The Bayesian Score Statistic

Frank Kleibergen* Richard Kleijn† Richard Paap‡

Econometric Institute Report EI2000-16/A

Abstract

We propose a novel Bayesian test under a (noninformative) Jeffreys' prior specification. We check whether the fixed scalar value of the so-called Bayesian Score Statistic (BSS) under the null hypothesis is a plausible realization from its known and standardized distribution under the alternative. Unlike highest posterior density regions the BSS is invariant to reparameterizations. The BSS equals the posterior expectation of the classical score statistic and it provides an exact test procedure, whereas classical tests often rely on asymptotic results. Since the statistic is evaluated under the null hypothesis it provides the Bayesian counterpart of diagnostic checking. This result extends the similarity of classical sampling densities of maximum likelihood estimators and Bayesian posterior distributions based on Jeffreys' priors, towards score statistics. We illustrate the BSS as a diagnostic to test for misspecification in linear and cointegration models.

1 Introduction

In applied statistical fields, like time series analysis and econometrics, models/hypotheses are typically compared with other competing (encompassing) models/hypotheses. In classical statistical analysis, it is common to use Wald, likelihood ratio, and score or Lagrange multiplier statistics for this purpose, see Engle (1984). The score statistic is often used as a diagnostic device to test for misspecification of a model since the computation of the score test statistic only requires the estimation of the model parameters under the null hypothesis, while for the likelihood ratio statistic both the model under the null and the alternative hypothesis have to be estimated. The Wald statistic can be calculated using the parameter estimates under the alternative hypothesis. As a consequence, both the score and the likelihood ratio statistic are invariant under reparameterizations of the model but the Wald statistic is not, see Dagenais and Dufour (1991).

In Bayesian statistics competing hypotheses are either compared using Bayes factors/posterior odds ratios, see for example Berger (1985), or tested using Highest Posterior Density (HPD) region based statistics, see Box and Tiao (1973). Model comparison using Bayes factors is

*Corresponding author. Department of Quantitative Economics, University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands, Email: kleiberg@fee.uva.nl. The authors thank the Netherlands Organisation for Scientific Research (N.W.O) for their financial support.

†Tinbergen Institute, Erasmus University Rotterdam, Email: rkleijn@few.eur.nl.

‡Rotterdam Institute for Business Economic Studies, Erasmus University Rotterdam, Email: paap@few.eur.nl.

quite different from testing using HPD region based statistics, see among others Poirier (1995). For example, HPD region based statistics do not require the specification of prior probabilities and proper prior densities for the parameters in the competing hypotheses, while Bayes factors do. Furthermore, in contrast with HPD region based tests, Bayes factors combine both testing and prediction criteria in model comparison. Finally, another important difference is that HPD region based statistics suffer from the same problem as the classical Wald statistic since they are not invariant with respect to reparameterizations of the model. Bayes factors however, are invariant with respect to reparameterizations. In this paper we propose a novel Bayesian test statistic to which we refer as the Bayesian Score Statistic (BSS), which among others overcomes the latter deficiency of the HPD region based statistics.

The BSS equals the posterior expectation of a quadratic form where the posterior results from a Jeffreys' prior and the likelihood. The quadratic form is such that the BSS is a random variable with a standardized distribution when evaluated under the alternative hypothesis and just a fixed scalar under the null hypothesis. We evaluate whether the realization of the BSS under the null hypothesis is a plausible realization from its distribution under the alternative hypothesis. Although both the posteriors under the null and alternative hypothesis are involved we only need the posterior under the null hypothesis, which follows from a Jeffreys' prior specification. The posterior under the alternative hypothesis is constructed such that the posterior under the null hypothesis is the conditional posterior of the parameters under the alternative hypothesis evaluated in the null hypothesis of interest, see Kleibergen (2000a). As the proposed Bayesian test statistic equals the posterior expectation of the classical score statistic and is computed using the posterior distribution under the null hypothesis, we refer to it as a Bayesian score statistic. The invariance of the BSS to reparameterizations follows from the invariance of the classical score statistic and the posterior based on a Jeffreys' prior to parameter transformations.

For several models, Jeffreys' priors lead to posteriors of the parameters that are identical in functional form to the sampling densities of maximum likelihood estimators, see for example Kleibergen and Zivot (1998) and Chao and Phillips (1998). It is therefore not surprising that they also allow for the construction of a Bayesian test statistic that is, in functional form, closely related to a likelihood based classical test statistic, see also Nicolaou (1993) and Tibshirani (1989). The interpretation of the classical and Bayesian test statistics is however quite different. The value of the BSS under the null hypothesis is compared with its exact distribution under the alternative hypothesis to analyze whether its value is a plausible realization from this distribution. Loosely speaking, we analyze whether the model under the null hypothesis is a plausible realization from the posterior under the alternative hypothesis. The classical score statistic has an asymptotic distribution under the null hypothesis and we analyze whether its value under the null hypothesis is a plausible realization from this asymptotic distribution. Hence, the distributions to evaluate the Bayesian and classical score statistics, stem from different hypotheses and are asymptotic for the classical score statistic and exact for the Bayesian score statistic.

The outline of the paper is as follows. In section 2, we define the BSS for a leading case, where we test for a parameter restriction in simple linear model. This model has orthogonal explanatory variables and normally distributed disturbances with identity covariance matrix. We define the assumptions and their resulting statistical properties that are needed for the construction of the BSS. We then extend the leading case to a general case and provide an expression for the BSS that is straightforward to compute. In section 3, we give examples of the BSS for some commonly tested hypotheses. These examples are tests for omitted variables in a standard linear regression model and tests for cointegration in a linear VAR model. Finally, the fourth section concludes.[1]

2 Bayesian Score Statistic

In this section we discuss the different steps involved in the construction of the BSS. The BSS uses an improper non-informative prior specification for the parameters of the nested model, that is the Jeffreys' prior. This prior is proportional to the square root of the determinant of the information matrix. We introduce the BSS statistic to test a simple nested model against a linear alternative model. We will call this the leading case. In the second part of this section, we show that all other representations of the BSS result from this leading case.

2.1 The Leading Case

Consider the model,

$$y = Xf(\varphi) + \varepsilon, \tag{1}$$

where y is a $T \times 1$ vector containing the dependent variables, X is a $T \times n$ matrix containing the non-random explanatory variables for which $X'X = I_n$, and ε is a $T \times 1$ vector of disturbances. The function $f(\varphi)$ is a $n \times 1$ known continuous differentiable vector function in the $k \times 1$ vector φ, whose parameter region is the $\mathbb{R}^k$, $k < n$. This model specifies our null hypothesis H_0. We test the model specification under H_0 against an alternative hypothesis specification H_1 which encompasses model (1)

$$y = X\pi + \varepsilon, \tag{2}$$

where π is an unrestricted $n \times 1$ vector of unknown parameters. Hence, the null hypothesis reads: $H_0 : \pi = f(\varphi)$, and the alternative hypothesis reads: $H_1 : \pi \neq f(\varphi)$. To test H_0 against H_1 we make four assumptions.

Assumptions:

1. The disturbances ε are conditional on σ^2 normally distributed, $\varepsilon|\sigma^2 \sim N(0, \sigma^2 I_T)$, and $\sigma^2 = 1$.

2. The prior on the parameters φ under H_0 is a Jeffreys' prior, that is proportional to the square root of the determinant of the conditional information matrix given $\sigma^2 = 1$, such that

$$p_{H_0}(\varphi|\sigma^2 = 1) \propto \left|I_{H_0}(\varphi|\sigma^2 = 1)\right|^{\frac{1}{2}} \propto \left|\left(\frac{\partial f}{\partial \varphi'}\right)' \frac{\partial f}{\partial \varphi'}\right|^{\frac{1}{2}}. \tag{3}$$

3. A parameterization of the model under H_1 exists such that,

$$\pi = f(\varphi) + g(\varphi)\lambda, \tag{4}$$

[1] To save on space, we often refer to the posterior of the parameters of the model under the null/alternative hypothesis as the posterior under the null/alternative hypothesis.

with λ a $m \times 1$ vector of unknown parameters, $n = k + m$, and $g(\varphi)$ a continuous differentiable $n \times m$ matrix function in the $k \times 1$ vector φ which is such that

$$\begin{aligned} f(\varphi)'g(\varphi) &\equiv 0, \\ \left(\frac{\partial f}{\partial \varphi'}\right)' g(\varphi) &\equiv 0, \\ g(\varphi)'g(\varphi) &\equiv I_m, \end{aligned} \tag{5}$$

and the relationship between π and (φ, λ) is invertible. Note that the number of restrictions on $g(\varphi)$, $k(m+1)$, exceeds the number of elements of $g(\varphi)$, km. As a consequence there exist models for which we cannot construct a function $g(\varphi)$ that satisfies all the conditions in (5).

