10-1-1992

Time Series Modeling with a Bayesian Frame of Reference: Concepts, Illustrations and Asymptotics

Peter C.B. Phillips

Werner Ploberger

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
AT YALE UNIVERSITY

Box 2125, Yale Station
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 1038

TIME SERIES MODELING WITH A BAYESIAN FRAME OF REFERENCE: CONCEPTS, ILLUSTRATIONS AND ASYMPTOTICS

Peter C. B. Phillips and Werner Ploberger

October 1992

TIME SERIES MODELING WITH
A BAYESIAN FRAME OF REFERENCE:
CONCEPTS, ILLUSTRATIONS AND ASYMPTOTICS *

by

Peter C. B. Phillips and Werner Ploberger

Cowles Foundation for Research in Economics, Yale University
and
University of Technology, Vienna

First Version: March, 1991
Revision: October, 1992

* The authors thank John Hartigan, Bent Sorensen, Lars Hansen and three referees for comments on the first version of this paper which was circulated in March 1991 under the title "Time Series Modelling with a Bayesian Frame of Reference: I. Concepts and Illustrations." The paper was written while Werner Ploberger was a visitor at the Cowles Foundation during 1990-1991. Phillips thanks the NSF for research support under Grant Nos. SES 8821180 and SES 9122142. Ploberger thanks the Fonds zur Förderung der wissenschaftlichen Forschung for supporting his stay at Yale with Schrödingerstipendium Nr. J0469-TEC. The authors' thanks go to Glena Ames for her skill and effort in keyboarding the manuscript.

1. INTRODUCTION

The Bayesian approach to modeling and inference in time series econometrics has become increasingly popular in recent years. Examples include the use of Bayesian priors to achieve economies in VAR parameterizations (Litterman, 1986; Doan, Litterman and Sims, 1984), Bayesian modeling of cyclical behavior in macroeconomic time series (Geweke, 1988) and Bayesian evaluations of the evidence in support of the presence of stochastic trends (DeJong and Whiteman, 1991; Schotman and Van Dijk, 1990; Phillips, 1991a, b). Advances in simulation-based technology (Kloek and Van Dijk, 1978; Geweke, 1989) and improvements in analytic devices like the Laplace approximation method (Phillips, 1983, 1991a; Tierney and Kadane, 1986; Tierney, Kass and Kadane, 1989) have both contributed to the successful implementation of Bayesian methods in time series applications.

Concurrent with the growing empirical use of Bayesian methods, there has been continued discussion of foundational issues, such as acceptance of the likelihood principle (Poirier, 1988) and the form of prior densities to represent the notion of "knowing little" in advance of data analysis (Phillips, 1991a; Zellner, 1984, 1990). Such matters are obviously of great importance and have, of course, been discussed in earlier literature (e.g. Barnard, Jenkins and Winsten, 1962; Basu, 1973; Hartigan, 1964). However, time series applications do raise issues that deserve further attention like the treatment of initial conditions, nonstationarity, high dimensional parameter spaces and even semiparametric model formulations.

Some econometricians, notably Sims (1988) and Sims and Uhlig (1991) have argued recently that time series models provide important examples where Bayesian and classical methods differ fundamentally. Phillips (1991a) showed that some aspects of the differences described in those papers, like the phenomena of disjoint classical confidence intervals in comparison to symmetric Bayesian confidence sets, are the result of the use of uniform priors, which Phillips argues are inappropriate in a time series context. However, not all of the apparent differences between classical and Bayesian methods in time series models can be explained in this way. For instance, in classical theory the Gaussian log-likelihood of an AR(1) model with a unit root cannot be asymptotically approximated uniformly by a quadratic without a change in the units of measurement (or equivalently, a random time change), since the sample variance of the data carries information about the autoregressive parameter and, upon standardization, has a limit that depends on this parameter and may even be random. By contrast, the likelihood principle that underpins Bayesian theory identifies the information content of the data with the likelihood function itself and, conditional on the given data, the Gaussian log-likelihood in this case is indeed quadratic for all sample sizes. The same result can be said to hold approximately in large samples for many non-Gaussian cases, as shown in Sims (1990) and Kim (1992). These additional differences between the classical and Bayesian approaches to inference arise because of the critical role of data conditioning in Bayesian analysis. They are every bit as fundamental as the question of which prior to use and they are especially significant in time series modeling where data conditioning has important implications.

