Research Report

Department of Statistics

No. 2001:2

Bayesian Assessment of Dimensionality in Multivariate Reduced Rank Regression

Jukka Corander
Mattias Villani

Department of Statistics, Stockholm University, SE-106 91 Stockholm, Sweden

BAYESIAN ASSESSMENT OF DIMENSIONALITY IN MULTIVARIATE REDUCED RANK REGRESSION

JUKKA CORANDER AND MATTIAS VILLANI

Abstract. We consider Bayesian inference about the dimensionality in the multivariate reduced rank regression framework, which encompasses several models such as MANOVA, simultaneous equations, factor analysis and cointegration models for multiple time series. The fractional Bayes approach is used to derive an approximation to the posterior distribution of the dimensionality. To investigate the finite sample properties of our solution, we have applied it to a wide variety of real and simulated data sets. The proposed approach compares favorably to other established estimators of the dimensionality.

1. Introduction

A common situation in multivariate analysis involves exploration of the relationships between sets of variables, either by explicit parametric models or by descriptive methods such as principal components and canonical correlations. Although it was early understood that these instances may be jointly represented in terms of multivariate regression obeying a so called reduced rank structure for certain parameters (see, *e.g.*, the pioneering work by Anderson, 1951), such an approach has only recently been fully appreciated by a larger statistical community.

One essential strength of the reduced rank regression (RRR) framework is its generality, as it encompasses several well-known models such that MANOVA, simultaneous equations, factor analysis and several popular models for multiple time series. For a thorough treatment concerning the last of the mentioned fields, see Ahn and Reinsel (1990), Johansen (1995), Velu, Reinsel and Wichern (1986) and for the others, see Anderson (1984, 1994) and the references therein. An excellent review of various issues may also be found in Reinsel and Velu (1998).

The typical model uncertainty in regular full rank multivariate regression is about the choice of relevant predictor variables. Several reasonable solutions are available for this latter model choice problem, see, *e.g.*, Brown et al. (1999) or George and Foster (2000). It has been more of a challenge to produce sensible inferences on the dimensionality of the subspace of regression coefficients for a *fixed* set of predictor variables. Geweke (1996) and Kleibergen and Paap (forthcoming) both propose computationally demanding semi-Bayesian approaches to assessing dimensionality in RRR. The methods employed in applied work are based on asymptotic approximations, such as the information theoretic criteria, *e.g.,* Akaike (1974), approximate logarithmic Bayes factor (Schwarz, 1978) or sequential tests (Anderson, 1951; Izenman, 1980). There have also been some proposals within a specific class of reduced rank models, see *e.g.* Chao and Phillips (1999) for a criterion tailored to cointegration models.

Key words and phrases. Cointegration, Fractional Bayes, Growth curves, Reduced Rank Regression.

The authors are grateful for financial support from the Academy of Finland, grant no. 50203 (Corander) and the Swedish Council of Research in Humanities and Social Sciences (HSFR) (Villani).

Among the mentioned approximations, only the approach of Schwarz (1978) aims at approximating the posterior distribution of the dimensionality. This is an important point as the posterior distribution forms an attractive representation of the uncertainty in the dimensionality inference. The Schwarz approximation is known to be rather rough, however, and it underestimates the underlying model dimension in many instances (Kass and Raftery, 1995).

We utilize the fractional marginal likelihood (FML) approach of O'Hagan (1995, 1997) to derive the approximate posterior distribution for the dimension of the RRR parameter structure. Our method provides an analytically tractable default solution, and is applicable without subjective input from the user. Its properties are investigated both theoretically and by applying it to both real and simulated data sets.

The present paper is structured as follows. In the next section we define formally the RRR model and in Section 3 we derive the posterior distribution for the dimensionality. Numerical results are provided in terms of examples with real data sets and simulations in the two sections thereafter, respectively. Some concluding remarks are given in the final section.

2. Reduced rank regression

Consider the following multivariate regression

$$y_i = \Pi x_i + \Gamma z_i + \varepsilon_i, \tag{2.1}$$

where x_i and z_i are q_1 and q_2-dimensional vectors of predictors, respectively, Π and Γ are $p \times q_1$ and $p \times q_2$ coefficient matrices and $\varepsilon_i \overset{iid}{\sim} N_p(0, \Sigma)$ are the model errors, for $i = 1, ..., n$.

The reduced rank regression introduced by Anderson (1951) allows for the possibility that Π has less than full rank, and the rank of Π will be termed the *regression dimensionality*. The other coefficient matrix Γ, could also be rank deficient (Velu, 1991), but we shall here retain the assumption of a full rank for Γ; see also Remark 2 below, however.

