COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
AT YALE UNIVERSITY

Box 2125, Yale Station
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 1035



# THE LARGE SAMPLE CORRESPONDENCE BETWEEN CLASSICAL HYPOTHESIS TESTS AND BAYESIAN POSTERIOR ODDS TESTS

Donald W. K. Andrews

November 1992

# THE LARGE SAMPLE CORRESPONDENCE BETWEEN CLASSICAL HYPOTHESIS TESTS AND BAYESIAN POSTERIOR ODDS TESTS

Donald W. K. Andrews[1]

Cowles Foundation for Research in Economics
Yale University

August 1992
Revised: November 1992

## ABSTRACT

This paper establishes a correspondence in large samples between classical hypothesis tests and Bayesian posterior odds tests for models without trends. More specifically, tests of point null hypotheses and one- or two-sided alternatives are considered (where nuisance parameters may be present under both hypotheses). It is shown that for certain priors the Bayesian posterior odds test is equivalent in large samples to classical Wald, Lagrange multiplier, and likelihood ratio tests for some significance level and vice versa.



JEL Classification No.: 211.

Keywords: Asymptotics, Bayesian, classical, hypothesis test, likelihood ratio, posterior odds, prior.

# 1. INTRODUCTION

This paper considers large sample approximations to Bayesian posterior odds test statistics for models without trends. For certain priors, the approximations turn out to be monotone functions of the standard Wald, Lagrange multiplier (LM), and likelihood ratio (LR) test statistics. In consequence, the posterior odds test for a given prior corresponds in large samples to a classical hypothesis test for some significance level $\alpha$. In turn, a classical hypothesis test with given significance level $\alpha$ corresponds in large samples to a variety of Bayesian posterior odds tests based on different priors. Thus, the choice of significance level for classical tests is seen to be analogous to the choice of prior for Bayesian posterior odds tests and vice versa.

The approximation results referred to above provide a Bayesian posterior odds interpretation of the Wald, LM, and LR tests that are commonly used in econometrics. This may be of interest to classical and Bayesian econometricians alike. In addition, Bayesian econometricians may find the approximations quite convenient from a computational perspective, because they eliminate high-dimensional integrations, especially with certain choices of priors. Bayesian econometricians also may find the approximations of interest because they illustrate certain robustness properties of posterior odds tests. The results show that posterior odds tests for a variety of different priors yield equivalent tests in large samples. We note that the approximations turn out to be exact for tests of linear restrictions in normal linear regression models with known variance.

There is considerable literature on the relationship between classical hypothesis tests and Bayesian posterior odds tests. Some of this literature focusses on the question of whether a $p$-value can be viewed as a posterior probability of the null hypothesis. In some cases where one is testing a one-sided null hypothesis against a one-sided alternative hypothesis concerning a scalar parameter, such an interpretation is possible (e.g., see Casella and Berger (1987), Berger (1985, pp. 147-148), and Pratt (1965)). In other cases, notably those with a point null hypothesis and one- or two-sided alternatives, the interpretation of a $p$-value as the posterior probability of the null hypothesis is not possible (e.g., see Berger (1985, pp. 148-151), Berger and Sellke (1987), and Edwards, Lindman, and Savage (1963)).

In this paper, we consider point null hypotheses with one- or two-sided alternatives (and allow for nuisance parameters under the null and the alternative). We are interested in the correspondence between classical hypothesis tests and Bayesian posterior odds tests, but we do not focus on the question of whether a $p$-value can be interpreted as a posterior probability of the null hypothesis. Rather, we ask the question of whether a Bayesian posterior odds test is equivalent to some classical test for some significance level. In a large sample sense, we find that the answer is yes for a variety of different priors.

Various results in the literature also are relevant to the question of whether a Bayesian posterior odds test is equivalent to some classical test for some significance level. A number of finite sample results show this for particular models and priors, see Jeffreys (1961, Chs. V, VI), Zellner and Siow (1979, 1980), and for further references Zellner (1984, Ch. 3.7, p. 285). In addition, there are some asymptotic results that are relevant, including those of Jeffreys (1961, pp. 246-7, 249-50), Lindley (1961), Schwarz (1978), Kass, Tierney, and Kadane (1988), Kass and Vaidyanathan (1992), and Phillips (1992). The relationship of our results to the latter asymptotic ones is discussed below (see Section 7).

In this paper, we also analyze classical and Bayesian tests that are designed to be impartial between the null hypothesis and a chosen alternative distribution. We find that the impartial classical and Bayesian tests are equivalent asymptotically for one-sided alternative hypotheses. For two-sided hypotheses, we find that they are not equivalent asymptotically, but are quite close to being so. We note that these results are established for models without trends. For example, they do not apply to tests of a unit root.

The remainder of this paper is organized as follows. Section 2 states the testing problem of interest and outlines the results of the paper. Section 3 specifies the parametric model under consideration and states various "high-level" assumptions that it is required to satisfy. These high-level assumptions include the assumptions that the normalized score function and sample information matrix satisfy a central limit theorem and law of large numbers, respectively, and that the maximum likelihood estimator is consistent. These assumptions are replaced by more primitive assumptions later in the paper. The high-level assumptions are employed, because they help to clarify the essential aspects of the parametric model that are needed for the results of the paper. Section 4 specifies the prior distributions that are considered in the paper. Section 5 states the main results and discusses their implications. Section 6 provides primitive conditions for the assumptions of Section 3 for the case of stationary nonlinear models. Section 7 discusses related results in the literature. An Appendix provides proofs of the results stated in the paper.

All limits below are taken "as $T \to \infty$," where $T$ is the sample size. We let wp $\to$ 1 denote "with probability that goes to one as $T \to \infty$," ~ denote "has the same distribution as," and $\bar{\pi}$ denote pi = 3.14... .

## 2. THE PROBLEM OF INTEREST AND OUTLINE OF RESULTS

The testing problem considered here is the following: Suppose we have a parametric model indexed by a parameter $\theta \in \Theta \subset R^s$. The parameter $\theta$ is of the form $\theta = (\beta', \delta')'$, where $\beta \in R^p$, $\delta \in R^q$, and $s = p+q$. We are interested in testing the null hypothesis

(2.1) $$H_0 : \beta = 0 \text{ (or equivalently, } \theta = \theta_0, \text{ where } \theta_0 = (0', \delta_0')' \text{ for some } \delta_0 \in R^q) .$$

In the classical testing scenario, $H_0$ is tested against the alternative hypothesis $H_1$ given by

(2.2) $$H_1 : \beta \neq 0 \text{ (or equivalently, } \theta = (\beta', \delta_0')', \text{ for some } \beta \in R^p,\ \beta \neq 0, \text{ and some } \delta_0 \in R^q).$$

In a Bayesian scenario, the alternative hypothesis $H_1$ is that the parameter $\beta$ has some distribution that is not pointmass at 0. For the case where $p = 1$, we also consider the one-sided alternative testing problem where the hypotheses are $H_0 : \beta = 0$ and $H_1 : \beta > 0$.

We note that, although (2.1) and (2.2) involve testing linear restrictions, the results given below apply more generally. One usually can reparametrize the model under consideration to convert nonlinear restrictions of the form $H_0 : h(\theta) = 0$ into linear restrictions of the form (2.1) in a transformed parameter space. The results given below can be applied with the transformed parameter space and then mapped back to the original parameter space. (For an alternative approach to posterior odds testing with nonlinear restrictions, see McCulloch and Rossi (1992).)

With classical methods, one can test the null hypothesis $H_0 : \beta = 0$ using a standard Wald, LM (score), or LR statistic given a significance level $\alpha$. Such tests have well-known asymptotic optimality properties, e.g., see Wald (1943). Under suitable regularity conditions, these statistics have a nuisance parameter free asymptotic distribution under $H_0$ and an asymptotically valid critical value, $k_{p,\alpha}$, can be obtained given a significance level $\alpha$.

Using Bayesian methods, one can carry out a posterior odds test of $H_0$ against $H_1$. To do so, one specifies a prior probability $\pi \in (0, 1)$ for the null hypothesis $H_0$ and prior distributions over the parameter values both under $H_0$ and under $H_1$. If $\pi$ is set equal to 1/2 (e.g., as advocated by Jeffreys (1961, pp. 246)), then the prior odds are even. In any event, the posterior odds statistic in favor of $H_1$ is given by the ratio of the posterior probability of $H_1$ to that of $H_0$. A posterior odds test rejects $H_0$ if the posterior odds statistic (in favor of $H_1$) is greater than 1 and otherwise accepts $H_0$.[2]

Let $W_T$, $LM_T$, and $LR_T$ denote the standard Wald, LM, and LR statistics respectively. (The statistic $LR_T$ equals -2 times the log of the likelihood ratio.) The main result of this paper is that for certain choices of prior, $Q_\mu$, the posterior odds ratio is approximately equal to a monotone function of $W_T$, $LM_T$, or $LR_T$ in large samples.

To define this monotone function, let $S_p$ denote the unit sphere $\{v \in R^p : v'v = 1\}$ for $p \geq 1$ and let $S_1^+$ denote the unit sphere in $R^+$, i.e., $S_1^+ = \{1\}$. Let $\mathcal{S}_p$ generically denote one of the unit spheres $S_p$ or $S_1^+$. The unit sphere $S_p$ arises with two-sided tests of the value of a $p$-dimensional vector $\beta$. The unit sphere $S_1^+$ arises with one-sided tests of the value of a scalar parameter $\beta$.

The posterior odds ratio in favor of $H_1$ is shown to be approximately equal to $PO(W_T, \mu)$, $PO(LM_T, \mu)$, and $PO(LR_T, \mu)$, where

$$PO(M, \mu) = \frac{1-\pi}{\pi}\int \exp(-r^2/2)g_p(Mr^2)d\mu(r) \ . \tag{2.3}$$

Here, $\mu(\cdot)$ is a probability distribution on $R^+ = \{r \in R : r \geq 0\}$ that depends on the prior $Q_\mu$ and $g_p(\cdot)$ is a monotone function defined by

$$g_p(\kappa) = \int_{\mathcal{S}_p} \exp(\kappa^{1/2}\mathrm{sgn}(\kappa)\xi'\underset{\sim}{1})dU(\xi)$$

(2.4)
$$= \begin{cases} \frac{1}{2}\exp(\kappa^{1/2}) + \frac{1}{2}\exp(-\kappa^{1/2}) & \text{for } \mathcal{S}_p = S_1 \\ 2\bar{\pi}^{p/2}\Gamma(p/2)\int_{-1}^{1}\exp(\kappa^{1/2}u)(1-u^2)^{(p-3)/2}du & \text{for } \mathcal{S}_p = S_p \text{ and } p \geq 2 \\ \exp(\kappa^{1/2}\mathrm{sgn}(\kappa)) & \text{for } \mathcal{S}_p = S_1^+ \end{cases}$$

for $\kappa \in R$, where $U(\cdot)$ denotes the uniform measure on $\mathcal{S}_p$, $\mathrm{sgn}(\kappa)$ denotes the sign of $\kappa$, $\underset{\sim}{1}$ denotes an arbitrary vector in $\mathcal{S}_p$, and $\Gamma(\cdot)$ denotes the gamma function. (The second equality for the case of $p \geq 2$ follows from Watson (1983, Appendix A, eqns. (1.1), (1.5), (1.6)).)

Of course, the approximate posterior probabilities of $H_0$ and $H_1$, denoted $PP(H_0)$ and $PP(H_1)$, respectively, can be obtained from the approximate posterior odds statistics via

(2.5)
$$PP(H_0) = \frac{1}{1 + PO(M, \mu)} \quad \text{and} \quad PP(H_1) = \frac{PO(M, \mu)}{1 + PO(M, \mu)}$$

for $M = W_T$, $LM_T$, or $LR_T$.

If the prior $Q_\mu$ is a certain multivariate normal distribution with variance proportional to a scalar $\tau > 0$ (or the absolute value of a normal distribution for the one-sided testing case), then the posterior odds statistics $PO(W_T, \mu)$ etc. simplify. The distribution $\mu = \mu_\tau$ that corresponds to $Q_\mu$ in this case equals the square root of $\tau$ times a chi-square random variable with $p$ degree of freedom ($\chi_p^2$), i.e., $\mu = \mu_\tau = \sqrt{\tau\chi_p^2}$ . The statistic $PO(M, \mu)$ of (2.3) simplifies in this case to

(2.6)
$$PO(M, \mu_\tau) = \frac{1-\pi}{\pi}(1+\tau)^{-p/2}\exp\left[\frac{1}{2}\,\frac{\tau}{1+\tau}M\right]$$

in the two-sided testing case for $p \geq 1$ and to

(2.7)
$$PO(M, \mu_\tau) = \frac{1-\pi}{\pi}(1+\tau)^{-1/2}\exp\left[\frac{1}{2}\,\frac{\tau}{1+\tau}M\right]2\Phi\left(\left(\frac{\tau}{1+\tau}|M|\right)^{1/2}\mathrm{sgn}(M)\right)$$

in the one-sided testing case, where $\Phi(\cdot)$ denotes the standard normal distribution function.[3]

Given the approximation of the posterior odds ratio by $PO(W_T, \mu)$, it is easy to see that there is a direct correspondence between a classical test and a posterior odds test. A classical Wald test rejects $H_0$ if $W_T > k_{p,\alpha}$. Here, $k_{p,\alpha}$ is the $(1-\alpha)$-th quantile of a $\chi_p^2$ distribution for two-sided tests with $p \geq 1$ and $k_{p,\alpha}$ is the $(1-2\alpha)$-th quantile of a $\chi_1^2$ distribution for one-sided tests. An approximate posterior odds test rejects $H_0$ if $PO(W_T, \mu) > 1$. Since $PO(M, \mu)$ is a strictly increasing function of $M$ (provided $\mu$ is not pointmass at 0), $PO(\cdot, \mu)$ has an inverse function $PO^{-1}(\cdot, \mu)$ and the approximate posterior odds test rejects if $W_T > PO^{-1}(1, \mu)$. Thus, the classical and approximate $PO$ tests are equivalent whenever $PO^{-1}(1, \mu) = k_{p,\alpha}$ or, equivalently, whenever $PO(k_{p,\alpha}, \mu) = 1$. For fixed $\alpha$ and $\mu$, one can always find $\pi$ such that equality holds. For fixed $\alpha$ and $\pi$, one can always find numerous different distributions $\mu$ such that equality holds. Alternatively, for fixed $\mu$ and $\pi$, one can always find $\alpha$ such that equality holds. It is in this sense that the present paper demonstrates a correspondence between classical and Bayesian tests of $H_0$ versus $H_1$.