The present paper seeks to explain and to reconcile these differences. Our analysis shows how traditional Bayesian inference that is based on the posterior distribution implicitly involves a change in the underlying probability measure, leading to a new Bayesian frame of reference for the data generating mechanism. When this change of reference measure is taken into account the symmetric Gaussian posterior density centered on the maximum likelihood estimate (MLE) is explained by the fact that the model to which inference that is based on the posterior relates is itself changed to a path dependent model where the coefficient is replaced by the MLE. Understanding this path dependent model, which we call the "Bayes model," and the measure associated with it is important if one is to properly interpret the results of traditional Bayesian inference based on the posterior density. We explore the consequences of this change of measure and model by studying several examples in detail. These, together with an analysis of some Bayes tests that we propose, are given in Sections 2 and 3 of the paper. Sections 4 and 5 outline a theory for the general case of a single parameter model where no assumptions concerning stationarity or rates of convergence are required. These sections also give a new proof that the likelihood function is asymptotically Gaussian and show how the Bayesian data density may be approximated asymptotically by a local exponential martingale that plays a big role in our theory. Section 6 concludes the paper, describes the results of some related research by the authors that uses the concepts of this paper, and offers some thoughts for further work.

The following notational conventions are employed in the paper. M_t is used to represent a continuous L_2 (i.e. square integrable) martingale, local martingale or semimartingale, and the square bracket $[M]_t = [M, M]_t$ denotes its quadratic variation process. Similar notation is employed in the case of a discrete time martingale M_n, and in this case we use $<M>_n = \langle M, M\rangle_n$ to denotes the conditional quadratic variation process. A_t (respectively, A_n) is often a shorthand notation for quadratic variation process (respectively, conditional quadratic variation). W_t and occasionally S_t denote standard Brownian motion which is signified by the symbolism "BM(1)". The symbol " $\equiv$ " signifies equivalence or equivalence in distribution and " $\ll$ " denotes the absolute continuity operator.

2. FIRST ORDER AUTOREGRESSION IN CONTINUOUS TIME

2.1. The Likelihood

We start our analysis with a continuous time diffusion model because this case will illustrate the main ideas of the paper in a simple way and thereby (we hope) make more accessible the general case to be discussed in Sections 4 and 5. Moreover, in our general discussion we will see how the discrete likelihood function admits an approximation in terms of continuous martingales that leads to an analysis which is similar to that of the simple diffusion model.

Specifically, our model in this section is the following stochastic differential equation for the Ornstein-Uhlenbeck process Y_t

$$(1) \qquad dY_t = \beta Y_t dt + dW_t\,, \quad t \geq 0$$

where $W_t \equiv$ BM(1). The processes Y_t and W_t are defined on a filtered sequence of measurable spaces $(\Omega, \mathcal{F}_t)$ with Y_t and W_t adapted to $\mathcal{F}_t$. Let P_t^β be the probability measure of Y_t given by (1) with parameter β on this filtered space and let us define $P_t = P_t^0$ which will subsequently serve as a reference measure. The probability measure P_t^β has density with respect to P_t given by the following Radon-Nikodym (hereafter, RN) derivative

$$(2) \qquad L_t = dP_t^\beta / dP_t = \exp\left\{\beta \int_0^t Y_s dY_s - (1/2)\beta^2 \int_0^t Y_s^2 ds\right\} .$$

The form of (2) is actually well known in the literature (e.g. Ibragimov and Has'minski, 1981, p. 16).

The log-likelihood corresponding to (2) is

(3) $$\Lambda_{\beta t} = \log(L_t) = \beta \int_0^t Y_s dY_s - (1/2)\beta^2 \int_0^t Y_s^2 ds = \beta V_t - (1/2)\beta^2 A_t \ , \quad \text{say}$$

from which we derive the maximum likelihood estimator (MLE)

$$\hat{\beta}_t = A_t^{-1} V_t = \left(\int_0^t Y_s^2 ds \right)^{-1} \left(\int_0^t Y_s dY_s \right) .$$

Observe that $\hat{\beta}_t$ is the usual continuous time least squares estimator of β in (1) i.e. the estimator that minimizes the formal "error sum of squares" functional $\int_0^t (\dot{y}_s - \beta y_s)^2 ds$.

The likelihood process (2) may now be written as

(4) $$L_t = \exp\{(1/2)\hat{\beta}_t^2 A_t\} \exp\{-(1/2)(\beta - \hat{\beta}_t)^2 A_t\} .$$

Only the second exponential factor of (4) depends explicitly on β and this is proportional to a $N(\hat{\beta}_t, A_t^{-1})$ density. In conventional Bayesian inference it is this latter factor that plays the key role in determining the shape of the posterior. The first factor, being independent of β, is traditionally ignored in the transition, via Bayes theorem, to the posterior. We shall have much more to say about this matter in the ensuing discussion.