4. The posterior $p_{H_0}(\varphi|\sigma^2 = 1, Y)$ is a proper density.

These assumptions are needed to define the BSS and to derive its distribution under H_1. The last assumption essentially results from the first two assumptions. Assumption 2 specifies the Jeffreys' prior for the parameter φ conditional on the nuisance parameter σ^2, which is equal to one in the leading case. Assumptions 2 and 3 allow us to use nesting arguments to derive specific statistical results to define the BSS. Assumption 3 implies that (*i.*) the transformation from π to (φ, λ) is properly defined, and (*ii.*) for a given φ, λ is the coordinate vector of $g(\varphi)\lambda$ relative to the ordered basis consisting of the columns of $g(\varphi)$, and (*iii.*) for any φ, the vector $g(\varphi)\lambda$ lies in the null (orthogonal) space of both $f(\varphi)'$ and $\left(\frac{\partial f}{\partial \varphi'}\right)'$. Although the conditional Jeffreys' prior (3) is zero when $f(\varphi)$ is degenerate, that is when it has identical values for different values of φ and $g(\varphi)$, assumptions 2 and 3 do not coincide. The conditions imposed by assumption 3 are stronger than the properties that result from the use of the Jeffreys' prior which can be constructed for any kind of model. Therefore, there exist nested models that result from restrictions on the parameters of linear models for which we cannot construct a parameter λ that coordinatizes the space orthogonal to $f(\varphi)$ and $\frac{\partial f}{\partial \varphi'}$, and for which the transformation from π to (φ, λ) is properly defined. We can construct for these models a conditional Jeffreys' prior H_1 as for any model, but assumption 3 is violated.

The BSS results from the expectation of a quadratic form of random variables. It is based on several statistical properties that result from the above assumptions and the specification of H_0 and H_1. Before we define the BSS, we discuss each of the statistical properties that are involved. These properties are:

A. The posterior under H_0 equals the conditional posterior under H_1 in $\lambda = 0$.

B. The posterior under H_1 can be specified as the product of the conditional density of λ given φ and the marginal density of φ.

C. Given the conditional density from B, under H_1, we can construct a quadratic form in λ. This quadratic form is stochastic independent from φ and has a standardized density such that also its expectation over φ has this density.

We will now discuss each property in detail.

A. Priors, Posteriors as Conditional Priors, Posteriors Assumptions 1 and 2 are such that the posterior of φ under H_0 is specified by

$$p_{H_0}(\varphi|\sigma^2=1,Y) \propto \left|\left(\frac{\partial f}{\partial \varphi'}\right)' \frac{\partial f}{\partial \varphi'}\right|^{\frac{1}{2}} \exp\left[-\frac{1}{2}\left(y - Xf(\varphi)\right)'\left(y - Xf(\varphi)\right)\right]. \tag{6}$$

The Jacobian of the transformation from π to (φ,λ) is determined by assumption 3

$$J(\pi,(\varphi,\lambda)) = \begin{pmatrix} \frac{\partial \pi}{\partial \varphi'} & \frac{\partial \pi}{\partial \lambda'} \end{pmatrix} = \begin{pmatrix} \frac{\partial f}{\partial \varphi'} + (\lambda' \otimes I_n)\frac{\partial vec(g(\varphi))}{\partial \varphi'} & g(\varphi) \end{pmatrix}, \tag{7}$$

where $vec(g(\varphi))$ is a vector that consists of the stacked columns of $g(\varphi)$. As $g(\varphi)'g(\varphi) \equiv I_m$,

$$\begin{array}{ll} \frac{\partial}{\partial \varphi'}\left(g(\varphi)'g(\varphi)\right) & = 0 \Leftrightarrow \\ \left(g(\varphi)' \otimes I_m\right)\left(\frac{\partial vec(g(\varphi)')}{\partial \varphi'}\right) + \left(I_m \otimes g(\varphi)'\right)\frac{\partial vec(g(\varphi))}{\partial \varphi'} & = 0 \Leftrightarrow \\ \left[\left(I_m \otimes g(\varphi)'\right) + \left(g(\varphi)' \otimes I_m\right)K_{k,m}\right]\frac{\partial vec(g(\varphi))}{\partial \varphi'} & = 0 \Leftrightarrow \\ \left[I_{m^2} + K_{m,m}\right]\left(I_m \otimes g(\varphi)'\right)\frac{\partial vec(g(\varphi))}{\partial \varphi'} & = 0 \end{array} \tag{8}$$

where $K_{k,m}$ is the $km \times km$ dimensional commutation matrix which is such that when A is a $k \times m$ matrix $vec(A') = K_{k,m}vec(A)$, and therefore

$$\left(I_m \otimes g(\varphi)'\right)\frac{\partial vec(g(\varphi))}{\partial \varphi'} = 0. \tag{9}$$

The Jacobian of the transformation from π to (φ,λ) can as a consequence be specified as

$$\begin{array}{rl} |J(\pi,(\varphi,\lambda))| = & |J(\pi,(\varphi,\lambda))'J(\pi,(\varphi,\lambda))|^{\frac{1}{2}} \\ = & \left|\begin{pmatrix} \left(\frac{\partial f}{\partial \varphi'} + (\lambda' \otimes I_n)\frac{\partial vec(g(\varphi))}{\partial \varphi'}\right)'\left(\frac{\partial f}{\partial \varphi'} + (\lambda' \otimes I_n)\frac{\partial vec(g(\varphi))}{\partial \varphi'}\right) & \left(\frac{\partial f}{\partial \varphi'} + (\lambda' \otimes I_n)\frac{\partial vec(g(\varphi))}{\partial \varphi'}\right)' g(\varphi) \\ g(\varphi)'\left(\frac{\partial f}{\partial \varphi'} + (\lambda' \otimes I_n)\frac{\partial vec(g(\varphi))}{\partial \varphi'}\right) & I_m \end{pmatrix}\right|^{\frac{1}{2}} \\ = & \left|\begin{pmatrix} \left(\frac{\partial f}{\partial \varphi'} + (\lambda' \otimes I_n)\frac{\partial vec(g(\varphi))}{\partial \varphi'}\right)'\left(\frac{\partial f}{\partial \varphi'} + (\lambda' \otimes I_n)\frac{\partial vec(g(\varphi))}{\partial \varphi'}\right) & 0 \\ 0 & I_m \end{pmatrix}\right|^{\frac{1}{2}} \\ = & \left|\left(\frac{\partial f}{\partial \varphi'} + (\lambda' \otimes I_n)\frac{\partial vec(g(\varphi))}{\partial \varphi'}\right)'\left(\frac{\partial f}{\partial \varphi'} + (\lambda' \otimes I_n)\frac{\partial vec(g(\varphi))}{\partial \varphi'}\right)\right|^{\frac{1}{2}}. \end{array} \tag{10}$$

Hence, the Jeffreys' prior under H_0 equals the Jacobian of the transformation from π to (φ,λ) evaluated in $\lambda=0$, $|J(\pi,(\varphi,\lambda))|_{\lambda=0}|$,

$$|J(\pi,(\varphi,\lambda))|_{\lambda=0}| = \left|\left(\frac{\partial f}{\partial \varphi'}\right)'\left(\frac{\partial f}{\partial \varphi'}\right)\right|^{\frac{1}{2}} = \left|I_{H_0}(\varphi|\sigma^2=1)\right|^{\frac{1}{2}}, \tag{11}$$

where $|_{\lambda=0}$ stands for evaluated in $\lambda=0$. Note that this is an important consequence of assumption 3.

The specification of λ is such that it unambiguously reflects the parameter restriction $\pi = f(\varphi)$ imposed by H_0 and nothing else. We can therefore consider the prior and posterior under H_0 as the unique conditional prior and posterior under H_1 given that $\lambda=0$ and consequently

$\pi = f(\varphi)$, see Kleibergen (2000a). When we cannot construct a parameter λ that accords with assumption 3, we cannot interpret the Jeffreys' prior as a Jacobian of a transformation and hence we can also not interpret the prior/posterior under H_0 as a conditional prior/posterior under H_1.

