

Bayesian encompassing specification test under not completely known partial observability

Carlos Almeida* and Michel Mouchart†

Abstract. This paper proposes the construction of a Bayesian specification test based on the encompassing principle for the case of partial observability of latent variables. A structural parametric model (*null model*) is compared against a non-parametric alternative (*alternative model*) at the level of latent variables. The *null extended model* is obtained by incorporating the non Euclidean parameter of the alternative model. This extension is defined through a Bayesian Pseudo-True Value, that makes the null model a reduction by sufficiency of the extended model. The same observability process is introduced in both the null and the alternative models; after integrating out the latent variables, a null and alternative statistical models are accordingly obtained. The comparison is made between the posterior measures of the non Euclidean parameter (of the alternative model) in the extended and in the alternative statistical models. The general development is illustrated with an example where only a linear combination of a latent vector is observed; in the example, the partial observability is known up to the vector defining the observed linear combination. Some identifiability issues are treated and the example shows the operationality and some pitfalls of the proposed test, through a numerical experiment.

Keywords: Bayesian encompassing; Bayesian specification test; Dirichlet prior; Partial observability.

1 Introduction

An example of *partial observability* can be found in covariance structure models (or, LISREL type models) involving ordinal variables. In these models, the normality hypothesis of the latent variables has an important role. As a 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; see Olsson (1979) or Jöreskog (1994) for the estimation of the identified parameters. If we consider models involving also ordinal variables, we should justify the normality of the latent variables supposedly generating, by discretization, these ordinal variables; for the use of the normality assumption in that class of models see Muthén (1983, 1984); Jöreskog, Sörbom, du Toit, and du Toit (2002) and for details on the discretization model see Almeida and Mouchart (2003a,b). As another example of partial observability, consider the observation of a linear combination of latent variables (with

*Institut de Statistique, Université catholique de Louvain, Belgiu,
<http://www.stat.ucl.ac.be/ISpersonnel/almeida/almeida.html>

†Institut de Statistique, Université catholique de Louvain, Belgiu,
<http://www.stat.ucl.ac.be/ISpersonnel/mouchart/>

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, 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 in the form of $X = g(\xi, \alpha) \doteq g_\alpha(\xi)$ where g is a known function, $\alpha \in \mathbf{R}^q$ is an unknown parameter, ξ is a vector of latent variables and X is a vector of manifest (or observable) 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 $\alpha \in \mathbf{R}^q$ is unknown, but the functional form of g is known, is called *not completely known partial observability* and is the subject matter of this paper.

When 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 when 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. This loss of identifiability due to partial observability requires a particular care for a correct interpretation of the hypotheses involved in a testing procedure.

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:

$$\mathcal{E}^0 : (\theta, \xi) \sim Q_{\theta, \xi}^0 = M_\theta^0 \otimes P_{\xi|\theta}^0 \quad (1)$$

$$\mathcal{E}^1 : (\psi, \xi) \sim Q_{\psi, \xi}^1 = M_\psi^1 \otimes P_{\xi|\psi}^1 \quad (2)$$

where θ and ψ are the parameters characterizing the respective models of the latent vector ξ , and the Markovian product $M_\theta^0 \otimes P_{\xi|\theta}^0$, between a marginal (prior) probability M_θ^0 and a (sampling) regular conditional probability $P_{\xi|\theta}^0$, is the unique probability on the product space (θ, ξ) generated by $(M_\theta^0 \otimes P_{\xi|\theta}^0)(A \times B) = \int_{\theta \in A} P_{\xi|\theta}^0(B) dM_\theta^0$, and similarly for \mathcal{E}^1 . Thus, when densities exist, \mathcal{E}^0 might be represented as $q^0(\theta, \xi) = m^0(\theta)p^0(\xi \mid \theta)$. Note however that in \mathcal{E}^1 the use of densities is inappropriate because the parameter space for ψ is infinite dimensional. A general nonparametric alternative is specified by $P_{\xi|\psi}^1 = \psi$.

In both models, $X = g(\xi, \alpha) \doteq g_\alpha(\xi)$ defines the same *partial observability process*, where g is a known bi-measurable function and α is an unknown finite dimensional parameter. The fact that the partial observability process is deterministic implies that:

$$(a) X \perp\!\!\!\perp \theta \mid \alpha, \xi; Q^0 \quad \text{and} \quad (b) X \perp\!\!\!\perp \psi \mid \alpha, \xi; Q^1, \quad (3)$$

and clearly: $P_{X|\alpha, \theta, \xi}^0 = P_{X|\alpha, \psi, \xi}^1 = \delta_{\{X=g(\xi, \alpha)\}}$, where $\delta_{\{\bullet\}}$ is the *unit mass measure*.