2.2. Bayesian Inference

Let $\pi(\beta)$ be a prior density for the parameter β in the model (1). This density need not be proper and could, for instance, be uniform. Combining the prior $\pi(\beta)$ with the likelihood as given in (4) we have the process

(5) $$\begin{aligned} \Pi_t &= \pi(\beta)(dP_t^\beta / dP_t) = \pi(\beta) L_t = \pi(\beta)\exp\{(1/2)\hat{\beta}_t^2 A_t\}\exp\{-(1/2)(\beta - \hat{\beta}_t)^2 A_t\} \\ &= \left[A_t^{-1/2}\exp\{(1/2)\hat{\beta}_t^2 A_t\}\right]\left[\pi(\beta) A_t^{1/2}\exp\{-(1/2)(\beta - \hat{\beta}_t)^2 A_t\}\right] \\ &= \left[\exp\{(1/2)\hat{\beta}_t^2 A_t - (1/2)\ln(A_t)\}\right]\left[\pi(\beta) A_t^{1/2}\exp\{-(1/2)(\beta - \hat{\beta}_t)^2 A_t\}\right] . \end{aligned}$$

When $\pi(\beta)$ is uniform, Π_t is proportional to the posterior density $N(\hat{\beta}_t, A_t^{-1})$ for β and might therefore be called the posterior process. The decomposition of the Π_t into the two factors in square brackets in (5) is important in what follows. As we shall see, the first factor is a local martingale and produces the density process that changes the measure to a Bayesian frame of reference. Since the first factor does not explicitly involve the parameter β the Bayesian posterior is, in effect, proportional to the second factor in square brackets in (5). Thus, in conventional Bayes inference, the transition from prior to posterior via Bayes theorem leads us to ignore the first factor as "irrelevant" for inferential purposes. We will find, however, that the first factor is not irrelevant from a conceptual standpoint.

Let us now take the case of a uniform prior $\pi(\beta) = \pi_0 = (2\pi)^{-1/2}$. The particular choice of the constant $\pi_0 = (2\pi)^{-1/2}$ in this prior will be explained later in Section 3. Integrating (5) with respect to β we define the measure Q_t by the RN derivative

$$(6) \qquad dQ_t/dP_t = \int_{\mathbf{R}} \pi(\beta)(dP_t^{\beta}/dP_t)d\beta = A_t^{-1/2}\exp\{(1/2)\hat{\beta}_t^2 A_t\} .$$

This expression gives the density (with respect to the reference measure P_t) of the data that is implied by the model (1) and a uniform prior on β. If the derivatives in (6) were taken instead with respect to Lebesgue measure (ν), we would have the usual Bayesian data density or mixture density

$$dQ_t/d\nu = \int_{\mathbf{R}} \pi(\beta)(dP_t^{\beta}/d\nu)d\beta .$$

For this reason we call Q_t the *Bayes measure*. It will be convenient to write the derivative process (6) in the alternate form

$$(7) \qquad Z_t = A_t^{-1/2}\exp\{(1/2)\hat{\beta}_t^2 A_t\} = A_t^{-1/2}\exp\{(1/2)V_t^2 A_t^{-1}\} .$$

Under P_t, we note that $V_t = \int_0^t Y_s dY_s$ is a martingale with quadratic variation process $A_t = \int_0^t Y_s^2 ds$.

A special case of (7) that is of some independent interest is

$$(8) \qquad Z_t' = t^{-1/2} \exp\{(1/2t)W_t^2\} ,$$

where $W_t \equiv BM(1)$. In fact, by a suitable time change that is achieved by setting

$$(9) \qquad \tau_t = \inf\{s : A_s \geq t\}$$

we may replace the continuous martingale V_t by a Brownian motion (e.g. Protter (1990), Theorem 41). Specifically, we have

$$V_{\tau_t} = W_t \text{ a.s.}, \ 0 \leq t < \infty$$

and then

$$(10) \qquad Z_{\tau_t} = Z_t' \quad \text{a.s.}$$

Thus, by using the time change (9) we may replace Z_t by the (apparently) simpler process Z_t'. Z_t' has some very interesting stochastic properties. In particular, Z_t' does not have finite expectation, as is immediately clear from the form of (8). However, it does have finite conditional expectation and, indeed, satisfies the martingale property that $E(Z_t'|\mathcal{F}_s) = Z_s'$ a.s. The process Z_t' is, in fact, a local martingale as we prove the following result.

THEOREM 2.1. *Z_t and Z_t' are continuous local L_2 martingales under the probability measure $P(\cdot|\mathcal{F}_\tau)$, where τ is a stopping time such that $\tau > 0$ a.s. (P).* □

REMARKS

(i) The reason for the use of the stopping time τ and the conditional measure $P(\cdot|\mathcal{F}_\tau)$ in Theorem 2.1 is that with the initialization at $t = 0$, $Z_0' = \infty$ a.s. (P). The simple conceptualization of this outcome is that at $t = 0$ we have no data and the improper prior measure (whose mass is infinity) dominates. Our use of $P(\cdot|\mathcal{F}_\tau)$ stresses the positivistic side of our theory: we will describe Bayes models only when there is given data. If we were to commence with a proper prior on β then this difficulty at the initial value would not arise. We examine this case below, where the proper prior is based on data accumulated over a preliminary or pre-sample period such as the interval $(0, t_0]$ for some fixed $t_0 > 0$.