The Jeffreys' prior under H_0 can be considered to result from the (flat) Jeffreys' prior under H_1,

$$p_{H_1}(\pi|\sigma^2 = 1) \propto |I_{H_1}(\pi|\sigma^2 = 1)|^{\frac{1}{2}} \propto |\sigma^{-2}X'X|^{\frac{1}{2}} = 1, \tag{12}$$

as it is proportional to the conditional Jeffreys' prior under H_1 given that λ is equal to 0,

$$\begin{aligned} p_{H_0}(\varphi|\sigma^2 = 1) &\propto p_{H_1}(\varphi, \lambda|\sigma^2 = 1)|_{\lambda=0} \\ &\propto p_{H_1}(\pi(\varphi, \lambda)|\sigma^2 = 1)|_{\lambda=0}|J(\pi, (\varphi, \lambda))|_{\lambda=0}| \\ &\propto |J(\pi, (\varphi, \lambda))|_{\lambda=0}| = |I_{H_0}(\varphi|\sigma^2 = 1)|^{\frac{1}{2}}. \end{aligned} \tag{13}$$

The Jeffreys' prior (12) for the parameters under H_1 leads to a normal posterior under H_1,

$$\begin{aligned} p_{H_1}(\pi|\sigma^2 = 1, Y) &\propto \exp\left[-\tfrac{1}{2}(y - X\pi)'(y - X\pi)\right] \\ &\propto \exp\left[-\tfrac{1}{2}\left(y'M_X y + (\pi - \hat{\pi})'(\pi - \hat{\pi})\right)\right], \end{aligned} \tag{14}$$

where $\hat{\pi} = (X'X)^{-1}X'y = X'y$ as $X'X = I_n$, $M_X = I_T - X(X'X)^{-1}X' = I_T - XX'$. Thus, the posterior under H_0 equals the conditional posterior under H_1 given that $\lambda = 0$,

$$\begin{aligned} p_{H_0}(\varphi|\sigma^2 = 1, Y) &\propto p_{H_1}(\varphi, \lambda|\sigma^2 = 1, Y)|_{\lambda=0} \\ &\propto p_{H_1}(\pi(\varphi, \lambda)|\sigma^2 = 1, Y)|_{\lambda=0}|J(\pi, (\varphi, \lambda))|_{\lambda=0}| \\ &\propto \left|\left(\tfrac{\partial f}{\partial \varphi'}\right)'\left(\tfrac{\partial f}{\partial \varphi'}\right)\right|^{\frac{1}{2}} \exp\left[-\tfrac{1}{2}(y - Xf(\varphi))'(y - Xf(\varphi))\right]. \end{aligned} \tag{15}$$

B. Posterior $\mathbf{H_1}$ as product of a conditional and marginal density The posterior of (φ, λ) under H_1 can be specified as

$$\begin{aligned} p_{H_1}(\varphi, \lambda|\sigma^2 = 1, Y) \propto\ & p_{H_1}(\pi(\varphi, \lambda)|\sigma^2 = 1, Y)|J(\pi, (\varphi, \lambda))| \\ \propto\ & \left|\left(\tfrac{\partial f}{\partial \varphi'} + (\lambda' \otimes I_n)\tfrac{\partial vec(g(\varphi))}{\partial \varphi'}\right)'\left(\tfrac{\partial f}{\partial \varphi'} + (\lambda' \otimes I_n)\tfrac{\partial vec(g(\varphi))}{\partial \varphi'}\right)\right|^{\frac{1}{2}} \\ & \exp\left[-\tfrac{1}{2}(y - X[f(\varphi) + g(\varphi)\lambda])'(y - X[f(\varphi) + g(\varphi)\lambda])\right]. \end{aligned} \tag{16}$$

The orthogonality conditions (5) imply that we can also directly solve φ from π without the involvement of λ since, see Kleibergen (2000b),

$$\left(\frac{\partial f}{\partial \varphi'}\right)'\pi = \left(\frac{\partial f}{\partial \varphi'}\right)'(f(\varphi) + g(\varphi)\lambda) = \left(\frac{\partial f}{\partial \varphi'}\right)'f(\varphi) = \left(\frac{\partial f}{\partial \varphi'}\right)'\pi(\varphi, \lambda)|_{\lambda=0}. \tag{17}$$

As a consequence, we can solve for φ from π by using

$$\left(\frac{\partial f}{\partial \varphi'}\right)'\pi = \left(\frac{\partial f}{\partial \varphi'}\right)'f(\varphi), \tag{18}$$

which, given a value of π, are k equations with the k elements of φ as the only unknown elements such that φ is exactly identified. We can solve for φ from (18) as it results from the first two orthogonality conditions from (5) that $f(\varphi)$ is spanned by $\frac{\partial f}{\partial \varphi'}$, *i.e.* $f(\varphi) = \left(\frac{\partial f}{\partial \varphi'}\right)q(\varphi)$

with $q(\varphi)$ a k-dimensional continuous differentiable function of φ, such that (18) has a unique solution φ. Equation (17) shows that when we solve for φ from $\pi(\varphi,\lambda)$, we first map $\pi(\varphi,\lambda)$ onto $\pi(\varphi,\lambda)|_{\lambda=0}$, which is equal to $f(\varphi)$, and then solve for φ from $\pi(\varphi,\lambda)|_{\lambda=0}$. The projection of $\pi(\varphi,\lambda)$ onto $\pi(\varphi,\lambda)|_{\lambda=0}$ is an orthogonal projection as the difference between $\pi(\varphi,\lambda)$ and $\pi(\varphi,\lambda)|_{\lambda=0}$, *i.e.* $g(\varphi)\lambda$, is orthogonal to $\pi(\varphi,\lambda)|_{\lambda=0}$. Only the projection onto $\pi(\varphi,\lambda)|_{\lambda=0}$ is an orthogonal projection as projections onto other values of λ, say $\lambda_0 \neq 0$, do not have the property that the difference between the original value and the projected value is orthogonal to the projected value, $(\pi(\varphi,\lambda)-\pi(\varphi,\lambda)|_{\lambda=\lambda_0})'\pi(\varphi,\lambda)|_{\lambda=\lambda_0} \neq 0$. When we have obtained the value of φ from π using (18), we can construct λ as

$$\lambda = g(\varphi)'\pi. \tag{19}$$

Equation (18) allows us to construct the marginal density of φ directly from the marginal density of π as λ is not involved when we solve φ from π. This is also reflected in (19) as (19) shows that, by construction, λ is stochastic independent of $\left(\frac{\partial f}{\partial \varphi'}\right)'\pi$, as $\left(\frac{\partial f}{\partial \varphi'}\right)$ and $g(\varphi)$ are orthogonal and π has a normal distribution with an identity covariance matrix, while $\left(\frac{\partial f}{\partial \varphi'}\right)'\pi$ is the random variable from which we obtain φ. All implicit values of λ in $\pi(\varphi,\lambda)$ in (18) lead to the same value of φ. This shows that λ operates in the space orthogonal to $\left(\frac{\partial f}{\partial \varphi'}\right)$ and doesnot influence the solution of φ from (18). Equation (18) therefore conducts an orthogonal projection of $\pi(\varphi,\lambda)$ onto $\pi(\varphi,\lambda)|_{\lambda=0} = f(\varphi)$ for all values of λ since $(\pi(\varphi,\lambda)-\pi(\varphi,\lambda)|_{\lambda=0})'\pi(\varphi,\lambda)|_{\lambda=0} = 0$. We can obtain the value of λ from $\pi(\varphi,\lambda)$ by using (19). Integrating the joint density of (φ,λ) over λ to obtain the marginal density of φ is thus identical to conditioning on the value of λ where all values of λ are mapped on using the orthogonal projection that we use to solve for φ, *i.e.* $\lambda = 0$. The marginal density of φ is therefore equal to the conditional density of φ given that λ is equal to zero, see Kleibergen (2000a,2000b),

$$\begin{array}{rl} p_{H_1}(\varphi|\sigma^2=1,Y) \propto & \int_{\mathbb{R}^m} p_{H_1}(\varphi,\lambda|\sigma^2=1,Y)d\lambda \\ \propto & \int_{\mathbb{R}^m} p_{H_1}(\pi(\varphi,\lambda)|\sigma^2=1,Y)|J(\pi,(\varphi,\lambda))|d\lambda \\ \propto & \int_{\mathbb{R}^m} [p_{H_1}(\pi(\varphi,\lambda)|\sigma^2=1,Y)|_{\lambda=0}|J(\pi,(\varphi,\lambda))|_{\lambda=0}|]\, p_{H_1}(\lambda|\varphi,\sigma^2=1,Y)d\lambda \\ \propto & [p_{H_1}(\pi(\varphi,\lambda)|\sigma^2=1,Y)|_{\lambda=0}|J(\pi,(\varphi,\lambda))|_{\lambda=0}|] \int_{\mathbb{R}^m} p_{H_1}(\lambda|\varphi,\sigma^2=1,Y)d\lambda \\ \propto & p_{H_1}(\pi(\varphi,\lambda)|\sigma^2=1,Y)|_{\lambda=0}|J(\pi,(\varphi,\lambda))|_{\lambda=0}| \\ \propto & p_{H_1}(\varphi,\lambda|\sigma^2=1,Y)|_{\lambda=0} \\ \propto & p_{H_0}(\varphi|\sigma^2=1,Y), \end{array} \tag{20}$$

which is again identical to the marginal posterior of φ under H_0.

