

# Bayesian Unit Root Tests and Marginal Likelihood

Aart F. de Vos and Marc K. Francke\*

VU University Amsterdam,

Department of Econometrics and Operation Researchs,

De Boelelaan 1105, NL-1081 HV Amsterdam

avos@feweb.vu.nl and mfrancke@feweb.vu.nl

February 1, 2008

---

\*Corresponding author

# Bayesian Unit Root Tests and Marginal Likelihood

## Abstract

Unit root tests based on classical marginal likelihood are practically uniformly most powerful (Francke and de Vos, 2007). Bayesian unit root tests can be constructed that are very similar, however in the Bayesian analysis the classical size is determined by prior considerations. A fundamental difference remains the link between the implied size and the number of observations.

To establish this correspondence, we get two intermediate results that may be important in a wider context. We prove that for inference on the covariance parameters in the general linear model classical and Bayesian versions of marginal likelihood are equivalent if Jeffreys' independence priors are used. Further we show equivalence between classical and Bayesian tests under some monotonicity conditions.

*Key words:* Hypothesis testing; Jeffreys' rule; Noninformative priors;

*JEL Classification:* C11, C12, C13, and C22.

# 1 INTRODUCTION

This article compares classical and Bayesian unit root testing in the regression model with first order autoregressive disturbances [AR(1)], where the location and scale parameters are considered as nuisance parameters. It is the Bayesian sequel of Francke and de Vos (2007), henceforth FV, where classical marginal likelihood ratio based tests are derived. We give a Bayesian analysis of the unit root problem inspired by a successful frequentist result in a context where existent Bayesian solutions are heavily disputed among Bayesians for various reasons.

The unit root tests based on marginal likelihood are almost uniformly most powerful invariant (UMPI), even in small samples. We show in this paper that there are Bayesian tests with the same property for any combination of priors (for the autoregressive parameter  $\rho$  that may be unity) and loss functions. Invariance appears to have a close connection with Jeffreys' independence priors for place and scale and the treatment of the initial observation. Problems in the Bayesian analysis of the unit root problem vanish in this setup. The UMPI property corresponds to a situation where classical and Bayesian tests coincide, apart from the fact that the classical (default 5%) size follows in the Bayesian analysis from priors and loss functions.

The results in FV rests on three pillars: the use of the marginal likelihood, proper model specification, and monotonicity of the marginal likelihood ratio (MLR) function.

In this article we show that the same three pillars are the basis of equivalent Bayesian results. Equivalent in the sense that they are practically the same tests, different however in the way the relevant size is determined.

The first pillar is the marginal likelihood. This term is used in both classical and Bayesian statistics. The classical concept is based on a transformation of the data to remove nuisance parameters, the Bayesian marginal likelihood is obtained by integrating nuisance parameters out. We show that in the general linear model for a specific noninformative (independence Jeffreys') prior both likelihoods contain the same information on the relevant parameters. This result is relevant for all inference on the covariance matrix in the general linear model. Moreover

we argue that classical marginal likelihood may directly be used in a Bayesian analysis, which may avoid problems by the use of improper marginalized likelihoods.

The second pillar is the the following specification of the regression model with first order autoregressive disturbances,

$$y_t = \mu + x_t' \beta + u_t, t = 1, \dots, n, \quad (1)$$

$$u_t = \rho u_{t-1} + \varepsilon_t, t = 2, \dots, n, \quad (2)$$

$$u_1 \quad \begin{cases} = \xi & \text{for } \rho = 1, \\ \sim N(0, \sigma^2/(1 - \rho^2)) & \text{for } |\rho| < 1, \end{cases} \quad (3)$$

where  $\varepsilon_t \sim N(0, \sigma^2)$  and  $x_t$  is a  $(k - 1) \times 1$  vector. The “parameter”  $\xi$  is the initial condition, central in Müller and Elliott (2003). We kept the same notation here, but it is meant to represent simply the limiting (for  $\rho \uparrow 1$ ) improper distribution of  $u_1$ . The initial condition is supposed to follow from the process described by the model, so if this is nonstationary there is no prior idea about the initial condition. In FV it is shown that the classical marginal likelihood is nonzero and finite in  $\rho = 1$  and continuous for  $\rho \uparrow 1$ . The transformation that is needed is simply taking first differences, and  $\mu$  drops out as well as the initial condition  $\xi$  in case  $\rho = 1$ . We will show that the same result can be obtained from a Bayesian perspective.

The third pillar is that a monotone marginal likelihood ratio leads to (one-sided) tests that are UMPI. We show that monotonicity is also the key to prove equivalence between Bayesian and classical one-sided tests. The Bayesian test procedure can be decomposed in a part that solely depends on the data (the Bayesian  $p$ -value), and a part that follows from prior considerations. The latter determines the size of the test. Once one agrees on the size of the test, Bayesian and classical unit root inference coincide. As in this case the marginal likelihood ratio test is UMPI, the same holds for the posterior odds test.

We show that Bayesian unit root tests are virtually (the MLR is approximately monotone) equal to classical marginal likelihood ratio tests, with a size that is determined by prior considerations and loss functions. We study in detail for the AR(1) model the conditions under

which the two approaches coincide. There remains a fundamental difference however. Suppose a Bayesian and a frequentist agree on a test for  $\alpha = 0.05$  and  $n = 100$ . For the Bayesian this is a size implied by priors and loss functions. Based on the same inputs however, the Bayesian will have a different implied size for other values of  $n$ . This result is specific for the unit root context. The cause is that marginal likelihood ratio has a limiting distribution under the null hypothesis in the local-to-unity format  $\gamma = n(1 - \rho)$ , while priors should logically not be specified in terms of  $\gamma$  but in terms of  $\rho$ .

There is a large amount of literature on Bayesian unit root testing and the differences with the classical approach. The many options in model specification, prior distribution and the treatment of the initial condition make this literature rather complex.

Three special issues are devoted to the many ways unit roots can be handled: the *Journal of Applied Econometrics* (1991), *Econometric Theory* (1994) and the *Journal of Econometrics* (1995). The power of classical unit root tests in small samples was questioned by Sims (1988) from a Bayesian point of view. Phillips (1991b) claims that the difference in results between classical and Bayesian inference is a result of the use of a flat prior for the autoregressive parameter. Sims and Uhlig (1991) have designed an experiment to compare Bayesian and classical inference, although in a model without a constant, and allowing for explosive values of the autoregressive parameters. Lubrano (1995) and Müller and Elliott (2003) show that the choice of the model and the treatment of the initial condition is essential.

Many of these contributions cannot be compared to our analysis. Often this is due to the model specification. The reduced form of (1)–(2) can be expressed as

$$y_t = \rho y_{t-1} + (1 - \rho)\mu + (x_t - \rho x_{t-1})'\beta + \varepsilon_t. \quad (4)$$

Other specifications like  $y_t = \rho y_{t-1} + \mu + x_t'\beta + \varepsilon_t$  lead to problems. Like Bhargava (1986), Phillips (1991a), Schotman and van Dijk (1991), Schmidt and Phillips (1992), and Harvey (2005) we argue that model (1)–(3) is a realistic specification because it has a coherent meaning for the expected level of the process  $(\mu + x_t'\beta)$  with  $|\rho| < 1$ .

Bayesian marginal likelihood has been used for the model (1)–(3), see for example Zellner (1971) and Lubrano (1995). However, due to the fact that the data are informative on  $\mu(1 - \rho)$ , Jeffreys’ rule leads to a prior  $\pi(\mu, \sigma^2 | \rho)$  containing a factor  $(1 - \rho)$  which leads, in combination with a proper prior for  $\rho$ , to a posterior that is zero for  $\rho = 1$ . The use of the independence Jeffreys’ prior removes this problem. An extensive survey of singularities at  $\rho = 1$  for different choices of priors, including the independence Jeffreys’ prior, called “flat” prior, model specifications, and initial conditions is provided by Bauwens, Lubrano, and Richard (1999, ch. 6).

The setup of this article is as follows. Section 2 considers the relation between Bayesian and classical marginal likelihood in the general linear model. In section 3 the correspondence between posterior odds and marginal likelihood ratio tests is treated. Section 4 compares classical and Bayesian inference on unit root tests. The size of the posterior odds test is studied for different priors and varying sample size. Section 5 concludes.