(ii) The translation of measure from P_t to Q_t that is effected by (6) is important in the interpretation of Bayesian inference. Using (6) we see that equation (5) for the process Π_t now has the form

$$\Pi_t = \pi(\beta)dP_t^\beta/dP_t = (dQ_t/dP_t)\left[(2\pi)^{-1/2}A_t^{-1/2}\exp\{-(1/2)(\beta - \hat{\beta}_t)^2A_t\}\right]$$

$$= (dQ_t/dP_t)N(\hat{\beta}_t, A_t^{-1}) \; . \qquad (11)$$

The Bayesian posterior process for β, viz. $N(\hat{\beta}_t, A_t^{-1})$, is now obtained from (11) by changing the reference measure from P_t to Q_t. We can write this as

$$\Pi_t^B = \Pi_t(dP_t/dQ_t) = \pi(\beta)(dP_t^B/dQ_t) = N(\hat{\beta}_t, A_t^{-1}) \; . \qquad (12)$$

We can interpret this translation of measure as saying that the reference frame for a flat prior Bayesian analysis is provided by the data dependent measure Q_t rather than the original reference measure P_t.

(iii) The posterior density Π_t^B in (12) is the traditional instrument of Bayesian inference about β. Since Π_t^B is Gaussian and the distribution is symmetric about the MLE $\hat{\beta}_t$, Bayesian inference based on Π_t^B appears to be much simpler than classical inference based on $\hat{\beta}_t$ (whose sampling distribution is asymmetric and complicated in analytic form). This observation has led some investigators, notably Sims (1988) and Sims-Uhlig (1991), to conclude that Bayesian inference is simpler and logically sounder than classical methods in this context. However, the transition from Π_t to Π_t^B and the simple Gaussian posterior $N(\hat{\beta}_t, A_t^{-1})$ is achieved by the implicit change of measure from P_t to Q_t as the above analysis shows. Under data conditioning and from a Bayesian perspective it is equivalent to work with either Π_t or Π_t^B since the factor of proportionality between these processes, viz. dQ_t/dP_t, is dependent only on the data and is absorbed into the constant of proportionality in the transition from the likelihood function to the posterior. However, only Π_t^B as given in (12) makes explicit the underlying reference measure Q_t that is implicit in the use of the Gaussian posterior density $N(\hat{\beta}_t, A_t^{-1})$. We will explore the consequences of this change of reference measure in terms of the implied probability model below.

(iv) Observe that the measure Q_t is induced by the local martingale Z_t from the relation $dQ_t/dP_t = Z_t$ given in (6). In consequence, Q_t is a σ-finite measure rather than a probability measure and $Q_t(\Omega) = \int_\Omega Z_t dP_t = E(Z_t) = \infty$. This is, in fact, the result of using an improper (diffuse) prior density on β in the construction of Π_t. Nevertheless, associated with Q_t are proper conditional densities $Q_t(\cdot|\mathscr{F}_{t_0})$ that follow from the ratio

$$(13) \qquad \frac{dQ_t/dP_t}{dQ_{t_0}/dP_{t_0}} = \frac{Z_t}{Z_{t_0}} = \exp\left\{(1/2)\int_{t_0}^t d[\hat{\beta}_s^2 A_s - \ln(A(s))]\right\}$$

for all $t \geq t_0$, where $t_0 > 0$ is some alternative initialization of the process. We can use this fact the deduce the probability model that is implied by the use of Q_t as the reference measure. First we will show in the lemmas that follow that we can write the density ratio R_t given in (13) in a more revealing form. The probability measure associated with the conditional density R_t is then given in Theorem 2.4. □

LEMMA 2.2

(14) $$(1/2)d[\hat{\beta}_t^2 A_t - \ln(A_t)] = \hat{\beta}_t dV_t - (1/2)\hat{\beta}_t^2 dA_t \ . \ \square$$

LEMMA 2.3. *The density ratio* R_t *in* (13) *may be written in the alternate form*

(15) $$R_t = \exp\{G_t - (1/2)[G, G]_t\}$$

where $G_t = \int_{t_0}^{t} \hat{\beta}_s dV_s$. *Moreover, we can write*

(16) $$R_t = 1 + \int_{t_0}^{t} R_s dG_s \ ,$$

so that R_t *is the Doléans exponential of* G_t. *Under* P_t, V_t *and hence,* G_t *are continuous martingales. If* $E[\exp\{(1/2)[G, G]_t\}] < \infty$, *then* R_t *is also a continuous martingale and*

(17) $$E(R_t) = 1 \ . \ \square$$

REMARKS

(i) The form of the density R_t given in (15) is called a Doléans exponential (cf. Meyer, 1989, p. 148 in the appendix to Emery, 1989). This exponential is especially interesting in the statistical theory of stochastic processes because it is known to represent the limit of the likelihood function for stochastic processes in very general situations (see Strasser, 1986, Theorem 1.15). When G_t is a continuous local martingale, then so too is the process R_t (e.g. Chung and Williams, 1990, Theorem 6.2). When G_t is a continuous L_2 martingale, as it is under P_t, and when $E(\exp\{(1/2)[G, G]_t\}] < \infty$ (which can always if necessary be arranged by the use of a suitable stopping time), then R_t is also a continuous L_2 martingale (see Ikeda and Watanabe, 1989, Theorem 5.3, p. 152). In this case R_t is known as a density martingale and it represents a proper probability density.