Since we can solve for φ from π in a way that doesnot involve λ, and therefore obtain the marginal posterior of φ, we can also construct the conditional posterior of λ given φ. Using (19), this conditional posterior of λ given φ results as

$$p_{H_1}(\lambda|\varphi,\sigma^2=1,Y) \propto \exp\left[-\frac{1}{2}(\lambda-\hat{\lambda})'(\lambda-\hat{\lambda})\right], \tag{21}$$

where $\hat{\lambda} = g(\varphi)'\hat{\pi} = g(\varphi)'X'y$, and is a normal density. The joint posterior of (φ,λ) (16) can thus also be represented by

$$\begin{array}{rl} p_{H_1}(\varphi,\lambda|\sigma^2=1,Y) \propto & p_{H_1}(\lambda|\varphi,\sigma^2=1,Y)p_{H_1}(\varphi|\sigma^2=1,Y) \\ \propto & p_{H_1}(\lambda|\varphi,\sigma^2=1,Y)\left[p_{H_1}(\varphi,\lambda|\sigma^2=1,Y)|_{\lambda=0}\right]. \end{array} \tag{22}$$

C. Quadratic form in λ that is stochastic independent from φ The specification of the density $p_{H_1}(\lambda|\varphi,\sigma^2=1,Y)$ (21) is such that the random $m\times 1$ vector τ, which we define as

$$\tau=\lambda-\hat{\lambda}, \tag{23}$$

is stochastic independent from φ and has a normal distribution with mean zero and covariance matrix I_m,

$$\tau\sim N(0,I_m). \tag{24}$$

The quadratic form of a normally distributed $m\times 1$ random vector with mean zero and identity covariance matrix has a $\chi^2(m)$ distribution,

$$\tau'\tau\sim\chi^2(m). \tag{25}$$

Because τ is stochastic independent from φ, also $\tau'\tau$ is stochastic independent from φ and the distribution of $\tau'\tau$ does not change when we take the expectation over φ,

$$\begin{aligned} E_{\varphi,\sigma^2=1}\left[\tau'\tau|H_1\right]= & \int_{\mathbb{R}^k}\left[\tau'\tau\right]p_{H_1}(\varphi|\sigma^2=1,Y)d\varphi \\ = & \int_{\mathbb{R}^k}\left[(\lambda-\hat{\lambda})'(\lambda-\hat{\lambda})\right]p_{H_1}(\varphi|\sigma^2=1,Y)d\varphi\sim\chi^2(m). \end{aligned} \tag{26}$$

The BSS results from the expectation of a quadratic form in random variables and uses the property derived above.

2.2 The Bayesian Score Statistic

Equation (26) shows that the posterior expectation, with respect to the marginal posterior of φ, of the quadratic form of τ has a $\chi^2(m)$ distribution that does not depend on any additional parameters. The marginal posterior in (26) is equal to the marginal posterior under H_0. When we want to evaluate the posterior expectation (26) under H_0, λ is equal to 0. As the marginal posteriors under H_0 and H_1 are identical, evaluating the posterior expectation (26) under H_0 only affects the expression of τ in (26),

$$\begin{aligned} E_{\varphi,\sigma^2=1}\left[\tau'\tau|H_0\right]= & \int_{\mathbb{R}^k}\left[\tau'\tau|_{\lambda=0}\right]p_{H_1}(\varphi|\sigma^2=1,Y)d\varphi \\ = & \int_{\mathbb{R}^k}\left[\tau'\tau|_{\lambda=0}\right]\left\{p_{H_1}(\varphi,\lambda|\sigma^2=1,Y)|_{\lambda=0}\right\}d\varphi \\ = & \int_{\mathbb{R}^k}[\hat{\lambda}'\hat{\lambda}]p_{H_0}(\varphi|\sigma^2=1,Y)d\varphi. \end{aligned} \tag{27}$$

We can now analyze whether the resulting value of the expectation is plausible to have been generated by its distribution under H_1, *i.e.* a $\chi^2(m)$ distribution. We use these expectations to define the Bayesian Score Statistic.

Definition 1 *Given assumptions 1-4 and the models under H_0 (1) and H_1 (2), the Bayesian Score Statistic (BSS) for testing the model under H_0 against the model under H_1, is under H_0 equal to*

$$\begin{aligned} BSS(H_0|H_1) &= E_{\varphi,\sigma^2=1}\left[\hat{\lambda}'\hat{\lambda}|H_0\right] \\ &= \int_{\mathbb{R}^k}[\hat{\lambda}'\hat{\lambda}]p_{H_0}(\varphi|\sigma^2=1,Y)d\varphi, \end{aligned} \tag{28}$$

and under H_1 equal to

$$\begin{aligned} BSS(H_1|H_1) &= E_{\varphi,\sigma^2=1}\left[(\lambda-\hat{\lambda})'(\lambda-\hat{\lambda})|H_1\right] \\ &= \int_{\mathbb{R}^k}[\tau'\tau]\,p_{H_1}(\varphi|\sigma^2=1,Y)d\varphi \\ &\sim \chi^2(m). \end{aligned} \tag{29}$$

BSS($H_0|H_1$) can be compared with a $\chi^2(m)$ distribution to analyze whether BSS($H_0|H_1$) is a plausible realization of BSS($H_1|H_1$). It therefore analyzes whether the model under H_0 is a plausible realization from the posterior of the parameters of the model under H_1. Because of the orthogonality conditions (5), $\hat{\lambda}$ is equal to the score ($=\frac{\partial L}{\partial\lambda}|_{H_0}$, where L is the log-likelihood) under H_0 evaluated in $(\varphi,\sigma^2=1)$. The BSS thus equals the posterior expectation under a Jeffreys' prior specification of the quadratic form of the score which explains its name. Since both the quadratic form of the score and the posterior using the Jeffreys' prior are invariant to parameter transformations, the BSS is an invariant test statistic.

In Nicolaou (1993) and Tibshirani (1989), it is shown that the use of Jeffreys' priors for globally orthogonal parameters, as defined by Cox and Reid (1987) and which corresponds with the last two conditions from (5), leads to highest posterior density regions of the resulting marginal posteriors with similar properties as classical confidence regions. For the class of models and hypotheses that we analyze, the BSS extends these results in several ways. First, the BSS shows how a highest posterior density region of a marginal posterior can be evaluated using a statistic that shows whether a certain parameter value lies in the highest posterior density region or not. Second, the BSS shows that by using unique conditional densities, the results of Nicolaou (1993) and Tibshirani (1989) can also be obtained by solely working from the perspective of the null hypothesis. These two extensions allow us to generalize the univariate results of Nicolaou (1993) and Tibshirani (1989) to a multivariate setting. Fourth, the BSS is based on exact finite sample arguments while the results of Nicolaou (1993) and Tibshirani (1989) are obtained in an asymptotic setting but their results therefore hold for a larger class of models and hypotheses then the linear models analyzed here.

The BSS constructed above is for the stylized case of orthogonal explanatory variables, unit variance and a parameter λ that coordinatizes the space orthogonal to φ and for which it holds that an invertible relationship between (λ,φ) and π exists. As the BSS equals the expectation of the classical score statistic with respect to the posterior that uses the Jeffreys' prior, it can straightforwardly be generalized to more complicated models and hypotheses when a (hypothetical) parameter λ exists that satisfies the above mentioned properties as both the classical score statistic as the posterior using the Jeffreys' prior are invariant with respect to parameter transformations.

2.3 The General Case

In the general case, the BSS can be used to test the model under H_0,

$$W(\theta)=Z(\theta)H(\varphi)+U, \tag{30}$$

where θ is a $l\times 1$ vector of unobserved components defined on the region B_θ conditional on which the model is linear, $W(\theta)$ is a $T\times p$ matrix of dependent variables that can depend on θ, $Z(\theta)$ is a $T\times n$ matrix of explanatory variables that can depend on θ, U is a $T\times p$ matrix of disturbances, $H(\varphi)$ is a continuous differentiable $n\times p$ dimensional function in the $k\times 1$

vector φ whose parameter region is the $\mathbb{R}^k$, $k < np$. The linear model under the alternative hypothesis H_1 reads,

$$W(\theta) = Z(\theta)\Gamma + U, \tag{31}$$

with Γ a $n \times p$ matrix of unknown parameters. The assumptions under which the BSS is constructed now become.

Assumptions:

1. The disturbances U are conditional on θ normally distributed, $vec(U)|\theta \sim N(0, \Sigma(\theta) \otimes \Omega(\theta))$ with $\Sigma(\theta)$ and $\Omega(\theta)$ $p \times p$ and $T \times T$ positive definite symmetric matrix functions of θ, respectively.