## 3. THE PARAMETRIC MODEL

In this section, we define the parametric model, state high-level assumptions that are sufficient for our results, and define the $W_T$, $LM_T$, and $LR_T$ statistics.

Let $Y_T$ denote the random data vector when the sample size is $T$ for $T = 1, 2$ .... Consider a parametric family $\{f_T(y_T, \theta) : \theta \in \Theta\}$ of densities of $Y_T$ with respect to some $\sigma$-finite measure $\gamma_T$, where $\Theta \subset R^s$. The likelihood function is given by $f_T(\theta) = f_T(Y_T, \theta)$.

In many cases, the likelihood function $f_T(\theta)$ can be written as a product of two terms, one that depends on $\theta$ and another that does not. Often the latter term is the product over $t = 1, ..., T$ of the conditional distribution of some weakly exogenous variables at time $t$ given all of the preceding variables (exogenous or not). In such cases, these conditional distributions of the weakly exogenous variables need not be known in order for one to construct the classical or posterior odds test statistics considered here. The results below hold for any such distributions for which the assumptions on $f_T(\theta)$ hold. See Section 6 below for a more explicit discussion of the factoring of $f_T(\theta)$ into known and unknown terms.

Let $\ell_T(\theta) = \log f_T(\theta)$. Let $D\ell_T(\theta)$ denote the $s$-vector of partial derivatives of $\ell_T(\theta)$ with respect to $\theta$. Let $D^2\ell_T(\theta)$ denote the $s \times s$ matrix of second partial derivatives of $\ell_T(\theta)$ with respect to $\theta$. We consider the standard case where the appropriate norming factors for $D\ell_T(\theta)$ and $D^2\ell_T(\theta)$ (so that each is $O_p(1)$ but not $o_p(1)$ as $T \to \infty$) are $T^{-1/2}$ and $T^{-1}$ respectively.

Let $\theta_0$ denote a value of $\theta$ in the null hypothesis. (Below we consider a pointmass prior distribution at $\theta_0$.) We say that a statement holds "under $\theta_0$" if it holds when the true density of $Y_T$ is $f_T(\theta_0)$ for $T = 1, 2, \ldots$.

The likelihood function/parametric model is assumed to satisfy:

ASSUMPTION 1: (a) $\theta_0$ *is an interior point of* $\Theta$.
(b) $f_T(\theta)$ *is twice continuously partially differentiable in* $\theta$ *for all* $\theta \in \Theta_0$ *with probability one under* $\theta_0$, *where* $\Theta_0$ *is some neighborhood of* $\theta_0$.
(c) $-T^{-1}D^2\ell_T(\theta) \xrightarrow{p} \mathcal{I}(\theta)$ *uniformly over* $\theta \in \Theta_0$ *under* $\theta_0$ *for some non-random* $s \times s$ *matrix function* $\mathcal{I}(\theta)$.
(d) $\mathcal{I}(\theta)$ *is uniformly continuous on* $\Theta_0$.
(e) $\mathcal{I} = \mathcal{I}(\theta_0)$ *is positive definite.*

ASSUMPTION 2: $T^{-1/2}D\ell_T(\theta_0) \xrightarrow{d} Z \sim N(0, \mathcal{I})$ *under* $\theta_0$.

We comment briefly on Assumptions 1 and 2. Assumptions 1(a), (b), (d), and (e) are fairly common maximum likelihood (ML) regularity conditions. Differentiability in $\theta$ is assumed for simplicity at the expense of some generality. As is well known, it is not needed for standard ML estimation results and undoubtedly could be relaxed here with some increase in complexity.

Assumption 1(c) is a high-level assumption that requires a uniform weak law of large numbers (WLLN) to hold (since $-T^{-1}D^2\ell_T(\theta_0)$ can be written as a normalized sum of random variables by factoring the likelihood function using conditional distributions). The "uniformity" in Assumption 1(c) can be established, e.g., by using the generic uniform convergence results of Andrews (1992). As stated, Assumption 1(c) allows one to be relatively agnostic regarding the temporal dependence and heterogeneity of the data. To verify 1(c), one needs to be more specific regarding these properties.

Assumption 2 is a second high-level assumption. It requires that the normalized score function satisfies a central limit theorem (CLT) (since $T^{-1/2}D\ell_T(\theta_0)$ can be written as a normalized sum of random variables that are mean zero under weak additional conditions).

Let $\hat{\theta}$ (= $\hat{\theta}_T$) be the unrestricted ML estimator of $\theta$. That is, $\hat{\theta}$ satisfies

$$\ell_T(\hat{\theta}) = \sup_{\theta \in \Theta} \ell_T(\theta) \quad \text{wp} \to 1 \text{ under } \theta_0 . \tag{3.1}$$

Let $\tilde{\theta}$ (= $\tilde{\theta}_T$) be the *restricted* ML estimator of $\theta$. That is, $\tilde{\theta}$ satisfies

$$\begin{aligned} &\tilde{\theta} \in \tilde{\Theta} = \left\{\theta \in \Theta : \theta = (0', \delta')' \text{ for some } \delta \in R^q\right\} \text{ and} \\ &\ell_T(\tilde{\theta}) = \sup_{\theta \in \tilde{\Theta}} \ell_T(\theta) \quad \text{wp} \to 1 \text{ under } \theta_0 . \end{aligned} \tag{3.2}$$

We now introduce two additional high-level assumptions. We assume that the parametric model is sufficiently regular that the ML and restricted ML estimators are consistent under the null hypothesis.

ASSUMPTION 3: $\hat{\theta} \xrightarrow{p} \theta_0$ *under* $\theta_0$.

ASSUMPTION 4: $\tilde{\theta} \xrightarrow{p} \theta_0$ *under* $\theta_0$.

Primitive sufficient conditions for Assumptions 1-4 are given in Section 6 below.

We add a final comment concerning Assumption 1(a) for the case of one-sided tests. This assumption requires that the parametric model is defined for two-sided alternatives even though $H_1$ is one-sided. It implies that $\hat{\theta} = (\hat{\beta}', \hat{\delta}')'$ is a two-sided unrestricted ML estimator. That is, $\hat{\beta}$ may take values greater than or less than zero. Assumption 1(a) can be restrictive in the one-sided case. For example, it precludes $\beta$ from being a variance parameter. On the other hand, it does cover many models of interest and it is an assumption that has been used frequently elsewhere in the literature. For example, it is imposed in the classic paper by Chernoff (1954) on large sample one-sided tests.

This completes the set of assumptions on the parametric model. We are now in a position to define the classical test statistics $W_T$, $LM_T$, and $LR_T$. For the case of two-sided tests with $p \geq 1$, define

$$
\begin{aligned}
W_T &= (HT^{1/2}\hat{\theta})'\left[H\mathcal{I}_T^{-1}(\hat{\theta})H'\right]^{-1}HT^{1/2}\hat{\theta}\ , \\
LM_T &= \left[T^{-1/2}D\ell_T(\tilde{\theta})\right]'\mathcal{I}_T^{-1}(\tilde{\theta})T^{-1/2}D\ell_T(\tilde{\theta})\ , \quad \text{and} \\
LR_T &= -2(\ell_T(\tilde{\theta}) - \ell_T(\hat{\theta}))\ , \quad \text{where} \\
H &= [I_p \ \vdots \ 0] \subset R^{p\times s} \quad \text{and} \quad \mathcal{I}_T(\theta) = -T^{-1}D^2\ell_T(\theta)\ .
\end{aligned}
\tag{3.3}
$$

Alternatively, one can define $\mathcal{I}_T(\theta)$ to be of outer product, rather than Hessian, form. Note that only the first $p$ elements of $D\ell_T(\tilde{\theta})$ are non-zero in the definition of $LM_T$, because $\frac{\partial}{\partial\delta}\ell_T(\tilde{\theta}) = 0$ by the first order conditions for the restricted estimator $\tilde{\theta}$ (wp $\rightarrow$ 1).

For the case of one-sided tests, let $W_T^*$, $LM_T^*$, and $LR_T^*$ denote the expressions on the right-hand side of (3.3) and define the test statistics $W_T$, $LM_T$ and $LR_T$ by

$$
W_T = W_T^*\text{sgn}(H\hat{\theta}),\ LM_T = LM_T^*\text{sgn}(H\mathcal{I}_T^{-1}(\tilde{\theta})D\ell_T(\tilde{\theta})),\ \text{and}\ LR_T = LR_T^*\text{sgn}(H\hat{\theta})\ . \tag{3.4}
$$

Note that a test based on $W_T$ is equivalent in this case to the standard large sample $t$-test, since $W_T$ just equals the $t$-statistic squared times its sign.

## 4. SPECIFICATION OF PRIORS

We take the prior probability of $H_0$ to be $\pi \in (0, 1)$ and that of $H_1$ to be $1-\pi$. We are able to obtain approximations that hold for any parameter vector in the null hypothesis. In consequence, we take the prior over $\theta$ in $H_0$ to be given by pointmass at $\theta_0$, where $\theta_0$ is an arbitrary parameter vector in $H_0$. By doing so, we avoid placing a prior over the nuisance parameter vector $\delta$. As is desirable, the results hold for any fixed value of the nuisance parameter.

Next we specify priors over $\theta$ in $H_1$. We consider priors that depend on the sample size $T$. We do so in order to obtain large sample approximations that hold under the marginal distributions of the data both under $H_0$ and under $H_1$ and that capture relatively detailed effects of the chosen prior. We do not envisage one changing the prior as $T$ changes in practice. Rather, in order to generate approximations for a fixed prior and fixed sample size, we find it useful theoretically to embed the prior in a sequence of priors that vary with $T$. The approach used here is analogous to the use of local alternatives in the analysis of the power of classical tests. If one does not change the prior with $T$, the large sample behavior of the posterior odds ratio in favor of $H_1$ is degenerate. It diverges to infinity under $H_1$ and converges to zero under $H_0$. Using such fixed prior asymptotics, the effect of the prior is captured only crudely (see the discussion and references in Section 7).

For $\theta$ in $H_1$, we write

$$\theta = \theta_0 + T^{-1/2}h , \tag{4.1}$$

where $\theta_0$ is as above and $h$ is some $R^s$ vector. We consider a prior $Q_\mu$ over vectors $h \in R^s$. $Q_\mu$ is fixed for all $T$. This corresponds to priors on $\theta$ that place greater mass on alternatives near $\theta_0$ as $T$ increases.

The prior $Q_\mu$ on $h$ is defined as follows. Let $V$ denote the linear subspace of $R^s$ defined by

$$V = \left\{\theta \in R^s : \theta = (0', \delta')' \text{ for some } \delta \in R^q\right\} . \tag{4.2}$$

The null hypotheses can be expressed as $H_0 : \theta \in \tilde{\Theta} = \Theta \cap V$. We consider a prior $Q_\mu$ over $h$ in $R^s$ that concentrates on the orthogonal complement of $V$ with respect to the inner product $\langle h, \ell\rangle_{\mathcal{I}} = h'\mathcal{I}\ell$ for $h, \ell \in R^s$; call it $V^\perp$. Since $V$ is a $q$ dimensional subspace of $R^s$, $V^\perp$ is a $p$ dimensional subspace of $R^s$. Let $\{a_1, ..., a_p\}$ be some basis of $V^\perp$ and define $A = [a_1, ..., a_p] \in R^{s\times p}$. For example, one can take

$$A = \begin{bmatrix} I_p \\ -\mathcal{I}_3^{-1}\mathcal{I}_2' \end{bmatrix}, \quad \text{where} \quad \mathcal{I} = \begin{bmatrix} \mathcal{I}_1 & \mathcal{I}_2 \\ \mathcal{I}_2' & \mathcal{I}_3 \end{bmatrix} \tag{4.3}$$

for $\mathcal{I}_1 \in R^{p\times p}$, $\mathcal{I}_2 \in R^{p\times q}$, and $\mathcal{I}_3 \in R^{q\times q}$. In consequence,

$$V^\perp = \left\{h \in R^s : h = A\lambda \text{ for some } \lambda \in R^p\right\} . \tag{4.4}$$

The prior $Q_\mu$ that we consider concentrates on $V^\perp$ and has contours given by certain ellipses. In other respects, the prior is arbitrary. Thus, the results given below apply for a wide range of priors. The ellipses over which $Q_\mu$ gives constant weight are the same as those considered by Wald (1943) in his demonstration of the property of asymptotically greatest weighted average power of classical Wald tests. The parameter vectors $\theta$ corresponding to different points on any such ellipse have the property that they are equally difficult to detect asymptotically -- no direction away from the null is favored over any other.