## 2 JEFFREYS’ RULE AND MARGINAL LIKELIHOOD

We consider classical and Bayesian marginal likelihood in the general linear model.

**Definition 1** The general linear model is provided by  $y = X\beta + u$ ,  $u \sim N(0, \sigma^2\Omega)$ , with  $\Omega$  a positive definite  $(n \times n)$  matrix depending on a  $n_\theta$  dimensional vector  $\theta$ , so  $\Omega = \Omega(\theta)$  and  $X$  an  $(n \times k)$  matrix of regressors with rank  $k$ .

We are interested in inference on  $\theta$ , and regard  $\beta$  and  $\sigma^2$  as nuisance parameters.

### 2.1 Classical marginal likelihood

The concept of classical marginal likelihood was introduced by Kalbfleisch and Sprott (1970). For the linear model it is used in the context of unbalanced incomplete block designs by Patterson and Thompson (1971), who refer to it as the likelihood of error contrasts. The use of

the classical marginal likelihood is limited to location and scale parameters and some other applications, which may explain that it remained relatively unknown.

The marginal likelihood is the likelihood of a transformation of the data, and it is independent of the nuisance parameters. The generally applicable transformation is  $y^* = A'y/\sqrt{y'AA'y}$ , leading to

$$L_{M_{\beta,\sigma}}(\theta) := f(y^*|\theta) = \frac{\frac{1}{2}\Gamma(\frac{m}{2})|X'X|^{1/2}}{\pi^{m/2}|X'\Omega^{-1}X|^{1/2}|\Omega|^{1/2}} \left( \frac{y'\Omega^{-1}M_X^\Omega y}{y'M_X y} \right)^{-m/2}, \quad (5)$$

an  $(m - 1)$  dimensional marginal likelihood, independent of  $\beta$  and  $\sigma^2$ , where  $A$  is an  $n \times m$  matrix,  $m = n - k$ ,  $A'X = 0$ ,  $r(A) = m$ ,  $A'A = I_m$ ,  $AA' = M_X^I = M_X$  and  $M_X^\Omega = I - X(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}$ , see King (1980).

The likelihood  $f(y|\theta, \sigma^2, \beta)$  is divided in two parts, the marginal likelihood and its complement,

$$f(y|\theta, \sigma^2, \beta) = f(y^*|\theta)f(B^{*'}y|\theta, \sigma^2, \beta). \quad (6)$$

where  $B^*$  is a regular transformation of  $X$ .

The basic assumption is that the marginal likelihood contains all information relevant for inference on  $\theta$  in absence of knowledge on the nuisance parameters, or equivalently that  $B^{*'}y$ , contains no information on  $\theta$ . As McCullagh and Nelder (1989) put it: “There appears to be no loss of information on  $\theta$  by using  $y^*$  in place of  $y$ , though it is difficult to give a totally satisfactory justification of this claim”. Anyhow, for  $n = k$  it seems obvious: the residuals are zero and can give no information on  $\theta$  (while  $\theta$  occurs in the profile likelihood).

A related argumentation is invariance. King (1980) shows that  $y^*$  is a maximal invariant under the group of transformations

$$y \rightarrow \eta_0 y + X\eta, \quad (7)$$

where  $\eta_0$  is a positive scalar and  $\eta$  is a  $k \times 1$  vector. The “principle of invariance” implies that we can treat the maximal invariant  $y^*$  as the observed random vector and (5) as its density

function, and therefore as a likelihood function for  $\theta$ , see Rahman and King (1997). Invariance principles are neither Bayesian nor classical. Gelman (1996) gives a plea to use them from a Bayesian perspective “at least before any specialized knowledge is added”.

In the absence of formal proofs we call the basis of marginal likelihood the “marginalization axiom”:

**Axiom 1** *Consider the general linear model as provided in definition 1. The  $(n - k - 1)$  dimensional marginal likelihood  $f(y^*|\theta)$  contains all information on  $\theta$  in absence of information on  $\beta$  and  $\sigma^2$ , where  $f(y^*|\theta)$  is provided by Eq. (5).*

In the next section we show that there is a unique Bayesian noninformative prior that is coherent with this axiom.

## 2.2 Bayesian marginal likelihood

In Bayesian inference the term marginal likelihood directly follows from the definitions of probability calculus. In the general linear model where  $\beta$  and  $\sigma^2$  are nuisance parameters, the Bayesian marginal likelihood is provided by

$$f(y|\theta) = \int \int f(y|\theta, \beta, \sigma^2) \pi(\beta, \sigma^2|\theta) d\beta d\sigma^2, \quad (8)$$

where  $\pi(\beta, \sigma^2|\theta)$  is a prior. The posterior of  $\theta$  follows from  $f(\theta|y) \propto f(y|\theta)\pi(\theta)$  : marginal posterior is proportional to marginal prior times marginal likelihood.

If Bayesians would claim the term “marginal likelihood” to prevent confusion, they would have a strong case. However, they should always add to marginal likelihood “for a given prior  $\pi(\beta, \sigma^2|\theta)$ ”. The use of the term marginal likelihood without reference to a prior should be reserved for the case  $\pi(\beta, \sigma^2|\theta)$  is noninformative. Unfortunately this is only possible for degenerate priors in which case  $f(y|\theta)$  is not a proper likelihood. Moreover the definition of noninformative is anything but settled.



## 2.3 Correspondence

The relation between the classical and Bayesian marginal likelihood depends on the specification of the noninformative prior  $\pi(\beta, \sigma^2|\theta)$ . Axiom 2.1, the marginalization axiom, appears to imply a unique choice.

The Bayesian version of the marginalization axiom is that, in absence of prior information on  $(\beta, \sigma^2)$

$$f(\theta|y) = f(\theta|y^*).$$

This condition is fulfilled for any prior  $\pi(\theta)$  if

$$f(y^*|\theta) \propto f(y|\theta),$$

as follows directly from Bayes' rule.

**Proposition 2.1** *In the general linear model defined by definition 1, within the class of conjugate priors, it holds that*

$$\pi_{IJ}(\beta, \sigma^2|\theta) \iff f(y^*|\theta) \propto f(y|\theta),$$

where  $\pi_{IJ}$  is the independence Jeffreys's prior given by

$$\pi_{IJ}(\beta, \sigma^2|\theta) = \pi(\beta, \sigma^2) \propto \sigma^{-2}. \quad (9)$$

**Proof** “ $\Rightarrow$ ”: straightforward.

“ $\Leftarrow$ ”: the class of conjugate priors is given by  $\pi_C(\beta, \sigma^2|\theta) \sim NIG(a(\theta), d(\theta), m(\theta), V(\theta))$ , so

$$\pi_C(\beta, \sigma^2|\theta) \propto \sigma^{-(d(\theta)+k+2)} \exp \left\{ -\frac{(\beta - m(\theta))' V(\theta)^{-1} (\beta - m(\theta)) + a(\theta)}{2\sigma^2} \right\}. \quad (10)$$

The Bayesian marginal likelihood follows from (8) and (10)

$$\begin{aligned}
f(y|\theta) &= \int \int f(y|\beta, \sigma^2, \theta) \pi_C(\beta, \sigma^2|\theta) d\beta d\sigma^2 \\
&\propto \int \int \sigma^{-(n+d(\theta)+k+2)} |\Omega|^{-1/2} \exp \left\{ -\frac{1}{2\sigma^2} (y - X\beta)' \Omega^{-1} (y - X\beta) \right\} \\
&\quad \times \exp \left\{ -\frac{1}{2\sigma^2} [(\beta - m(\theta))' V(\theta)^{-1} (\beta - m(\theta)) + a(\theta)] \right\} d\beta d\sigma^2 \\
&\propto |V^*|^{1/2} |\Omega|^{-1/2} (y' \Omega^{-1} y - m^{*'} (V^*)^{-1} m^* + a(\theta) + m(\theta) V(\theta)^{-1} m(\theta))^{-\frac{n+d(\theta)}{2}},
\end{aligned}$$

where  $V^* = (X' \Omega^{-1} X + V(\theta)^{-1})^{-1}$  and  $m^* = V^* (X' \Omega^{-1} y + V(\theta)^{-1} m(\theta))$ . It follows that  $f(y|\theta) \propto f(y^*|\theta)$  only for  $a(\theta) = 0$ ,  $d(\theta) = -k$ ,  $V(\theta)^{-1} = 0$  (zero prior precision). ■

The proposition provides a direct way to avoid improper priors for the nuisance parameters  $\beta$  and  $\sigma^2$  in a Bayesian analysis, and no problems with a degenerate likelihood arise. One simply uses the likelihood of the transformed data  $y^*$ .