A reduced rank r of Π can be modelled explicitly by writing $\Pi = \Psi\Lambda$, where Ψ and Λ are $p \times r$ and $r \times q_1$ matrices, respectively. Inserting this decomposition into (2.1) gives

$$y_i = \Psi\Lambda x_i + \Gamma z_i + \varepsilon_i, \tag{2.2}$$

where Λx_i may now be considered as a new set of r predictors with regression coefficient matrix Ψ. One would expect that Λ is related to the canonical variates between y_i and x_i, corrected for z_i, and that Ψ is related to the corresponding (partial) canonical correlations, and this is indeed so (see Reinsel and Velu, 1998). In the sequel, we let $r_{\max} = \min(p, q_1)$ denote the maximal rank of Π.

As $\Psi U U^{-1} \Lambda = \Psi\Lambda$, for any $r \times r$ non-singular matrix U, the model (2.2) is not identified. In this paper we impose exactly identifying restrictions on Λ, see, *e.g.,* Johansen (1995), for a simple set of possible restrictions. It will be evident that our proposed inference tool is invariant to the choice of restrictions.

3. Posterior distribution of the regression dimensionality

The question of dimensionality in RRR is essentially a model inference problem. Given a class of available models, Bayesian paradigm provides in principle a solution to any model assessment problem, by quantifying the model uncertainty in terms of posterior

probabilities (see Bernardo and Smith, 1994), and has been successfully applied to a wide range of problems; for a general review, see Robert and Casella (1999). More specifically, let $M_1, ..., M_h \in \mathcal{M}$ denote the models under comparison, x the available data, $L_i(x|\theta_i)$ and $\pi_i(\theta_i)$ the likelihood and prior given M_i, respectively. The posterior distribution over $\mathcal{M}$ is then given by

$$p(M_i|x) \propto m(M_i)p(M_i)$$

where $m(M_i) = \int L_i(x|\theta_i)\pi_i(\theta_i)d\theta_i$ is the marginal likelihood of model M_i and $p(M_i)$ is the prior probability of the ith model.

However, the intrinsic complexity of the parameter structure in general RRR models brings two impediments to a successful practical implementation of the Bayes solution. First, Bayesian inference about dimensionality is a very laborious computational exercise, even in the presence of modern computing facilities (see Geweke, 1996 and Kleibergen and Paap, forthcoming). Secondly, the reduced rank restriction makes the elicitation of prior distributions for the model parameters far from straight-forward.

The fractional Bayes approach was proposed by O'Hagan (1995, 1997) as a default method to handle Bayesian model inference based on improper, non-subjective, priors. Fractional Bayes approximates the usual marginal likelihood with the marginal likelihood conditional on a 'idealized' training sample. The resulting quantity is called the *fractional marginal likelihood* (FML) and is defined as

$$m_b(M_i) = \frac{\int L_i(x|\theta_i)\pi_i(\theta_i)d\theta_i}{\int L_i(x|\theta_i)^b\pi_i(\theta_i)d\theta_i} \tag{3.1}$$

where $0 < b < 1$ is the fraction of the data used implicitly to 'train' the improper prior $\pi_i(\theta_i)$ into a proper distribution. It will be assumed here that b is *minimal*, *i.e.* $b = m/n$, where m is the smallest possible number of observations necessary to convert $\pi_i(\theta_i)$ into a proper distribution for the *largest* model in $\mathcal{M}$. For further details on the fractional Bayes approach, we refer to O'Hagan (1995, 1997).

Conditional on Λ, the RRR model is a regular full rank multivariate regression. This proves to be a good starting point as the FML of a full rank multivariate regression, stated in the next lemma, is a simple expression. In the following lemma, let

$$S_{xz} = n^{-1}\sum_{i=1}^{n} x_i z_i'.$$

Lemma 1. *The fractional marginal likelihood of the multivariate regression $y_i = \Pi x_i + \varepsilon_i$, where y_i is a p-dimensional response vector, x_i a d-dimensional vector of predictors, given the prior $\pi(\cdot) \propto |\Sigma|^{-(p+1)/2}$, is*

$$m_b(d) = \frac{\Gamma_p(n-d)}{\Gamma_p(m-d)}\left|\hat{\Sigma}\right|^{-(n-m)/2} \tag{3.2}$$

where $\Gamma_p(a) = \Pi_{i=1}^{p}\Gamma[(a-i+1)/2]$, for integer $a \geq p$, $\hat{\Sigma} = S_{yy} - S_{yx}S_{xx}^{-1}S_{xy}$ is the ML estimate of Σ and $m = p + d$ is the size of the minimal sample. A constant which does not depend on x_i has been discarded in (3.2).

Proof. This lemma is a trivial extension of Theorem 3.1 in Villani (2001). □

Inference about lag length in multivariate autoregressive processes based on the FML was in Villani (2001) shown to yield consistently more reliable results than the various asymptotic default criteria proposed in the statistical literature.

If Λ in (2.2) was known, the model would be an ordinary multivariate regression with predictors Λx_i and z_i, and consequently, Lemma 1 could be applied directly to compute the FML. Since Λ is generally unknown, a naive approach would be to replace Λ by an estimate $\hat{\Lambda}$, and use Lemma 1 with $d = q_2 + r$. This entirely ignores the uncertainty in Λ is therefore easily seen to favor smaller r.