Let us try to localize the contribution of this paper within the literature on Bayesian testing in general and, more specifically, within the literature on the encompassing principle. It is difficult to organize an abundant, and still growing, literature on Bayesian

testing in a systematic framework. After Florens and Mouchart (1993) it may nevertheless be useful to gather contributions focusing on the posterior probabilities of hypotheses (null (H_0) and alternative (H_1)) under different forms, as posterior odds, Bayes factors or p-values, see for example Berger (1985), Zellner (1984), Bayarri and Berger (2000), and for relations with a sampling framework see Bernardo (1980), DeGroot (1973), Berger and Delampady (1987). These approaches typically consider H_0 and H_1 within a unique Bayesian model and the null hypothesis as a restriction over the alternative hypothesis. In another approach different Bayesian models are associated to H_0 and H_1 and attention is focused on how different the inferences are relative to a parameter of interest, often defined in a more decision oriented framework; the Bayesian encompassing test permits this approach. For approaches based on Bayes factors and dealing with specification testing see Berger and Guglielmi (2001), Verdinelli and Wasserman (1998), see also Florens, Richard, and Rolin (2003) for some criticisms when, in the alternative model, a Dirichlet process is used as a prior distribution on the space of probability measures.

In line with seminal papers by Cox (1961, 1962) about testing non-nested hypotheses the encompassing principle has been developed for comparing two experiments sharing the same sample space. The main idea is to compare the inference made on the parameter of the second model using the first one and the inference on the same parameter using the second model directly. 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 a sampling theory approach and in a parametric framework see Mizon (1984); Mizon and Richard (1986). In a Bayesian framework, encompassing testing has been sketched in (Florens, Mouchart, and Rolin 1990, section 3.5) and developed in a general setup and applied to parametric models by Florens and Mouchart (1989, 1993) and in time series in Florens, Larribeau, and Mouchart (1994).

Let us be more specific on the Bayesian encompassing approach and consider the two statistical models on the same sample space, namely $\{P_{\xi|\theta}^0 : \theta \in \Theta\}$ and $\{P_{\xi|\psi}^1 : \psi \in \Psi\}$, endowed with prior distributions (M_θ^0 and M_ψ^1 respectively). In order to include the parameter of the second model in the, so-called, *extended model*, a *Bayesian Pseudo-True Value* (BPTV) is defined 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. The extended model is accordingly written as:

$$Q^{0,*} = M_\theta^0 \otimes P_{\xi|\theta}^0 \otimes M_{\psi|\theta} \quad \text{under } \xi \perp\!\!\!\perp \psi \mid \theta; Q^{0,*} \text{ the BPTV condition.}$$

The comparison is made between the posterior distributions of ψ in the extended and in the alternative models, namely: $M_{\psi|\xi}^{0,*}$ and $M_{\psi|\xi}^1$. A test statistic is constructed through a distance or a divergence between these two posterior distributions. Since, in general the distribution of this statistics is not known, it will be calibrated against the predictive measure in the null model, P_ξ^0 . Florens and Mouchart (1993) suggest as the specification of the BPTV the sampling expectation, in the null model, of the posterior

measure in the alternative one, namely:

$$M_{\psi|\theta} = \int M_{\psi|\xi}^1 dP_{\xi|\theta}^0 \quad (= E^0[M_{\psi|\xi}^1 | \theta]). \quad (4)$$

This suggestion is motivated by the idea that the meaning of ψ if the null model were preferred might be obtained as the null predictive expectation of the alternative posterior distribution; this suggestion for the specification of the BPTV may accordingly be viewed as a Bayesian adaptation of the suggestion by [Cox \(1961\)](#).

More recently, [Florens, Richard, and Rolin \(2003\)](#) developed an operational specification test. They use as null hypothesis a parametric specification of the sampling null model and as alternative a nonparametric specification with a Dirichlet process as prior distribution. With this specification and using the BPTV specified as in (4), they show that the posterior measure $M_{\psi|\xi}^{0,*}$ is a mixture of Dirichlet processes. They use direct simulation of Dirichlet process, as developed in [Rolin \(1992\)](#) or in [Sethuraman \(1994\)](#), in order to compute the test statistic and for its calibration against the predictive measure in the null model, P_{ξ}^0 . For a nonparametric alternative they suggest to focus attention on the two posterior distributions of finite dimensional functionals of the parameter in the alternative model. Later, in [Almeida and Mouchart \(2005\)](#) the encompassing specification test has been extended to the a case of partial observability when the function defining the partial observability is completely known.

A comparison of the present paper with a recent paper, [Berger and Guglielmi \(2001\)](#), to be referred as BG, may be illuminating. Firstly, BG is basically oriented toward Bayes factors. Thus the comparison of two models is made on the sample space through the respective predictive distributions. In BG, model fit is crucial and there is a concern to neutralize the role of the prior distribution by making a privileged use of non-informative prior distributions. Furthermore, the strictly positive probability of ties when simulating the predictive distribution under a Dirichlet process specification creates difficulties when the sampling distribution is assumed to be continuous. BG resort to a mixture of Polya tree processes to face this issue. In this paper the model comparison is based on the posterior distributions and therefore operates on the parameter space; thus ties in the predictive distributions do not raise difficulties. Here the emphasis is on interpretation (i.e. how far can we interpret a model and its parameters) in the light of the other model and the BPTV is the key for carrying out that comparison. Secondly, BG embed the parameter of the null model into the alternative model and use, for that purpose, Polya trees. In this paper we embed the alternative model into an extended version of the null model and this embedding is operated by making use of the BPTV. Therefore BG interpret θ in the alternative model whereas we interpret ψ in the (extended) null model.