We refer to (9) as the independence Jeffreys' prior, because it implies that

$$\pi(\beta, \sigma^2, \theta) \propto \pi(\beta) \pi(\sigma^2) \pi(\theta) \propto \sigma^{-2} \pi(\theta). \quad (11)$$

It differs from the prior following from Jeffreys' rule, that says that the prior is proportional to the square root of the determinant of the Fisher information matrix associated with the likelihood function of the model. This rule is usually applied in univariate cases and its application in multivariate cases sometimes yields unwanted results. The strict use of Jeffreys' rule in the simple linear regression model without covariance structure would lead to a prior  $\pi(\beta, \sigma^2) \propto \sigma^{-(k+2)}$ . As this has implausible consequences, Jeffreys assumed a priori independence between  $\beta|\sigma^2$  and  $\sigma^2$  to obtain  $\pi(\beta, \sigma^2) \propto \sigma^{-2}$ , according to Bernardo and Smith (1994, p. 361) an *ad hoc* recommendation (italics from Bernardo and Smith).

The application of Jeffreys' rule separately to  $\sigma^2$  and  $(\beta, \theta)$ , may also lead to problems. Bauwens, Lubrano, and Richard (1999) show for the AR(1) model (1)–(3) that the posterior for  $\rho$  is zero when Jeffreys' rule is applied separately to  $\sigma^2$  and  $(\mu, \beta, \rho)$ . This problem is solved

when the independence Jeffreys' prior (9) is used, as will be shown in section 4.1.

Reference analysis, see Berger and Bernardo (1992), does not lead to unambiguous results. Fernández and Steel (1999) show that for inference on scale and location parameters the reference prior equals  $\pi(\beta, \sigma^2) \propto \sigma^{-2}$ . However, when  $\theta$  is included, reference analysis becomes complicated and ambiguous. We discuss this in section 4.2 for the AR(1) model.

A point of attention is that Eq. (9) refers to priors that are meant to be noninformative with respect to inference on  $\theta$ . It should not automatically be used for other inference, as the purpose of inference may matter for the choice of reference priors, see Stone and Dawid (1972).

It may have some advantages, even for Bayesians, to use the classical marginal likelihood (5) including the proportionality constants  $|X'X|^{1/2}$  and  $(y'M_X y)^{m/2}$ , because it is a well defined probability function and avoids errors that are easily made using improper prior distributions. The conclusion is that classical marginal likelihood can be used for inference on  $\theta$  in an “objective” Bayesian framework.

### 3 BAYES FACTORS AND CLASSICAL TESTS

The proportionality of the Bayesian and classical marginal likelihood in the general linear model appears to be very useful when comparing Bayesian and classical hypothesis testing. The parameter of interest is the scalar  $\theta$ , and the parameters  $\beta$  and  $\sigma^2$  are nuisance. We only consider the one-sided test  $H_0 : \theta = \theta_0$ , against the alternative  $H_1 : \theta > \theta_0$  (or  $\theta < \theta_0$ ). This is the setting for the unit root tests.

In general there are important differences between the Bayesian and classical hypothesis testing approaches. In contrast to the classical approach, Bayesian hypothesis testing procedures requires a prior distribution on  $\theta$ , and a loss function. In this one-sided test context, we have to use a mixed prior discrete - continuous probability distribution for  $\theta$ . We restrict ourselves to the use of proper priors for  $\theta$ , thus avoiding Bartlett's (1957) paradox.

Despite the major differences between the two approaches, we will show in this section that in case of a marginal likelihood depending on only one parameter, and a monotone marginal

likelihood ratio, the only difference between the Bayesian and classical approach is the size of the test.

### 3.1 Classical hypothesis testing

Classical hypothesis testing can be based on the likelihood ratio. The likelihood ratio contains the nuisance parameters  $\beta$  and  $\sigma^2$ . In section 2 it is shown that the marginal likelihood  $L_{M_{\beta,\sigma}}(\theta)$  can be treated as the likelihood function of  $\theta$ . As the marginal likelihood contains only one parameter and the marginal likelihood ratio (MLR) is monotone in some statistic  $S$ , the test that rejects  $H_0$  if  $S(y) > \kappa_0^*$ , is uniformly most powerful invariant (UMPI), see Lehmann (1986). In general this is the case if  $S$  is a sufficient statistic. The marginal likelihood ratio test evaluated in the maximum likelihood estimator of  $\theta$  is also UMPI, because  $\text{MLR}(\hat{\theta}_{\text{ML}})$  is a function of  $S$ . The MLR test has the format: reject  $H_0$  if

$$\text{MLR}(\hat{\theta}_{\text{ML}}) = \frac{f(y^*|\hat{\theta}_{\text{ML}})}{f(y^*|\theta_0)} > \kappa_0, \quad (12)$$

where  $\kappa_0$  is chosen such that

$$P_{y^*|H_0}(\text{MLR}(\hat{\theta}_{\text{ML}}) > \kappa_0) = \alpha, \quad (13)$$

and  $\alpha$  is the predetermined size of the test.

Alternatively, the  $p$ -value can be calculated as

$$p = P_{y^*|H_0}(\text{MLR}_{y^*}(\hat{\theta}_{\text{ML}}) > \text{MLR}_{\underline{y}^*}(\hat{\theta}_{\text{ML}})), \quad (14)$$

where  $\text{MLR}_{\underline{y}^*}$  is the observed marginal likelihood ratio. The null hypothesis is rejected if  $p < \alpha$ .

### 3.2 Bayesian hypothesis testing

Bayesian tests are based on posterior odds. The posterior odds ratio can be expressed as prior odds times Bayes factor (BF),

$$\frac{f(H_1|y)}{f(H_0|y)} = \frac{\pi(H_1)}{\pi(H_0)} \times \text{BF}, \quad (15)$$

where the Bayes Factor is given by

$$\text{BF} = \frac{\int_{H_1} f(y|\theta)\pi(\theta|H_1)d\theta}{\int_{H_0} f(y|\theta)\pi(\theta|H_0)d\theta}. \quad (16)$$

In one-sided tests the Bayes factor simply is  $\int_{H_1} f(y|\theta)/f(y|\theta_0)\pi(\theta|H_1)d\theta$ , a weighted average of likelihood ratio's.

A full Bayesian motivation of the choice between  $\theta = \theta_0$  and the alternative thus requires the specification of a prior  $\pi(\theta|H_1)$ , prior odds and a loss function such that the Bayes Factor may be used to decide whether the decision  $\theta = \theta_0$  is better than the alternative.

The link to the classical choice between hypotheses is provided by the observation that a Bayesian decision rule has the format “Choose  $H_1$  (reject  $H_0$ ) if  $\text{BF} > \kappa$ ”. Given a loss function  $L(i, j)$  when model  $i$  is chosen while model  $j$  is true, one must choose  $H_1$  if

$$\text{BF} = \frac{f(y|H_1)}{f(y|H_0)} > \frac{\pi(H_0)}{\pi(H_1)} \frac{L(1, 0)}{L(0, 1)} = \kappa. \quad (17)$$

Formally, the Bayes factor is not defined when an improper prior  $\pi(\beta, \sigma^2|\theta)$  is used to derive the marginal likelihood  $f(y|\theta)$ : the marginal likelihood is known up to a proportionality constant. This problem can easily be circumvented by using the classical marginal likelihood directly to obtain the Bayes factor,

$$\text{BF} = \int_{H_1} \frac{f(y^*|\theta)}{f(y^*|\theta_0)} \pi(\theta|H_1)d\theta, \quad (18)$$

as  $f(y^*|\theta)$  is a well defined density function and proportional to  $f(y|\theta)$  for the prior (9). A more general approach is given by Fernández, Ley, and Steel (2001), which is formalized by

Strachan and van Dijk (2005), who develop classes of improper priors from which well defined Bayes Factors result.