2. The conditional prior on φ given θ is a Jeffreys' prior,

$$p_{H_0}(\varphi|\theta) \propto \left| \left(\frac{\partial vec(H(\varphi))}{\partial \varphi'} \right)' \left[\Sigma(\theta)^{-1} \otimes Z(\theta)'\Omega(\theta)^{-1}Z(\theta) \right] \frac{\partial vec(H(\varphi))}{\partial \varphi'} \right|^{\frac{1}{2}}. \tag{32}$$

3. A parameterization of the model under H_1 exists such that, for a given θ,

$$\begin{aligned} F(\varphi|\theta) &= \left(Z(\theta)'\Omega(\theta)^{-1}Z(\theta)\right)^{\frac{1}{2}} H(\varphi)\Sigma(\theta)^{-\frac{1}{2}}, \\ \Pi(\Gamma|\theta) &= \left(Z(\theta)'\Omega(\theta)^{-1}Z(\theta)\right)^{\frac{1}{2}} \Gamma\Sigma(\theta)^{-\frac{1}{2}}, \end{aligned} \tag{33}$$

   where $\Pi(\Gamma|\theta) = \Pi$ is unrestricted, and a parameterization of Π exists such that

$$\Pi = F(\varphi|\theta) + G_1(\varphi|\theta)\lambda G_2(\varphi|\theta), \tag{34}$$

   with λ a $(n-q) \times (p-q)$ matrix of unknown parameters, $G_1(\varphi|\theta)$ and $G_2(\varphi|\theta)$ are $n \times (n-q)$ and $(p-q) \times p$ dimensional continuous differentiable matrix functions of φ, $k + (n-q)(p-q) = pn$,

$$\begin{aligned} (G_2(\varphi|\theta)' \otimes G_1(\varphi|\theta))\, vec(F(\varphi|\theta)) &\equiv 0, \\ (G_2(\varphi|\theta)' \otimes G_1(\varphi|\theta))\, \frac{\partial vec(F(\varphi|\theta))}{\partial \varphi'} &\equiv 0, \\ (G_2(\varphi|\theta)' \otimes G_1(\varphi|\theta))' \, (G_2(\varphi|\theta)' \otimes G_1(\varphi|\theta)) &\equiv I_{p-q} \otimes I_{n-q} \end{aligned} \tag{35}$$

   and an invertible relation between Π and (φ, λ) exists.

4. The prior $p_{H_0}(\theta) = p_{H_1}(\theta)$ does not depend on φ. Furthermore, it is such that the posterior $p_{H_0}(\varphi, \theta|Y, Z)$ is a proper density.

The assumptions 1-4 are such that $F(\varphi|\theta)$ corresponds with $f(\varphi)$ in (1) and we can then also, given θ, specify the model under H_0 (30) as,

$$Y(\theta) = X(\theta)F(\varphi|\theta) + \varepsilon, \tag{36}$$

where $Y(\theta) = \Omega(\theta)^{-\frac{1}{2}}W(\theta)\Sigma(\theta)^{-\frac{1}{2}}$, $\varepsilon = \Omega(\theta)^{-\frac{1}{2}}U\Sigma(\theta)^{-\frac{1}{2}}$, $X(\theta) = \Omega(\theta)^{-\frac{1}{2}}Z(\theta)(Z(\theta)'\Omega(\theta)^{-1} Z(\theta))^{-\frac{1}{2}}$ such that $X(\theta)'X(\theta) = I_n$; and under H_1 as

$$Y(\theta) = X(\theta)\Pi + \varepsilon, \tag{37}$$

with Π a $n \times p$ matrix of unknown parameters that results from (34). The conditional Jeffreys' prior (32) then becomes

$$p_{H_0}(\varphi|\theta) \propto \left| \left(\frac{\partial vec(F(\varphi|\theta))}{\partial \varphi'} \right)' \left(\frac{\partial vec(F(\varphi|\theta))}{\partial \varphi'} \right) \right|^{\frac{1}{2}}. \tag{38}$$

Given θ, the resulting models under H_0 (36) and H_1 (37) and the conditional prior (38) satisfy all assumptions from the leading case which shows that the general model can be translated to the leading case in a straightforward way. The invariance of both the conditional prior (38) and the classical score statistic to parameter transformations plays a crucial role in this respect. From definition 1 and the invariance of both the score statistic and the posterior using the Jeffreys' prior we can then define the BSS for the general case to test the model under H_0 (30) against the model under H_1 (31).

Definition 2 *Given assumptions 1-4 and the models under H_0 (30) and H_1 (31), the Bayesian Score Statistic (BSS) for testing the model under H_0 against the model under H_1, is under H_0 equal to*

$$\begin{aligned} BSS(H_0|H_1) = & \int_{B_\theta} \int_{\mathbb{R}^k} [tr(\hat{\lambda}'\hat{\lambda})] p_{H_0}(\varphi, \theta|Y) d\varphi d\theta \\ = & \int_{B_\theta} \int_{\mathbb{R}^k} \left\{ vec\left(\Omega(\theta)^{-\frac{1}{2}} W(\theta) \Sigma(\theta)^{-\frac{1}{2}}\right)' \left[M_{\left(\Sigma(\theta)^{-\frac{1}{2}} \otimes \Omega(\theta)^{-\frac{1}{2}} Z(\theta)\right) \frac{\partial vec(H(\varphi))}{\partial \varphi'}} \right.\right. \\ & \left.\left. - M_{\left(\Sigma(\theta)^{-\frac{1}{2}} \otimes \Omega(\theta)^{-\frac{1}{2}} Z(\theta)\right)} \right] vec\left(\Omega(\theta)^{-\frac{1}{2}} W(\theta) \Sigma(\theta)^{-\frac{1}{2}}\right) \right\} p_{H_0}(\varphi, \theta|Y) d\varphi d\theta, \end{aligned} \tag{39}$$

where $M_A = I_T - A(A'A)^{-1}A'$. The BSS is under H_1 a $\chi^2((n-q)(p-q))$ distributed random variable.

We directly stated the BSS (39) and have left out all the tedious manipulations that are needed to show that it corresponds with (28). Although the BSS (39) looks complicated, it is straightforward to construct when a posterior simulator to sample from the posterior under H_0 is available and can for some models under H_0 (linear) even be constructed analytically. In case of an available posterior simulator, the BSS is a convenient diagnostic tool that can be used to test the model under H_0 against various alternatives in one run of the posterior simulator as the BSS is constructed under H_0 only. Note, however, that we still need to check whether assumption 3 holds before we use the BSS (39). Models namely exist for which the BSS (39) can be computed but that do not satisfy assumption 3 and for which the BSS is then not properly defined. This results as the assumption of a proper transformation from Π to (φ, λ) is a more stringent condition then the computability of the BSS (39). The computability condition namely only requests that we need to be able to characterize the space orthogonal to $\frac{\partial F(\varphi|\theta)}{\partial \varphi'}$ while it also needs to be orthogonal to $F(\varphi)$ to satisfy assumption 3.

BSS($H_0|H_1$) (39) is equal to the posterior expectation of the classical score statistic for a given value of θ that tests (30) against (31) when we use the prior (32). The presence of the unobserved components (nuisance parameters) θ is the main difference between the general and the leading case. For a given value of θ, it can be shown that, after a transformation, the leading and general case coincide. The leading case then shows that the distribution of the BSS under H_1 for a given value of θ is the same and stochastic independent from θ for all values of θ. This distribution does not change when we take the expectation of the BSS with respect to the posterior of θ under H_1. The only difference between the BSS in the leading and general case is therefore the unobserved components (nuisance parameters) θ that are integrated out. These nuisance parameters can take all sorts of forms and can represent, for example, parameters of other explanatory variables in which the model is linear, covariance parameters, unobserved components as in state space or probit/tobit models, mixing parameters for the disturbances when these are mixtures of normal random variables like Student-t random variables etc..

In the next sections we give examples of a linear model and an error correction cointegration model for which we construct the BSS to test them against alternative linear encompassing model specifications.

3 Examples

The definition 2 allows us to compute the BSS for some commonly tested hypotheses in a straightforward way. Although the BSSs can be computed straightforwardly, we still need to verify whether assumption 3 is satisfied. We therefore discuss two examples of the BSS and analyze whether assumption 3 is satisfied. In the first example we test for omitted variables in a standard linear model and in the second example we test for cointegration in an error correction model.