This paper is organized as follows. The next section exposes the theoretical framework that describes the impact of partial observability on an encompassing test of a parametric hypothesis against a non-parametric alternative hypothesis. Section 3 applies the general framework for the particular case of observing a linear combination of latent variables and gives a summary of the numerical results of an example, that also provides some details on numerical procedures. Some concluding remarks complete the

paper.

2 General Results

The encompassing principle. This principle may be viewed as a possible approach for switching from an “old” theory to 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. Here, we extend that idea for a case where only partial observation is available.

When completing the structural models (1) and (2), in order to incorporate α and X , we shall also assume a separation between the structural and the observability models in the sense that the sampling distributions of ξ should not depend on α ; more precisely:

$$(a) \alpha \perp\!\!\!\perp \xi \mid \theta; Q^0 \quad \text{and} \quad (b) \alpha \perp\!\!\!\perp \xi \mid \psi; Q^1. \quad (5)$$

Condition (5a) also means that in the structural null model, θ is a *sufficient* parameter and therefore conditionally on θ , the ξ 's are not informative about α (i.e. the prior and the posterior distributions of α given θ are the same). Note that θ being *minimal sufficient* also means that θ is *identified*. Similarly for (5b) in the alternative model. For the approach of sufficiency on the parameter space see Barankin (1960) and, in a Bayesian framework, (Florens, Mouchart, and Rolin 1990, subsection 2.3.3, and for the link with identification as minimal sufficiency the results in sections 4.4 and 4.6).

The *complete null model* \mathcal{E}^0 and the *complete alternative model* \mathcal{E}^1 are therefore

$$\begin{aligned} \mathcal{E}^0 : \quad (\theta, \alpha, \xi, X) &\sim Q_{\theta, \alpha, \xi, X}^0 = M_{\theta, \alpha}^0 \otimes P_{\xi \mid \theta}^0 \otimes P_{X \mid \xi, \alpha}^0 \\ \mathcal{E}^1 : \quad (\psi, \alpha, \xi, X) &\sim Q_{\psi, \alpha, \xi, X}^1 = M_{\psi, \alpha}^1 \otimes P_{\xi \mid \psi}^1 \otimes P_{X \mid \xi, \alpha}^1. \end{aligned}$$

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* $\mathcal{E}^{0,*}$ makes use of an *extended Bayesian Pseudo-True Value (BPTV) condition*, namely:

$$\psi \perp\!\!\!\perp \alpha, \xi \mid \theta; Q^{0,*}. \quad (6)$$

This condition actually aggregates two conditions, namely:

$$(a) \psi \perp\!\!\!\perp \alpha \mid \xi, \theta; Q^{0,*} \quad \text{and} \quad (b) \psi \perp\!\!\!\perp \xi \mid \theta; Q^{0,*} \quad (7)$$

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. Note that (5a) and (7a) are equivalent to:

$$\alpha \perp\!\!\!\perp (\xi, \psi) \mid \theta; Q^{0,*}. \quad (8)$$

The *extended statistical model*, obtained after integrating the latent variable ξ , can be written as:

$$Q_{\theta, \alpha, \psi, X}^{0,*} = M_{\theta, \alpha}^0 \otimes P_{X \mid \theta, \alpha}^0 \otimes M_{\psi \mid \theta},$$

where the conditional probability $M_{\psi|\theta}$ represents a BPTV satisfying (6).

Since the structural process acts independently of the partial observability process, we assume that α and θ are a priori independent, more specifically:

$$(a) \alpha \perp\!\!\!\perp \theta; Q^0 \quad \text{and} \quad (b) \alpha \perp\!\!\!\perp \psi; Q^1. \quad (9)$$

The next theorem shows that (9a), along with the extended BPTV condition (6), implies in the extended complete model, a complete separation between the structural model and the parameter of the observability process.

Theorem 1. *Using the extended BPTV condition (6), we have under (9a), that:*

$$\alpha \perp\!\!\!\perp (\xi, \psi, \theta); Q^{0,*} \quad (10)$$

Proof. Clearly (10) is equivalent to (a) $\alpha \perp\!\!\!\perp \xi, \theta$ and (b) $\alpha \perp\!\!\!\perp \psi \mid \xi, \theta$. Property (a) is equivalent to (9a) and $\alpha \perp\!\!\!\perp \xi \mid \theta$, implied by (8), and property (b) is also implied by (8). ■

The encompassing principle consists in evaluating $d_1(X) = d^*(M_{\psi|X}^{0,*}, M_{\psi|X}^1)$ as a test statistic, where d^* is a discrepancy or a divergence between probability measures which must be calibrated against P_X^0 , the predictive distribution under the model \mathcal{E}^0 . In a decision theoretic framework, loss functions, and therefore decision procedures, are not likely to depend on the complete specification of the functional parameter ψ . It is instead more likely that a finite dimensional functional $\lambda = h(\psi)$ represents the parameter of actual interest. In such a case, the test statistic takes the form $d^*(M_{\lambda|X}^{0,*}, M_{\lambda|X}^1)$ rather than $d^*(M_{\psi|X}^{0,*}, M_{\psi|X}^1)$. [Florens, Richard, and Rolin \(2003\)](#) have developed an operational version of the encompassing test statistic when the parameter of interest is λ and the observation is complete, *i.e.* when $\xi = X$.