We now consider the circumstances under which the Bayesian decision rule “Choose  $H_1$ ” corresponds to the classical decision “Reject  $H_0$ ”. The outcome of the Bayesian decision rule (17) corresponds to a classical test with size  $\alpha$ , if

$$\alpha = P_{y^*|H_0} \left( \int_{H_1} \frac{f(y^*|\theta)}{f(y^*|\theta_0)} \pi(\theta|H_1) d\theta > \kappa \right). \quad (19)$$

So  $\alpha$  is the implied (by  $\kappa$  and  $\pi(\theta|H_1)$ ) size of the corresponding classical test. It does *not* depend on the data, and can be reconstructed analytically or by simulation.

Dual to the classical hypothesis testing procedure it is possible to compute a “ $p$ -value”. Define  $\text{BF}_{\underline{y}^*}$  as the observed Bayes factor and  $\text{BF}_{y^*}$  as a possible outcome when  $y^*$  is generated under the null hypothesis. We define the Bayesian  $p$ -value as

$$p = P_{y^*|H_0}(\text{BF}_{y^*} > \text{BF}_{\underline{y}^*}). \quad (20)$$

Note that there are different definitions of Bayesian  $p$ -values in the literature, see for example Gelman (2005). Under the null (20) has a uniform  $U(0;1)$  distribution. Although both  $\text{BF}_{y^*}$  and  $\text{BF}_{\underline{y}^*}$  depend on the prior  $\pi(\theta|H_1)$ , the  $p$ -value does not. It only depends on the data.

**Proposition 3.1** *In case of a monotone marginal likelihood ratio function, depending on only 1 parameter, it holds that*

$$P_{y^*|H_0}(\text{BF}_{y^*} > \text{BF}_{\underline{y}^*}) = P_{y^*|H_0}(\text{MLR}_{y^*}(\hat{\theta}_{ML}) > \text{MLR}_{\underline{y}^*}(\hat{\theta}_{ML})). \quad (21)$$

**Proof** The marginal likelihood ratio  $\text{MLR}(\theta)$  can be expressed as  $f(\theta, S)$ , where  $f$  is monotone in  $S = S(y^*)$ , and therefore  $g(S) = \int_{H_1} f(\theta, S) \pi(\theta|H_1) d\theta$  is monotone in  $S$  as well. As as

consequence the  $p$ -value (20) can be expressed as

$$\begin{aligned} p &= P_{y^*|H_0}(g(S(y^*)) > g(S(\underline{y}^*))) = P_{y^*|H_0}(S(y^*) > S(\underline{y}^*)) \\ &= P_{y^*|H_0}(f(\theta, S(y^*)) > f(\theta, S(\underline{y}^*))) = P_{y^*|H_0}(\text{MLR}_{y^*}(\hat{\theta}_{\text{ML}}) > \text{MLR}_{\underline{y}^*}(\hat{\theta}_{\text{ML}})). \quad \blacksquare \end{aligned} \quad (22)$$

We can now formulate the Bayesian  $p$ -value decision rule as “Choose  $H_1$  (reject  $H_0$ ) if  $p < \alpha$ ”. In a classical analysis the size  $\alpha$  of the test is predetermined. In a Bayesian analysis it implicitly follows from the prior  $\pi(\theta|H_1)$  and  $\kappa$ , see Eq. (19). In case of a monotone likelihood ratio it is possible to derive the implied value of  $\kappa$ , given  $\alpha$  and  $\pi(\theta|H_1)$ . The relation between  $\kappa$  and  $\alpha$  strongly depends on  $\pi(\theta|H_1)$ . From the monotonicity of  $g$  it follows that

$$P_{y^*|H_0}(g(S) > g(\psi_\alpha)) = P_{y^*|H_0}(S > \psi_\alpha), \quad (23)$$

where  $\psi_\alpha$  is the  $\alpha$ -critical value of  $S$  under  $H_0$ . Consequently,

$$\kappa = \int f(\theta, \psi_\alpha) \pi(\theta|H_1) d\theta, \quad (24)$$

the integral over the  $\alpha$ -quantiles of the marginal likelihood ratio function. In section 4 this representation will be used to illustrate the link between  $\alpha$  and  $\kappa$  in the AR(1) model.

Proposition 3.1 is related to Andrews (1994), who showed, under more general assumptions, that for certain priors the Bayesian posterior odds tests is equivalent in large samples to classical likelihood ratio tests with a size determined by prior considerations. Noteworthy is Andrews’ statement (p. 1208) that his results “do not apply to tests of a unit root”. However, in the next section it is shown that the same results for unit root tests holds approximately, though not based on asymptotics.

A final remark is that as the marginal likelihood ratio test is UMPI, by Eq. (22) the same holds for the Bayesian posterior odds test.

### 3.3 The use of the $p$ -value

In the previous subsection two different Bayesian decision rules were presented: the standard approach using Bayes factors, and the “ $p$ -value” approach. Table 1 shows both representations of the posterior odds tests, the  $\kappa$  representation (17) and the  $\alpha$  representation (20).

Bayesians normally do not compute  $p$ -values. In case of a monotone likelihood ratio and a likelihood containing only one parameter, we think there are good reasons to do it. Choosing the  $p$ -value as the test statistic to communicate, has a number of advantages. It facilitates the discussion with frequentists, and even for a Bayesian it might be interesting to derive the probability of the “type I” error.

A more fundamental advantage is that the evidence from the data and prior considerations are separated, which is useful for sensitivity analysis. The  $p$ -value provides all relevant information from the data. The discussion on the appropriate priors and loss function to determine whether the statistic is sufficiently informative to decide against the null is a separate and subjective matter, where readers can make different choices.

On the relevant value of  $\alpha$  there is substantial discussion in the literature. It is well known that for  $\kappa = 1$  (a default choice for most Bayesians) sizes of at least about 0.25 are needed, see Berger (2003) for references. This choice of  $\kappa$  however is in no way compulsory and may differ from situation to situation, depending on prior beliefs and loss functions. One may just as well argue that, in absence of a context, the default choice of  $\alpha = 0.05$  has, notwithstanding many cases where it is inappropriate or interpreted badly, proven to be reasonable.

In our simple setting, the Bayesian view is that there is no such thing as a prechosen  $\alpha$ -level – it varies among settings and subjective judgments – while the  $p$ -value is relevant in all cases. For more complex settings, the results of Andrews (1994) suggest that this will approximately be true.



# 4 CLASSICAL AND BAYESIAN TESTING TESTING FOR A UNIT ROOT

## 4.1 Marginal likelihood and the initial condition

In proposition 2.1 it is shown that classical and Bayesian marginal likelihood are proportional when the independence Jeffreys' prior is used, and in proposition 3.1 that the marginal likelihood ratio and the posterior odds test use the data in a similar way in case of a monotone marginal likelihood ratio depending on only one parameter. Both results provide the basis to show a strong analogy between classical and Bayesian unit root testing in the linear model with first order autoregressive disturbances, as provided in (1)–(3). We investigate this for the AR(1) model with only a constant, and for the model with constant and linear trend, denoted by the superscripts  $\mu$  and  $\tau$  respectively.

The classical marginal likelihood for the model with constant is provided by

$$L_{M_{\beta,\sigma}}^{\mu}(\rho) = \frac{\frac{1}{2}\Gamma(\frac{n-1}{2})}{\pi^{(n-1)/2}} \left( \frac{n(1+\rho)}{n(1-\rho)+2\rho} \right)^{1/2} \left( \frac{\text{RSS}^{\mu}(\rho)}{\text{RSS}^{\mu}(0)} \right)^{-(n-1)/2} \quad (25)$$

where

$$\text{RSS}^{\mu}(\rho) = (1-\rho^2)y_1^2 + \sum_{t=2}^n (y_t - \rho y_{t-1})^2 - \frac{1-\rho}{n-(n-2)\rho} \left( y_1 + (1-\rho) \sum_{t=2}^{n-1} y_t + y_n \right)^2. \quad (26)$$

The marginalization axiom implies directly that this likelihood may be used for Bayesian inference on  $\rho$ . From Eq. (25) it is clear that the marginal likelihood is well defined in  $\rho = 1$ . A general expression of the marginal likelihood for model (1)–(3) can be found in FV.