3.1 Linear Model

Consider the linear model under H_0,

$$y = X\beta + \varepsilon, \tag{40}$$

where y is a $T \times 1$ vector of dependent variables, X is a $T \times k$ matrix of explanatory variables, β a $k \times 1$ vector of unknown parameters and $\varepsilon|\theta \sim N(0, \Omega(\theta))$ with $\Omega(\theta)$ a $T \times T$ positive definite symmetric matrix. The standard linear model results when $\Omega(\theta) = \sigma^2 I_T$, such that $\theta = \sigma^2$. The parameter θ may however also contain mixing and degrees of freedom parameters, which allows us to have independent student t distributed errors, see Geweke (1993). The model under the alternative hypothesis H_1 reads,

$$y = X\beta + Z\gamma + \varepsilon, \tag{41}$$

where Z is a $T \times m$ matrix of additional explanatory variables and γ a $m \times 1$ vector of unknown elements. Under H_0, we specify a Jeffreys' prior on β given θ,

$$p_{H_0}(\beta|\theta) \propto \left|X'\Omega(\theta)^{-1}X\right|^{\frac{1}{2}}, \tag{42}$$

and the prior on θ is independent of β. The prior on θ is assumed to be such that $p_{H_0}(\beta, \theta|y)$ is a proper density.

When $(X\ Z)$ has full rank, the specification of the prior and the models under H_0 and H_1 are such that they satisfy assumptions 1 to 4 and we can use the BSS to test H_0 against H_1. Assumption 1 and 2 are satisfied through the specification of the prior and the distribution of the disturbances. Later on we show that assumption 3 is satisfied. The specification of $\text{BSS}(H_0|H_1)$ results from definition 2 (39) when we use $Z(\theta) = (X\ Z)$, $W(\theta) = y$, $\Sigma(\theta) = 1$, and $H(\varphi) = (\beta' \ 0)'$ such that $\frac{\partial vec(H(\varphi))}{\partial \varphi'} = (I_k \ 0)'$,

$$\begin{aligned}
BSS(H_0|H_1) = \ & E_{\beta,\theta}\left[y'\Omega(\theta)^{-\frac{1}{2}\prime}\left[M_{\Omega(\theta)^{-\frac{1}{2}}X} - M_{\Omega(\theta)^{-\frac{1}{2}}(X \ \ Z)}\right]\Omega(\theta)^{-\frac{1}{2}}y|H_0\right] \\
= \ & \int_{B_\theta}\left[y'\Omega(\theta)^{-\frac{1}{2}\prime}M_{\Omega(\theta)^{-\frac{1}{2}}X}\Omega(\theta)^{-\frac{1}{2}}Z\left(Z'\Omega(\theta)^{-\frac{1}{2}\prime}M_{\Omega(\theta)^{-\frac{1}{2}}X}\Omega(\theta)^{-\frac{1}{2}}Z\right)^{-1}\right. \\
& \left. Z'\Omega(\theta)^{-\frac{1}{2}\prime}M_{\Omega(\theta)^{-\frac{1}{2}}X}\Omega(\theta)^{-\frac{1}{2}}y\right]p_{H_0}(\theta|Y)d\theta, \\
BSS(H_1|H_1) \sim \ & \chi^2(m).
\end{aligned} \tag{43}$$

When $\Omega(\theta) = \sigma^2 I_T$ and we take a flat prior for σ^2, that is $p_{H_0}(\sigma^2) \propto 1$, expression (43) simplifies to

$$\begin{aligned}
BSS(H_0|H_1) &= E_{\beta,\sigma^2}\left[\tfrac{1}{\sigma^2}y'(M_X - M_{(X \ \ Z)})y|H_0\right] \\
&= E_{\sigma^2}\left[\tfrac{1}{\sigma^2}y'M_X Z\left(Z'M_X Z\right)^{-1}Z'M_X y|H_0\right] \\
&= \left(y'M_X Z\left(Z'M_X Z\right)^{-1}Z'M_X y\right)E_{\sigma^2}\left[\tfrac{1}{\sigma^2}|H_0\right] \\
&= \left(y'M_X Z\left(Z'M_X Z\right)^{-1}Z'M_X y\right)\int_0^\infty \tfrac{1}{\sigma^2}p_{H_0}(\sigma^2|y)d\sigma^2 \\
&= (T-2)\tfrac{y'M_X Z(Z'M_X Z)^{-1}Z'M_X y}{y'M_X y}, \\
BSS(H_1|H_1) &\sim \chi^2(m),
\end{aligned} \tag{44}$$

where we use that $\int_0^\infty \frac{1}{\sigma^2}p_{H_0}(\sigma^2|y)d\sigma^2 = \frac{T-2}{y'M_X y}$ as the posterior of σ^2 reads

$$p_{H_0}(\sigma^2|y) = 2^{-\frac{1}{2}(T-2)}\frac{1}{\Gamma(\frac{1}{2}(T-2))}(y'M_X y)^{\frac{1}{2}(T-2)}(\sigma^2)^{-\frac{1}{2}T}\exp\left[-\frac{1}{2\sigma^2}y'M_X y\right]. \tag{45}$$

The BSSs (43) and (44) differ from classical test statistics in several respects. First, their distributions result from the alternative hypothesis which is unlike classical test statistics where the distributions result from the null hypothesis. Second, both BSSs are exact test statistics and exactly have the specified distribution under H_1. Many classical test statistics only have asymptotic distributions and their distributions are thus essentially only valid in infinite samples from H_0. The distribution of the BSS (44) also differs from the distribution of the exact classical test statistic for this hypothesis which has a F distribution. This difference results as the BSS under H_1 is stochastic independent from the nuisance parameters. Unlike exact classical test statistics, the distribution of the BSS therefore does not change when nuisance parameters are added which explains why the BSSs (43) and (44) have the same distribution under H_1.

The Jeffreys' prior implies that, conditional on θ, both β and γ are obtained from independent random variables, say φ and λ, which have a normal distribution with an identity covariance matrix. To show this in the case when $\Omega(\theta) = \sigma^2 I_T$, consider that

$$\begin{aligned}
\begin{pmatrix}\beta \\ \gamma\end{pmatrix} &= \left(\begin{pmatrix}X & Z\end{pmatrix}'\begin{pmatrix}X & Z\end{pmatrix}\right)^{-\frac{1}{2}}\begin{pmatrix}\varphi \\ \lambda\end{pmatrix}\sigma \\
&= \begin{pmatrix}(X'X)^{-\frac{1}{2}} & -\left(X'X\right)^{-1}X'Z\left(Z'Z\right)^{-\frac{1}{2}} \\ 0 & (Z'Z)^{-\frac{1}{2}}\end{pmatrix}\begin{pmatrix}\varphi \\ \lambda\end{pmatrix}\sigma \\
&= \begin{pmatrix}(X'X)^{-\frac{1}{2}}\varphi - \left(X'X\right)^{-1}X'Z\left(Z'Z\right)^{-\frac{1}{2}}\lambda \\ (Z'Z)^{-\frac{1}{2}}\lambda\end{pmatrix}\sigma,
\end{aligned} \tag{46}$$

where

$$\begin{pmatrix} \varphi \\ \lambda \end{pmatrix} \sim N\left(\begin{pmatrix} \hat{\varphi} \\ \hat{\lambda} \end{pmatrix}, I_{k+m}\right), \tag{47}$$

with $\begin{pmatrix} \hat{\varphi} \\ \hat{\lambda} \end{pmatrix} = \left(\begin{pmatrix} X & Z \end{pmatrix}' \begin{pmatrix} X & Z \end{pmatrix}\right)^{-\frac{1}{2}} \begin{pmatrix} X & Z \end{pmatrix}' y\sigma^{-1}$, and we have used that

$$\begin{pmatrix} X & Z \end{pmatrix}' \begin{pmatrix} X & Z \end{pmatrix} = \begin{pmatrix} (X'X)^{\frac{1}{2}} & (X'X)^{-\frac{1}{2}} X'Z \\ 0 & (Z'Z)^{\frac{1}{2}} \end{pmatrix}' \begin{pmatrix} (X'X)^{\frac{1}{2}} & (X'X)^{-\frac{1}{2}} X'Z \\ 0 & (Z'Z)^{\frac{1}{2}} \end{pmatrix} \tag{48}$$

and

$$\begin{pmatrix} (X'X)^{\frac{1}{2}} & (X'X)^{-\frac{1}{2}} X'Z \\ 0 & (Z'Z)^{\frac{1}{2}} \end{pmatrix}^{-1} = \begin{pmatrix} (X'X)^{-\frac{1}{2}} & -(X'X)^{-1} X'Z (Z'Z)^{-\frac{1}{2}} \\ 0 & (Z'Z)^{-\frac{1}{2}} \end{pmatrix}. \tag{49}$$

This shows that $\lambda = 0$ implies both that β is equal to $(X'X)^{-\frac{1}{2}} \varphi\sigma$ and that γ is equal to zero. As φ and λ are stochastic independent, β and γ are therefore, conditional on θ, (locally) stochastic independent when $\gamma = 0$.

