\!\perp X_1^n \mid \gamma_X; Q^{0,*}, \tag{63}$$

or, equivalently:

$$P^0_{\tilde{X}_1^n|X_1^n} = \int P^0_{\tilde{X}_1^n|\gamma_X} \, dM^0_{\gamma_X|X_1^n} \tag{64}$$

Then we obtain the posterior distribution of ω_X, as follows:

$$\begin{aligned} M^{0,*}_{\omega_X|X_1^n} &= E^0[\, M_{\omega_X|\gamma_X} \mid X_1^n \,] \qquad \text{by BPTV property (62)} \\ &= E^0[\, E^0[\, M^1_{\omega_X|\tilde{X}_1^n} \mid \gamma_X \,] \mid X_1^n \,] \qquad \text{by (61)} \\ &= E^0[\, M^1_{\omega_X|\tilde{X}_1^n} \mid X_1^n \,] \qquad \text{by (63)} \end{aligned} \tag{65}$$

Thus, by (53) and (65), both in the null and in the alternative models the posterior distribution of the identified parameter ω_X are mixtures of Dirichlet processes. The algorithm for Bayesian encompassing testing used in Florens *et al.* (2003), with a simple adaptation, can be applied in this case; this algorithm, based in the direct simulation of trajectories of Dirichlet process as developed in Rolin (1992) and Sethuraman (1994), can be summarized as follows:

Computation and calibration of the test statistic Consider $\lambda = h(\omega_X) \in \mathbb{R}^\ell$, a finite dimensional sub-parameter of ω_X. As in Florens *et al.* (2003), λ may correspond to the first two moments only. The simulation of the posterior distributions $M^1_{\lambda|X_1^n}$ and $M^{0,*}_{\lambda|X_1^n}$ may be obtained through the following steps.

Step 1 *Simulation of* $\lambda_1^1, \ldots, \lambda_{N_1}^1$ *given* X_1^n *in* $\mathcal{E}^1$, where N_1 is the simulation size.

For each $i = 1, \ldots, N_1$, let α_i be a simulated sample from $M^1_{\alpha|X_1^n}$ from (60) and $(\omega_X)_i$ a simulated trajectory of $M^1_{\omega_X|X_1^n,\alpha_i}$ from (52), then $(\omega_X)_i$ is a simulated trajectory of $M^1_{\omega_X|X_1^n}$.

Compute $\lambda_i^1 = h[(\omega_X)_i]$ for $i = 1, \ldots, N_1$

Step 2 *Simulation of* $\lambda_1^0, \ldots, \lambda_{N_0}^0$ *given* X_1^n *in* $\mathcal{E}^{0,*}$ where N_0 is the simulation size.

For each $i = 1, \ldots, N_0$, simulate $(\tilde{X}_1^n)_i$ from $P^0_{\tilde{X}_1^n|X_1^n}$ through (64). We simulate again $(\omega_X)_i$ from $M^1_{\omega_X|\tilde{X}_1^n}$, using the same device as in step 1. From (65), this sequence is distributed as $M^{0,*}_{\omega_X|X_1^n}$.

Compute $\lambda_i^0 = h[(\omega_X)_i]$ for $i = 1, \ldots, N_0$

Step 3 *Computation of the test statistics* $d(X_1^n)$

From the iid samples $\lambda_1^1, \ldots, \lambda_{N_1}^1$ from step 1 and $\lambda_1^0, \ldots, \lambda_{N_0}^0$ from step 2, estimate $d(X_1^n)$ by means of a discrepancy between the two posterior distributions of $\lambda \mid X_1^n$; the algorithm developed in Wang *et al.* (2005) is used for this estimation.

Step 4 *Simulation of* $(\tilde{X}_1^n)_\ell$, $\ell = 1, \ldots, N_C$ *for calibration*

Simulate $(\alpha, \theta)_\ell$ $\ell = 1, \ldots, N_C$ from (45), for each $(\alpha, \theta)_\ell$, simulate $(\tilde{X}_1^n)_\ell \mid (\alpha, \theta)_\ell$ from (45). Thus $(\tilde{X}_1^n)_\ell \sim P_{X_1^n}^0$.