(ii) Let us now consider the case where we wait until time t_0 for a minimal amount of information about the process to accumulate. In this case we can use the data over the period $[0, t_0]$ to construct a suitable prior density for β. A natural choice is to take the posterior densities for β given in (12) based on information up to t_0. That is, we set

$$\pi(\beta) = \Pi_{t_0}^{B} = N(\hat{\beta}_{t_0}, A_{t_0}^{-1}) \ .$$

Now combine this prior with the likelihood for the data over the interval $[t_0, t]$. Using an obvious notation in which the second subscript t_0 indicates the new initialization we have the joint density process

$$\text{(18)}\qquad \Pi_{t,t_0} = \pi(\beta)L_{t,t_0} = \Pi^B_{t_0}L_{t,t_0}$$

$$= \Pi^B_{t_0}\exp\left\{\beta \int_{t_0}^t Y_s dY_s - (1/2)\beta^2 \int_{t_0}^t Y_s^2 ds\right\}$$

$$= \Pi^B_{t_0}(L_t/L_{t_0})$$

$$= [(2\pi)^{-1/2}L_{t_0}/Z_{t_0}][(2P)^{1/2}Z_t N(\hat{\beta}_t,\ A_t^{-1})/L_{t_0}]$$

$$= [Z_t/Z_{t_0}]N(\hat{\beta}_t,\ A_t^{-1})$$

$$\text{(19)}\qquad = R_t N(\hat{\beta}_t,\ A_t^{-1})\ .$$

Now $R_t = dQ_t/dP_t|_{\mathcal{F}_{t_0}} = Z_t/Z_{t_0}$ is the conditional data density given information in the start up period $[0,\ t_0]$. Note that R_t depends only on the data and not on β. Note also that the joint process Π_{t,t_0} in (18) leads to the same posterior process, viz. $N(\hat{\beta}_t,\ A_t^{-1})$, from (19) as would have been obtained directly from (12), where the initialization in the process was set at $t = 0$. In this sense there is coherent Bayesian updating, so that inferences made at time t will agree, irrespective of the initialization, provided that the prior information is properly updated.

(iii) Observe that the joint density process given in (18) of $(\beta,\ Y_t)$ conditional on $\mathcal{F}_{t_0}$ can be written as

$$\text{(20)}\qquad \Pi_{t,t_0} = \pi(\beta)\frac{dP_t^{\beta}/dP_t}{dP_{t_0}^{\beta}/dP_{t_0}} = R_t N(\hat{\beta}_t,\ A_t^{-1})\ ,$$

where

$$R_t = \frac{Z_t}{Z_{t_0}} = \frac{dQ_t/dP_t}{dQ_{t_0}/dP_{t_0}}\ .$$

It follows that the posterior density process for β at time t has the form

$$\text{(21)}\qquad \Pi^B_{t,t_0} = \pi(\beta)\frac{dP_t^{\beta}/dQ_t}{dP_{t_0}/dQ_{t_0}} = N(\hat{\beta}_t,\ A_t^{-1})\ ,$$

where inferences about β are centered on the MLE $\hat{\beta}_t$. Comparing (20) and (21), we see that the reference measure for the construction of the likelihood function changes from the measure P to the measure Q. The situation is analogous to that described earlier when the initial condition was set at

$t = 0$ (see Remark (i) following Theorem 2.1). In effect, the frame of reference for a Bayesian posterior analysis based on $\Pi^B_{t,t_0} = N(\hat{\beta}_t, A_t^{-1})$ is provided by the data dependent measure Q_t and the associated density process R_t. □

It is now of interest to find an explicit form of the model that corresponds to the density process R_t. This model turns out to be trajectory dependent and it provides the frame of reference in a Bayesian analysis of the "classical" model (1). The model is given by $(1)^B$ in our next result.

THEOREM 2.4. (a) *Under a uniform prior for* β *in the model* (1) *the process* Π^B_t *determines a sequence of Gaussian posterior distributions* $N(\hat{\beta}_t, A_t^{-1})$, *i.e. normal with mean* $\hat{\beta}_t$ *and variance* A_t^{-1}. *The same posterior density* $\Pi^B_t = N(\hat{\beta}_t, A_t^{-1})$ *applies if an initialization at* $t_0 > 0$ *is selected and the prior for* β *is chosen as the posterior* $\Pi^B_{t_0}$ *from the earlier period* $[0, t_0]$.