To show that assumption 3 is satisfied, we note that $(\beta' \ \gamma')' = \left(\begin{pmatrix} X & Z \end{pmatrix}' \begin{pmatrix} X & Z \end{pmatrix}\right)^{-\frac{1}{2}} \pi\sigma^{-1}$, where

$$\pi = \begin{pmatrix} I_k \\ 0 \end{pmatrix} \varphi + \begin{pmatrix} 0 \\ I_m \end{pmatrix} \lambda, \tag{50}$$

such that $g(\varphi) = (0 \ I_m)'$ is orthonormal and orthogonal to $f(\varphi) = (\varphi' \ 0)'$ and $\frac{\partial f}{\partial \varphi} = \begin{pmatrix} I_k & 0 \end{pmatrix}'$.

3.2 Error Correction Cointegration Model

The error correction cointegration model is defined by, see *e.g.* Engle and Granger (1987) and Johansen (1991),

$$\Delta y_t = \alpha'\beta' y_{t-1} + \varepsilon_t, \qquad t = 1, \dots, T, \tag{51}$$

where y_t is a k dimensional vector time series, α' and β are $k \times r$ dimensional matrices that contain the unknown parameters and to identify the parameters we normalize β as $\beta = (I_r \ -\beta_2')'$, where β_2 is a $(k-r) \times r$ dimensional matrix of unknown parameters, ε_t is a $k \times 1$ vector that contains the disturbances which are independent normal with mean zero and covariance matrix Σ. For a Bayesian analysis of the error correction cointegration model, we refer to Kleibergen and Paap (1998).

We can denote the error correction cointegration model (51) in matrix notation by,

$$\Delta Y = Y_{-1}\beta\alpha + \varepsilon, \tag{52}$$

where $\Delta Y = (\Delta y_1 \cdots \Delta y_T)'$, $Y_{-1} = (y_0 \cdots y_{T-1})'$, $\varepsilon = (\varepsilon_1 \cdots \varepsilon_T)'$. We consider the error correction model (52) as the model under H_0 that we test against the linear model under H_1,

$$\Delta Y = Y_{-1}\Theta + \varepsilon, \tag{53}$$

where Θ is a unrestricted $k \times k$ matrix that contains the unknown parameters. The restriction imposed by the model under H_0 (52) is a reduced rank restriction as the rank of $\beta\alpha$ is r while the rank of Θ is k. This restriction satisfies assumption 3 and we can, as we show later on, also construct the different functions and parameters that are used in assumption 3, see Kleibergen and Paap (1998). The prior that we specify under H_0 is a conditional prior for (α, β) given Σ and is a Jeffreys' prior, see Kleibergen and van Dijk (1994),

$$\begin{aligned} p_{H_0}(\alpha, \beta_2|\Sigma) &\propto |I(\alpha, \beta_2|\Sigma)|^{\frac{1}{2}} \propto \left| \left(\frac{\partial vec(\beta\alpha)}{(\partial vec(\alpha)' \ \partial vec(\beta_2)')} \right)' \left(\Sigma^{-1} \otimes Y_{-1}' Y_{-1} \right) \frac{\partial vec(\beta\alpha)}{(\partial vec(\alpha)' \ \partial vec(\beta_2)')} \right|^{\frac{1}{2}} \\ &\propto \left| \left(\begin{array}{cc} I_k \otimes \beta & \alpha' \otimes \left(\begin{array}{c} 0 \\ I_{k-r} \end{array} \right) \end{array} \right)' \left(\Sigma^{-1} \otimes Y_{-1}' Y_{-1} \right) \left(\begin{array}{cc} I_k \otimes \beta & \alpha' \otimes \left(\begin{array}{c} 0 \\ I_{k-r} \end{array} \right) \end{array} \right) \right|^{\frac{1}{2}}, \end{aligned} \tag{54}$$

and the prior on Σ is diffuse and leads to an integrable posterior $p_{H_0}(\alpha, \beta, \Sigma|Y)$. It is important to note here that, as the prior and the posterior have to result from nesting arguments in order to apply the BSS, see section 2, we do not take the expectation over the data in the Jeffreys' priors that we use. As a Bayesian we treat the data as fixed and given both in the prior and the posterior. The involved Jeffreys' priors thus violate the likelihood principle. Efficient posterior simulators to generate drawings from the posteriors of the parameters of reduced rank models are given in Kleibergen and Paap (1998) for general classes of priors, see also Kleibergen and Zivot (1998).

Assumptions 1 to 4 are satisfied and we can construct the BSS to test H_0 (52) against H_1 (53) using the BSS from definition 3 (39) when we use that $W(\theta) = \Delta Y$, $Z(\theta) = Y_{-1}$, $\Omega(\theta) = I_T$, $\Sigma(\theta) = \Sigma$, $H(\varphi) = \beta\alpha$ such that $\frac{\partial vec(H(\varphi))}{\partial \varphi'} = \left(\begin{array}{cc} I_k \otimes \beta & \alpha' \otimes \left(\begin{array}{c} 0 \\ I_{k-r} \end{array} \right) \end{array} \right)$.

After these substitution we obtain that,

$$\begin{aligned} & BSS(H_0|H_1) \\ = \ & E_{\alpha,\beta,\Sigma} \left[vec(\Delta Y \Sigma^{-\frac{1}{2}})' \left[M_{(\Sigma^{-\frac{1}{2}} \otimes Y_{-1}) \frac{\partial vec(H(\varphi))}{\partial \varphi'}} - M_{(\Sigma^{-\frac{1}{2}} \otimes Y_{-1})} \right] vec(\Delta Y \Sigma^{-\frac{1}{2}}) | H_0 \right] \\ = \ & E_{\alpha,\beta,\Sigma} \Big[tr \left(\Sigma^{-1} \Delta Y' Y_{-1} \left(Y_{-1}' Y_{-1} \right)^{-1} Y_{-1}' \Delta Y \right) - tr \left(\Sigma^{-1} \Delta Y' Y_{-1} \beta \left(\beta Y_{-1}' Y_{-1} \beta \right)^{-1} \beta Y_{-1}' \Delta Y \right) \\ & - tr \left(\alpha \Sigma^{-1} \alpha' \right)^{-1} \alpha \Sigma^{-1} \Delta Y' M_{Y_{-1}\beta} Y_{-1,2} \left(Y_{-1,2}' M_{Y_{-1}\beta} Y_{-1,2} \right) Y_{-1,2}' M_{Y_{-1}\beta} \Delta Y \Sigma^{-1} \alpha' | H_0 \Big] \\ = \ & E_{\alpha,\beta,\Sigma} \Big[tr \Big(\left(\Sigma^{-1} - \Sigma^{-1} \alpha' \left(\alpha \Sigma^{-1} \alpha' \right)^{-1} \alpha \Sigma^{-1} \right) \Delta Y_{-1}' M_{Y_{-1}\beta} Y_{-1,2} \\ & \Delta Y' M_{Y_{-1}\beta} Y_{-1,2} \left(Y_{-1,2}' M_{Y_{-1}\beta} Y_{-1,2} \right) Y_{-1,2}' M_{Y_{-1}\beta} \Delta Y \Sigma^{-1} \alpha' \Big) | H_0 \Big] \\ = \ & E_{\alpha,\beta,\Sigma} \left[tr \left(\left(\Sigma^{-\frac{1}{2}\prime} M_{\Sigma^{-\frac{1}{2}} \alpha'} \Sigma^{-\frac{1}{2}} \right) \Delta Y' \left(M_{Y_{-1}\beta} - M_{Y_{-1}} \right) \Delta Y \right) | H_0 \right] \\ = \ & \int_{B_\Sigma} \int_{\mathbb{R}^{rk}} \int_{\mathbb{R}^{(k-r)r}} \left[tr \left(\left(\Sigma^{-\frac{1}{2}\prime} M_{\Sigma^{-\frac{1}{2}} \alpha'} \Sigma^{-\frac{1}{2}} \right) \Delta Y' \left(M_{Y_{-1}\beta} - M_{Y_{-1}} \right) \Delta Y \right) \right] p_{H_0}(\alpha, \beta_2, \Sigma|Y) d\alpha d\beta_2 d\Omega, \\ & \\ & BSS(H_1|H_1) \sim \chi^2((k-r)^2), \end{aligned} \tag{55}$$

where $Y_{-1} = (Y_{-1,1} \ Y_{-1,2})$, $Y_{-1,1} : T \times r$, $Y_{-1,2} : T \times (k-r)$. The BSS has under H_1 a $\chi^2((k-r)^2)$ distribution as the number of restrictions imposed by H_0 on the parameters equals $(k-r)^2$.