For each $(\tilde{X}_1^n)_\ell$, compute the test statistic $d[(\tilde{X}_1^n)_\ell]$ repeating the steps 1 to 3. Finally estimate the $p - value$ as follows:

$$\widehat{p - value} = \frac{1}{N_C} \sum_{1 \leq \ell \leq N_C} \mathbb{I}_{\{d((\tilde{X}_1^n)_\ell) \, > \, d(X_1^n)\}} \tag{66}$$

Remark. The simulation of the distribution of $\alpha \mid X_1^n$ can be simplified once the predictive probability that all observations are different is equal to one, namely:

$$P_{X_1^n}^0(\forall\, i, j,\ X_i \neq X_j) = 1. \tag{67}$$

This is indeed the case once, in the null model, the sampling distribution is a continuous one. Therefore the density of $X_1^n \mid \alpha; \mathcal{E}^1$, as given in (59), may be simplified, with probability one, into:

$$\frac{dP_{X_1^n|\alpha}^1}{dL_*^{(n)}}(X_1^n) \ \propto \prod_{\{1 \leq i \leq n\}} f_0^{1,\alpha}(X_i). \tag{68}$$

which corresponds to an iid sample of $f_0^{1,\alpha}$. In other words, when the sample space is explored in order to find the predictive distribution of the test statistic under the null model, we are going to fall, with probability one, in the region where the sample can be considered as a iid sample related to the alternative model, conditionally on α.

Numerical illustration For the sake of simplicity, we specify in this exercise a same prior distribution for α*i.e.* $M_\alpha^0 = M_\alpha^1$. A simple specification of that prior distribution may be obtained through an opportune reparametrization. Consider now the null model obtained after integrating θ *i.e.* the probability measure $Q_{\alpha, X_1^n}^0$. Because of the integration of θ, the $(X_i \mid \alpha)$'s are not mutually independents, but for a sample of size one, characterized by $Q_{\alpha, X}^0$, the minimal

sufficient parameter may be easily described, namely $\alpha_X = \alpha'(A_0 + B_0)\alpha$ (*i.e.* $\alpha_X \subset \alpha$ is identified by X in $Q^0_{\alpha,X}$). Let us reparametrize α into (α_X, τ) as follows:

$$\alpha = \sqrt{\alpha_X} R_0^{-1} \begin{pmatrix} \cos\tau \\ \sin\tau \end{pmatrix}; \qquad \tau \in [\tau_0, \tau_1) \qquad \alpha_X \geq 0 \tag{69}$$

where R_0 is a 2×2-matrix is a such that: $(A_0 + B_0) = R_0' R_0$ and $[\tau_0, \tau_1)$ is specified in such way that $\alpha_2 \geq 0$ with prior probability one. For the prior distribution let us consider:

$$\begin{aligned} &\alpha_X \perp\!\!\!\perp \beta \\ &\alpha_X \sim Gamma(a, b) \\ &\beta \sim \mathcal{U}_{[\tau_0,\tau_1)}. \end{aligned} \tag{70}$$

Let us specify in the null model (44) the variances covariances matrices as follows:

$$A_0 = B_0 = \begin{pmatrix} 1 & .5 \\ .5 & 1 \end{pmatrix}, \tag{71}$$

and the parameters for the prior distribution of α, in (70) are $a = b = 1$.

The purpose of this exercise is to evaluate how far a sample generated from a distribution in the alternative sampling model would be likely associated to a value of the test statistic relatively far in the tail of the null predictive distribution. Consider accordingly the following specification:

$$\begin{aligned} &\phi = (\phi_1, \phi_2)'; \qquad \phi_1 \perp\!\!\!\perp \phi_2; \qquad \phi_i \sim \chi^2_2 \\ &\tilde{\xi} = (\tilde{\xi}_1, \tilde{\xi}_2); \qquad \tilde{\xi}_1 \perp\!\!\!\perp \tilde{\xi}_2 \mid \phi; \qquad \tilde{\xi}_i \sim \text{Expo}\left(\frac{1}{\phi_i}\right) \\ &\xi = D_0 \left(\tilde{\xi} - \begin{pmatrix} 2 \\ 2 \end{pmatrix}\right); \qquad D_0 \text{ is a } 2 \times 2 \text{ matrix,} \end{aligned}$$

where we choose D_0 in such a way that 12 $D_0 D_0' = A_0 + B_0$; so-doing ensures that the predictive expectation and variances are the same as in the null model.

For the test statistic we compute the Kullback-Leibler divergence

$$d_3(X_1^n) = d_{KL}(M^{0,*}_{\lambda|X_1^n}, M^1_{\lambda|X_1^n}) \tag{72}$$

obtained by means of Wang *et al.* (2005) procedure.