(b) *Bayes methods that are based on the posterior* Π^B_t *imply a replacement of the underlying reference measure* P_t *in* (1) *with the Bayes measure* Q_t *defined by* (6), *i.e. the likelihood function on which Bayes inference is based relies on* dP^β_t/dQ_t *not* dP^β_t/dP_t *as in* (2).

(c) *The Bayes measure* Q_t *has conditional density process* R_t *given by* (13) *for some* $t_0 > 0$. R_t *is the likelihood ratio or density process of the output* $\{Y_s\}^t_{t_0}$ *for any* t_0 *of the nonlinear stochastic differential equation*

$$(1)^B \qquad dY_t = \hat{\beta}_t Y_t dt + dW_t\ , \quad t>0$$

in which the parameter β *that appears in the model* (1) *is replaced by the trajectory dependent value* $\hat{\beta}_t = \int_0^t Y_s dY_s / \int_0^t Y_s^2 ds$. □

In a Bayesian analysis of model (1) with a uniform prior on β there is no commitment to a particular value of β, i.e. there is no concept of a true dynamic model (1) with a true value of β. Instead, in such a Bayesian analysis the underlying reference measure P_t (i.e. the probability measure of the standard Brownian motion that drives (1) and for which $Y_t = W_t$ when $\beta = 0$) is replaced by what we have called the Bayes measure Q_t. This measure Q_t is trajectory dependent and thus, to a Bayesian, the reference model evolves according to the recorded history of the process (on which all Bayesian inference is conditioned). The Bayesian reference model is the nonlinear stochastic differential equation $(1)^B$ and is an evolving parameter model. Theorem 2.5 tells us that Bayesian inference based on the posterior density process Π_t^B relates to a trajectory based version of the original model, i.e. $(1)^B$, rather than the true original model. In effect, a Bayesian approach to inference on model (1), whereby one conditions on the available data, involves the implicit adoption of a path dependent measure which treats the parameter of interest as evolving in such a way that it is continuously updated as new data becomes available.

What we argue is that the simplicity of the posterior density $\Pi_t^B = N(\hat{\beta}_t, A_t^{-1})$ needs to be tempered by the fact that this posterior is associated with a path dependent reference measure Q_t and a model, $(1)^B$, that is itself path dependent and very different from the classical model (1). In the context of the new model $(1)^B$ the simple Gaussian posterior $N(\hat{\beta}_t, A_t^{-1})$ centered on $\hat{\beta}_t$ seems quite logical. However, we emphasize that inference about β that is based on the $N(\hat{\beta}_t, A_t^{-1})$ posterior needs to be interpreted in the light of the new model $(1)^B$ not the classical model (1).

3. AUTOREGRESSIONS IN DISCRETE TIME, BAYES MODEL TESTS, POSTERIOR ODDS AND SOME MODEL EXTENSIONS

3.1. The AR(1) Model and its Gaussian Likelihood

Our model in this section is the Gaussian AR(1)

$$H_\alpha : Y_t = \alpha Y_{t-1} + u_t \, , \quad u_t \equiv \text{iid } N(0, \sigma^2) \tag{22}$$

where $\sigma^2 = 1$ and the process is initialized at $t = 0$ with Y_0 any $\mathcal{F}_0$-measurable variable. Y_t and u_t are defined on a filtered sequence of measurable spaces $(\Omega, \mathcal{F}_t)$ with Y_t and u_t adapted to $\mathcal{F}_t$. We use P_n^α to represent the probability measure of $Y^n = \{Y_t\}_1^n$ conditional on Y_0. So when $\alpha = 0$ we have the measure P_n^0 and when $\alpha = 1$ we have the random walk H_1 with measure $P_n^1 = P_n$, which will serve as our reference measure.

The log-likelihood of H_α, given H_1 as the reference model and conditional on Y_0, is

$$\Lambda_{hn} = \ln(dP_n^\alpha/dP_n) = \ln[(dP_n^\alpha/dP_n^0)(dP_n^0/dP_n)]$$

$$= -(1/2)\Sigma_1^n(Y_t - \alpha Y_{t-1}^2 + (1/2)\Sigma_1^n(Y_t - Y_{t-1})^2$$

(23) $$= h\Sigma_1^n Y_{t-1}\Delta Y_t - (1/2)h^2\Sigma_1^n Y_{t-1}^2$$

where $h = \alpha$-1. Since H_1 is our reference model it will be convenient in what follows to work with the deviation h as our parameter rather than α, just as in (23) above.

The likelihood process is given by $L_n = dP_n^\alpha/dP_n = \exp(\Lambda_{hn})$, and the score function process is

$$N_n = \partial\Lambda_{hn}/\partial h = \Sigma_1^n Y_{t-1}\Delta Y_t - h\Sigma_1^n Y_{t-1}^2 = V_n - hA_n \text{ , say,}$$

giving the MLE

(24) $$\hat{h}_n = A_n^{-1}V_n = (\Sigma_1^n Y_{t-1}^2)(\Sigma_1^n Y_{t-1}\Delta Y_t) \ .$$

We can now write the likelihood process in the form

(25) $$L_n = \exp\{(1/2)\hat{h}_n^2 A_n\}\exp\{-(1/2)(h - \hat{h}_n)^2 A_n\} \ ,$$

analogous to the continuous time case.