Identification and Encompassing testing. As mentioned before, 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 is a sufficient statistic in both models, see [Mouchart and Oulhaj \(2003\)](#).

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 γ_X is a measurable function of θ and α , and in the alternative model the parameter identified by the statistical model $P_{\xi|\psi}^1 \circ g_\alpha^{-1}$ is $\omega_X = \psi \circ g_\alpha^{-1}$, a measurable function of ψ and α . These identified parameters are such that:

$$(a) X \perp\!\!\!\perp \theta, \alpha \mid \gamma_X; Q^0 \quad (b) X \perp\!\!\!\perp \psi, \alpha \mid \omega_X; Q^1. \quad (11)$$

This raises the question whether $d_1(X)$ 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 legitimate to concentrate the encompassing test on the identified part of ψ , i.e. to evaluate

$$d_2(X) = d^*(M_{\omega_X|X}^{0,*}, M_{\omega_X|X}^1)$$

instead of $d_1(X)$. 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 not depend on the data. This is clearly the case for the alternative model in view of condition (11b). The next theorem gives a condition under which a similar property holds in the extended null model and provides as a Corollary, i.e. equation (13), a more operational structure.

Theorem 2. *Under (6), the extended model $\mathcal{E}^{0,*}$ satisfies:*

$$X \perp\!\!\!\perp \psi, \theta, \alpha \mid \gamma_X; Q^{0,*} \tag{12}$$

Proof. Indeed, in $Q^{0,*}$, we have that (6) implies $\psi \perp\!\!\!\perp X \mid \theta, \alpha$, which jointly with (11a) implies (12). ■

Then, under condition (6) we have $X \perp\!\!\!\perp \omega_X \mid \gamma_X; Q^{0,*}$ and we may write:

$$Q_{\theta, \alpha, \psi, X}^{0,*} = \left[M_{\gamma_X}^0 \otimes M_{\omega_X|\gamma_X}^{0,*} \otimes P_{X|\gamma_X}^0 \right] \otimes M_{\alpha, \theta, \psi|\gamma_X}^{0,*}. \tag{13}$$

Equation (13) means that in the extended complete model, the prior conditional distribution $M_{\alpha, \theta, \psi|\gamma_X}^{0,*}$ need not be explicitly specified for the construction of the encompassing test.

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

Theorem 3. *Under the extended BPTV condition (6), if*

$$\psi \perp\!\!\!\perp \gamma_X \mid \omega_X; Q^{0,*} \tag{14}$$

then

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

Proof. Clearly (15) is implied by $\psi \perp\!\!\!\perp X, \gamma_X \mid \omega_X$ which is equivalent to (14) along with $\psi \perp\!\!\!\perp X \mid \gamma_X, \omega_X$ implied by (12). ■

Theorem 3 permits us to write: $M_{\psi|X}^{0,*} = \int M_{\psi|\omega_X}^{0,*} dM_{\omega_X|X}^{0,*}$. If, by specification, we impose the coherence condition:

$$M_{\psi|\omega_X}^{0,*} = M_{\psi|\omega_X}^1, \tag{16}$$

a convenient distance for the posterior distributions of ψ can be defined through a distance between posterior distributions of ω_X , as in $d_2(X)$.

It should be emphasized that the conditions (14) and (16) are restrictions on the prior specification only and are accordingly not testable as they do not bear on the sampling neither on the structural process. Furthermore, these conditions depend on the specification of $M_{\alpha,\theta,\psi|\gamma_X}^{0,*}$ that leaves the testing procedure invariant, as a result of equation (13). For given M_α^0 and M_θ^0 , these conditions restrict the specification of the BPTV $M_{\psi|\theta}$ that provides the interpretation of ψ (the natural parameter in the alternative model) in the (extended) null model. The motivation of these restrictions is to give explicit conditions that justify formally that the testing procedure compares the posterior distributions of ω_X , the identified part of (α, ψ) , rather than the complete posterior distribution, a rather natural requirement.

Summarizing, under the conditions (6), (14) and (16), d_2 is an adequate and more operational substitute of d_1 . Because of (13), conditions (14) and (16) are restrictions on the unspecified $M_{\alpha,\theta,\psi|\gamma_X}^{0,*}$. In a decision theoretic framework, we assume also that the parameter of interest λ is a function of ω_X rather than ψ , *i.e.* $\lambda = h(\omega_X)$.