The prior distributions are assumed to satisfy:

ASSUMPTION 5: (a) *The prior probabilities of $H_0$ and $H_1$ are $\pi \in (0, 1)$ and $1-\pi$ respectively.*
(b) *The prior distribution of $\theta$ under $H_0$ is pointmass at $\theta_0$, where $\theta_0$ is the null parameter vector considered in Assumptions* 1-4.
(c) *The prior distribution of $\theta$ under $H_1$ is given by $\theta = \theta_0 + T^{-1/2}h$ and $h \sim Q_\mu$, where $\theta_0$ is as in part* (b). *The distribution $Q_\mu$ of $h$ is such that $h/\|h\|_{\mathcal{I}} \sim A(A'\mathcal{I}A)^{-1/2}\xi$, where $\xi$ is a random vector that is uniformly distributed on the $p$ dimensional unit sphere $\mathcal{S}_p$, and $h/\|h\|_{\mathcal{I}}$ and $\|h\|_{\mathcal{I}}$ are independent.*

Let $\mu$ denote the prior distribution of $\|h\|_{\mathcal{I}}$. Assumption 5 allows $\mu$ to be arbitrary (provided it is not pointmass at 0).

When $p = 1$ and $H_1$ is one-sided, the unit sphere $\mathcal{S}_p$ referred to in Assumption 5(c) is a pointmass at 1. In this case, Assumption 5(c) places no restriction on the distribution of $h$ other than that it lies in $V^\perp$. When $p = 1$ and $H_1$ is two-sided, the unit sphere $\mathcal{S}_1$ equals $\{-1, 1\}$. In this case, Assumption 5(c) requires the prior on $h$ to satisfy a symmetry property. If $\mathcal{I}_2 = 0$ (i.e., if the information matrix is diagonal between the parameters $\beta$ and $\delta$), the symmetry property just requires that $h$ and $-h$ are given equal prior density (or prior mass) for all $h$. When $p > 1$, the unit sphere $\mathcal{S}_p$ is non-degenerate and Assumption 5(c) requires the prior $Q_\mu$ to have contours given by certain ellipses, as noted above.

For particular choices of distribution $\mu$ on $\|h\|_{\mathcal{I}}$, the distribution $Q_\mu$ and the approximation to the posterior odds ratio simplify. More specifically, suppose $\mu$ is the distribution of the square root of $\tau$ times a chi-square random variable with $p$ degrees of freedom (i.e.,

$\mu \sim \sqrt{\tau\chi_p^2}$ ) for some $\tau > 0$. Then, in the two-sided testing case, $h$ and $\theta$ have prior distributions given by

$$h \sim Q_\mu = N(0, \tau\Sigma) \quad \text{and} \quad \theta \sim N\left(\theta_0, (\tau/T)\Sigma\right), \quad \text{where}$$

(4.5)
$$\Sigma = A(A'\mathcal{I}A)^{-1}A' = \begin{bmatrix} \left(\mathcal{I}_1 - \mathcal{I}_2\mathcal{I}_3^{-1}\mathcal{I}_2'\right)^{-1} & -\left(\mathcal{I}_1 - \mathcal{I}_2\mathcal{I}_3^{-1}\mathcal{I}_2'\right)^{-1}\mathcal{I}_2\mathcal{I}_3^{-1} \\ -\mathcal{I}_3^{-1}\mathcal{I}_2'\left(\mathcal{I}_1 - \mathcal{I}_2\mathcal{I}_3^{-1}\mathcal{I}_2'\right)^{-1} & \mathcal{I}_3^{-1}\mathcal{I}_2'\left(\mathcal{I}_1 - \mathcal{I}_2\mathcal{I}_3^{-1}\mathcal{I}_2'\right)^{-1}\mathcal{I}_2\mathcal{I}_3^{-1} \end{bmatrix}$$

and $N(0, \Sigma)$ denotes a multivariate normal distribution with mean 0 and covariance matrix $\Sigma$ (possibly singular). With some algebraic manipulations, one can show that the above prior on $\theta$ corresponds to the following prior on $\beta$:

(4.6)
$$\beta \sim N\left(0, (\tau/T)H\mathcal{I}^{-1}H'\right).$$

Note that $(1/T)H\mathcal{I}^{-1}H'$ is the asymptotic variance of the unrestricted ML estimator $\hat{\beta}$.

In the one-sided testing case, the above prior $\mu \sim \sqrt{\tau\chi_1^2}$ on $\|h\|_\mathcal{I}$ yields priors on $h$, $\theta$, and $\beta$ given by

(4.7)
$$h \sim Q_\mu = |N(0, \tau\Sigma)|, \quad \theta \sim \theta_0 + |N(0, (\tau/T)\Sigma)|, \quad \text{and}$$
$$\beta \sim |N(0, (\tau/T)H\mathcal{I}^{-1}H')|$$

for $\Sigma$ as in (4.5), where $|N(0, \Sigma)|$ denotes the distribution of the absolute value of a random variable with $N(0, \Sigma)$ distribution.

For convenience, we refer to the above cases as Assumption 5*:

ASSUMPTION 5*: *Assumption 5 holds with the distribution $Q_\mu$ of $h$ given by $N(0, \tau\Sigma)$ for tests of two-sided alternatives and by $|N(0, \tau\Sigma)|$ for tests of one-sided alternatives, for some constant $\tau > 0$.*

Under Assumption 5*, the formula (2.3) for the approximate posterior odds ratio can be simplified:

LEMMA 1: *Under Assumption 5*, the expression for $PO(M, \mu)$ given in* (2.3) *simplifies to that given in* (2.6) *for two-sided tests and to that given in* (2.7) *for one-sided tests.*

# 5. MAIN RESULTS

In this section, we state the main approximation results for the posterior odds statistic and discuss their interpretation.

## 5.1. *Statement of the Main Results*

To start, we define the posterior odds statistic $PO_T(Q_\mu)$ given the priors defined in Assumption 5:

$$PO_T(Q_\mu) = \frac{1-\pi}{\pi}\int f_T(\theta_0 + T^{-1/2}h)dQ_\mu(h)/f_T(\theta_0) \ . \tag{5.1}$$

Note that for convenience we have defined $PO_T(Q_\mu)$ to be the posterior odds ratio *in favor of* $H_1$, not in favor of $H_0$. (The latter is just the reciprocal of $PO_T(Q_\mu)$.)

Below we say that probabilistic results hold "under $H_0$" and "under $H_1$" if they hold under the predictive densities of the data under $H_0$ and under $H_1$ respectively (i.e., under the marginal distribution of the data $Y_T$ that is determined by the parametric model and the prior on $\theta$ under $H_0$ and under $H_1$, as described in Assumption 5). Thus, "under $H_0$" is equivalent to "under $\theta_0$," whereas "under $H_1$" depends on $\theta_0$ and $Q_\mu$.

The main result of the paper is the following:

THEOREM 1: *Suppose Assumptions* 1-5 *hold. Then, under both* $H_0$ *and* $H_1$, *we have:* (a) $PO_T(Q_\mu) - PO(W_T, \mu) \xrightarrow{p} 0$, (b) $PO_T(Q_\mu) - PO(LM_T, \mu) \xrightarrow{p} 0$, *and* (c) $PO_T(Q_\mu) - PO(LR_T, \mu) \xrightarrow{p} 0$.

COMMENTS: 1. Theorem 1 holds not just for a single vector $\theta_0$, but for all null parameter vectors $\theta_0$ for which Assumptions 1-5 hold, since $PO_T(W_T, \mu)$, ..., $PO(LR_T, \mu)$ do not depend on $\theta_0$.

2. The result of Theorem 1 suggests approximating $PO_T(Q_\mu)$ by $PO(W_T, \mu)$, $PO(LM_T, \mu)$, or $PO(LR_T, \mu)$. At least in some cases, this approximation is quite good. For example, part (a) of Theorem 1 holds exactly (i.e., $PO_T(Q_\mu) = PO(W_T, \mu)$) in the case of a linear regression model with regression parameter $\theta$, iid normal $(0, \sigma^2)$ errors, $\sigma^2$ known, and weakly exogenous regressors.[4]

*5.2. Interpretation of Theorem 1*

Theorem 1 shows that the posterior odds test, which rejects $H_0$ when $PO_T(Q_\mu) > 1$, is approximately equal to the test that rejects $H_0$ when $PO(W_T, \mu) > 1$, or equivalently, when

$$W_T > PO^{-1}(1, \mu) , \tag{5.2}$$

where $PO^{-1}(\cdot, \mu)$ is the inverse function of the strictly increasing function $PO(\cdot, \mu)$. (The same holds with $W_T$ replaced by $LM_T$ or $LR_T$.) On the other hand, a classical Wald test of asymptotic significance level $\alpha$ rejects $H_0$ when

$$W_T > k_{p,\alpha} , \tag{5.3}$$

where $k_{p,\alpha}$ is as defined in Section 2. In consequence, the posterior odds test is approximately equal to a classical test, and vice versa, whenever $\alpha$, $\mu$, and $\pi$ are such that

$$PO(k_{p,\alpha}, \mu) = 1 . \tag{5.4}$$

It can be shown that for any significance level $\alpha \in (0, 1)$ there exist pairs of priors $(\pi, \mu)$ such that (5.4) holds. In fact, there are many such $(\pi, \mu)$ pairs corresponding to a given $\alpha$. Conversely, given any pair of priors $(\pi, \mu)$, there exists a (unique) significance level $\alpha$ such that (5.4) holds. In fact, given any pair from the triplet $(\alpha, \pi, \mu)$, there exists a value of the third element such that (5.4) holds. Thus, for the special case of even prior odds ($\pi = 1/2$), given any significance level $\alpha$ there exists a prior $\mu$ such that (5.4) holds and vice versa. These results imply that for any classical test there exist equivalent approximate posterior odds tests and vice versa. The mapping of posterior odds tests to classical tests is many-to-one.

The discussion above indicates that in large samples there are numerous posterior odds tests that are approximately equal. Different pairs of priors $(\pi, \mu)$ that yield the same value of $PO^{-1}(1, \mu)$ generate posterior odds tests that are approximately equal. This is a useful robustness property for posterior odds tests: The result of a posterior odds test holds not just for a single pair of priors $(\pi, \mu)$, but for the whole family of priors that generate the same value of $PO^{-1}(1, \mu)$. Note that even if the prior probability $\pi$ of $H_0$ is fixed, say at 1/2, there is still a whole family of priors $\mu$ that generate the same value of $PO^{-1}(1, \mu)$.

The $p$-value of a test based on the statistic $W_T$, say, is a monotone decreasing function of

$W_T$. In consequence, the approximate posterior odds ratio is a monotone decreasing function of the $p$-value. In turn, the posterior probability of $H_0$ is a monotone increasing function of the $p$-value. (It will not equal the $p$-value in general.) More specifically, let $P_{W_T}$ denote the $p$-value of the test based on $W_T$. Then, by definition, $k_{p,P_{W_T}} = W_T$. Hence, the approximate posterior odds statistic equals $PO(k_{p,P_{W_T}}, \mu)$, which is a monotone decreasing function of $P_{W_T}$.

We now analyze the correspondence between classical and posterior odds tests more closely for two particular families of prior distributions $\mu$ on $\|h\|_{\mathcal{I}}$. The first family corresponds to priors $Q_\mu \sim N(0, \tau\Sigma)$ on $h$ for different $\tau > 0$, or equivalently, to priors $\theta \sim N(\theta_0, (\tau/T)\Sigma)$. The prior distributions $\mu$ in this case equal $\sqrt{\tau\chi_p^2}$ . Equation (5.4) holds for such priors if

$$(5.5)\qquad (1+\tau)^{-p/2}\exp\left(\frac{1}{2}\ \frac{\tau}{1+\tau}k_{p,\alpha}\right) = \frac{\pi}{1-\pi}$$

for the case of two-sided tests with $p \geq 1$, and if

$$(5.6)\qquad (1+\tau)^{-1/2}\exp\left(\frac{1}{2}\ \frac{\tau}{1+\tau}k_{1,\alpha}\right)2\Phi\left(\left(\frac{\tau}{1+\tau}k_{1,\alpha}\right)^{1/2}\right) = \frac{\pi}{1-\pi}$$

for the case of one-sided tests.

Table 1 provides the values of $\tau$ that solve (5.5) and (5.6) when $\pi = 1/2$ for a variety of different values of $\alpha$ and $p$. The table shows that as $\alpha$ increases (so that the classical test rejects more frequently), the value of $\tau$ that yields an (approximately) equivalent posterior odds test decreases.