The integrals in (3.1) with respect to Λ seem intractable for any imaginable prior. Therefore, we suggest an approximate FML of rank r, using (3.2) with a correction for the degrees of freedom lost in the estimation of Λ. Notice that d in (3.2) equals the number of parameters in each response equation. The number of parameters in each equation of the reduced rank regression depends on the arbitrarily chosen identifying restrictions and cannot thus be directly used as a correction term. However, as the number of free parameters in Λ is $(q_1 - r)r$, the average number of parameters in each equation of (2.2) is $c_r = q_2 + r + (q_1 - r)(r/p)$, which is well defined and a natural proposal as the correction term. Note that the extra penalty for the uncertainty in Λ compared to the naive approach is $(q_1 - r)(r/p)$. This quantity is always positive for $r = 1, ..., r_{\max} - 1$ and zero for $r = 0$ and $r = r_{\max}$, which is sensible as the models with $r = 0$ and $r = r_{\max}$ are full rank multivariate regressions, and Lemma 1 holds exactly.

By applying established results in multivariate theory (*e.g.* Anderson, 1984), it can be shown that $\left|\hat{\Sigma}\right| = \left|S_{yy} - S_{yz}S_{zz}^{-1}S_{zy}\right| \prod_{i=1}^{r}(1 - \rho_i^2)$, where ρ_i is the ith largest partial canonical correlation between y and x, corrected for z, and the first factor is independent of r. Using the above, make the following definition.

Definition 1. The approximate fractional marginal likelihood of rank r is proportional to

$$\frac{\Gamma_p(n - c_r)}{\Gamma_p(m - c_r)} \prod_{j=1}^{r}(1 - \rho_j^2)^{-(n-m)/2}, \tag{3.3}$$

where ρ_j is the jth largest partial canonical correlation between y_i and x_i, $c_r = q_2 + r + (q_1 - r)(r/p)$, $m = p + q_1 + q_2$.

Remark 1. The approximate FML of the rank generalizes immediately to a joint assessment criterion for r and the set of relevant predictors in x and z. If z varies across the models under consideration, the factor $\left|S_{yy} - S_{yz}S_{zz}^{-1}S_{zy}\right|$ in the expression for $\left|\hat{\Sigma}\right|$ must be retained as it depends on z. The aspects of joint assessment will be illustrated in the numerical examples.

Remark 2. Several generalizations of the RRR model in (2.2) have been presented in the literature, see Reinsel and Velu (1998) for the relevant references. Velu (1991), for example, allows also Γ to be rank deficient. The way we have defined the approximate FML in Definition 1 makes it straightforward to cover such extensions. The fact that dimensionality inference based on likelihood ratio tests is intractable in many of the extended models makes this point especially important.

The asymptotic properties of (3.3) are established in the following theorem.

Theorem 1. *The posterior mode estimator of the regression dimensionality based on the approximate FML in Definition 1 is consistent. Furthermore, (3.3) is an $O_p(1)$ approximation of the log marginal likelihood if and only if $c_r = q_2 + r + (q_1 - r)(r/p)$, otherwise it is an $O_p(\log n)$ approximation.*

Proof. In general, the marginal likelihood of a model M_i with likelihood $L_i(\cdot|\theta_i)$, where k_i is the number of free parameters in θ_i, can be written (Gelfand and Dey, 1994)

$$\log m(M_i) = \log L_i(\cdot|\hat{\theta}_i) - \frac{k_i \log n}{2} + O_p(1), \tag{3.4}$$

where $\hat{\theta}_i$ is the maximum likelihood estimator of θ_i. A straightforward modification of Theorem 3.2 in Villani (2001) gives

$$\log m_b(c_r) = -\frac{n}{2}\sum_{i=1}^{r}(1-\rho_i^2) - \frac{c_r p \log n}{2} + O_p(1). \tag{3.5}$$

As $-\frac{n}{2}\sum_{i=1}^{r}(1-\rho_i^2)$ is proportional to the maximized log likelihood and $c_r p = pq_2 + pr + (q_1 - r)r$ is the number of free parameters in the rank r model, standard Bayesian asymptotics (Gelfand and Dey, 1994) show the consistency. The proof of the last statement of the theorem is immediate by comparing (3.4) and (3.5). □

Theorem 1 deserves two comments. First, the leading term of $\log m(M_i)$ is of order n so that the approximation error of order 1 diminishes in importance as the sample size increases. Secondly, the Schwarz (1978) approximation is also $O_p(1)$, but with a general tendency to support too small models (Kass and Raftery, 1995). Our approximation seems to improve on the one in Schwarz (1978), see Section 5.