3 An Example

Let us illustrate the computations implied by the test developed so far by examining a simple example. We first sketch a parametric (fully 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 necessary steps for evaluating the posterior distributions of ω_X , the parameter identified by the data in the alternative model, from which we evaluate the two posterior distributions of $\lambda = h(\omega_X)$, the parameter of interest, and finally simulate the distribution of the test statistic $d^*(M_{\lambda|X}^{0,*}, M_{\lambda|X}^1)$. In both models, the simulation of the posterior distribution of ω_X requires, as a preliminary step, the simulation of the posterior distribution of α . Under a continuous specification of the prior distribution of α , the posterior distribution of α given a sample $X_1^n = (X_1, \dots, X_n)$ can be simulated by acceptance rejection methods. Finally, we control the operationality of the proposed procedure through a numerical exercise.

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

$$X_i = \beta' \zeta_i + \sigma_x \varepsilon_{x_i}, \quad Z_i = \zeta_i + \Sigma_{zz}^{\frac{1}{2}} \varepsilon_{z_i} \quad (X_i, Z_i)' \in \mathbf{R}^{1+k}, \quad \zeta_i \sim \text{ind } \mathcal{N}_{(k)}(\mu_\zeta, \Sigma_{\zeta\zeta})$$

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

$$W_i = A \xi_i^* \quad \text{where} \quad A = \begin{pmatrix} \beta' & \sigma_x & 0' \\ I_k & 0 & \Sigma_{zz}^{\frac{1}{2}} \end{pmatrix} \quad \text{and} \quad \xi_i^* = (\zeta_i', \varepsilon_{x_i}, \varepsilon_{z_i}')'$$

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: $X_i = g_\alpha(\xi_i) = \alpha' \xi_i$, 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 nonparametric alternative model both satisfying (5) and (9), namely: $\alpha \perp\!\!\!\perp (\xi, \theta); Q^0$ and $\alpha \perp\!\!\!\perp (\xi, \psi); Q^1$; more explicitly:

$$\mathcal{E}^0 : \begin{cases} \xi | \theta, \alpha \sim \mathcal{N}_{(2)}(\theta, A_0) \\ \theta | \alpha \sim \mathcal{N}_{(2)}(0, B_0) \\ \alpha \sim M_\alpha^0 \end{cases} \quad \mathcal{E}^1 : \begin{cases} \xi | \psi, \alpha \sim \psi \\ \psi | \alpha \sim \mathcal{Di}(n_0^1 \mathcal{N}_{(2)}(0, C_0)) \\ \alpha \sim M_\alpha^1, \end{cases} \quad (17)$$

where A_0, B_0 and C_0 are positive definite matrices, $n_0 > 0$ and \mathcal{Di} denotes a Dirichlet process. The corresponding statistical models are:

$$\mathcal{E}^0 : \begin{cases} X | \theta, \alpha \sim \mathcal{N}(\alpha' \theta, \alpha' A_0 \alpha) \\ \theta | \alpha \sim \mathcal{N}_{(2)}(0, B_0) \\ \alpha \sim M_\alpha^0 \end{cases} \quad \mathcal{E}^1 : \begin{cases} X | \psi, \alpha \sim \omega_X \\ \omega_X | \alpha \sim \mathcal{Di}(n_0^1 \mathcal{N}(0, \alpha' C_0 \alpha)) \\ \alpha \sim M_\alpha^1. \end{cases} \quad (18)$$

where, as in Section 2, $\omega_X = \psi \circ g_\alpha^{-1}$ is the parameter of \mathcal{E}^1 identified by X .

Posterior distributions of α . In the null model, the posterior distribution of $(\alpha | X_1^n)$ may be obtained as follows: For an n -size sample, the joint distribution of X_1^n and θ conditionally on α is:

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

with:

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

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

Using a continuous prior distribution, $m^0(\alpha)$, the posterior will be also continuous, moreover: $m^0(\alpha | X_1^n) \propto m^0(\alpha) p^0(X_1^n | \alpha)$, where $p^0(X_1^n | \alpha)$ is the density function of a multivariate normal distribution with zero means and variance matrix Σ_n .

In the alternative model \mathcal{E}^1 , Dirichlet process properties imply that: $X | \alpha; \mathcal{E}^1 \sim \mathcal{N}(0, \alpha' C_0 \alpha)$ and, conditionally on α , the posterior distribution of the identified parameter ω_X is a Dirichlet process:

$$\omega_X | \alpha, X_1^n; \mathcal{E}^1 \sim \mathcal{Di}(n_0^1 \mathcal{N}(0, \alpha' C_0 \alpha) + n \hat{F}_n), \quad (19)$$

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:

$$M_{\omega_X|X_1^n}^1 = \int M_{\omega_X|\alpha, X_1^n}^1 dM_{\alpha|X_1^n}^1. \quad (20)$$

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

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

where \hat{F}_j is the empirical distribution using the first j observations (X_1^j) and \otimes stands for the Markovian product as defined in (1). A Radon-Nikodym derivative of $P_{X_1^n|\alpha}^1$ is easily obtained relative to a measure $L_*^{(n)}$ to be defined. Let L be the Lebesgue measure on the real numbers and denote the following Radon-Nikodym derivatives:

$$f_0^{1, \alpha} = \frac{d\mathcal{N}(0, \alpha C_0 \alpha)}{dL} \quad \text{and} \quad \hat{f}_i = \frac{d\hat{F}_i}{d\hat{F}_n}.$$