To illustrate the use of Table 1, consider a situation where the upper-left $p \times p$ block of $\mathcal{I}^{-1}$ equals $I_p$. Then, a posterior odds test with priors $\pi = 1/2$ and $\beta \sim N(0, 41/T)$, where $T$ is the sample size, corresponds (approximately) to a two-sided classical test with significance level $\alpha = .05$. A posterior odds test with $\pi = 1/2$ and $\beta \sim |Z_T|$ for $Z_T \sim N(0, 50/T)$ corresponds to a one-sided classical test with significance level $\alpha = .05$.

Next, we consider the family of prior distributions $\mu$ on $\|h\|_{\mathcal{I}}$ that equal pointmass at $r_*$ for different $r_* > 0$. Such distributions correspond to the prior on $\theta$ being given by the distribution of $\theta_0 + T^{-1/2}A(A'\mathcal{I}A)^{-1}\xi r^*$ and the prior on $\beta$ being given by the distribution of

$T^{-1/2}(H\mathcal{I}^{-1}H')^{-1/2}\xi r_*$, where $\xi$ is uniformly distributed on $\mathcal{S}_p$. (Note that the variance of $\beta$ under this prior is $(r_*^2/T)H\mathcal{I}^{-1}H'$, which equals $r_*^2$ times the asymptotic variance of the ML estimator $\hat{\beta}$.) Equation (5.4) holds for such priors if

$$\exp(-r_*^2/2)g_p(k_{p,\alpha}r_*^2) = \frac{\pi}{1-\pi} . \tag{5.7}$$

In particular, for the case of one-sided tests, this reduces to

$$\exp(-r_*^2/2 + \sqrt{k_{1,\alpha}}\ r_*) = \frac{\pi}{1-\pi} . \tag{5.8}$$

Equation (5.8) can be solved analytically to yield $r_* = (k_{1,\alpha})^{1/2}$ $\pm (k_{1,\alpha} - 2\log(\pi/(1-\pi)))^{1/2}$ provided $r_* > 0$. Thus, for one-sided tests with $\pi = 1/2$, we have $r_* = 2\sqrt{k_{1,\alpha}}$. Here, $\sqrt{k_{1,\alpha}}$ equals the critical value for the Wald $t$-statistic. Thus, one needs the prior on $\beta$ to put pointmass at twice the critical value for the $t$-statistic times $\sigma_\beta = (H\mathcal{I}^{-1}H'/T)^{1/2}$ in order for the posterior odds test and significance level $\alpha$ tests to be equivalent (approximately), where $\sigma_\beta$ is the asymptotic standard error of the ML estimator $\hat{\beta}$. In particular, for $\alpha = .01, .05, .10$, and $.25$, one has $r_*$ equal to $\sigma_\beta$ times 4.65, 3.29, 2.56, and 1.35, respectively.

### *5.3. Bayesian Versus Classical Tests That Are Impartial Between the Null and a Given Alternative*

Suppose one wants a test that treats the null and a particular alternative distribution impartially. A Bayesian test can be constructed in this case by specifying a pointmass prior distribution at the alternative distribution of interest under $H_1$ and by taking the prior probability of $H_1$ to equal that of $H_0$, i.e., $\pi = 1/2$. In contrast, a classical test can be constructed that has the property that its probability of type I error (significance level) equals its probability of type II error for the alternative distribution of interest.

How do these Bayesian and classical tests compare? Using the results above and those of Andrews (1989), we find that for one-sided alternatives the two tests are asymptotically equivalent and for two-sided tests they are not asymptotically equivalent, but are quite close.

First, consider one-sided alternatives. Suppose the alternative distribution of interest has

$\beta$ equal to $\beta_*$. Then, for a Bayesian posterior odds test, a pointmass prior distribution on $\beta_*$ under $H_1$ corresponds to the distribution $\mu$ introduced above being pointmass at $r_* = \beta_*/\sigma_\beta$, where $\sigma_\beta = (H\mathcal{I}^{-1}H'/T)^{1/2}$. For this prior, the (approximate) posterior odds test rejects $H_0$ if $W_T > k^*_{1,\alpha}$, where $k^*_{1,\alpha}$ solves (5.8) with $\pi = 1/2$, i.e., $k^*_{1,\alpha} = (r_*/2)^2 = \beta_*^2/(4\sigma_\beta^2)$.

On the other hand, asymptotic inverse power results in Andrews (1989) give the (approximate) magnitude of $\beta$ for which the probability of type II error of a Wald, LM, or LR test of level $\alpha$ equals $\alpha$. The magnitude is $2z_\alpha\sigma_\beta$, where $z_\alpha$ is the $(1-\alpha)$-th quantile of the standard normal distribution. In consequence, for a test to have significance level equal to the probability of type II error against $\beta = \beta^*$, one needs $\alpha$ to satisfy $\beta_* = 2z_\alpha\sigma_\beta$ or $z_\alpha = \beta_*/(2\sigma_\beta)$. A Wald $t$-test rejects if the $t$-statistic exceeds $z_\alpha = \beta_*/(2\sigma_\beta)$, or equivalently, if $W_T$ (which equals the squared $t$-statistic times the sign of the $t$-statistic) exceeds $(\beta_*^2/(4\sigma_\beta^2)$. Thus, the impartial posterior odds and classical tests are equivalent asymptotically.

Next, for two-sided tests, similar calculations can be made, but one does not find that the impartial posterior odds and classical tests are exactly equivalent asymptotically. Nevertheless, they do not differ greatly. For example, suppose $p = 1$ and $\beta_*$ and $\sigma_\beta$ are such that $\beta_*/\sigma_\beta = 3.605$. Then, by Table 1 of Andrews (1989), the classical test with size equal to one minus power at $\beta_*$ has $\alpha = .05$. This asymptotic test rejects if $W_T > 3.84$, using the $\chi_1^2$ table. For the impartial posterior odds test, on the other hand, $r_* = \beta_*/\sigma_\beta = 3.605$ corresponds to a critical value $k_{1,\alpha} = 3.98$, according to equation (5.7). That is, the approximate posterior odds test rejects if $W_T$ (or $LM_T$ or $LR_T$) exceeds 3.98. Since the critical values 3.98 and 3.84 are quite close, the posterior odds and classical tests are quite close asymptotically. If $\beta_*$ and $\sigma_\beta$ are such that $r_* = \beta_*/\sigma_\beta = 4.902$ (or 2.926), then the impartial classical test must have $\alpha$ equal to .01 (.10 respectively) and the asymptotic critical values of the impartial posterior odds and classical tests are 6.72 and 6.63 (2.89 and 2.71 respectively). In each case, the impartial posterior odds and classical tests are quite close asymptotically.

We now add several caveats to the results described in this subsection. First, the results are established only for standard asymptotic scenarios, where the ML estimator is asymptotically normal. They also should apply if the asymptotic distribution of the ML estimator is a location shift family of symmetric distributions, such as mixed normal distributions. On the other hand, they cannot be expected to hold for tests involving unit root parameters in models with stochastic trends. Second, the results depend on the difference between the prior on the nuisance parameter $\delta$ under $H_0$ and its prior under $H_1$ going to zero as $T \to \infty$. If one had different priors on $\delta$ asymptotically under the two hypotheses, then the results would not hold. (In finite samples, this cause of a difference between impartial Bayesian and classical tests can be illustrated by considering tests of $H_0 : Y \sim N(0, \sigma_1^2)$ versus $H_1 : Y \sim N(\beta, \sigma_2^2)$ where $\sigma_1^2 \neq \sigma_2^2$). Third, for two-sided tests, the results depend on the restrictions on the prior on $\beta$ under $H_1$ (as specified in Assumption 5(c)).

### *5.4. Tests Based on Expected Posterior Losses*

In this section, we show that the results above, which show a correspondence between Bayesian posterior odds tests and classical tests, also provide a correspondence between Bayesian posterior expected loss tests and classical tests.

Let $L(\theta, H_0)$ (resp., $L(\theta, H_1)$) denote the loss when $\theta$ is the true parameter and $H_0$ (resp. $H_1$) is chosen. By assumption, $L(\theta, H_j) \geq 0$ and $L(\theta, H_j) = 0 \ \forall\theta$ in $H_j$ for $j = 0, 1$. A Bayesian posterior expected loss test rejects $H_0$ (i.e., chooses $H_1$) if the posterior expected loss of $H_1$ is less than that of $H_0$, or equivalently, if the ratio of the posterior expected loss of $H_0$ over that of $H_1$, denoted $REPL_T$, is greater than 1. A Bayesian posterior odds test is a special case for which $L(\theta, H_1) = L(\theta', H_0)$ $\forall\theta$ in $H_0$ and $\forall\theta'$ in $H_1$.

Let $L(\theta_0, H_1) = 1$ without loss of generality. Suppose one takes $L(\theta, H_0)$ and the prior on $\theta$ under $H_1$ to be such that their product equals the prior on $\theta$ under $H_1$ specified in Assumption 5 above. Then, under the other assumptions above, $REPL_T - PO(W_T, \mu) \xrightarrow{p} 0$ under $H_0$ and under $H_1$, and likewise for $LM_T$ and $LR_T$, where $PO(W_T, \mu)$ is as defined in (2.3) except with $(1-\pi)/\pi$ replaced by $\int L(\theta_0 + T^{-1/2}h, H_0)dQ_\mu(h)(1-\pi)/\pi$. This result gives a direct correspondence between classical tests and certain posterior expected loss tests.[5]

## 6. NONLINEAR MODELS

In this section, we consider nonlinear dynamic models. We provide primitive assumptions that are sufficient for Assumptions 1-4 of Section 3. For simplicity, we consider strictly stationary $m$-th order Markov models. With some additional complexity in the assumptions, the results could be extended to allow for non-Markov models with non-stationary non-trending random variables.

The sample of observations is given by

$$Y_T = \{(S_t, X_t) : t \le T\} , \tag{6.1}$$

where $\{S_t : t \le T\}$ are endogenous variables and $\{X_t : t \le T\}$ are weakly exogenous variables.[6] Let

$$\{g_t(\theta) : \theta \in \Theta\} = \{g_t(S_t|S_1, ..., S_{t-1}; X_1, ..., X_t) : \theta \in \Theta\} \tag{6.2}$$

denote a parametric family of conditional densities (with respect to some measure $\lambda$) of $S_t$ given $S_1, ..., S_{t-1}, X_1, ..., X_t$ evaluated at the random variables $S_1, ..., S_t, X_1, ..., X_t$, where $\Theta \subset R^s$. Let

$$h_t = h_t(X_t|S_1, ..., S_{t-1}; X_1, ..., X_{t-1}) \tag{6.3}$$

denote the conditional density (with respect to some measure) of $X_t$ given $S_1, ..., S_{t-1}, X_1, ..., X_{t-1}$ evaluated at the random variables $S_1, ..., S_{t-1}, X_1, ..., X_t$. By the assumption of weak exogeneity, $h_t$ does not depend on $\theta$.

The likelihood function $f_T(\theta)$ and log likelihood function $\ell_T(\theta)$ are given by

$$f_T(\theta) = \prod_{t=1}^{T} g_t(\theta) \cdot \prod_{t=1}^{T} h_t \quad \text{and} \quad \ell_T(\theta) = \Sigma_1^T \log g_t(\theta) + \Sigma_1^T \log h_t , \tag{6.4}$$

where $\Sigma_1^T$ denotes $\Sigma_{t=1}^T$. The function $\mathcal{I}(\theta)$ equals $-E \frac{\partial^2}{\partial\theta\partial\theta'} \log g_t(\theta)$.

We consider the case where $\{(S_t, X_t) : t \ge 1\}$ is part of a doubly infinite strictly stationary ergodic sequence $\{(S_t, X_t) : t = ..., 0, 1, ...\}$ and $\{S_t : t = ..., 0, 1, ...\}$ is $m$-th order Markov for some integer $m \ge 0$. By definition, $\{S_t : t = ..., 0, 1, ...\}$ is $m$-th *order Markov* if the conditional distribution of $S_t$ given $\mathcal{F}_{t-1} = \sigma(..., S_{t-2}, S_{t-1}; ..., X_{t-1}, X_t\}$ equals the conditional distribution of $S_t$ given $S_{t,m} = (S_{t-m}, ..., S_{t-1})$ and $X_{t,m} = (X_{t-m}, ..., X_t)$ for all $t$. The Markov assumption yields the simplification that the summands $\log g_t(\theta)$ in the log-likelihood function are strictly stationary and ergodic for $t > m$. Without the Markov assumption this would not be the case, because the number of relevant observed variables in the conditioning set would vary with $t$.

The following assumption provides primitive sufficient conditions for Assumptions 1-4 of Section 3:

ASSUMPTION NL: (a) *$\Theta$ is compact and $\theta_0$ lies in the interior of $\Theta$.*

(b) *$\{(S_t, X_t) : t = ..., 0, 1, ...\}$ is strictly stationary and ergodic and $\{S_t : t = ..., 0, 1, ...\}$ is $m$-th order Markov under $\theta_0$.*

(c) *$g_t(\theta)$ is continuous in $\theta$ on $\Theta$ and twice continuously partially differentiable in $\theta$ on $\Theta_0$ with probability one under $\theta_0$, where $\Theta_0$ is some compact set that contains a neighborhood of $\theta_0$.*

(d) *$g_t(\theta) \neq g_t(\theta_0)$ with positive probability under $\theta_0$ $\forall\theta \in \Theta$ with $\theta \neq \theta_0$.*

(e) $E \sup_{\theta\in\Theta} |\log g_t(\theta)| < \infty$, $E \sup_{\theta\in\Theta_0} \| \frac{\partial}{\partial\theta}\log g_t(\theta)\| < \infty$, $E\|\frac{\partial}{\partial\theta}\log g_t(\theta_0)\|^2 < \infty$, *and*

$E \sup_{\theta\in\Theta_0} \| \frac{\partial^2}{\partial\theta\partial\theta'}\log g_t(\theta)\| < \infty$.