4. Numerical examples with real data

Example 1. *Assessing dimensionality in growth curve analysis.*

Growth curve analysis, or GMANOVA, was introduced by Potthoff and Roy (1964) as a framework for analyzing repeated measurements on a response variable over time. The basic model is of the form

$$y_i = \Psi \Lambda x_i + \varepsilon_i, \tag{4.1}$$

where x_i is a set of time-invariant predictor variables. The model in (4.1) is exactly the RRR model without z-predictors. In the original setting of Porthoff and Roy (1964), Ψ was assumed to be a *known* matrix, completely specified in terms of a parametric function of time, *e.g.* a polynomial of degree $r-1$, which determines the general shape of the mean response profiles over time. Here, Λx_i are the coefficients of the parametric function for the ith individual. A natural extension of the model is to let Ψ be any unknown matrix and to infer the dimensionality of Ψ from data as in the RRR model (Reinsel and Velu, forthcoming). As Reinsel and Velu point out, the columns of Ψ then represent unknown basis functions for the mean response profiles over time.

To illustrate dimensionality determination in the context of growth curve analysis, we use a bioassay data set from Volund (1980), subsequently analyzed in Reinsel and Velu (forthcoming). The response comprises measurements of blood sugar concentration on $n = 36$ rabbits at $1, 2, 3, 4$ and 5 hours after the administration of an insulin dose; the response is thus 5-dimensional. The 36 rabbits were divided in four balanced groups according to a 2×2 design with factors *insulin type* and *dose level*. Four predictors were used: Indicators for insulin type and dose level, initial blood sugar concentration and an interaction variable between dose level and initial concentration. A constant term was also added to the model.

In this and all other examples we use a uniform prior for the regression dimensionality. The other candidate default methods are those widely used in the statistical

literature, AIC (Akaike, 1974), SBC (Schwarz, 1978) and the sequential likelihood ratio test (Anderson, 1951).

Using definition 1, the posterior distribution of the rank is $p(r|y,x) = .000,\ .117,\ .808,\ .073,\ .002$ and .001, which clearly indicates that the most likely value of r on the basis of data is 2, although there is a non-negligible degree of uncertainty related to the choice of the rank. As a comparison, the p-values for asymptotic likelihood ratio tests of $H_0 : r = 0, 1, ..., r_{\max} - 1$ against $H_1 : r = p$ are $.000, .022, 0.473, 0.919$ and 0.680, respectively. Thus, 1% and 5% significance levels lead to choices $r = 1$ and $r = 2$, respectively. The other default criteria choose $r = 1$ (SBC) and $r = 2$ (AIC). Using SBC to approximate the posterior distribution gives $p_{SBC}(r|y,x) = .000,\ .587,\ .408,\ .005,\ .000$ and .000. As expected, the SBC approximation puts more probability mass on smaller r.

Reinsel and Velu (forthcoming) also raise the question of whether any of the predictor variables is redundant for the response. In particular, there was reasons for believing in the absence of an insulin type effect *a priori*. Let A_j denote the event that the jth predictor in x is redundant. The answer to whether or not a variable is redundant is given by the posterior probability of this event

$$p(A_j|y,x) = \sum_{r=1}^{r_{\max}} p(A_j|r,y,x)p(r|,y,x),$$

where $p(A_j|r,y,x)$ is obtained from the posterior distribution over the models with different subsets of predictors conditional on rank r and $p(r|y,x)$ is the posterior distribution of the rank stated above. In the blood sugar data, $p(A_j|y,x) = .000,\ .444,\ .023,\ .003,\ .152$ for $j = 1, 2, ..., 5$ (first predictor is the constant). Thus, the predictors *insulin type effect* and *interaction between dose level and initial concentration* have non-negligible redundancy probability, whereas *dose level* and *initial concentration* seem to be important predictors, which is in accordance with the results of Reinsel and Velu (forthcoming).

Example 2. *Assessing dimensionality for cointegrated time series.*

Next we consider the case of the cointegrated vector autoregressive process

$$\Delta w_t = \Psi \Lambda w_{t-1} + \sum_{i=1}^{k} \Gamma_i \Delta w_{t-i} + \Phi d_t + \varepsilon_t, \qquad t = 1, ..., n, \tag{4.2}$$

where w_t is an observation on a p-dimensional process at time t, Δ is the time difference operator, d_t is a vector of deterministic trends and $\varepsilon_t \overset{iid}{\sim} N(0, \Sigma)$. The r rows of Λ are the cointegration vectors which determine r stationary linear combinations between the otherwise non-stationary components of w_t. Here, Ψ is a $p \times r$ matrix of adjustment coefficients. The model in (4.2) can be put in RRR form by defining $y_t = \Delta w_t$, $x_t = w_{t-1}$, $z_t = (w'_{t-1}, ..., w'_{t-k}, d'_t)'$ and $\Gamma = (\Gamma_1, ..., \Gamma_k, \Phi)$.

It has been customary to consider five different forms for the deterministic trends in d_t, depending on the desired trending behavior of both the original process and the cointegration relations, Λw_{t-1}. The first trend model has a quadratic trend in the original series and linear trend in the cointegration relations, whereas the second trend model has a linear trend in both the original series and the cointegration relations and so on to the fifth and last model where both processes has neither constant nor trend.