As the measures L and \hat{F}_n are mutually singular, we may write:

$$\begin{aligned} \frac{dP_{X_i|X_1^{i-1}, \alpha}^1}{d(L + \hat{F}_n)}(x) &= \frac{n_0^1 f_0^{1, \alpha}(x) + (i-1) \hat{f}_{i-1}(x)}{n_0^1 + (i-1)} \\ \frac{dP_{X_1^n|\alpha}^1}{dL_*^{(n)}}(x_1^n) &= \prod_{\{1 \leq i \leq n\}} \frac{n_0^1 f_0^{1, \alpha}(x_i) + (i-1) \hat{f}_{i-1}(x_i)}{n_0^1 + (i-1)} \end{aligned} \quad (21)$$

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

$$m^1(\alpha | X_1^n) \propto m^1(\alpha) \frac{dP_{X_1^n|\alpha}^1}{dL_*^{(n)}}(X_1^n). \quad (22)$$

Encompassing. Let us now build an encompassing test under assumptions (14) and (16) where the BPTV has the following structure:

$$M_{\omega_X|\gamma_X} = \int M_{\omega_X|\tilde{X}_1^n}^1 dP_{\tilde{X}_1^n|\gamma_X}^0 = E^0[M_{\omega_X|\tilde{X}_1^n}^1 | \gamma_X] \quad (23)$$

where $\tilde{X}_1^n = (\tilde{X}_1, \dots, \tilde{X}_n)$ is a virtual sample generated from $P_{\tilde{X}_1^n|\gamma_X}^0$. This specification ensures the BPTV condition, i.e. $\omega_X \perp\!\!\!\perp X_1^n | \gamma_X; Q^{0,*}$.

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

$$\tilde{X}_1^n \perp\!\!\!\perp X_1^n | \gamma_X; Q^{0,*}, \text{ i.e. } P_{\tilde{X}_1^n|X_1^n}^0 = \int P_{\tilde{X}_1^n|\gamma_X}^0 dM_{\gamma_X|X_1^n}^0 \quad (24)$$

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

$$\begin{aligned}
 M_{\omega_X|X_1^n}^{0,*} &= E^0[M_{\omega_X|\gamma_X} | X_1^n] && \text{by BPTV condition} \\
 &= E^0[E^0[M_{\omega_X|\tilde{X}_1^n}^1 | \gamma_X] | X_1^n] && \text{by (23)} \\
 &= E^0[M_{\omega_X|\tilde{X}_1^n}^1 | X_1^n] && \text{by (24)}
 \end{aligned} \tag{25}$$

By (25) and (20), both in the null and in the alternative models the posterior distributions of the identified parameter ω_X are mixtures of Dirichlet processes. The algorithm for Bayesian encompassing testing developed in Florens, Richard, and Rolin (2003) can be applied in this case with a simple adaptation; this algorithm is based on a direct simulation of trajectories of the Dirichlet process as developed in Rolin (1992) and in Sethuraman (1994).

For the sake of illustration, we have selected for λ the first two moments, making our exercise comparable with that of Florens, Richard, and Rolin (2003). The simulation of the posterior distributions $M_{\lambda|X_1^n}^1$ and $M_{\lambda|X_1^n}^{0,*}$ may be obtained as follows:

Step 1 *Simulation of $\lambda_1^1, \dots, \lambda_{M_1}^1$ given X_1^n in \mathcal{E}^1 .* For each $i = 1, \dots, M_1$, let α_i be a simulated sample from $M_{\alpha|X_1^n}^1$ from (22) and $(\omega_X)_i$ a simulated trajectory of $M_{\omega_X|X_1^n, \alpha_i}^1$ from (19), then $(\omega_X)_i$ is a simulated trajectory of $M_{\omega_X|X_1^n}^1$. Compute $\lambda_i^1 = h[(\omega_X)_i]$ for $i = 1, \dots, M_1$

Step 2 *Simulation of $\lambda_1^0, \dots, \lambda_{M_0}^0$ given X_1^n in $\mathcal{E}^{0,*}$.* For each $i = 1, \dots, M_0$, simulate $(\tilde{X}_1^n)_i$ from $P_{\tilde{X}_1^n|X_1^n}^0$ through (24). We simulate again $(\omega_X)_i$ from $M_{\omega_X|\tilde{X}_1^n}^1$ as in step 1. From (25), this sequence is distributed as $M_{\omega_X|X_1^n}^{0,*}$. Compute $\lambda_i^0 = h[(\omega_X)_i]$ for $i = 1, \dots, M_0$