3.2. The Bayes Posterior Process, Bayes Model and Bayes Measure

Suppose we have given a prior density $\pi(h)$ on $h = \alpha$-1. The joint density process for (β, Y^n) is then

$$\Pi_n = \pi(h)(dP_n^\alpha/dP_n) = \pi(h)L_n = \pi(h)\exp\{(1/2)\hat{h}_n^2 A_n\}\exp\{-(1/2)(h-\hat{h}_n)^2 A_n\}$$

(26) $$= \left[A_n^{-1/2}\exp\{(1/2)\hat{h}_n^2 A_n\}\right]\left[\pi(h)A_n^{1/2}\exp\{-(1/2)(h-\hat{h}_n)^2 A_n\}\right] .$$

As in the continuous case, we shall take the case of a uniform prior on h with the improper density $\pi(h) = (2\pi)^{-1/2}$. Integrating (26) over h, we define the *discrete Bayes measure* Q_n by its derivative with respect to P_n, i.e.

(27) $$dQ_n/dP_n = \int_{\mathbf{R}}\pi(h)(dP_n^\alpha/dP_n)dh = A_n^{-1/2}\exp\{(1/2)h_n^2 A_n\}$$

(28) $$= A_n^{-1/2}\exp\{(1/2)V_n^2 A_n^{-1}\} = Z_n \text{ , say.}$$

Note that $V_n = \Sigma_1^n Y_{t-1}\Delta Y_t = \Sigma_1^n Y_{t-1}u_t$ under H_1 so that V_n is a P_n-martingale. Its conditional quadratic variation process is

$$\langle V_n \rangle = \Sigma_1^n Y_{t-1}^2 E[(\Delta Y_t)^2 | \mathcal{F}_{t-1}] = \Sigma_1^n Y_{t-1}^2 = A_n \ , \quad \text{say.}$$

Then Z_n in (28) has a general form in terms of the martingale V_n and its quadratic variation process A_n. This form, like the continuous version given in (7), is important in generalizing our ideas and methods beyond the simple models considered here and in Section 2. Note the special case where V_n is a random walk with quadratic variation process $A_n = n$, leading to

$$Z_n' = n^{-1/2}\exp\{(1/2n)V_n^2\} \ .$$

The process Z_n' (like Z_t' in (8)) satisfies the martingale property $E(Z_n'|\mathcal{F}_m) = Z_m'$ a.s. $(m < n)$ even though Z_n' itself is not integrable.

Using (27) we now write (26) in the form

$$\begin{aligned} \Pi_n &= \pi(h)(dP_n^\alpha/dP_n) \\ &= (dQ_n/dP_n)[(2\pi)^{-1/2}A_n^{-1/2}\exp\{-(1/2)(h-\hat{h}_n)^2A_n\}] \\ &= (dQ_n/dP_n)N(\hat{h}_n, A_n^{-1}) \ . \qquad (29) \end{aligned}$$

Thus, as in the continuous case, the Bayesian posterior process for h, viz. $N(\hat{h}_n, A_n^{-1})$, is obtained from (29) by changing the reference measure from P_n to Q_n, i.e.

$$\Pi_n^B = \Pi_n(dP_n/dQ_n) = \pi(\beta)dP_n^\alpha/dQ_n = N(\hat{h}_n, A_n^{-1}) \ . \qquad (30)$$

This translation of measure has the same interpretation as in the continuous case: Bayesian inference based on the Gaussian posterior $\Pi_n^B = N(\hat{h}_n, A_n^{-1})$ involves an implicit change of reference measure to the path dependent measure Q_n.

Like Q_t, Q_n is σ-finite, since $Q_n(\Omega) = \int_{\mathbf{R}} Z_n dP_n = E(Z_n) = \infty$, but the measure does give rise to proper conditional densities $Q_n(\cdot|\mathcal{F}_{n_0})$ for some $n_0 \geq 1$ from the ratio

$$\begin{aligned} \frac{dQ_n/dP_n}{dQ_{n_0}/dP_{n_0}} &= \frac{Z_n}{N_{n_0}} = [A_n/A_{n_0}]^{-1/2}\exp\{(1/2)V_n^2A_n^{-1} - (1/2)V_{n_0}^2A_{n_0}^{-1}\} \qquad (31) \\ &= R_n \ , \quad \text{say.} \end{aligned}$$

The following lemma gives an alternative representation of R_n that is very important in understanding the model associated with the measure Q_n.