The different trend types are obtained by adding either 1 or t, or both, to w_{t-1} and d_t, see Johansen (1995) for a detailed description.

The dimensionality determination in cointegrated processes thus involves a wider variety of quantities: the number of lags in the process ($k = 0, 1, ...$) the number of cointegration relationships ($r = 0, 1, ..., r_{\max}$) and the choice of trend type ($s = 1, 2, ..., 5$). As an illustration we use the five variable data from Johansen (1995), which consists of log prices in Australia and US, the exchange rate and a bond rate from each country. Data are in quarterly observations during the period 1972:1 to 1991:1.

To keep the range of different models at a reasonable level, we restrict the number of lags to be no larger than four. If this upper limit should prove to be too small, the posterior mass would be concentrated on the largest values, and the model class could thereby be extended after the initial analysis. The (k, r, s)-combinations with highest posterior probabilities are listed in Table 1.

TABLE 1
THE MOST PROBABLE (k, r, s) COMBINATIONS IN THE JOHANSEN DATA.

k	1	1	1	1	1	1	1	1	1	1
r	3	4	5	5	4	3	2	3	4	4
s	1	2	2	1	1	2	1	5	5	4
$p(k, r, s\|y, x, z)$	.156	.154	.147	.147	.115	.090	.040	.018	.017	.017

First, we notice that the posterior of the number of lags in the process is concentrated on $k = 1$, suggesting that there is no need to enlarge the initial class of models in this respect. It is also clear that the uncertainty about the cointegration rank is prominent. For comparison, the asymptotic criteria choose the following models: SBC ($k = 0$, $r = 2$, $s = 5$), AIC ($k = 1$, $r = 5$, $s = 2$), which further illustrates the uncertainty related to the choice of a model. Guided by sequential likelihood testing and residual diagnostics, Johansen (1995) decided to use $(k = 1, r = 2, s = 3)$; the test of $r = 2$ against $r = 3$ was almost significant on the 5% level, making $r = 3$ a possible alternative for the number of cointegration relations. See Johansen (1995) for parameter estimates conditional on $(k = 1, r = 2, s = 3)$ and further analyses.

5. Simulation results

A small scale simulation study was conducted to learn more about the properties of the approximate FML in Definition 1. To be able to compare our results with methods used in typical applications, we confine our investigation to point estimates of the dimensionality. The posterior mode estimator is chosen as the Bayesian estimator.

First, we consider models where $p = q_1$, $q_2 = 0$ and $r = 0$ or 1, given as

$$y_i = \Pi x_i + \varepsilon_i,$$

where $x_i \overset{iid}{\sim} N_{q_1}(0, I_{q_1})$. The reduced rank of Π is made explicit by the singular value decomposition of Π,

$$\Pi = u\lambda v',$$

where u and v are p-dimensional vectors of unit length and $\lambda \geq 0$ is the single non-zero singular value of Π, *i.e.* $\text{rank}(\Pi) \leq 1$. This parametric setup enables us to systematically investigate the behavior of model assessment criteria as a function of the informativeness

INSERT FIGURE HERE

FIGURE 1. Relative frequency of choice $r = 1$ in 10.000 simulations as a function of λ for three samples sizes, $n = 25, 50, 100$. Graphs in the first row correspond to the model with $p = 2$ and graphs in the second row to $p = 4$. FML $(- \bullet -)$, SBC $(-\Box-)$, LRT $\alpha = .01$ (—), LRT $\alpha = .05$ $(\cdots)$ and AIC (- - -).

INSERT FIGURE HERE

FIGURE 2. Relative frequency of choice $r = 2$ in 5.000 simulations as a function of the sample size. FML $(- \bullet -)$, SBC $(-\Box-)$, LRT $\alpha = .01$ (—), LRT $\alpha = .05$ $(\cdots)$ and AIC (- - -).

of the sample data, which increases with both λ and n. To make the results less dependent of specific choices of parameter values, we let u and v be uniformly distributed on the unit p-sphere (see *e.g.* Muirhead, 1982, p. 38) and let the error covariance Σ have an outspread Wishart distribution $W_p(I_p, p)$.

Figure 1 below shows the relative frequencies of choosing $r = 1$ in 10.000 simulated configurations as a function of λ. Each subgraph corresponds to a (n, p)-pair as indicated in the figure. A striking feature of Figure 1 is the bad performance of AIC, especially when data are relatively informative. The posterior mode estimator compares favorably to the other estimators in general and to SBC in particular. This latter point is important as SBC is the only competing approximation of the posterior distribution. The largest difference between posterior mode estimator and other methods appears for $p = 4$ and $n = 25$. The sample size $n = 25$ should in no way be regarded as unimportant from an applied viewpoint, it is probably a fair representation of the uncertainty in many real world applications where the sample size is substantially larger but the data are much less tidy than those resulting from our generating model.