Step 3 *Computation of the test statistics $d_2(X_1^n)$.* From the iid samples $\lambda_1^1, \dots, \lambda_{M_1}^1$ from step 1 and $\lambda_1^0, \dots, \lambda_{M_0}^0$ from step 2, estimate $d_2(X_1^n)$ by means of a discrepancy between the two posterior distributions of $\lambda | X_1^n$; the simplest algorithm, developed in Wang, Kulkarni, and Verdú (2005), is used for this estimation. This algorithm uses a data driven partition for integration. In fact when, for example, the Kullback-Leibler discrepancy between two equivalent probability measures, P and Q is estimated through generating samples X_1^n from P and Y_1^n from Q we compute:

$$\begin{aligned}
 d_{KL}(P, Q) &= \int \log \left(\frac{dP}{dQ} \right) dP \\
 &= \int \frac{dP}{dQ} \log \left(\frac{dP}{dQ} \right) dQ \\
 &\simeq \sum_{\{i \in \mathcal{P}\}} \frac{\Delta P_i}{\Delta Q_i} \log \left(\frac{\Delta P_i}{\Delta Q_i} \right) \Delta Q_i
 \end{aligned}$$

where \mathcal{P} is a partition of the space, ΔP_i and ΔQ_i are the number of points of the samples X_1^n and Y_1^n respectively in the set i of the partition. This partition is chosen such that the quantities (ΔQ_i) are greater than a given positive value.

Step 4 Simulation of $(X_1^n)_\ell^*$, $\ell = 1, \dots, N_C$ for calibration. Simulate $(\alpha, \theta)_\ell$, $\ell = 1, \dots, N_C$ from (18), for each $(\alpha, \theta)_\ell$, simulate $(X_1^n)_\ell^* \mid (\alpha, \theta)_\ell$ from (18). Thus $(X_1^n)_\ell^* \sim P_{X_1^n}^0$. For each $(X_1^n)_\ell^*$, compute the test statistic $d_2[(X_1^n)_\ell^*]$ repeating the steps 1 to 3. Finally estimate the p -value as follows:

$$p\text{-value} = \frac{1}{N_C} \sum_{1 \leq \ell \leq N_C} \mathbf{I}_{\{d_2((X_1^n)_\ell^*) > d_2(X_1^n)\}}$$

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

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

This is indeed the case when, in the null model, the sampling distribution is continuous. Therefore the density of $X_1^n \mid \alpha; \mathcal{E}^1$, as given in (21), may be simplified, with probability one, into:

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

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 an iid sample relative to the alternative model, conditionally on α . ■

Numerical illustration. For the sake of simplicity, we specify in this exercise the same prior distribution for α i.e. $M_\alpha^0 = M_\alpha^1$. A suitable specification of that prior distribution may be obtained through the following reparametrization. The minimal sufficient parameter may be easily described, namely $\alpha_X = \alpha'(A_0 + B_0)\alpha$. Let us reparametrize α into (α_X, τ) as follows:

$$\alpha = \sqrt{\alpha_X} R_0^{-1} \begin{pmatrix} \cos \tau \\ \sin \tau \end{pmatrix}; \quad \tau \in [0, 2\pi) \quad \alpha_X \geq 0$$

where R_0 is a 2×2 -matrix such that: $(A_0 + B_0) = R_0' R_0$. For the prior distribution let us consider:

$$\alpha_X \perp\!\!\!\perp \tau; \quad \alpha_X \sim \text{Gamma}(1, 1), \quad \tau \sim \mathcal{U}_{[0, 2\pi)}.$$

The variances covariances matrices are fixed to $A_0 = B_0 = \begin{pmatrix} 1 & .5 \\ .5 & 1 \end{pmatrix}$. The purpose of this exercise is to evaluate to which extent a sample generated by a distribution in the alternative sampling model is likely to be associated to a value of the test statistic relatively far in the tail of the null predictive distribution.

Consider accordingly the following specification:

$$\begin{aligned} \phi &= (\phi_1, \phi_2)'; & \phi_1 &\perp\!\!\!\perp \phi_2; & \phi_i &\sim \chi_2^2 \\ \xi &= D_0 \left(\tilde{\xi} - \begin{pmatrix} 2 \\ 2 \end{pmatrix} \right); & \tilde{\xi} &= (\tilde{\xi}_1, \tilde{\xi}_2); & \tilde{\xi}_1 &\perp\!\!\!\perp \tilde{\xi}_2 \mid \phi; & \tilde{\xi}_i &\sim \text{Expo} \left(\frac{1}{\phi_i} \right) \end{aligned}$$

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 consider the Kullback-Leibler divergence numerically computed by means of a procedure suggested by Wang, Kulkarni, and Verdú (2005).

If one considers the rule “Reject \mathcal{E}^0 if $p\text{-value} \leq 0.05$ ”, one may define the empirical coverage as the proportion of the rejected samples because the sample is generated in the alternative region. The empirical coverage is expected to be higher than 0.05 and increase with the sample size. Table 3 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	0.10	0.14
50	0.17	0.13	0.11
100	0.25	0.20	0.18
200	0.28	0.23	0.24

Table 1: Coverage rates

One may notice that, with a slight exception in trial 3, the coverage rate increases monotonically with the sample size and is always higher than 0.05.