If one considers the rule "Reject $\mathcal{E}^0$ if $\widehat{p-value} \leq 0.5$", one may define the empirical coverage as the proportion of the rejecting sample. For data generated from the alternative model, as done in this exercise, the empirical coverage should be higher than 0.05 and increase with the sample size. Table 1 gives the observed results for four different sample sizes. ($n = 10, 50, 100, 200$) and 3 trials.

n	trial 1	trial 2	trial 3
10	0.14	0.100	0.14
50	0.17	0.128	0.11
100	0.25	0.204	0.18
200	0.28	0.226	0.24

Table 1: Coverage rates

What we observe and learn from this numerical illustration may be summmarised as follows:
(i.) The Bayesian encompassing test, as developed above, under reasonable hypotheses, is indeed operational even under partial observability: we have described completely an algorithm and proved its operationality in an example.
(ii.) The computational burden increases substantially with the sample size: In the present case, from $20'$ for $n = 10$ to $130'$ for $n = 200$, for a simple run of the algorithm (with $N_1 = 200$, $N_2 = 200$ and $N_C = 500$). The 3 columns of table 1 give the results for 3 trials with identical parameters (N_1, N_2, N_C) and increasing sample size n. The similarity of the results suggests that from a numerical point of view the problem is reasonable well-conditioned.
(iii.) The problem of partial observability is a pervading issue in statistical practice. Common sense suggests that partial observability deteriorates the power, or the efficiency, of statistical procedures, when compared with complete observability. In the proposed illustration we have chosen a not too-favorable situation: the observability process reduces the dimension of the data from 2 to 1 and the predictive distributions of the two models are not far from each other as the distributions of the latent variable share identical expectations and variances and same prior distribution for the parameter of the partial observability equation.

4 Conclusions

This paper demonstrates the operationallity of a Bayesian encompassing test in the framework of partial observability even if numerical issues require powerful computations. We show that

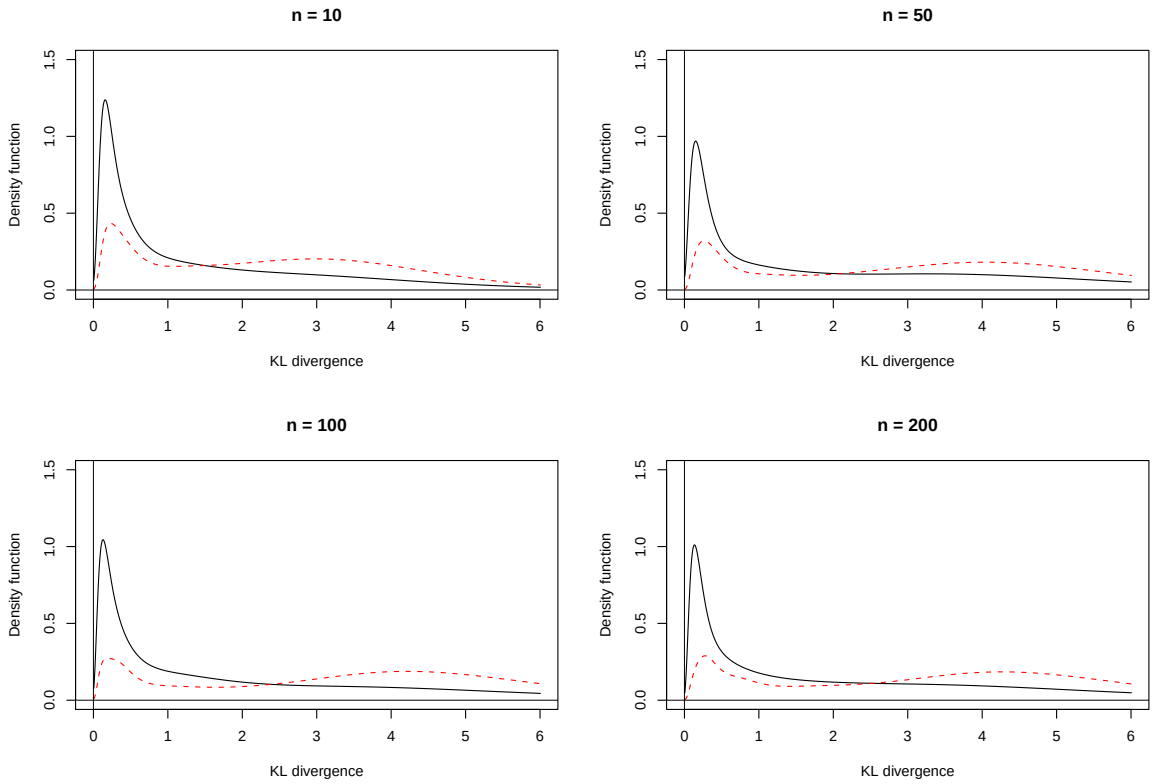

Figure 1: Density of the test statistics, null model (continuous line) and alternative model (dashed line)