To illustrate the rank assessment in more detail for a specific parameter setting, as a function of n, we consider an RRR model with $p = 3, q_1 = 4, q_2 = 0$ and $r = 2$, specified by

$$y_i = \begin{pmatrix} 1 & 0 \\ .5 & 1 \\ .5 & 2 \end{pmatrix} \begin{pmatrix} 1 & 0 & -1 & -1 \\ 0 & 1 & 0 & -1 \end{pmatrix} + \varepsilon_i$$

where the entries of x_i are independent standard normal variates and the error terms follow the $N(0, \Sigma)$ distribution with

$$\Sigma = \begin{pmatrix} 3 & \cdot & \cdot \\ 2 & 3 & \cdot \\ .5 & 1 & 1 \end{pmatrix}$$

Here we adopt a similar strategy as O'Hagan (1995, p.108). Data sets of size 200 observations were generated from the model and the rank was determined using $n = 20, ..., 200$ first observations. The rank assessment process was replicated 5.000 times for each n, to estimate the (relative frequency) probabilities of different rank choices. The relative frequency of choosing the true rank for $n = 20, ..., 70$ is shown in Figure 2. This range of sample sizes captures the essential differences between the methods, while keeping details in the plot distinguishable.

We see here that, for small sample sizes, the sequential likelihood ratio (LR) tests do not choose the right dimensionality very often and as n increases, their error probabilities

approach the chosen significance level α. AIC yields slightly higher probabilities of choosing the true rank than the posterior mode estimator for the smallest sample sizes, but it also overestimates r with relatively high probability as n grows. For instance, AIC gives probability .155 to the choice $r = 3$, even for $n = 200$ (not shown in the figure). The relative frequency of $r = 2$ for the posterior mode estimator and SBC was very close to one for $n = 200$, thus confirming their consistency.

6. Concluding remarks

We have presented a default Bayesian approach to dimensionality assessment in reduced rank regression, *i.e.* with a minimum of subjective elicitation of prior opinions on the part of the user.

The proposed approximate Bayes approach gave reasonable results in two real data sets and performed well in a simulation study. In particular it seems to offer an improvement of the well known approximate Bayes solution derived by Schwarz (1978). It was also demonstrated how the method can be extended to cover simultaneous inferences of the regression dimensionality and other features of the model, such as the set of relevant predictors.

Finally, we want to stress the importance of a proper quantification of the uncertainty in the rank inference. A Bayesian approach, approximate or not, delivers, through the posterior distribution over the regression dimensionality, exactly the relevant information. This is in sharp contrast to non-Bayesian approaches which at best offer a rough and hard-to-interpret description of the uncertainty in a form which is often unsuitable for further analyses, *e.g.* prediction exercises or parametric hypothesis testing.

References

Ahn, S. K. and Reinsel, G. C. (1990). Estimation for partially nonstationary multivariate autoregressive models, *J. Amer. Stat. Assoc.* **85**, 813-23.

Akaike, H. (1974). A new look at the statistical model identification, *IEEE Trans. Automatic Control*, **19**, 716-727.

Anderson, T. W. (1951). Estimating linear restriction on regression coefficients for multivariate normal distributions, *Ann. Stat. Math. Stat.*, **22**, 327-51.

Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*, New York: Wiley.

Anderson, T. W. (1994). Inference in linear models, in *Multivariate Analysis and Its Applications* (Anderson, T. W., Fang, K. T. and Olkin, L, ed.), IMS Lecture Notes - Monograph Series, **24**, 1-20.

Bernardo, J. M. and Smith, A. F. M. (1994). *Bayesian theory*. Chichester: Wiley.

Brown, P. J., Fearn, T. and Vannucci, M. (1999). The choice of variables in multivariate regression: A non-conjugate Bayesian decision theory approach. *Biometrika*, **86**, 635-648.

Chao, J. C. and Phillips, P. C. B. (1999). Model selection in partially nonstationary vector autoregressive processes with reduced rank structure, *Jo. Econometrics*, **91**, 227-271.

Gelfand, A. E. and Dey, D. K. (1994). Bayesian model choice: Asymptotics and exact calculations. *J. Roy. Statist. Soc.*, **B 56**, 501-514.

George, E. I. and Foster, D. P. (2000). Calibration and empirical Bayes variable selection. *Biometrika*, **87**, 731-747.

Geweke, J. (1996). Bayesian reduced rank regression in econometrics, *J. Econometrics*, **75**, 121-146.

Johansen , S. (1995). *Likelihood-based Inference in Cointegrated Vector Autoregressive Models*, Oxford University Press, New York.

Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *J. Amer. Stat. Assoc.* **90**, 773-795.

Kleibergen, F. and Paap, R. (forthcoming). Priors, posteriors and Bayes factors for a Bayesian analysis of cointegration, *J. Econometrics*.

Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory*. New York: Wiley.