The BSS (55) tests the null hypothesis of cointegration against the alternative hypothesis of a full rank model. In classical statistical analysis, these test statistics have asymptotic distributions that are functionals of Brownian motions, see *e.g.* Johansen (1991). Contrary to the distributions of the classical test statistics, the distribution of the BSS results from the alternative hypothesis instead of the null hypothesis. As a consequence, the distribution of the BSS to test for cointegration differs in two respects from the distributions of the classical test statistics. First, the distribution of the BSS is exact while the distributions of the classical test statistics are asymptotic. Second, the distribution of the BSS is a χ^2 distribution and not a function of Brownian motions like the asymptotic distributions of the classical test statistics. This results since the model under H_1 has a normal posterior leading to the χ^2 critical values. In classical statistical analysis, the limiting distributions of the test statistics result from H_0 where random walks, which converge to Brownian motions, are the main feature of the model.

To show that the Jeffreys' prior implies that the parameters which represent the difference between the models under H_0 (52) and H_1 (53) satisfy the orthogonality conditions from assumption 3, we construct the independent random variables where the whole analysis is based on. We therefore conduct a transformation of random variables from (β_2, α) to (γ_2, Φ) such that

$$(Y_{-1}'Y_{-1})^{\frac{1}{2}}\beta\alpha\Sigma^{-\frac{1}{2}} = \Gamma\Phi, \tag{56}$$

where $\Gamma = (I_r \ -\gamma_2')'$, $\gamma_2 : (k-r) \times r$, $\Phi : r \times k$. The Jacobian of this transformation is such that the Jeffreys' prior for (γ_2, Φ) results as,

$$\begin{aligned} p_{H_0}(\Phi, \gamma_2|\Sigma) \propto\ & p_{H_0}(\alpha(\Phi,\gamma_2), \beta_2(\Phi,\gamma_2)|\Sigma)\,|J\left((\alpha,\beta_2),(\gamma_2,\Phi)\right)| \\ \propto\ & \left|\begin{pmatrix} I_k \otimes \Gamma & \Phi' \otimes \begin{pmatrix} 0 \\ I_{k-r} \end{pmatrix} \end{pmatrix}' \begin{pmatrix} I_k \otimes \Gamma & \Phi' \otimes \begin{pmatrix} 0 \\ I_{k-r} \end{pmatrix} \end{pmatrix}\right|^{\frac{1}{2}}, \end{aligned} \tag{57}$$

since the Jeffreys' prior is invariant to parameter transformations. The Jeffreys' prior (57) equals the Jacobian of the transformation from Π to $(\gamma_2, \Phi, \lambda)$ evaluated in $\lambda = 0$, see Kleibergen and Paap (1998),

$$p_{H_0}(\Phi, \gamma_2|\Sigma) \propto |J\left(\Pi, (\gamma_2, \Phi, \lambda)\right)|_{\lambda=0}| \propto \left|\begin{pmatrix} I_k \otimes \Gamma & \Phi' \otimes \begin{pmatrix} 0 \\ I_{k-r} \end{pmatrix} & \Phi_\perp' \otimes \Gamma_\perp \end{pmatrix}\right|, \tag{58}$$

where $\Pi = (Y_{-1}'Y_{-1})^{\frac{1}{2}}\Theta\Sigma^{-\frac{1}{2}}$, $\lambda : (k-r) \times (k-r)$, and

$$\Pi = \Gamma\Phi + \Gamma_\perp\lambda\Phi_\perp, \tag{59}$$

with $\Gamma_\perp'\Gamma \equiv 0$, $\Gamma_\perp'\Gamma_\perp \equiv I_{k-r}$, $\Phi_\perp\Phi' \equiv 0$, $\Phi_\perp\Phi_\perp' \equiv I_{k-r}$. A singular value decomposition, see Golub and van Loan (1989), can be used to show the validity of (59) and to obtain the values of γ_2, Φ and λ from Π, see Kleibergen and Paap (1998). As both $(\Phi_\perp' \otimes \Gamma_\perp)'vec(\Gamma\Phi) = 0$, $(\Phi_\perp' \otimes \Gamma_\perp)'vec(\frac{\partial(\Gamma\Phi)}{\partial vec(\gamma_2)'\,\partial vec(\Phi)'}) = 0$ and $(\Phi_\perp' \otimes \Gamma_\perp)'(\Phi_\perp' \otimes \Gamma_\perp) = (I_{k-r} \otimes I_{k-r})$, assumption 3 is satisfied. The specification of the parameters Π, (γ_2, Φ) and λ corresponds with the specification of π, φ and λ from Section 2, respectively. The parameter λ unambiguously reflects the restriction imposed by the null hypothesis on the parameters of the model under the alternative hypothesis. This implies that the BSS can be constructed.

4 Conclusions

We have constructed a Bayesian statistic that equals the posterior expectation of the classical score statistic where the posterior results from a Jeffreys' prior. We refer to this statistic as the Bayesian Score Statistic (BSS). The BSS can be constructed both under the null hypothesis where it is a scalar and under the alternative hypothesis where it is a random variable with a standardized and known distribution. We consider the BSS under the null hypothesis as a realization from its distribution under the alternative hypothesis and analyze whether this realization is plausible. Hence, we consider the model under the null hypothesis as a realization from the posterior under the alternative hypothesis. We evaluate the plausibility of this realization by analyzing whether it lies in the tail of this posterior. Unlike classical test statistics, the distribution of the BSS results from the alternative hypothesis. Other differences with classical test statistics are that the distribution under the alternative hypothesis is exact and insensitive to nuisance parameters like unobserved components, mixing parameters and so forth.

In several models, a similarity between classical statistical analysis and Bayesian analysis using a Jeffreys' prior exists with respect to the functional form of the sampling density of the maximum likelihood estimator and the density of the posterior. The BSS shows that this similarity further extends to likelihood based test statistics like score statistics.

Since the BSS equals the posterior expectation of the classical score statistic, it is straightforward to compute when an efficient posterior simulator is available. It can be used as a diagnostic tool to test for misspecification in a model of consideration. Furthermore, as the computation of the BSS that tests against an encompassing linear model only involves the posterior under the null hypothesis, we can compute several diagnostics in one run of the posterior simulator.

References

[1] Berger, J.O. *Statistical Decision Theory and Bayesian Inference*. Springer-Verlag (New York), (1985).

[2] Box, G.E.P. and G.C. Tiao. *Bayesian Inference in Statistical Analysis*. Addison-Wesley (Reading, MA), (1973).

[3] Chao, J.C., and P.C.B. Phillips. Posterior distributions in limited information analysis of the simultaneous equations models using the Jeffreys' Prior. *Journal of Econometrics*, **87**:49–86, 1998.

[4] Cox, D.R, and N. Reid. Parameter orthogonality and approximate conditional inference. *Journal of the Royal Statistical Society B*, **49**:1–39, 1987.

[5] Dagenais, M.G. and J.-M. Dufour. Invariance, Nonlinear Models, and Asymptotic Tests. *Econometrica*, **59**:1601–1615, 1991.

[6] Engle, R.F. Wald, Likelihood Ratio, and Lagrange Multiplier Tests in Econometrics. In Z. Griliches and M.D. Intrilligator, editor, *Handbook of Econometrics, Volume 2*. Elsevier Science (Amsterdam), 1984.

[7] Engle, R.F. and C.W.J. Granger. Co-integration and error correction: Representation, estimation and testing. *Econometrica*, **55**:251–276, 1987.

[8] Geweke, J. Bayesian Treatment of the Independent Student-t Linear Model. *Journal of Applied Econometrics*, **8**:519–540, (1993).

[9] Golub, G.H. and C.F. van Loan. *Matrix Computations*. The John Hopkins University Press (Baltimore), (1989).

[10] Johansen, S. Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica*, **59**:1551–1580, (1991).

[11] Kleibergen, F. Conditional Densities in Statistical Modelling. Working paper, University of Amsterdam, 2000a.

[12] Kleibergen, F. Exact Test Statistics and Distributions of Maximum Likelihood Estimators that result from Orthogonal Parameters with applications to the Instrumental Variables Regression Model. Working paper, Department of Quantitative Economics, University of Amsterdam, 2000b.

[13] Kleibergen F. and E. Zivot. Bayesian and Classical Approaches to Instrumental Variable Regression. Econometric Institute Report 9835/A, 1998.

[14] Kleibergen, F. and H.K. van Dijk. On the Shape of the Likelihood/Posterior in Cointegration Models. *Econometric Theory*, **10**:514–551, (1994).

[15] Kleibergen, F. and R. Paap. Priors, Posteriors and Bayes Factors for a Bayesian Analysis of Cointegration. Econometric Institute Report 9821/A, 1998.

[16] Nicolaou, A. Bayesian intervals with good frequentist behaviour in the presence of nuisance parameters. *Journal of the Royal Statistical Society B*, **55**:377–390, 1993.

[17] Poirier, D.J. *Intermediate Statistics and Econometrics: A Comparative Approach*. MIT Press, (Cambridge, MA), (1995).

[18] Tibshirani, R. Noninformative priors for one parameter of many. *Biometrika*, **76**:604–608, 1989.