(f) $\mathcal{I} = -E\frac{\partial^2}{\partial\theta\partial\theta'}\log g_t(\theta_0)$ *is positive definite.*

Assumption NL constitutes a fairly standard set of ML regularity conditions for stationary and ergodic situations.

LEMMA 2: *Assumption* NL *implies Assumptions* 1-4.

Thus, Assumptions NL and 5 are sufficient for the result of Theorem 1 to hold.

# 7. RELATED RESULTS IN THE LITERATURE

Here we discuss some asymptotic results for posterior odds tests due to Jeffreys (1961, pp. 246-47, 249-50), Lindley (1961), Schwarz (1978), Kass and Vaidyanathan (1992), and Phillips (1992). In particular, we focus on the relationship between these results and the results presented above.

Kass and Vaidyanathan (1992, eqn. (2.3)) (hereafter denoted KV) provide a family of approximations to the PO statistic that depend on the choice of certain functions $b_0$ and $b$. Their approximations do not yield the approximate PO statistic to be a monotone function of $W_T$, $LM_T$, or $LR_T$. Their approximations are closest to being such a function, however, if one takes $b_0$ and $b$ to equal the priors on the parameters under $H_0$ and under $H_1$ respectively.

In this case, their approximate PO statistic in favor of $H_1$ is given by

$$\text{(7.1)} \qquad \frac{1-\pi}{\pi}\,\frac{\pi_1(\hat{\theta})}{\pi_0(\tilde{\delta})}(2\pi)^{p/2}\frac{\det(T[\mathcal{I}_T(\tilde{\theta})]_{22})^{1/2}}{\det(T\mathcal{I}_T(\hat{\theta}))^{1/2}}\exp\left(\frac{1}{2}LR_T\right),$$

where $\pi_1(\cdot)$ is the prior on $\theta$ under $H_1$, $\pi_0(\cdot)$ is the prior on $\delta$ under $H_0$, $\pi$, $\hat{\theta}$, and $\tilde{\theta}$ are as above, $[\mathcal{I}_T(\tilde{\theta})]_{22}$ is the lower right $q \times q$ sub-matrix of $\mathcal{I}_T(\tilde{\theta})$, and $\det(A)$ denotes the determinant of the matrix $A$. This formula reduces to that given by Lindley (1961) for the case where $p = 1$. It is similar to, but different from, the approximation given by Jeffreys (1961) for the case $p = 1$.

KV show that the approximation (7.1) is valid to within a multiplicative error of $O(1/T)$ (i.e., the ratio of exact to approximate PO equals $1 + O(1/T)$) with probability one under $H_0$ and under $H_1$ using asymptotics that employ the same prior for all $T$.

A direct comparison of KV's approximations to those given in this paper is not straightforward, because different asymptotics are employed. Under KV's asymptotics, the PO statistic has limit zero or infinity depending on whether $H_0$ or $H_1$ is true. In consequence, multiplicative approximation errors are considered and the prior only affects the approximation in a relatively crude fashion. With the asymptotics used in this paper, the PO statistic has a nongenderate limit as $T \to \infty$, so additive approximation errors are considered and the whole prior affects the approximation.

KV's multiplicative approximation errors are $O(1/T)$, which is of second order and is quite desirable. This does not necessarily translate into additive approximation errors of $o(1)$ under $H_1$, however, because the true PO statistic diverges to infinite very quickly under $H_1$ using their asymptotics. On the other hand, the additive approximation errors of the approximations given in this paper are necessarily $o_p(1)$ under $H_0$ and under $H_1$ using our asymptotics, but the corresponding multiplicative approximation errors are not $O_p(1/T)$ in general. It seems likely that the approximations given here are more accurate than KV's in some cases and less accurate in others. The former is known to be true, since the approximations given here are exact for linear regression models with iid normal errors and known variance for a wide variety of priors, whereas KV's approximations are not exact in these cases.

Next, we consider Schwarz's (1978) asymptotic results for Bayesian model selection procedures. His results yield the following approximation to the PO statistic in favor of $H_1$:

$$T^{-p/2}\exp\left(\frac{1}{2}LR_T\right) . \tag{7.2}$$

The same formula applies for all prior probabilities $\pi \in (0, 1)$ of $H_0$, for all (proper) priors over the values of $\theta$ in $H_0$, and for all priors over those values of $\theta$ in $H_1$. Obviously, his approximation can be quite crude given its extremely broad range of applicability. (One can make the PO statistic take any value in $(0, \infty)$ by suitable choice of $\pi$. Thus, it is clear that Schwarz's approximation does not hold uniformly in $\pi$ and can have arbitrarily poor accuracy.) This crudeness of approximation is both a virtue and a drawback. It is a virtue, if one wants to construct an asymptotic procedure that is independent of the prior, as does Schwarz. It is a drawback, however, if one wants a reasonably accurate approximation of the PO statistic, as is desired here.

Schwarz's results apply to a subclass of the models considered in this paper, viz., iid linear exponential models. On the other hand, his results apply to a broader class of hypotheses. In addition to considering hypotheses of the form $H_0 : \beta = 0$ versus $H_1 : \beta \neq 0$, which correspond to choosing between nested models of different dimensions, he also considers choosing between non-nested iid linear exponential models. The type of asymptotics considered by Schwarz differs from that considered here. He fixes the values of the (normalized) sufficient statistics of the linear exponential model for all $T$. Thus, in his asymptotics, nothing is random.

Schwarz's approximations are valid in the sense that the ratio of the true PO statistic to the approximate posterior odds statistic (6.2) is bounded away from zero and infinity for all $T$. Obviously, this is a very weak approximation result. For the purposes of obtaining reasonably accurate approximations, one would like the ratio to have limit one as $T \to \infty$, as with KV's approximations.

Last, we discuss some results of Phillips (1992, Remark 3.2(ii)). Phillips' results are similar to those given here in that they establish an asymptotic correspondence between Bayesian posterior odds and classical tests. His results differ, however, in terms of the priors considered, the models considered, and the choice of distributions under which the asymptotics hold. Thus, his results and those given here are more complements than substitutes.

More specifically, Phillips' results and ours differ as follows: (i) Phillips discusses the asymptotic correspondence when the asymptotics are derived under the null, whereas we consider the asymptotic correspondence under the null and also under the alternative, (ii) Phillips considers a non-informative (improper) prior, whereas we consider classes of proper priors, and (iii) Phillips considers linear vector autoregressive models that are stationary under the null and possibly non-stationary under the alternative, whereas we consider nonlinear models with random variables that are nontrending under the null and the alternative.[7]

The discussion above shows that there are several useful asymptotic approximations in the literature for PO statistics. Hopefully, it also indicates that the approximation results of Sections 2 and 5 provide value-added beyond those results currently in the literature.

## APPENDIX

Without loss of generality, we set the prior probability $\pi$ of $H_0$ to be 1/2 throughout the Appendix. This simplifies notation, since the multiplicative factor $(1-\pi)/\pi$ reduces to 1. The proofs of Lemmas 1 and 2 follow those of Theorem 1. We start by stating an assumption, several definitions, and four lemmas that are used in the proof of Theorem 1.

ASSUMPTION 1′: $T^{-1/2}D\ell_T(\theta_0) = O_p(1)$ *under* $\theta_0$.

Assumption 1′ is implied by Assumption 2. Assumption 1′ is introduced, so that it is evident below where the full strength of Assumption 2 is used in the proofs.

The unrestricted ML estimator $\hat{\theta}$, suitably shifted and scaled, can be approximated under $\theta_0$ by the score function $D\ell_T(\theta_0)$ suitable scaled. We refer to the latter as the *approximate* ML estimator $\bar{\theta}$. By definition,

$$\bar{\theta} = \mathcal{I}^{-1}T^{-1/2}D\ell_T(\theta_0) \; . \tag{A.1}$$

LEMMA A-1: *Suppose Assumptions* 1, 1′, *and* 3 *hold. Then,* $T^{1/2}(\hat{\theta} - \theta_0) - \bar{\theta} \xrightarrow{p} 0$ *under* $\theta_0$.

Next, we define an unobserved large sample approximation to the posterior-odds statistic $PO_T(Q_\mu)$. This approximation is based on the approximate ML estimator $\bar{\theta}$. Let

$$\overline{PO}_T(Q_\mu) = \int \exp\left[-\frac{1}{2}(\bar{\theta}-h)'\mathcal{I}(\bar{\theta}-h)\right]dQ_\mu(h)/\exp\left[-\frac{1}{2}\bar{\theta}'\mathcal{I}\bar{\theta}\right] . \tag{A.2}$$

LEMMA A-2: *Suppose Assumptions* 1, 1′ *and* 3 *hold. Then,* $PO_T(Q_\mu) - \overline{PO}_T(Q_\mu) \xrightarrow{p} 0$ *under* $\theta_0$.

For the case of two-sided tests, we define an *approximate Wald* statistic $\bar{W}_T$ based on the approximate ML estimator $\bar{\theta}$ by

$$\bar{W}_T = (H\bar{\theta})'\left(H\mathcal{I}^{-1}H'\right)^{-1}H\bar{\theta} \; . \tag{A.3}$$

For the case of one-sided tests, we let $\bar{W}_T^*$ equal the right-hand side of (A.3) and we define $\bar{W}_T$ to equal $\bar{W}_T^* \operatorname{sgn}(H\bar{\theta})$.

The approximate posterior-odds statistic $\overline{PO}_T(Q_\mu)$ simplifies to a simple function of the approximate Wald statistic:

LEMMA A-3: *Suppose Assumption 5 holds. Then,* $\overline{PO}_T(Q_\mu) = PO(\bar{W}_T, \mu)$.

Combining Lemmas A-2 and A-3, one sees that the results of Theorem 1 under $\theta_0$ are equivalent to

$$PO(\bar{W}_T, \mu) - PO(W_T, \mu) \xrightarrow{p} 0 \quad \text{under} \quad \theta_0 \tag{A.4}$$

and likewise with $W_T$ replaced by $LM_T$ and $LR_T$. The latter results are relatively straightforward to establish under Assumptions 1-5 (see below).

To obtain the results of Theorem 1 under $H_1$, it suffices to show that they hold under $\theta_0$ and that the alternative marginal densities $\{\int f_T(\theta_0 + T^{-1/2}h)dQ_\mu(h) : T \geq 1\}$ of the data vectors $\{Y_T : T \geq 1\}$, under the parametric model $\{f_T(\theta) : \theta \in \Theta\}$ and the prior $Q_\mu$, are contiguous to the null densities $\{f_T(\theta_0) : T \geq 1\}$ of $\{Y_T : T \geq 1\}$.

LEMMA A-4: *Suppose Assumptions 1-5 hold. Then, the alternative marginal densities* $\{\int f_T(\theta_0 + T^{-1/2}h)dQ_\mu(h) : T \geq 1\}$ *of the data vectors* $\{Y_T : T \geq 1\}$ *are contiguous to the null densities* $\{f_T(\theta_0) : T \geq 1\}$.