O'Hagan, A. (1995). Fractional Bayes factors for model comparisons. *J. Roy. Statist. Soc.* **B 57**, 99-138.

O'Hagan, A. (1997). Properties of intrinsic and fractional Bayes factors, *Test*, **6**, 101-118.

Potthoff, R. F. and Roy, S. N. (1964). A generalized multivariate analysis of variance model useful especially for growth curve problems, *Biometrika*, **51**, 313-26.

Reinsel, G. C. and Velu, R. P. (1998). *Multivariate reduced-rank regression: Theory and applications*. New York: Springer-Verlag.

Reinsel, G. C. and Velu, R. P. (forthcoming). Reduced-rank growth curve models, *J. Statist. Plann. Inference*.

Robert, C. P. and Casella, G. (1999). *Monte Carlo statistical methods*. New York: Springer-Verlag.

Schwarz, G. (1978). Estimating the dimension of a model. *Ann. Stat.*, **6**, 461-464.

Velu, R. P. (1991). Reduced rank models with two sets of regressors, *Appl. Statist.*, **40**, 159-170.

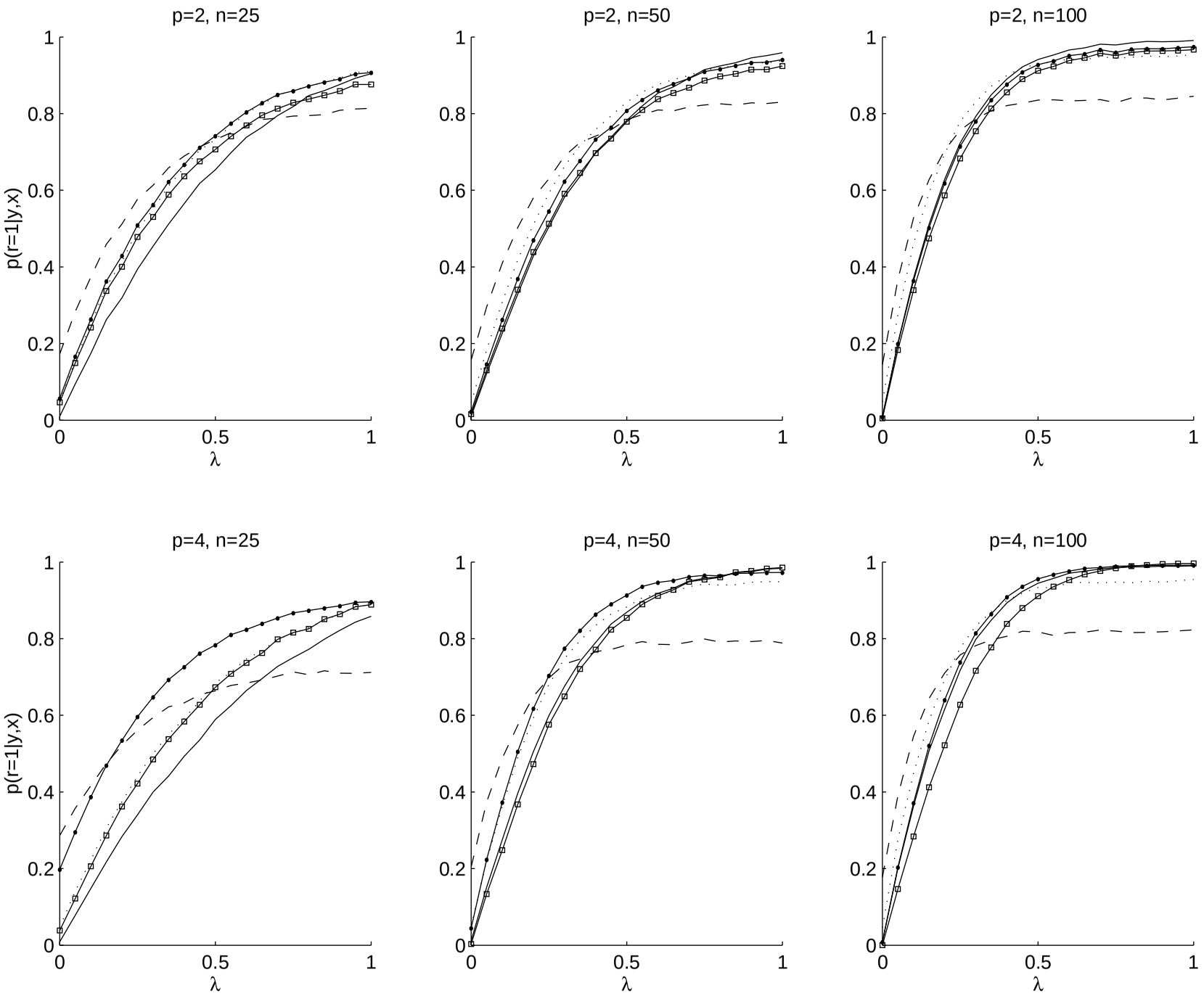
Velu, R. P., Reinsel, G. C. and Wichern, D. W. (1986). Reduced rank models for multiple time series, *Biometrika*, **73**, 105-118.