The link with the noninformative prior is somewhat more complicated than simply using Jeffreys' independence prior. This is due to the initial condition, denoted as  $\xi$  in (3). From (1)–(3) it follows that for  $|\rho| < 1$  the place parameter is  $(\mu, \beta')'$ , while for  $\rho = 1$  it is  $(\mu + \xi, \beta')'$ . For  $|\rho| < 1$  the Bayesian marginal likelihood is proportional to the classical one,  $f(y^*|\rho) \propto f(y|\rho)$ .

For  $\rho = 1$  the Bayesian marginal likelihood  $f(y|\rho)$  (with respect to  $\mu$ ,  $\beta$  and  $\sigma$ ) depends on  $\xi$  and cannot straightforwardly be compared to  $f(y|\rho)$  for  $|\rho| < 1$ . The principle of place invariance however is applicable for both  $|\rho| < 1$  and  $\rho = 1$ . This is an example of our claim that it may have advantages to use  $f(y^*|\rho)$  directly.

The Bayesian marginal likelihood was derived by Lubrano (1995, Eq. 22). For  $|\rho| < 1$  this result is equal to Eq. (25), though with a unspecified constant term. He shows that this expression has a finite limit for  $\rho \rightarrow 1$ . However, different from (25) the marginal likelihood in  $\rho = 1$  is not defined as his initial condition  $y_1 \sim N(\mu + x_1'\beta, \sigma^2/(1 - \rho^2))$  is degenerate in  $\rho = 1$ .

Another Bayesian derivation was provided by Zivot (1994, Eq. 58). Using the unobserved component format he derives the posterior of  $\rho$  in the model with constant under flat priors for  $\mu$ ,  $\sigma$  and  $\rho$ . The limiting case for  $|\rho| < 1$  ( $S \rightarrow \infty$ , implying covariance stationarity) of that posterior is proportional to the marginal likelihood. However, when  $\rho = 1$  the likelihood of the initial observation disintegrates.

As appears from the sequel in Bauwens, Lubrano, and Richard (1999, ch. 6) the prior  $\pi(\mu, \beta, \sigma^2|\rho) \propto \sigma^{-2}$  is not undisputed. Following Zellner (1971) Jeffreys' rule for the multi-parameter case applied separately to  $\sigma^2$  and the other parameters, leads to  $\pi(\mu, \beta, \sigma^2|\rho) \propto (1 - \rho)\sigma^{-2}$ . The term  $(1 - \rho)$  implies in combination with a proper prior for  $\rho$ , a posterior that is zero for  $\rho = 1$ . Bauwens, Lubrano, and Richard (1999) give a survey of singularities at  $\rho = 1$  for different choices of priors, model specifications, and initial conditions.

## 4.2 Reference priors for $\rho$

That reference priors do not unambiguously lead to the same results, appears from Ghosh and Heo (2003). They derived reference priors  $\pi(\mu, \beta, \sigma^*, \rho)$  for inference on  $\rho$  in the model specified by (1)–(3), where  $\sigma^* = \sigma^2(1 - \rho^2)^{-1/n}$ . It turns out to matter whether  $\sigma^*$  and  $(\mu, \beta)'$  are treated simultaneously or sequentially, in their notation  $\pi_{R2}$  and  $\pi_{R3}$ , respectively. The reference priors

are provided by

$$\pi_{R2}(\mu, \beta, \sigma^*, \rho) = \sigma^{*-3/2}(1 - \rho^2)^{-1} \sqrt{n(1 - \rho^2) + 2\rho^2}, \quad (27)$$

$$\pi_{R3}(\mu, \beta, \sigma^*, \rho) = \sigma^{*-1}(1 - \rho^2)^{-1} \sqrt{n(1 - \rho^2) + 2\rho^2}. \quad (28)$$

It can be deduced that the conditional reference prior  $\pi(\mu, \beta, \sigma^*|\rho)$  corresponds to the independence Jeffreys' prior (9) only for  $\pi_{R3}$ . If we would use the prior

$$\pi_{MR3}(\rho) \propto \sqrt{n(1 - \rho^2) + 2\rho^2}, \quad (29)$$

in combination with the marginal likelihood (25) we would obtain their posterior (Eq. 15). One might expect that (29) equals the marginal reference prior  $\pi_{R3}(\rho)$ , but this is not the case as  $\pi_{R3}(\rho) \propto (1 - \rho^2)^{-1} \pi_{MR3}(\rho)$ . Neither it is true that the reference prior derived from the marginal likelihood (25) equals (29). We did not derive this analytically but by simulation. The reason for the difference in priors may be that Ghosh and Heo (2003) use a different transformation of the parameters. We did not pursue this further. Our approach is to concentrate on the marginal likelihood and to investigate the role of priors  $\pi(\rho)$  separately, which will be done in the next subsection

### 4.3 Unit root tests

In this section we compare the power of marginal likelihood ratio tests with Bayesian posterior odds tests in the AR(1). The marginal likelihood ratio is provided by

$$\text{MLR}^i(\rho) = \frac{L_{M\beta,\sigma}^i(\rho)}{L_{M\beta,\sigma}^i(1)}, \quad (30)$$

for  $i = \mu, \tau$ .

For the model with constant the marginal likelihood ratio is a linear combination of more than one statistic, with weights that depend on  $\gamma = n(1 - \rho)$ , even asymptotically. However,

under the null hypothesis,  $\text{MLR}^\mu(\rho)$  is almost a monotone function of  $\text{MLR}^\mu(\hat{\rho}_{\text{ML}})$  for values of  $\rho$  not too close to  $\rho = 1$ , for details see FV. As the marginal likelihood ratio only depends on  $\rho$  and is approximately monotone in  $\hat{\rho}_{\text{ML}}$ , we might expect that the marginal likelihood ratio test and the posterior odds test have the same power function. In this subsection this is investigated for the AR(1) model with constant and constant and trend. We expect that for the ARX(1) model similar results will apply, but we did not pursue this further.

A Bayesian unit root test ( $H_0 : \rho = 1$ , and  $H_1 : |\rho| < 1$ ) is based on the Bayes factor

$$\text{BF} = \frac{f(y^*|H_1)}{f(y^*|H_0)} = \int \frac{f(y^*|\rho)}{f(y^*|\rho = 1)} \pi(\rho|H_1) d\rho. \quad (31)$$

If  $\text{BF} > \kappa$  the alternative is chosen. Examples can be found in Conigliani and Spezzaferri (2007), who choose  $\kappa = 1$  in an experiment with different priors using generalized fractional Bayes factors. Apart from the use of training samples for the prior of  $\rho$  their approach is very similar to ours: in a model containing a level  $\mu$  they also use the data in first differences.

The classical test based on the marginal likelihood ratio has the format: reject  $H_0$  if

$$\text{MLR}(\hat{\rho}_{\text{ML}}) = \frac{f(y^*|\rho = \hat{\rho}_{\text{ML}})}{f(y^*|\rho = 1)} > \kappa_0, \quad (32)$$

where  $\kappa_0$  is chosen such that  $P_{y^*|H_0}(\text{MLR}(\hat{\rho}_{\text{ML}}) > \kappa_0) = \alpha$ , and  $\alpha$  is the predetermined size of the test.