PROOF OF LEMMA A-1: All probability calculations in this proof are made "under $\theta_0$." By Assumptions 1(a), 1(b), and 3 and the definition of $\hat{\theta}$, $D\ell_T(\hat{\theta}) = 0$ wp $\rightarrow$ 1. Hence, by one-term Taylor expansions of the elements of $D\ell_T(\hat{\theta})$ about $\theta_0$ we get, wp $\rightarrow$ 1,

$$\begin{aligned} 0 &= T^{-1/2}D\ell_T(\hat{\theta}) = T^{-1/2}D\ell_T(\theta_0) - \mathcal{I}_{1T}T^{1/2}(\hat{\theta} - \theta_0)\,, \quad \text{where} \\ \mathcal{I}_{1T} &= -\int_0^1 T^{-1}D^2\ell_T(\theta_0 + \lambda(\hat{\theta} - \theta_0))d\lambda\,. \end{aligned} \tag{A.5}$$

The matrix $\mathcal{I}_{1T}$ satisfies

$$\|\mathcal{I}_{1T} - \mathcal{I}\| \le \left\| \int_0^1 [-T^{-1}D^2\ell_T(\theta_0 + \lambda(\hat{\theta} - \theta_0)) - \mathcal{I}(\theta_0 + \lambda(\hat{\theta} - \theta_0))]d\lambda \right\|$$

(A.6)
$$+ \left\| \int_0^1 [\mathcal{I}(\theta_0 + \lambda(\hat{\theta} - \theta_0)) - \mathcal{I}]d\lambda \right\|$$

$$= o_p(1) ,$$

where the equality holds using Assumptions 1(c) and 3 for the first term and Assumptions 1(d) and 3 for the second term. Equation (A.6) and Assumption 1(e) yield $\|\mathcal{I}_{1T}^{-1} - \mathcal{I}^{-1}\| = o_p(1)$. This result, (A.5), and Assumptions 1(e) and 1′ give

(A.7)
$$o_p(1) = \|T^{1/2}(\hat{\theta} - \theta_0) - \mathcal{I}_{1T}^{-1}T^{-1/2}D\ell_T(\theta_0)\|$$
$$= \|T^{1/2}(\hat{\theta} - \theta_0) - \mathcal{I}^{-1}T^{-1/2}D\ell_T(\theta_0)\| + o_p(1) . \quad \square$$

PROOF OF LEMMA A-2: All probability calculations in this proof are made "under $\theta_0$." For $0 < M < \infty$, define

(A.8)
$$PO_{TM} = \int_{\|h\| \le M} f_T(\theta_0 + T^{-1/2}h)dQ_\mu(h)/f_T(\theta_0) \quad \text{and}$$

(A.9)
$$\overline{PO}_{TM} = \int_{\|h\| \le M} \exp\left( -\frac{1}{2}(\bar{\theta}-h)'\mathcal{I}(\bar{\theta}-h) \right) dQ_\mu(h)/\exp\left( -\frac{1}{2}\bar{\theta}'\mathcal{I}\bar{\theta} \right) .$$

For any $\varepsilon > 0$,

(A.10)
$$P(|PO_T(Q_\mu) - \overline{PO}_T(Q_\mu)| > \varepsilon) \le P(|PO_T(Q_\mu) - PO_{TM}| > \varepsilon)$$
$$+ P(|PO_{TM} - \overline{PO}_{TM}| > \varepsilon) + P(|\overline{PO}_T(Q_\mu) - \overline{PO}_{TM}| > \varepsilon) .$$

Hence, it suffices to show that (1) given any $\eta > 0$ we can choose $T^* < \infty$ and $M < \infty$ sufficiently large so that $P(|PO_T(Q_\mu) - PO_{TM}| > \varepsilon) < \eta$ and $P(|\overline{PO}_T(Q_\mu) - \overline{PO}_{TM}| > \varepsilon) < \eta$ for all $T \ge T^*$ and (2) $PO_{TM} - \overline{PO}_{TM} \xrightarrow{p} 0 \ \forall 0 < M < \infty$.

We show (1) first. We have

$$(A.11)\quad \begin{aligned} &P(|PO_T(Q_\mu) - PO_{TM}| > \varepsilon) \le \varepsilon^{-1}E|PO_T(Q_\mu) - PO_{TM}| \\ &= \varepsilon^{-1}E\int_{|h|>M} [f_T(\theta_0 + T^{-1/2}h)/f_T(\theta_0)]dQ_\mu(h) = \varepsilon^{-1}\int_{|h|>M} dQ_\mu(h) , \end{aligned}$$

where the second equality holds by Fubini's Theorem and the fact that $E[f_T(\theta_0 + T^{-1/2}h)/f_T(\theta_0)] = 1 \quad \forall h$. The right-hand side (rhs) of (A.11) can be made arbitrarily small for all $T$ by taking $M$ large.

Next, we have

$$(A.12)\quad \begin{aligned} |\overline{PO}_T(Q_\mu) - \overline{PO}_{TM}| &= \exp\left[\frac{1}{2}\bar{\theta}'\mathcal{I}\bar{\theta}\right]\int_{|h|>M} \exp\left[-\frac{1}{2}(\bar{\theta}-h)'\mathcal{I}(\bar{\theta}-h)\right]dQ_\mu(h) \\ &\le \exp\left[\frac{1}{2}\|T^{-1/2}D\ell_T(\theta_0)\|^2\cdot\|\mathcal{I}^{-1}\|\right]\int_{|h|>M} dQ_\mu(h) , \end{aligned}$$

where the inequality uses Assumption 1(e). The first term on the rhs of (A.12) is $O_p(1)$ by Assumptions 1(e) and 1′ and the second term on the rhs can be made arbitrarily small by taking $M$ large.

We now establish (2). A two term Taylor series expansion gives

$$(A.13)\quad \ell_T(\theta_0 + T^{-1/2}h) - \ell_T(\theta_0) = h'T^{-1/2}D\ell_T(\theta_0) + \frac{1}{2}h'T^{-1}D^2\ell_T(\theta_0)h + r_{1T}(h) ,$$

where the remainder term $r_{1T}(h)$ satisfies

$$(A.14)\quad \begin{aligned} \sup_{|h|\le M} |r_{1T}(h)| &\le M^2 \sup_{\theta:\|T^{1/2}(\theta-\theta_0)\|\le M} \|T^{-1}D^2\ell_T(\theta) - T^{-1}D^2\ell_T(\theta_0)\| \\ &\le M^2 \sup_{\theta\in\Theta_0}\|T^{-1}D^2\ell_T(\theta) - \mathcal{I}(\theta)\| + M^2 \sup_{\theta:\|T^{1/2}(\theta-\theta_0)\|\le M} \|\mathcal{I}(\theta) - \mathcal{I}(\theta_0)\| \\ &\quad + M^2\|T^{-1}D^2\ell_T(\theta_0) + \mathcal{I}(\theta_0)\| \\ &= o_p(1) , \end{aligned}$$

where the equality uses Assumptions 1(c) and (d). In addition,

$$(A.15)\quad h'T^{-1}D^2\ell_T(\theta_0)h = -h'\mathcal{I}h + r_{2T}(h) , \text{ where } \sup_{h:|h|\le M} |r_{2T}(h)| = o_p(1) ,$$

by Assumption 1(c). It follows from (A.14) and (A.15) that

$$\text{(A.16)} \qquad \exp[r_{1T}(h) + r_{2T}(h)] = 1 + s_T(h)\,, \quad \text{where} \quad \sup_{h:\|h\|\le M} |s_T(h)| = o_p(1)\,.$$

Combining (A.13) and (A.15) and using the definition of $\bar{\theta}$ yields

$$\text{(A.17)} \qquad \begin{aligned} \ell_T(\theta_0 + T^{-1/2}h) - \ell_T(\theta_0) &= h'\mathcal{I}\bar{\theta} - \frac{1}{2}h'\mathcal{I}h + r_{1T}(h) + r_{2T}(h) \\ &= \frac{1}{2}\bar{\theta}'\mathcal{I}\bar{\theta} - \frac{1}{2}(\bar{\theta}-h)'\mathcal{I}(\bar{\theta}-h) + r_{1T}(h) + r_{2T}(h)\,. \end{aligned}$$

Combining (A.7), (A.8), (A.16), and (A.17) gives

$$\text{(A.18)} \qquad \begin{aligned} PO_{TM} &= \int_{\|h\|\le M} \exp\left[\ell_T(\theta_0 + T^{-1/2}h) - \ell_T(\theta_0)\right] dQ_\mu(h) \\ &= \int_{\|h\|\le M} \exp\left[\frac{1}{2}\bar{\theta}'\mathcal{I}\bar{\theta} - \frac{1}{2}(\bar{\theta}-h)'\mathcal{I}(\bar{\theta}-h)\right](1 + s_T(h))dQ_\mu(h) \\ &= \overline{PO}_{TM} + o_p(1)\,, \end{aligned}$$

where the third equality uses $\overline{PO}_{TM} = O_p(1)$, which follows from a close analogue to (A.12). This completes the proof. □

PROOF OF LEMMA A-3: Let $P$ and $P^\perp$ denote the projection matrices onto $V$ and $V^\perp$, respectively, with respect to $\langle\cdot,\cdot\rangle_{\mathcal{I}}$. The projection matrix $P^\perp$ is given by $P^\perp = AH$, where $A$ and $H$ are as in the text. To see this, note that $HA = I_p$ and, hence, $(AH)(AH) = AH$ and $AH$ is an oblique projection matrix. Furthermore, for $v = (0', \delta')' \in V$, $AHv = 0$, so $AH$ projects onto a space orthogonal to $V$. On the other hand, for $m = (m_1', m_2')' \in V^\perp$, $v'\mathcal{I}m = 0$ $\forall v \in V$ iff $[0 \,\vdots\, I_q]\,\mathcal{I}m = 0$ iff $[\mathcal{I}_2' \,\vdots\, \mathcal{I}_3]m = 0$ iff $m_2 = -\mathcal{I}_3^{-1}\mathcal{I}_2'm_1$ iff $m = Am_1$. In consequence, $AHm = AHAm_1 = Am_1 = m$ $\forall m \in V^\perp$. That is, $AH$ projects onto the entire orthogonal complement of $V$ with respect to $\langle\cdot,\cdot\rangle_{\mathcal{I}}$.

Now, the integrand of $\overline{PO}_T(Q_\mu)$ can be rewritten as follows:

$$\text{(A.19)} \qquad \begin{aligned} \frac{1}{2}\bar{\theta}'\mathcal{I}\bar{\theta} - \frac{1}{2}(h-\bar{\theta})'\mathcal{I}(h-\bar{\theta}) &= -\frac{1}{2}h'\mathcal{I}h + h'\mathcal{I}\bar{\theta} = -\frac{1}{2}h'\mathcal{I}h + h'\mathcal{I}P^\perp\bar{\theta} \\ &= \frac{1}{2}(P^\perp\bar{\theta})'\mathcal{I}P^\perp\bar{\theta} - \frac{1}{2}(h - P^\perp\bar{\theta})'\mathcal{I}(h - P^\perp\bar{\theta})\,, \end{aligned}$$

where the second equality holds because $h \in V^\perp$ implies $h'\mathcal{I}P\bar{\theta} = 0$.

Define $Z = (A'\mathcal{I}A)^{1/2}H\bar{\theta}$. Let $\xi$ be a random $p$-vector with uniform ($U$) distribution on the unit sphere, as in Assumption 5. Let $R$ be a real random variable independent of $\xi$ with distribution $\mu(\cdot)$. Let $\tilde{h} = A(A'\mathcal{I}A)^{-1/2}\xi$. Let $\tilde{Q}(\cdot)$ denote the distribution of $\tilde{h}$. Note that $\|\tilde{h}\|_{\mathcal{I}} = 1$. Define $h = R\tilde{h}$. Then, $h \sim Q_\mu(\cdot)$, as desired.

Using these definitions, (A.19), and $P^\perp = AH$, we obtain

$$\overline{PO}_T(Q_\mu) = \int \exp\left[\frac{1}{2}(AH\bar{\theta})'\mathcal{I}AH\bar{\theta} - \frac{1}{2}(h - AH\bar{\theta})'\mathcal{I}(h - AH\bar{\theta})\right]dQ_\mu(h)$$

$$= \int\int \exp\left[\frac{1}{2}Z'Z - \frac{1}{2}(rA(A'\mathcal{I}A)^{-1/2}\xi - AH\bar{\theta})'\mathcal{I}(rA(A'\mathcal{I}A)^{-1/2}\xi - AH\bar{\theta})\right]dU(\xi)d\mu(r)$$

$$(\text{A.20}) \quad = \int\int \exp\left[\frac{1}{2}Z'Z - \frac{1}{2}(r\xi - Z)'(r\xi - Z)\right]dU(\xi)d\mu(r)$$

$$= \int\int \exp\left[-\frac{1}{2}r^2\xi'\xi + r\xi'Z\right]dU(\xi)d\mu(r)$$

$$= \int \exp\left[-\frac{1}{2}r^2\right]\int \exp[\|Z\|r\xi'(Z/\|Z\|)]dU(\xi)d\mu(r)$$

$$= PO(X, \mu)\ ,$$

where $X = \|Z\|^2$ for two-sided tests and $X = \|Z\|^2\text{sgn}(HZ)$ for one-sided tests. The last equality of (A.20) holds by the definitions of $g_p(\kappa)$ and $PO(\cdot, \mu)$.

It remains to show that $\|Z\|^2 = \bar{W}_T$ for two-sided tests and $\|Z\|^2\text{sgn}(HZ) = \bar{W}_T$ for one-sided tests. Since $\|Z\|^2 = (H\bar{\theta})'A'\mathcal{I}AH\bar{\theta}$ and $\text{sgn}(HZ) = \text{sgn}(H\bar{\theta})$, the latter results hold provided $A'\mathcal{I}A = (H\mathcal{I}^{-1}H')^{-1}$. By simple algebra, $A\mathcal{I}A = \mathcal{I}_1 - \mathcal{I}_2\mathcal{I}_3^{-1}\mathcal{I}_2'$. On the other hand, $(H\mathcal{I}^{-1}H')^{-1}$ equals the inverse of the upper $p \times p$ submatrix of $\mathcal{I}^{-1}$, which equals $\mathcal{I}_1 - \mathcal{I}_2\mathcal{I}_3^{-1}\mathcal{I}_2'$ by the formula for a partitioned inverse. □

Proof of Lemma A-4: The following assertion is verified below using results of Strasser (1985). If (i) $PO_T(Q_\mu) \xrightarrow{d} PO(X, \mu)$ under $\theta_0$ for some random variable $X$ and (ii) $E[PO(X, \mu)] = 1$, then the result of Lemma A-4 holds. Condition (i) is obtained as follows: By Lemmas A-2 and A-3, $PO_T(Q_\mu) - PO(\bar{W}_T, \mu) \xrightarrow{p} 0$ under $\theta_0$. Next, $PO(\bar{W}_T, \mu) \xrightarrow{d} PO(X, \mu)$ under $\theta_0$, where $X = Z'Z$ for two-sided tests, $X = Z^2\text{sgn}(Z)$ for one-sided tests, and $Z \sim N(0, I_p)$, by Assumptions 1(e) and 2, continuity of $PO(\cdot, \mu)$, and the continuous mapping theorem. Condition (ii) is obtained as follows: For $X$ as above, by (A.20),

$$\text{(A.21)}\quad \begin{aligned} E[PO(X,\ \mu)] &= E\int\int \exp\left[-\frac{1}{2}r^2 + r\xi'Z\right]dU(\xi)d\mu(r) \\ &= \int\int \exp\left[-\frac{1}{2}r^2\right]E\ \exp[r\xi'Z]dU(\xi)d\mu(r) = \int\int \exp\left[-\frac{1}{2}r^2\right]\exp\left[\frac{1}{2}r^2\xi'\xi\right]dU(\xi)d\mu(r) = 1\ , \end{aligned}$$

where the second equality holds by Fubini's Theorem and the third uses the expression for the standard normal moment generating function.