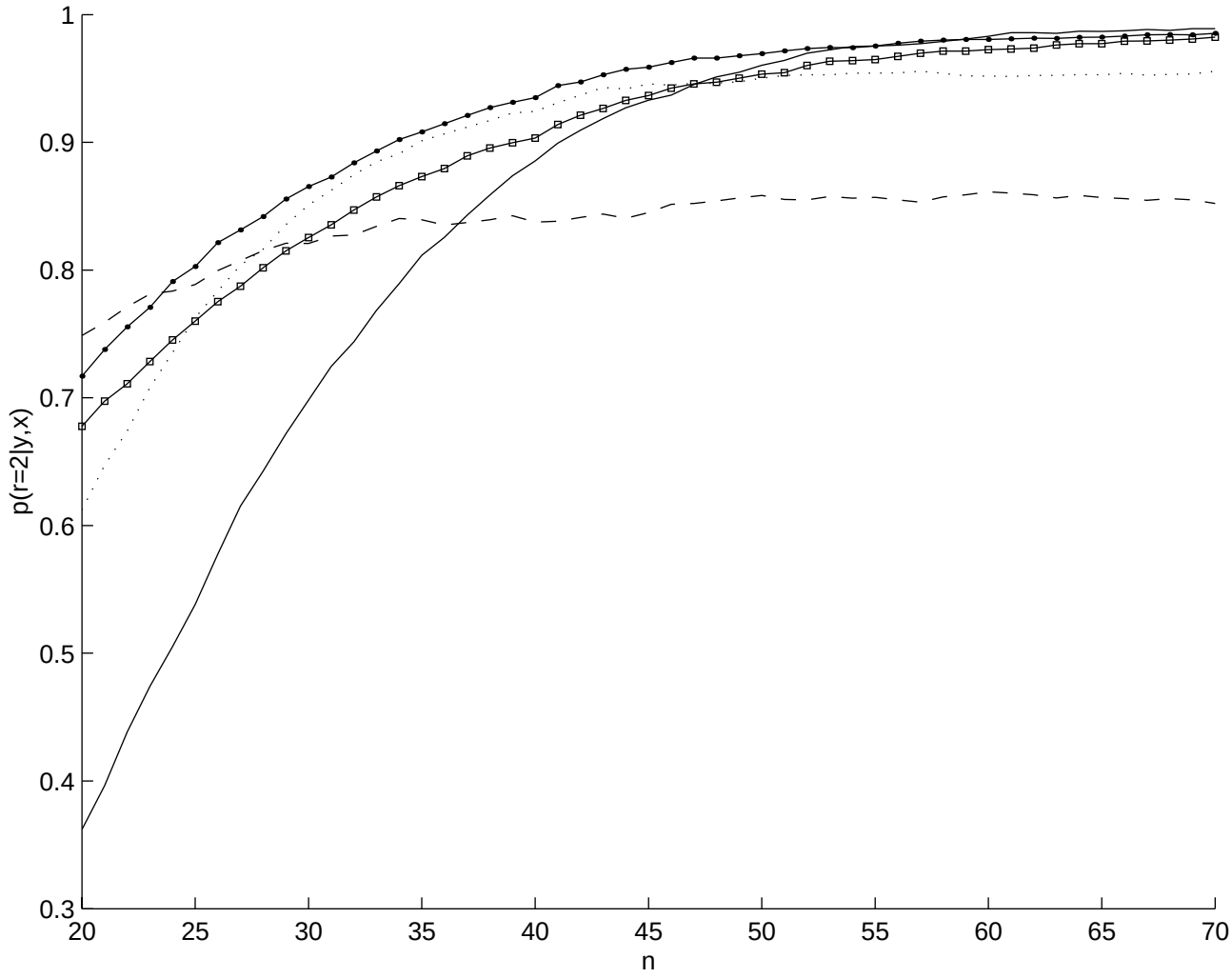
Villani, M. (2001). Fractional Bayesian lag length inference in multivariate autoregressive processes, *J. Time Ser. Anal.*, **22**, 67-86.

Volund, A. (1980). Multivariate bioassay, *Biometrics*, **36**, 225-236.

Rolf Nevanlinna Institute, P.O. Box 4, Fin-00014, Finland
E-mail address: jukka.corander@rni.helsinki.fi

Department of Statistics, Stockholm University, S-106 91 Stockholm, Sweden
E-mail address: mattias.villani@stat.su.se


p=2, n=25
p=2, n=50
p=2, n=100
p=4, n=25
p=4, n=50
p=4, n=100
p(r=1|y,x)
λ
0
0.2
0.4
0.6
0.8
1
0.5


1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
p(r=2|y,x)
20
25
30
35
40
45
50
55
60
65
70
n