We give some numerical examples of the equivalence between the Bayesian test in terms of  $\kappa$  and the classical marginal likelihood test in terms of  $\alpha$ . As explained in section 3 this correspondence depends on  $\pi(\rho|H_1)$ . Table 2 compares power functions for three tests for  $n = 100$  (100,000 replications): two Bayesian tests, based on a uniform and exponential prior and one classical marginal likelihood ratio test. The priors are

$$\pi_1(\rho) = U(0.5; 1), \quad (33)$$

$$\pi_2(\rho) = 6.9133 \times \exp(-20/3(1 - \rho)), \quad (34)$$

for  $0.5 \leq \rho \leq 1$ , with  $\kappa_j^i$  such that  $P_{y^*|H_0}(\text{BF}^i > \kappa_j^i) = \alpha = 0.05$  for  $i = \mu, \tau$  and  $j = 1, 2$ .  $\kappa_0^i$  follows from  $P_{y^*|H_0}(\text{MLR}^i(\hat{\rho}_{\text{ML}}) > \kappa_0^i) = \alpha = 0.05$ . Note that  $\kappa_0^i$  is an upper bound for values of  $\kappa$  obtained for different priors as it is based on the prior giving most weight to the alternative given the data:  $\pi(\hat{\rho}_{\text{ML}}) = 1$  for  $\rho = \hat{\rho}_{\text{ML}}$ .

From table 2 it can be concluded that the power functions for the three tests are indistinguishable and very close to the power envelope. The difference in values of  $\kappa$  for different priors when  $\alpha$  is fixed, illustrates the tension between Bayesian and classical analysis. In the next subsection the relation between  $\kappa$  and  $\alpha$  is analyzed further.

Note again that this relation is independent from data that are actually observed. The consequence of a sensitivity analysis for the prior that incorporates both  $\kappa$  and  $\alpha$  as valuable inputs is a choice of an  $\alpha$ -level. This can be used to judge the  $p$ -value of the data at hand:  $p = P_{y^*|H_0}(\text{MLR}_{y^*}(\hat{\theta}_{\text{ML}}) > \text{MLR}_{\underline{y}^*}(\hat{\theta}_{\text{ML}}))$ , where  $\text{MLR}_{\underline{y}^*}$  is the observed likelihood ratio.

#### 4.4 The relation between the prior, $\kappa$ , $\alpha$ and $n$

The distribution of the marginal likelihood ratio has a limiting distribution under the null hypothesis in terms of  $\gamma = n(1 - \rho)$ . This asymptotic distribution gives a remarkable good approximation in finite samples, even as small as  $n = 25$ , see FV. Consequently, priors formulated in terms of  $\gamma$  imply an almost fixed relation between  $\kappa$  and  $\alpha$  values for different values of  $n$ . The figures 1 and 2 provide this relation for the model with constant, and constant and trend, respectively. The relations between  $\kappa$  and  $\alpha$  for  $n = 100$  are almost indistinguishable from that obtained for  $n = 1,000$ , when the priors  $\pi_1(\gamma) \sim U(0; 50)$ , and  $\pi_2(\gamma) \sim 1/14.47 \exp(-\gamma/15)$  are used. For  $n = 100$  these priors correspond to the priors (33) and (34).

From Eq. (24) it follows that the relation between  $\kappa$  and  $\alpha$  strongly depends on the prior  $\pi(\gamma)$ . In the AR(1) model where  $\theta := \gamma$  and  $S := \hat{\gamma}_{\text{ML}}$ , the relation (24) holds approximately. The use of this relation is illustrated by the figures 3 and 4, where the  $\alpha$ -quantiles are given for the asymptotic distribution of  $\text{MLR}^i$ , as a function of  $\gamma$ , for  $i = \mu, \tau$ , respectively. Let us consider the 5% quantile function to explain the figures. In figure 3 the function has its

maximum in  $\gamma = 11$ , and a maximum marginal likelihood ratio of about 6, corresponding to a marginal loglikelihood of 1.8. In figure 4 the function has its maximum in  $\gamma = 16$ , and a maximum marginal likelihood ratio of about 5.3, corresponding to a marginal loglikelihood of 1.7.

The  $\alpha$ -quantile functions provide all necessary information to compute the value of  $\kappa^i$  corresponding to  $\alpha$  for any  $\pi(\gamma)$ . An interesting example is provided by the uniform priors  $\gamma \sim U(0; K^i)$  with  $K^\mu \geq 30$  and  $K^\tau \geq 36$ . For a uniform prior with  $K^\mu = 30$  and  $K^\tau = 36$  it can be calculated that a 5% size is obtained for  $\kappa^\mu = 2.6$  and  $\kappa^\tau = 2.6$ . As for  $K^\mu > 30$  and  $K^\tau > 36$  the  $\alpha$ -quantile is virtually zero, the 5% size is obtained for

$$\kappa \approx 2.6 \times \frac{36}{K^i}. \quad (35)$$

Note that for  $K^i = 50$  a 5 percent size is obtained for  $\kappa^\mu = 1.58$  and  $\kappa^\tau = 1.89$ , corresponding to the values in table 2. For other priors than the uniform, the relation between  $\kappa$  and  $K$  is more complex. Low values of  $\kappa$  corresponding to  $\alpha = 0.05$  are obtained for priors with much probability mass for  $\gamma > 25$  and/or near  $\gamma = 0$ . For this reason noninformative priors which are infinite at  $\gamma = 0$  seem to be less appropriate.

Normally priors are formulated in terms of  $\rho$  instead of  $\gamma$ . Priors should indeed be formulated in terms of  $\rho$ . We argue this as follows. If one formulates a prior for a series, logically it may not depend on the frequency with which this series will be measured. We define interpolation of order  $k$  as adding  $k - 1$  points between each two points. The number of data points changes from  $n$  to  $k \times (n - 1) + 1$ , the autoregressive parameter changes from  $\rho$  to  $\rho^{1/k}$ . A prior in terms of  $\rho$  implies a different prior for  $\rho^{1/k}$ . To be coherent a prior refers to a specific frequency of measurement, and the prior for any other frequency is the implied transformation. For a given frequency of measurement the prior is thus in terms of  $\rho$ , and not in terms of  $\gamma$ .

This has interesting consequences. The relation between  $\kappa$  and  $\alpha$  then depends on  $n$ . A clear illustration of this dependence on  $n$  is given by the uniform prior  $\rho \sim U(0.5; 1)$ . In terms of  $\gamma = n(1 - \rho)$  this prior corresponds to  $\gamma \sim U(0; 50)$  for  $n = 100$ , and for  $n = 1,000$  to

$\gamma \sim U(0; 500)$ . For  $n = 100$  this prior implies that  $\kappa^\mu = 1.58$  and  $\kappa^\tau = 1.89$ , corresponding to  $\alpha = 0.05$ . For  $n = 1,000$ ,  $\kappa^i$  follows from (35), so  $\kappa^\mu = 0.158$  and  $\kappa^\tau = 0.189$ . Table 3 gives the implied size as a function of  $n$  for the priors  $\pi_1(\rho)$ , and  $\pi_2(\rho)$  with  $\kappa^i$  chosen such that for  $n = 100$ , the size is 5 percent.

Thus, if one specifies uniform priors in terms of  $\rho$ , the size of the corresponding test is of order  $\alpha/n$  (as table 3 shows this is less clear for the exponential prior). Whether this is a desirable thing, is a matter of taste. The intuitive notion that nonstationarity should show in the long run is confirmed if the prior is formulated in terms of  $\rho$ , for which we provided a compelling reason.

## 5 CONCLUSION

The main goal of this paper was to establish connections between the strong results we obtained in FV for unit root tests using classical marginal likelihood and Bayesian unit root inference. This has led to a variety of results, often unexpected. We choose to confine ourselves to the highlights, well aware that many complications may arise in more complex models. The duality between classical and Bayesian marginal likelihood in the general linear model by a plausible modification of Jeffreys’ rule was the first result. In the unit root case it leads to a marginal likelihood, which is neither degenerate nor zero in  $\rho = 1$ . It may be that this duality only makes sense for inference on the parameters of the covariance matrix. The “marginalization paradox” from Stone and Dawid (1972) suggests that being noninformative in a Bayesian context may depend on the goal of the analysis.

Next we showed for the case that there is only one parameter in the covariance matrix, a monotone marginal likelihood ratio implies an exact correspondence of a common Bayesian decision rule and a classical (one-sided) test. And that it is possible to report the Bayesian rule differently: as a comparison of a  $p$ -value that only depends on the data and a size that depends on prior information and loss functions.

The analysis of the unit root test in the AR(1) model, showed that the “almost perfect”

classical unit root tests, obtained in FV, indeed have Bayesian counterparts for any prior on  $\rho$ . This prior appears to be very important for the size of the corresponding classical test.