It remains to verify the assertion above. Let $(\Omega, \mathcal{A})$ be a measurable space. Let $P_T$ and $Q_T$ be a null distribution and an alternative distribution on $(\Omega, \mathcal{A})$ for $T \geq 1$. Let $E_T = (\Omega, \mathcal{A}, (P_T, Q_T))$. $E_T$ is called a (binary) *experiment* and $\{E_T : T \geq 1\}$ is a sequence of experiments. One can define equivalent experiments and one can put a metric $\Delta_2$ on the space $\mathcal{E}_2/\sim$ of equivalence classes of experiments, see Strasser (1985, pp. 74, 75). By Theorem 18.11 of Strasser (1985), if $\Delta_2(E_T, E) \to 0$ as $T \to \infty$ for some experiment $E = (\Omega, \mathcal{A}, (P, Q))$, then $\{Q_T : T \geq 1\}$ is contiguous to $\{P_T : T \geq 1\}$ if and only if $Q$ is absolutely continuous with respect to $P$, i.e., if and only if $\int_0^\infty x\mu_E(dx) = 1$, where $\mu_E$ is the distribution of the likelihood ratio $dQ/dP$ under $P$, $\mathcal{L}(dQ/dP\,|P)$.

Also, by Theorem 16.8 of Strasser (1985), $(\mathcal{E}_2/\sim, \Delta_2)$ and $(\mathcal{M}, \mathcal{T})$ are homeomorphic, where $\mathcal{M}$ is the set of all probability measures $\mu$ on $[0, \infty)$ with $\int_0^\infty x\mu(dx) \leq 1$ and $\mathcal{T}$ is the topology of weak convergence, with homeomorphism $T$ defined by $T(\dot{E}) = \mathcal{L}(dQ/dP\,/P)$, where $\dot{E}$ is the equivalence class of experiments that contains $E$ and $E = (\Omega, \mathcal{A}, (P, Q))$. In consequence, for any experiment $E = (\Omega, \mathcal{A}, (P, Q))$ and any sequence of experiments $\{E_T : T \geq 1\} = \{(\Omega, A, (P_T, Q_T)) : T \geq 1\}$, $\Delta_2(E_T, E) \to 0$ as $T \to \infty$ if and only if $\mathcal{L}(dQ_T/dP_T/P_T) \Rightarrow \mathcal{L}(dQ/dP\,/P)$ as $T \to \infty$, where "$\Rightarrow$" denotes weak convergence (or equivalently, convergence in distribution). This result and the result of the previous paragraph establish the assertion above. □

PROOF OF THEOREM 1: Suppose parts (a)-(c) of Theorem 1 hold under $\theta_0$. Then, the probability of the set $\{|PO_T(Q_\mu) - PO(W_T, \mu)| > \varepsilon\}$ converges to zero as $T \to \infty$ under $\theta_0$ $\forall\varepsilon > 0$. By contiguity (Lemma A-4), its probability also converges to zero under $H_1$. The same holds with $W_T$ replaced by $LM_T$ and $LR_T$. Thus, it remains to establish parts (a)-(c) of Theorem 1 under $\theta_0$.

Given Lemmas A-2 and A-3, to establish parts (a)-(c) of Theorem 1 under $\theta_0$, it suffices to establish (A.4). Since $PO(\cdot, \mu)$ is a continuous function, (A.4) holds if

$$\text{(A.22)} \qquad (a)\, \bar{W}_T - W_T \xrightarrow{p} 0,\ (b)\, W_T - LM_T \xrightarrow{p} 0, \text{ and } (c)\, LM_T - LR_T \xrightarrow{p} 0 \text{ under } \theta_0.$$

For two-sided tests, part (a) of (A.22) holds because $HT^{1/2}\hat{\theta} - H\bar{\theta} \xrightarrow{p} 0$ under $\theta_0$ by Lemma A-2,

$$\text{(A.23)} \qquad \|\mathcal{I}_T(\hat{\theta}) - \mathcal{I}\| \leq \sup_{\theta \in \Theta_0} \|\mathcal{I}_T(\theta) - \mathcal{I}(\theta)\| + \|\mathcal{I}(\hat{\theta}) - \mathcal{I}\| = o_p(1)\ ,$$

and $\mathcal{I}$ is positive definite (Assumption 1(e)). The inequality in (A.23) holds wp $\rightarrow$ 1 using Assumption 3 and the equality in (A.23) holds by Assumptions 1(c), 1(d), and 3. For one-sided tests, part (a) holds for the above reasons plus the fact that $P(\text{sgn}(H\bar{\theta}) = \text{sgn}(H\hat{\theta})) \rightarrow 1$ under $\theta_0$, since $H\bar{\theta} - HT^{1/2}\hat{\theta} \xrightarrow{p} 0$ and $H\bar{\theta} \xrightarrow{d} N(0, H\mathcal{I}^{-1}H')$ under $\theta_0$.

Part (b) of (A.23) is established as follows. By (A.5) with $\hat{\theta}$ replaced by $\tilde{\theta}$, we obtain

$$\text{(A.24)} \qquad T^{-1/2}D\ell_T(\tilde{\theta}) = T^{-1/2}D\ell_T(\theta_0) - \mathcal{I}_{1T}T^{1/2}(\tilde{\theta} - \theta_0)\ ,$$

where $\mathcal{I}_{1T}$ is defined with $\tilde{\theta}$ in place of $\hat{\theta}$. Equation (A.6) holds with $\mathcal{I}_{1T}$ so defined using Assumption 4. In addition, as established below, $T^{-1/2}D\ell_T(\tilde{\theta}) = O_p(1)$. In consequence,

$$\begin{aligned} & H\mathcal{I}_T^{-1}(\tilde{\theta})T^{-1/2}D\ell_T(\tilde{\theta}) = H\mathcal{I}_{1T}^{-1}T^{-1/2}D\ell_T(\tilde{\theta}) + o_p(1) \\ \text{(A.25)} \qquad & = H\mathcal{I}_{1T}^{-1}T^{-1/2}D\ell_T(\theta_0) + HT^{1/2}(\tilde{\theta} - \theta_0) + o_p(1) \\ & = H\mathcal{I}^{-1}T^{-1/2}D\ell_T(\theta_0) + o_p(1) = HT^{1/2}(\hat{\theta} - \theta_0) + o_p(1) = HT^{1/2}\hat{\theta} + o_p(1)\ , \end{aligned}$$

where the third equality uses $HT^{1/2}\tilde{\theta} = HT^{1/2}\theta_0 = 0$ and the fourth equality holds by Lemma A-1. For two-sided tests, (A.25) yields $W_T - LM_T^0 \xrightarrow{p} 0$ under $\theta_0$, where

$$\text{(A.26)} \qquad LM_T^0 = (T^{-1/2}D\ell_T(\tilde{\theta}))'\mathcal{I}_T^{-1}(\tilde{\theta})H'\left(H\mathcal{I}_T^{-1}(\tilde{\theta})H'\right)^{-1}H\mathcal{I}_T^{-1}(\tilde{\theta})T^{-1/2}D\ell_T(\tilde{\theta})\ .$$

Now, wp $\rightarrow$ 1, $T^{-1/2}D\ell_T(\tilde{\theta}) = -H'\lambda$ for some random $p$-vector of Lagrange multipliers $\lambda$. In consequence, $LM_T = LM_T^0$ wp $\rightarrow$ 1 and part (b) is established. For one-sided tests,

$P\left(\text{sgn}(H\hat{\theta}) = \text{sgn}(H\mathcal{I}_T^{-1}(\tilde{\theta})D\ell_T(\tilde{\theta}))\right) \xrightarrow{p} 1$ under $\theta_0$ for the same reasons as for the analogous result in part (a).

It remains to show that $T^{-1/2}D\ell_T(\tilde{\theta}) = O_p(1)$ for part (b). By (A.24), $H\mathcal{I}_{1T}^{-1}T^{-1/2}D\ell_T(\tilde{\theta}) = O_p(1)$, since $HT^{1/2}(\tilde{\theta} - \theta_0) = 0$. In turn, this yields $H\mathcal{I}_{1T}^{-1}H'\lambda = O_p(1)$, $\lambda = O_p(1)$, and $T^{-1/2}D\ell_T(\tilde{\theta}) = -H'\lambda = O_p(1)$ under $\theta_0$.

Last, we establish part (c) of (A.22). A two-term Taylor expansion of $\ell_T(\tilde{\theta})$ about $\hat{\theta}$ gives

$$\ell_T(\tilde{\theta}) = \ell_T(\hat{\theta}) + (\tilde{\theta} - \hat{\theta})'D\ell_T(\hat{\theta}) + \frac{1}{2}(\hat{\theta} - \tilde{\theta})'D^2\ell_T(\theta^\dagger)(\hat{\theta} - \tilde{\theta}) , \tag{A.27}$$

where $\theta^\dagger$ satisfies $\theta^\dagger \xrightarrow{p} \theta_0$ under $\theta_0$. Since $D\ell_T(\hat{\theta}) = 0$ wp $\to 1$, and $T^{1/2}(\hat{\theta} - \tilde{\theta}) = O_p(1)$ (which follows from (A.30) below)), we obtain for two-sided tests:

$$LR_T = (T^{1/2}(\hat{\theta} - \tilde{\theta}))'\mathcal{I}_T(\tilde{\theta})T^{1/2}(\hat{\theta} - \tilde{\theta}) + o_p(1) . \tag{A.28}$$

One term Taylor expansions, as in (A.5), give

$$T^{-1/2}D\ell_T(\tilde{\theta}) = T^{-1/2}D\ell_T(\hat{\theta}) - \mathcal{I}_{1T}T^{1/2}(\tilde{\theta} - \hat{\theta}) , \tag{A.29}$$

where $\mathcal{I}_{1T}$ is as defined in (A.5) with $\hat{\theta} - \theta_0$ replaced by $\tilde{\theta} - \hat{\theta}$. In consequence,

$$T^{1/2}(\hat{\theta} - \tilde{\theta}) = \mathcal{I}_{1T}^{-1}T^{-1/2}D\ell_T(\tilde{\theta}) + o_p(1) = \mathcal{I}_T^{-1}(\tilde{\theta})T^{-1/2}D\ell_T(\tilde{\theta}) + o_p(1) . \tag{A.30}$$

Substituting this result in (A.28) yields part (c) of (A.22) for two-sided tests. Part (c) holds for one-sided tests by the same argument plus the fact that $P\left(\text{sgn}(H\hat{\theta}) = \text{sgn}(H\mathcal{I}_T^{-1}(\tilde{\theta})D\ell_T(\tilde{\theta}))\right) \xrightarrow{p} 1$ under $\theta_0$. □

PROOF OF LEMMA 1: Let $\tilde{M} = M^{1/2}\,\text{sgn}(M)$. By definitions (2.3) and (2.4),

$$\begin{aligned} PO(M, \mu) &= \int \exp\left[-\frac{1}{2}r^2\right] \int_{\mathcal{S}_p} \exp[\tilde{M}r\xi'\underset{\sim}{1}]dU(\xi)d\mu(r) \\ &= \int\int_{\mathcal{S}_p} \exp\left[\frac{1}{2}\tilde{M}'\tilde{M} - \frac{1}{2}(r\xi - \tilde{M}\underset{\sim}{1})'(r\xi - \tilde{M}\underset{\sim}{1})\right]dU(\xi)d\mu(r) . \end{aligned} \tag{A.31}$$

Now, in the two sided case with $R \sim \mu = \sqrt{\tau\chi_p^2}$, we have $R\xi = \omega \sim N(0, \tau I_p)$. Thus, using (A.31), we obtain

$$
\begin{aligned}
PO(M, \mu) &= \int_{R^p} \exp\left[\frac{1}{2}\tilde{M}'\tilde{M} - \frac{1}{2}(\omega-\tilde{M})'(\omega-\tilde{M})\right](2\pi\tau)^{-p/2}\exp\left[-\frac{1}{2\tau}\omega'\omega\right]d\omega \\
&= (1+\tau)^{-p/2}\int_{R^p}\left(2\pi\frac{\tau}{1+\tau}\right)^{-p/2}\exp\left[-\frac{1}{2}\left(\frac{1+\tau}{\tau}\left(\omega-\frac{\tau}{1+\tau}\tilde{M}\right)'\left(\omega-\frac{\tau}{1+\tau}\tilde{M}\right)\right)\right]d\omega \\
&\quad \times \exp\left[\frac{1}{2}\frac{\tau}{1+\tau}\tilde{M}'\tilde{M}\right] \\
&= (1+\tau)^{-p/2}\exp\left[\frac{1}{2}\frac{\tau}{1+\tau}M\right],
\end{aligned}
\tag{A.32}
$$

where the last equality holds because the integral of a normal density equals one.