the test, from a theoretical point of view, is feasible, but we also give evidence of practical problems to be faced. The proposed procedure might therefore be adapted to a wider class of problems.

Theoretically, the specifications of a model involving partial observability, as defined in this paper, can therefore be tested under reasonable hypotheses. The main hypotheses for using the encompassing test is the *extended BPTV* condition, which involves the *BPTV* condition at the level of latent variables and the neutrality of the model defining the partial observability process. Under these hypotheses, the *BPTV* condition at the level of manifest variables and using only identified parameters is fulfilled (Theorem 2.2).

A second result provides a sufficient condition for specifying the BPTV through the statistical models involving manifest variables and identified parameters only, but this condition should be justified from a contextual framework (Theorem 2.4) as it is not empirically controllable.

The generally substantial simplification of the procedure is due to the neutrality of the partial observability process: equations (9) and (20) bin the null model implying (22) in the extended model and equations (10) and (21) in the alternative model. Fortunately enough such

a structure seems natural in most real-life situations.

From a numerical point of view, the computations are made easier thanks to the possibility of direct simulations for the trajectories of a Dirichlet process (making use of Rolin-Sethuraman representation). If this were not the case, *e.g.* for other non parametric alternatives, recourse to heavier, and numerically more problematic, procedures, such as those based on MCMC, could probably not be avoided.

A Appendix: Bayesian Encompassing

The encompassing principle has been developed originally for comparing two experiments, in line with seminal papers by Cox (1961, 1962) about non-nested hypotheses. The main idea is to compare the inference made on parameter of the second model using the first one and the inference using the second model directly.

In a sampling framework, the comparison of these two inferences is operated though the concept of *Pseudo-True Value*, *i.e.* the probability limit in the first model of the maximum likelihood estimator of the parameter in the second model; more explicitly, if we have two statistical models on the same sample space: $\{P^0_{\xi|\theta} : \theta \in \Theta\}$ and $\{P^1_{\xi|\psi} : \psi \in \Psi\}$, the *classical* Pseudo-True Value (PTV) is a function of θ defined by:

$$PTV(\theta) = p \lim_{n\to\infty} \hat{\psi}_n(\theta) \Leftrightarrow \lim_{n\to\infty} P^0_{\xi|\theta}(|\hat{\psi}_n - PTV(\theta)| > \varepsilon) \longrightarrow 0, \tag{73}$$

where $\hat{\psi}_n$ is the maximum likelihood estimator in the second model. Therefore the comparison can be made between the inference on ψ both in the first model, through the PTV, and in the second model. Thus, the comparison by encompassing takes into account the idea that the preferred model explains in a "good" level the inference made with the non-preferred one.

In a Bayesian framework, the encompassing principle was sketched in Florens *et al.* (1990) and developed in Florens and Mouchart (1993) in the framework of a sufficiency principle on the parametric space. The parametric spaces are endowed with prior distributions (M^0_θ and M^1_ψ respectively); in order to include the parameter of the second model in the, so-called, *extended model*, a *Bayesian Pseudo-True Value* (BPTV) must be defined; this definition is made through a probability transition. The use of a conditional independence condition (*BPTV condition*) permits us to interpret the first model as the marginalization by sufficiency of the extended model, in the notation of Bayesian experiments as Markovian product, the extended model can

be written as:

$$Q^{0,*} = M^0_\theta \otimes P^0_{\xi|\theta} \otimes M_{\psi|\theta} \qquad \text{under } \xi \perp\!\!\!\perp \psi \mid \theta; Q^{0,*} \text{ the BPTV condition.} \tag{74}$$

Thus, the comparison is made between the posterior distributions of ψ both in the extended and in the alternative models, namely: $M^{0,*}_{\psi|\xi}$ and $M^1_{\psi|\xi}$. A statistics of test is a distance or divergence between these two posterior distributions. As, in general the distribution of this statistics is not known, this will be calibrated against the predictive measure in the null model, P^0_ξ.