So a Bayesian and a frequentist might agree on a unit root test. This is not true however with respect to the role of  $n$ , the number of datapoints. We argued that priors should be formulated in terms of  $\rho$  and that consequently the implied classical size changes with  $n$ .

## References

- Andrews, D. W. K. (1994). The large sample correspondence between classical hypothesis tests and Bayesian posterior odds tests. *Econometrica* 62, 1207–1232.
- Bartlett, M. S. (1957). A comment on D.V. Lindley’s statistical paradox. *Biometrika* 44, 533–534.
- Bauwens, L., M. Lubrano, & J.-F. Richard (1999). *Bayesian Inference in Dynamic Econometric Models*. Oxford University Press, Oxford.
- Berger, J. O. (2003). Could Fisher, Jeffreys and Neyman have agreed on testing? *Statistical Science* 18, 1–32.
- Berger, J. O. & J. M. Bernardo (1992). On the development of reference priors (with discussion). In J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith (Eds.), *Bayesian Statistics 4*, pp. 25–60. Oxford University Press, Oxford.
- Bernardo, J. M. & A. F. M. Smith (1994). *Bayesian Theory*. John Wiley, New York.
- Bhargava, A. (1986). On the theory of testing for unit roots in observed time series. *Review of Economic Studies* 53, 137–160.
- Conigliani, C. & F. Spezzaferri (2007). A robust Bayesian approach for unit root testing. *Econometric Theory* 23, 440–462.
- Fernández, C., A. Ley, & M. F. J. Steel (2001). Benchmark priors for Bayesian model averaging. *Journal of Econometrics* 100, 381–427.
- Fernández, C. & M. F. J. Steel (1999). Reference priors for the general location-scale model. *Statistics and Probability Letters* 43, 377–384.
- Francke, M. K. & A. F. de Vos (2007). Marginal likelihood and unit roots. *Journal of Econometrics* 137, 708–728.
- Gelman, A. (1996). Bayesian model-building by pure thought: Some principles and examples. *Statistica Sinica* 6, 215–232.
- Gelman, A. (2005). Comment: Fuzzy and Bayesian  $p$ -values and  $u$ -values. *Statistical Science* 20, 380–381.
- Ghosh, M. & J. Heo (2003). Default Bayesian priors for regression models with first-order autoregressive residuals. *Journal of Time Series Analysis* 24, 269–282.
- Harvey, A. (2005). A unified approach to testing for stationarity and unit roots. In D. Andrews, J. Powell, P. Ruud, and J. Stock (Eds.), *Identification and Inference for Econometric Models. A Festschrift for Tom Rothenberg*, pp. 403–425. Cambridge University Press, Cambridge.
- Kalbfleisch, J. D. & D. A. Sprott (1970). Application of likelihood methods to models involving large numbers of parameters. *Journal of the Royal Statistical Society B* 32, 175–208.



- King, M. L. (1980). Robust tests for spherical symmetry and their application to least squares regression. *The Annals of Statistics* 8, 1630–1638.
- Lehmann, E. L. (1986). *Testing Statistical Hypotheses* (2 ed.). John Wiley, New York.
- Lubrano, M. (1995). Testing for unit roots in a Bayesian framework. *Journal of Econometrics* 69, 81–109.
- McCullagh, P. & J. A. Nelder (1989). *Generalized Linear Models* (2 ed.). London: Chapman & Hall.
- Müller, U. & G. Elliott (2003). Tests for unit roots and the initial condition. *Econometrica* 71, 1269–1286.
- Patterson, H. D. & R. Thompson (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika* 58, 545–554.
- Phillips, P. C. B. (1991a). Bayesian routes to unit roots: De rebus prioribus semper est disputandum. *Journal of Applied Econometrics* 6, 435–474.
- Phillips, P. C. B. (1991b). To criticize the critics: An objective Bayesian analysis of stochastic trends. *Journal of Applied Econometrics* 6, 333–364.
- Rahman, S. & M. L. King (1997). Marginal-likelihood score-based tests of regression disturbances in the presence of nuisance parameters. *Journal of Econometrics* 82, 81–106.
- Schmidt, P. & P. C. B. Phillips (1992). LM tests for a unit root in the presence of deterministic trends. *Oxford Bulletin of Economics and Statistics* 54, 257–287.
- Schotman, P. & H. K. van Dijk (1991). A Bayesian analysis of the unit root in real exchange rates. *Journal of Econometrics* 49, 195–238.
- Sims, C. A. (1988). Bayesian skepticism on unit root econometrics. *Journal of Economic Dynamics and Control* 12, 463–474.
- Sims, C. A. & H. Uhlig (1991). Understanding unit rooters: A helicopter tour. *Econometrica* 59, 1591–1599.
- Stone, M. & A. P. Dawid (1972). Un-Bayesian implications of improper Bayes inference in routine statistical problems. *Biometrika* 59, 369–375.
- Strachan, R. W. & H. K. van Dijk (2005). Weakly informative priors and well behaved Bayes Factors. Technical Report Econometric Institute Report EI 2005-40, Erasmus University Rotterdam.
- Zellner, A. (1971). *An Introduction to Bayesian Inference in Econometrics*. John Wiley, New York.
- Zivot, E. (1994). A Bayesian analysis of the unit root hypothesis within an unobserved components model. *Econometric Theory* 10, 552–578.

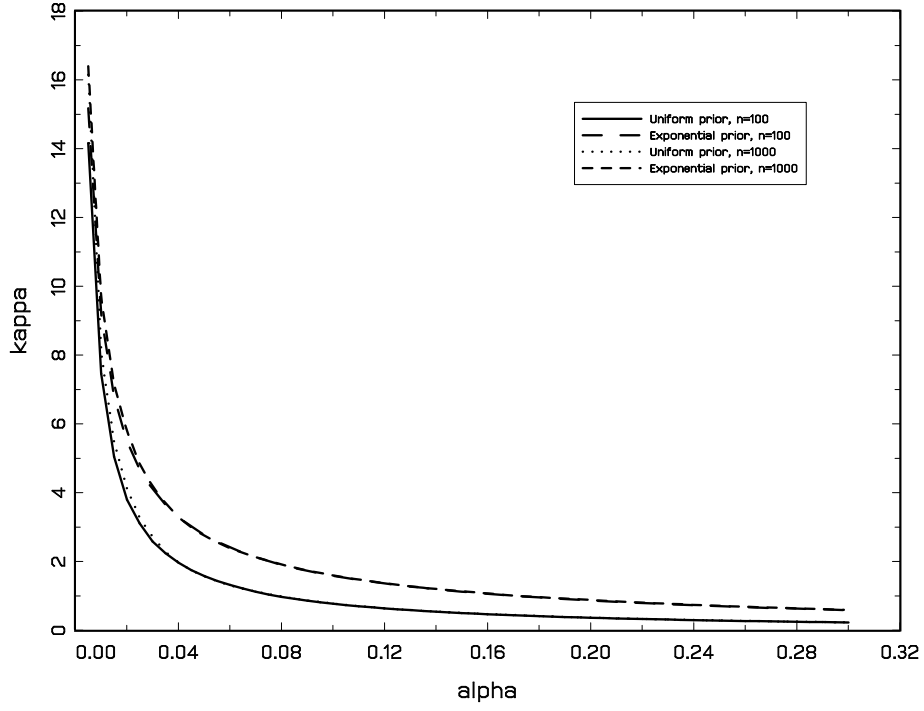


Figure 1: The relation between  $\kappa$  and  $\alpha$  values in the AR(1) model with constant.