Next, in the one-sided case with $R \sim \mu = \sqrt{\tau\chi_1^2}$, we have $R\xi \sim |N(0, \tau)|$. Let $\omega \sim N(0, \tau)$ and let $R\xi = |\omega|$. Let $Z \sim N(0, 1)$. Then, using (A.31), we have

$$
\begin{aligned}
PO(M, \mu) &= \int_{R} \exp\left[\frac{1}{2}\tilde{M}^2 - \frac{1}{2}(|\omega| - \tilde{M})^2\right](2\pi\tau)^{-1/2}\exp\left[-\frac{1}{2\tau}\omega^2\right]d\omega \\
&= 2\int_0^\infty (2\pi\tau)^{-1/2}\exp\left[-\frac{1}{2}\left(\frac{1+\tau}{\tau}\omega^2 - 2\tilde{M}\omega\right)\right]d\omega \\
&= (1+\tau)^{-1/2}\exp\left[\frac{1}{2}\frac{\tau}{1+\tau}\tilde{M}^2\right]2\int_0^\infty\left(2\pi\frac{\tau}{1+\tau}\right)^{-1/2}\exp\left[-\frac{1}{2}\left(\frac{1+\tau}{\tau}\left(\omega-\frac{\tau}{1+\tau}\tilde{M}\right)^2\right)\right]d\omega \\
&= (1+\tau)^{-1/2}\exp\left[\frac{1}{2}\frac{\tau}{1+\tau}M\right]2P\left(\sqrt{\frac{\tau}{1+\tau}}Z + \frac{\tau}{1+\tau}\tilde{M} > 0\right) \\
&= (1+\tau)^{-1/2}\exp\left[\frac{1}{2}\frac{\tau}{1+\tau}M\right]2\Phi\left(\left(\frac{\tau}{1+\tau}|M|\right)^{1/2}\mathrm{sgn}(M)\right). \quad \square
\end{aligned}
\tag{A.33}
$$

PROOF OF LEMMA 2: Assumption 1(a) holds by Assumption NL(a). 1(b) holds by NL(c). 1(c) holds with $\mathcal{I}(\theta) = -E\frac{\partial^2}{\partial\theta\partial\theta'}\log g_t(\theta)$ provided a uniform WLLN can be established. The Markov property (NL(b)) ensures that $\left\{\frac{\partial^2}{\partial\theta\partial\theta'}\log g_t(\theta) : t > m\right\}$ is part of a doubly infinite stationary and ergodic sequence. Thus, using NL(b) and (e), the ergodic theorem implies that

$-T^{-1}D^2\ell_T(\theta) \xrightarrow{p} \mathcal{I}(\theta) \quad \forall\theta \in \Theta_0$. A generic uniform WLLN (e.g., Assumptions TSE-1D, BD, DM, and P-WLLN and Theorem 4 of Andrews (1992)) strengthens this result to uniform convergence over $\Theta_0$ using NL(b), (c), and (e).

Assumption 1(d) holds, because $\mathcal{I}(\theta)$ is continuous on $\Theta_0$ by the dominated convergence theorem using NL(c) and (e) and $\Theta_0$ is compact. 1(e) holds by NL(f).

To verify Assumption 2, note that $\left\{\left(\frac{\partial}{\partial\theta}\log g_t(\theta_0), \mathcal{F}_{t-1}\right) : t > m\right\}$ is a martingale difference sequence (MDS), because

(A.34)
$$E\left(\frac{\partial}{\partial\theta}\log g_t(\theta_0)|\mathcal{F}_{t-1}\right) = E\left(\frac{\partial}{\partial\theta}\log g_t(\theta_0)|S_{t,m}, X_{t,m}\right)$$
$$= \int\frac{\partial}{\partial\theta}g_t(\theta_0)d\lambda(s_t) = \frac{\partial}{\partial\theta}\int g_t(\theta_0)d\lambda(s_t) = 0 ,$$

where the third equality holds by the dominated convergence theorem using NL(c) and (e). Using the Cramer-Wold device, Assumption 2 now follows from the univariate CLT for stationary ergodic square-integrable MDS with positive variances (e.g., see Brown (1971)).

Sufficient conditions for Assumption 3 are: (i) $\Theta$ is compact, (ii) $\log g_t(\theta)$ is continuous in $\theta$ on $\Theta$ with probability one under $\theta_0$, (iii) $\sup_{\theta\in\Theta}|\frac{1}{T}\Sigma_1^T(\log g_t(\theta) - E \log g_t(\theta)| \xrightarrow{p} 0$ under $\theta_0$, and (iv) $E \log g_t(\theta)$ is uniquely maximized over $\Theta$ at $\theta_0$ (e.g., see Amemiya (1985, Thm. 4.1.1, pp. 106-107). Parts (i) and (ii) hold by NL(a) and (c) respectively. Part (iii) holds by the same argument as for 1(c) above. To obtain part (iv), note that for $\theta \neq \theta_0$,

(A.35)
$$E \log g_t(\theta) - E \log g_t(\theta_0) = E \log[g_t(\theta)/g_t(\theta_0)] < \log Eg_t(\theta)/g_t(\theta_0) = 0 ,$$

where the inequality is an application of Jensen's inequality and is strict by NL(d).

Assumption 4 holds by the same argument as for Assumption 3 with $\tilde{\Theta}$ in place of $\Theta$. □

# FOOTNOTES

[1]The author thanks Peter Phillips, Peter Rossi, Chris Sims, and Arnold Zellner for helpful comments. He also gratefully acknowledges the research support of the National Science Foundation via grant number SES-9121914.

[2]Bayesian tests also can be based on ratios of posterior expected losses (e.g., see Zellner (1971, pp. 295-6)). Such tests are more general than posterior odds tests. The application of the results of this paper to tests based on posterior expected losses is discussed briefly in Section 5 below.

[3]It is sometimes of interest to a Bayesian from a robustness/sensitivity perspective to compute the maximum of the PO statistic over certain classes of priors, e.g., see Edwards, Lindman, and Savage (1963), Berger (1985, Sec. 4.3.3), and Berger and Sellke (1987). If one considers the class of multivariate normal priors referred to above, then for two-sided tests the maximum of the approximate PO statistic (2.6) over $\tau > 0$ is given by

$$\frac{1-\pi}{\pi}\left(\frac{M}{p}\right)^{-p/2}\exp\left(\frac{1}{2}[M+p]\right) .$$

[4]To see this, one can go through the proof of Theorem 1 in the Appendix and verify that all of the $o_p(1)$ terms are 0 in this case, provided one defines $\mathcal{I} = -T^{-1}D^2\ell_T(\theta_0)$.

[5]The proof of this result holds by the same argument as used to establish Theorem 1 with the prior on $\theta$ under $H_1$ replaced by the product of $L(\theta, H_1)$ and the prior on $\theta$ under $H_1$ divided by its integral over $\theta$ in $H_1$ (i.e., by $\int L(\theta_0 + T^{-1/2}h)dQ_\mu(h)$) in order to ensure that the product integrates to one, and hence, is a proper distribution.

[6]Weak exogeneity of $\{X_t : t \leq T\}$, defined in Engle, Hendry, and Richard (1983), implies that the likelihood function for $Y_T$ can be factored into two pieces, one of which contains conditional distributions of $S_t$ and depends on $\theta$ and the other of which contains conditional distributions of $X_t$ and does not depend on $\theta$, see below.

[7]Note that Phillips (1992) considers general non-stationary linear models in most of his paper, but the section dealing with an asymptotic correspondence between Bayesian and classical tests considers a null hypothesis under which the random variables are stationary.

# REFERENCES


Amemiya, T. (1985): *Advanced Econometrics*. New York: Harvard University Press.

Andrews, D. W. K. (1989): "Power in Econometric Applications," *Econometrica*, 57, 1059-1090.

______ (1992): "Generic Uniform Convergence," *Econometric Theory*, 8, 241-257.

Berger, J. O. (1985): *Statistical Decision Theory and Bayesian Analysis*, second edition. New York: Springer-Verlag.

Berger, J. O. and T. Sellke (1987): "Testing a Point Null Hypothesis: The Irreconcilability of *P* Values and Evidence," *Journal of the American Statistical Association*, 82, 112-122.

Brown, B. M. (1971): "Martingale Central Limit Theorems," *Annals of Statistics*, 42, 59-66.

Casella, G. and R. L. Berger (1987): "Reconciling Bayesian and Frequentist Evidence in the One-sided Testing Problem," *Journal of the American Statistical Association*, 82, 106-111.

Chernoff, H. (1954): "On the Distribution of the Likelihood Ratio," *Annals of Mathematical Statistics*, 54, 573-578.

Edwards, W., H. Lindman, and L. J. Savage (1963): "Bayesian Statistical Inference for Psychological Research," *Psychological Review*, 70, 193-242. Reprinted in *Robustness of Bayesian Analysis*, 1984, ed. by J. Kadane, Amsterdam: North-Holland.

Engle, R. F., D. F. Hendry, and J.-F. Richard (1983): "Exogeneity," *Econometrica*, 51, 277-304.

Jeffreys, H. (1961): *Theory of Probability*, third edition. Oxford: Clarendon.

Kass, R. E., L. Tierney, and J. B. Kadane (1988): "Asymptotics in Bayesian Computation," in *Bayesian Statistics*, Vol. 3, ed. by J. M. Bernardo, M. H. DeGroot, D. V. Lindley, and A. F. M. Smith. Oxford, England: Oxford University Press.

Kass, R. E. and S. K. Vaidyanathan (1992): "Approximate Bayes Factors and Orthogonal Parameters, with Application to Testing Equality of Two Binomial Proportions," *Journal of the Royal Statistical Society*, Series B, 54, 129-144.

Lindley, D. V. (1961): "The Use of Prior Probability Distributions in Statistical Inference and Decisions," *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, 1, 453-468. Berkeley: University of California Press.

McCulloch, R. E. and P. E. Rossi (1992): "Bayes Factors for Nonlinear Hypotheses and Likelihood Distributions," *Biometrika*, 79, forthcoming.

Phillips, P. C. B. (1992): "Bayes Methods for Trending Multiple Time Series with an Empirical Application to the U.S. Economy," Cowles Foundation Discussion Paper No. 1025, Yale University.

Pratt, J. W. (1965): "Bayesian Interpretation of Standard Inference Statements," *Journal of the Royal Statistical Society*, Ser. B, 27, 169-203.

Schwarz, G. (1978): "Estimating the Dimension of a Model," *Annals of Statistics*, 6, 461-464.

Strasser, H. (1985): *Mathematical Theory of Statistics: Statistical Experiments and Asymptotic Decision Theory*. New York: de Gruyter.

Wald, A. (1943): "Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations Is Large," *Transactions of the American Mathematical Society*, 54, 426-482.

Watson, G. S. (1983): *Statistics on Spheres*. New York: Wiley.

Zellner, A. (1971): *An Introduction to Bayesian Inference in Econometrics*. New York: Wiley.

______ (1984): *Basic Issues in Econometrics*. Chicago: University of Chicago Press.

Zellner, A. and A. Siow (1979): "On Posterior Odds Ratios for Sharp Null Hypotheses and One-sided Alternatives," H. G. B. Alexander Research Foundation Discussion Paper, Graduate School of Business, University of Chicago.

______ (1980): "Posterior Odds Ratios for Selected Regression Hypotheses," in *Bayesian Statistics: Proceedings of the First International Meeting Held in Valencia, May 28 to June 2, 1979*, ed. by J. M. Bernardo, M. H. DeGroot, D. V. Lindley, and A. F. M. Smith. Valencia, Spain: University Press.

TABLE 1

Values of the Prior Parameter $\tau$ for Which Classical Tests of Significance Level $\alpha$ Are Equivalent to Approximate Posterior Odds Tests When $\pi = 1/2$

| $p$ \ $\alpha$ | .01 | .05 | .10 | .25 | .5[a] |
|---|---|---|---|---|---|
| 1 (one-sided) | 881. | 50. | 14. | 1.6 | 0. |
| 1 (two-sided) | 750. | 41. | 11. | .79 | -- |
| 2 | 94. | 16. | 6.3 | 1.0 | -- |
| 3 | 38. | 9.5 | 4.5 | .95 | -- |
| 4 | 23. | 6.9 | 3.6 | .88 | -- |
| 5 | 16. | 5.6 | 3.0 | .81 | -- |

[a]For $\alpha$ sufficiently large, there is no distribution $\mu$ of the form $\sqrt{\tau\chi_p^2}$ that solves (5.5) or (5.6). This accounts for the dashes in the table.