4 Conclusions

This paper demonstrates the operationality of a Bayesian encompassing test in the framework of partial observability even if numerical issues require powerful computations. We show that the test, from a theoretical point of view, is feasible. The proposed procedure might therefore be adapted to a wider class of problems.

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 the Rolin-Sethuraman representation). If this were not the case, e.g. for other nonparametric alternatives, recourse to heavier, and numerically more problematic, procedures, such as those based on MCMC, could probably not be avoided.

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. <http://www.stat.ucl.ac.be/ISpub/dp/2003/dp0321.pdf>

- (2003b). “A note on a copula approach to polychoric correlations.” Discussion Paper 03-22, Institut de statistique, Université catholique de Louvain. <http://www.stat.ucl.ac.be/ISpub/dp/2003/dp0322.pdf>
- (2005). “Bayesian encompassing test under partial observability.” Discussion Paper 05-14, Institut de statistique, Université catholique de Louvain. <http://www.stat.ucl.ac.be/ISpub/dp/2005/dp0514.pdf>
- Barankin, E. (1960). “Sufficient parameters: Solution of the minimal dimensionality problem.” *Ann. Inst. Stat. Math.*, 12: 91–118.
- Bayarri, M. J. and Berger, J. O. (2000). “P-values for Composite Null Models.” *Journal of the American Statistical Association*, 95(452): 1127–1142.
- Berger, J. O. (1985). *Statistical decision theory and Bayesian analysis*. 2nd ed.. Springer Series in Statistics. New York: Springer-Verlag.
- Berger, J. O. and Delampady, M. (1987). “Testing Precise Hypotheses.” *Statistical Science*, 2(3): 317–335.
- Berger, J. O. and Guglielmi, A. (2001). “Bayesian and Conditional Frequentist Testing of a Parametric Model Versus Nonparametric Alternatives.” *Journal of the American Statistical Association*, 94: 174–184.
- Bernardo, J. M. (1980). “A Bayesian analysis of classical hypothesis testing.” *Bayesian Statistics, Proc. 1st Int. Meet., Valencia 1979*, 605–647. Valencia: University Press.
- Cox, D. R. (1961). “Tests of separate families of hypotheses.” *Proc. Fourth Berkeley Symp. Math. Stat. Probab. 1*, 105–123. Berkeley: University of California Press.
- (1962). “Further Results on Tests of Separate Families of Hypotheses.” *Journal of the Royal Statistical Society. Series B (Methodological)*, 24: 406–424.
- DeGroot, M. H. (1973). “Doing What Comes Naturally: Interpreting a Tail Area as a Posterior Probability or as a Likelihood Ratio.” *Journal of the American Statistical Association*, 68(344): 966–969.
- Florens, J.-P, Larribeau, S., and Mouchart, M. (1994). “Bayesian encompassing test of a unit root Hypothesis.” *Econometric Theory*, 10: 747–763.
- Florens, J.-P and Mouchart, M. (1989). “Bayesian Specification Test.” In Cornet, B. and Tulkens, H. (eds.), *Contributions in Operational Research and Economics*, 467–460. Cambridge:MIT Press.
- (1993). “Bayesian Testing and Testing Bayesians).” In Maddala, G., Rao, C., and Vinod, H. (eds.), *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 specification test of a parametric model against a nonparametric alternative.” *Unpublished mimeo*, revised version of Discussion Paper 96-08, Institut de Statistique, Université catholique de Louvain. <http://www.stat.ucl.ac.be/ISpub/dp/1996/dp9608.ps>
- Jöreskog, K. (1994). “On the estimation of polychoric correlations and their asymptotic covariance matrix.” *Psychometrika*, 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. (1984). “The encompassing approach in econometrics.” In Hendry, D. F. and Wallis, K. F. (eds.), *Econometrics and Quantitative Economics*, 135 – 172. Oxford: Basil Blackwell.
- Mizon, G. E. and Richard, J.-F. (1986). “The Encompassing Principle and its Application to Testing Non-Nested Hypotheses.” *Econometrica*, 54(3): 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.
- (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. <http://www.stat.ucl.ac.be/ISpub/dp/1992/dp9202.ps>
- Sethuraman, J. (1994). “A constructive definition of Dirichlet priors.” *Statist. Sinica*, 4: 639–650.
- Verdinelli, I. and Wasserman, L. (1998). “Bayesian Goodness-Of-Fit Testing Using Infinite-Dimensional Exponential Families.” *The Annals of Statistics*, 26(4): 1215–1241.
- Wang, Q., Kulkarni, S. R., and Verdú, S. (2005). “Divergence Estimation of Continuous Distributions Based on Data-Dependent Partitions.” *IEEE Transactions on Information Theory*, 51: 3064 – 3074.
- Zellner, A. (1984). “Posterior odds ratios for regression hypotheses: General considerations and some specific results.” In Zellner, A. (ed.), *Basic Issues in Econometrics*, 275–304. Chicago: Univ. of Chicago Press.

Acknowledgments

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. Without implication, comments by Jean-Marie Rolin, an anonymous referee and an associate editor have been highly useful for improving the presentation of this paper.