In Florens and Mouchart (1993), the authors suggest a possibility for the specification of the BPTV, motivated by the corresponding definition in the *classical* framework; this is the sampling expectation in the first model of the posterior measure in the second one, namely:

$$M_{\psi|\theta} = \int M^1_{\psi|\xi} dP^0_{\xi|\theta} \qquad (\, = \; E^0[M^1_{\psi|\xi} \mid \theta\,]\,) \tag{75}$$

Florens *et al.* (2003) develop an operational specification test. They use as null hypothesis a parametric specification of the sampling model and as alternative a non parametric one; in the alternative model, a Dirichlet process is used as prior distribution. With this specification and using the BPTV specified as in (75) they show that the posterior measure $M^{0,*}_{\psi|\xi}$ is a mixture of Dirichlet processes and they use direct simulation of Dirichlet process, as developed in Rolin (1992) or in Sethuraman (1994), in order to compute the statistics of test and for its calibration against the predictive measure in the null model, P^0_ξ. For a non parametric alternative they suggest to focus the attention on the two posterior distributions of finite dimensional functionals of the parameter in the alternative model.

References

ALMEIDA, C. and MOUCHART, M. (2003a). Identification of polychoric correlations: A copula approach. Discussion Paper 03-21, Institut de statistique, Université catholique de Louvain.

ALMEIDA, C. and MOUCHART, M. (2003b). A note on a copula approach to polychoric correlations. Discussion Paper 03-22, Institut de statistique, Université catholique de Louvain.

ALMEIDA, C. and MOUCHART, M. (2005). Bayesian encompassing test under partial observability. Discussion Paper 05-14, Institut de statistique, Université catholique de Louvain.

COX, D.R. (1961). Tests of separate families of hypotheses. In: *Proc. 4th Berkeley Symp. Math. Stat. Probab. 1, 105-123* .

COX, D.R. (1962). Further Results on Tests of Separate Families of Hypotheses. *Journal of the Royal Statistical Society. Series B (Methodological)*, 24, 406–424.

FLORENS, J.P. and MOUCHART, M. (1993). Bayesian Testing and Testing Bayesians, (Chapter 11). In: G. Maddala, C. Rao, and H. Vinod, editors, *Handbook in Statistics*, 303–333. Amsterdam:North-Holland.

FLORENS, J.P., MOUCHART, M., and ROLIN, J.M. (1990). *Elements of Bayesian statistics*. New York: Marcel Dekker.

FLORENS, J.P., RICHARD, J.F., and ROLIN, J.M. (2003). Bayesian Encompassing Specfication test of a parametric model against a onparametric alternative. *Unpublished mimeo*.

JÖRESKOG, K. (1994). On the estimation of polychoric correlations and their asymptotic covariance matrix. *Psychometrike*, 59, 381–389.

JÖRESKOG, K., SÖRBOM, D., DU TOIT, S., and DU TOIT, M. (2002). *LISREL8: New Statistical Features*. SSI Scientific International Inc.

MIZON, G.E. and RICHARD, J.F. (1986). The Encompassing Principle and its Application to Testing Non-Nested Hypotheses. *Econometrica: Journal of the Econometric Society*, 54, 657–678.

MOUCHART, M. and OULHAJ, A. (2003). A Note on Partial Sufficiency with connection to the Identification Problem. *Metron*, LXI, 267–283.

MUTHÉN, B. (1983). Latent Variable Structural Equation Modeling with Categorical Data. *Journal of Econometrics*, 22, 43–65.

MUTHÉN, B. (1984). A General Structural Equation Model with dichotomous, ordered categorical variables, and continuous latent variables indicators. *Psychometrika*, 49, 115–132.

OLSSON, U. (1979). Maximum Likelihood of the polychoric correlation coefficient. *Psychometrika*, 44, 443–460.

ROLIN, J.M. (1992). Some Useful Properties of the Dirichlet Process. Discussion Paper 92-02, Institut de statistique, Université catholique de Louvain.

SETHURAMAN, J. (1994). A constructive definition of Dirichlet priors. *Statist. Sinica*, 4, 639–650.

WANG, Q., KULKARNI, S.R., and SERGIO, V. (2005). Divergenvce Estimation of Continuous Distribitions Based on Data-Dependent Partitions. *IEEE Transactions on Information Theory*, 3064 – 3074.