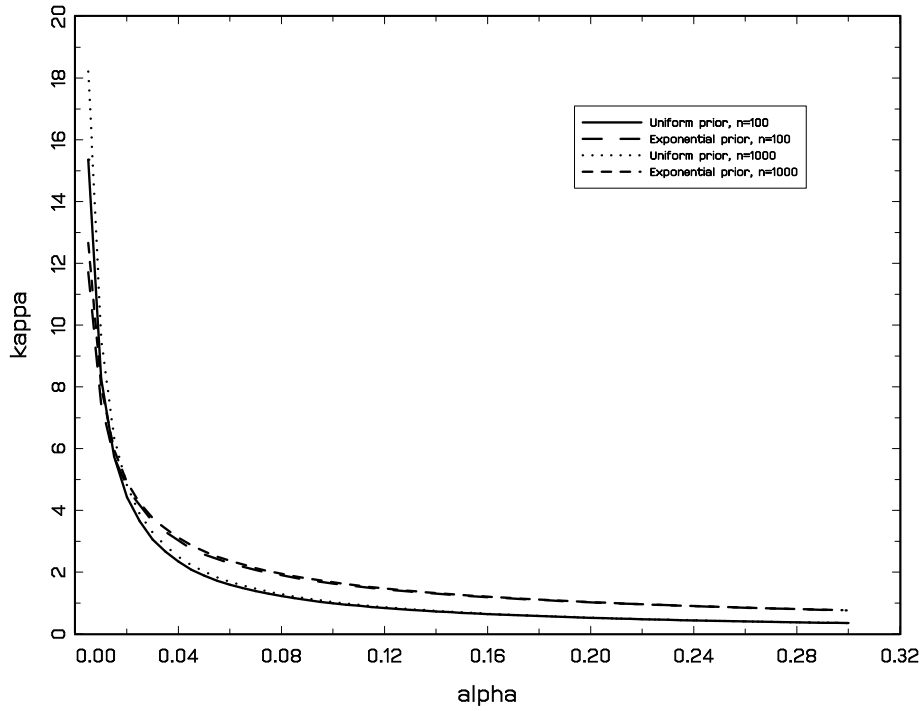


Figure 2: The relation between  $\kappa$  and  $\alpha$  values in the AR(1) model with constant and trend.

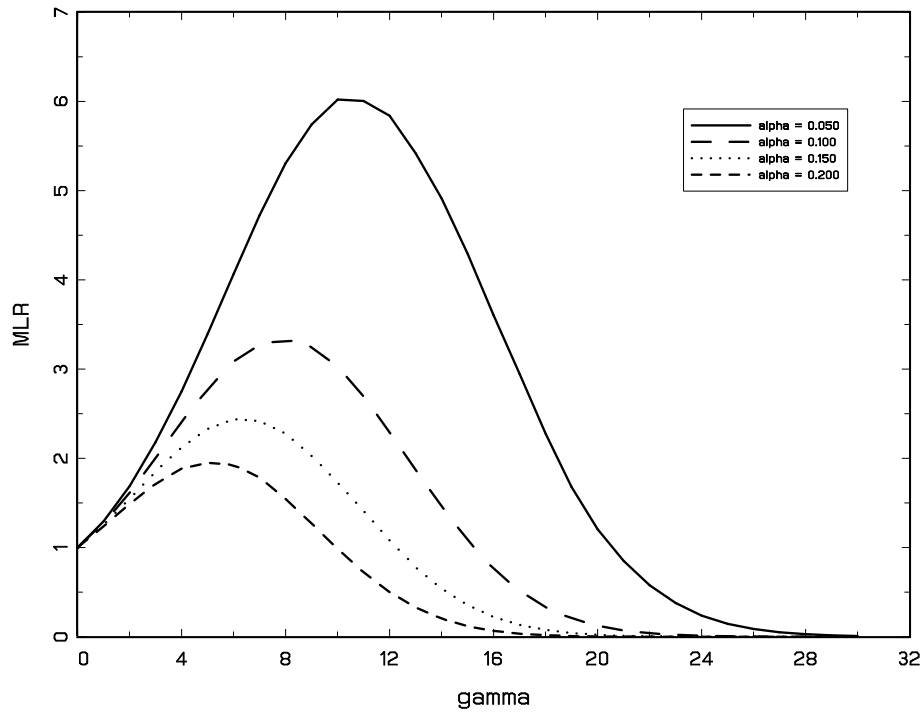


Figure 3: Quantiles of the MLR of the AR(1) model with constant.

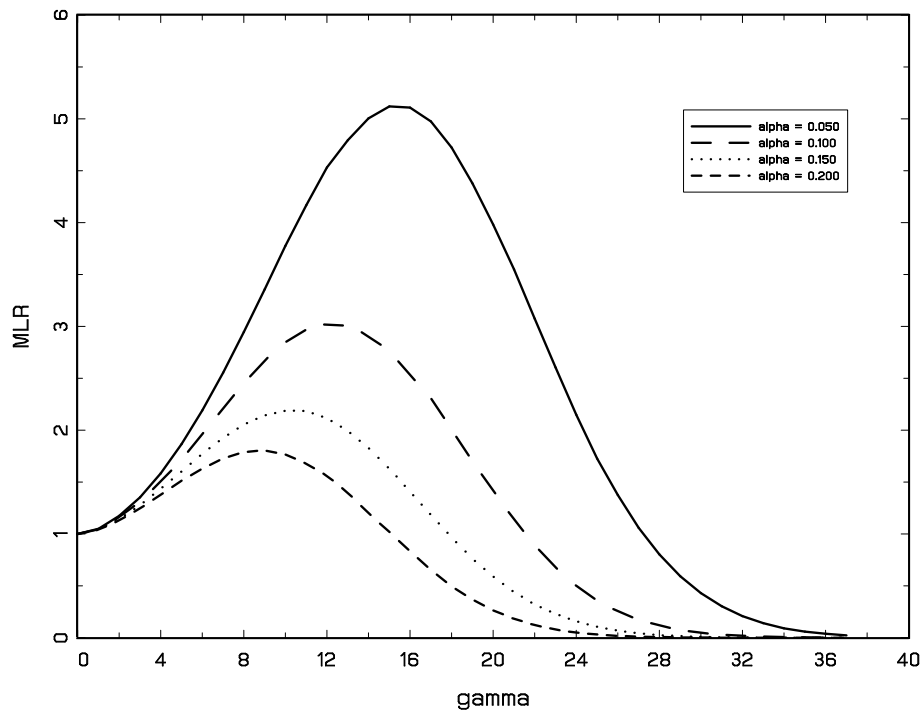


Figure 4: Quantiles of the MLR of the AR(1) model with constant and trend.

Table 1: The relation between the  $\kappa$  and  $\alpha$  analysis.

	statistic	threshold
$\kappa$ analysis	$\text{BF}(y^*, \pi(\theta H_1))$	$\kappa$
$\alpha$ analysis	$P_{y^* H_0}(\text{BF}_{y^*} > \text{BF}_{y^*})$	$\alpha(\kappa, \pi(\theta H_1))$

Table 2: Power functions for the MLR test and Bayes Factors in the AR(1) model for  $n = 100$ .

$\rho$	1.00	0.95	0.90	0.85	0.80
<b>constant</b>					
$P(\text{MLR}^\mu > \kappa_0^\mu = 6.04)$	0.050	0.191	0.515	0.832	0.968
$P(\text{BF}_1^\mu > \kappa_1^\mu = 1.58)$	0.050	0.190	0.513	0.834	0.970
$P(\text{BF}_2^\mu > \kappa_2^\mu = 2.78)$	0.050	0.190	0.514	0.833	0.970
Power envelope for n=100	0.050	0.196	0.521	0.838	0.973
<b>constant and trend</b>					
$P(\text{MLR}^\tau > \kappa_0^\tau = 5.54)$	0.050	0.099	0.252	0.515	0.779
$P(\text{BF}_1^\tau > \kappa_1^\tau = 1.89)$	0.050	0.099	0.251	0.514	0.779
$P(\text{BF}_2^\tau > \kappa_2^\tau = 2.58)$	0.050	0.099	0.252	0.516	0.779
Power envelope for n=100	0.050	0.101	0.256	0.519	0.782

Table 3: Implied sizes of Bayesian tests for different  $n$ , when  $\alpha = 0.05$  for  $n = 100$ .

	constant		trend	
$n$	$\pi_1(\rho)$	$\pi_2(\rho)$	$\pi_1(\rho)$	$\pi_2(\rho)$
25	0.167	0.029	0.118	0.000
50	0.092	0.030	0.099	0.043
<b>100</b>	<b>0.050</b>	<b>0.050</b>	<b>0.051</b>	<b>0.051</b>
250	0.020	0.028	0.021	0.034
1000	0.005	0.009	0.005	0.011