LEMMA 3.1

$$(32) \qquad R_n = \prod_{t=n_0+1}^{n} \frac{(1/2\pi f_t)^{1/2}\exp\{-(1/2f_t)(\Delta f_t - \hat{h}_{t-1}Y_{t-1})^2\}}{(1/2\pi)^{1/2}\exp\{-(1/2)(\Delta Y_t)^2\}}$$

where $f_t = 1 + Y_{t-1}^2/A_{n-1}$. □

THEOREM 3.2

$$(33) \qquad dQ_n/dQ_{n_0} = \prod_{t=n_0+1}^{n} (1/2\pi f_t)^{1/2}\exp\{-(1/2f_t)(\Delta Y_t - \hat{h}_{t-1}Y_{t-1})^2\} \ . \quad \square$$

REMARKS

(i) Expression (33) gives the conditional density of the measure Q_n given $\mathcal{F}_{n_0}$. Note that

$$dQ_n/dQ_{n_0} = (dQ_n/d\nu)/(dQ_{n_0}/d\nu) \ ,$$

so that (33) is in fact the conditional density with respect to Lebesgue measure (ν) of Q_n given $\mathcal{F}_{n_0}$. As we see below in Theorem 3.3, the form of (33) reveals the nature of the model associated with the measure Q_n.

(ii) Let n_0 be interpreted as a time when a minimal amount of information about the process Y_t has accumulated. For instance, if $n_0 = 1$ there is just enough data to estimate h by $\hat{h}_1 = Y_0\Delta Y_1 = \Delta Y_1/Y_0$. For values of $n_0 \geq 1$ we can use the period $0 \leq t \leq n_0$ to construct the prior $\pi(h) = \Pi_{n_0}^B = N(\hat{h}_{n_0}, A_{n_0}^{-1})$. Combining this prior with the likelihood based on data over $n_0 < t \leq n$ we get (as in the case of (18))

$$\Pi_{n,n_0}^B = N(\hat{h}_n, A_n^{-1})$$

which is the same as (30). Thus, there is coherent updating in the prior and posterior when the initial condition is shifted from $t = 0$ to $t = n_0$. □

THEOREM 3.3

(a) *Under a uniform prior* $\pi(h)$, *the posterior process* Π_n^B *is Gaussian with distribution* $N(\hat{h}_n, A_n^{-1})$ *at time n. The same posterior applies when a different initialization* n_0 *is chosen and the prior is updated to* $\pi(h) = \Pi_{n_0}^B$.

(b) *Bayes methods that are based on the posterior* Π_n^B *imply the use of the discrete Bayes measure* Q_n *as the reference measure in constructing the likelihood.*

(c) *The model to which* Q_n *refers is the time varying parameter model*

$$(35)\qquad H_{\hat{\alpha}_{n-1}} : Y_n = \hat{\alpha}_{n-1} Y_{n-1} + v_n\ , \quad n \geq n_0 + 1$$

where the evolving parameter $\hat{\alpha}_{n-1}$ *is the path dependent MLE*

$$\hat{\alpha}_{n-1} = 1 + \hat{h}_{n-1} = \Sigma_1^{n-1} Y_t Y_{n-1} / \Sigma_1^{n-1} Y_{t-1}^2$$

and $v_n \equiv$ iid $N(0, f_n)$ *with*

$$(36)\qquad f_n = 1 + Y_{n-1}^2 / A_{n-1} = 1 + Y_{n-1}^2 / \Sigma_1^{n-1} Y_{t-1}^2\ .$$

In particular, the conditional density of Q_n *given* $\mathcal{F}_{n_0}$ *is identical to the conditional density of data generated by the model* (35) *given* $\mathcal{F}_{n_0}$. □

REMARKS

(1) As in the continuous case, traditional Bayes inference converts the concept of a true model (here H_1, with reference measure P_n) to a Bayes model (here $H_{\hat{\alpha}_{n-1}}$, with reference measure Q_n) in which the parameters evolve according to the observed trajectory of the process. The form of the "Bayes model" (35) follows directly from the conditional density dQ_n/dQ_{n_0} given in Theorem 3.2 by (33). Setting $n_0 = n-1$ in the latter expression we have

$$(37)\qquad dQ_n/dQ_{n-1} = dQ_n/d\nu|_{\mathcal{F}_{n-1}} = (1/2\pi f_n)^{1/2} \exp\{-(1/2f_n)(\Delta Y_n - \hat{h}_{n-1} Y_{n-1})^2\} \equiv N(\hat{h}_{n-1} Y_{n-1}, f_n)$$

which is the conditional density of the most recent observation Y_n given the past history of the process to Y_{n-1} (i.e. given $\mathcal{F}_{n-1}$).

(ii) The Bayes model (35) is identical to a classical prediction model for Y_n given $\mathcal{F}_{n-1}$. Note that $\hat{\alpha}_{n-1} Y_{n-1}$ is the MLE of the Wiener-Kolmogorov predictor $E(Y_n | \mathcal{F}_{n-1}) = \alpha Y_{n-1}$ in the model

(1) where the expectation is with respect to P^{α} measure. The variance, $f_n = 1 + Y_{n-1}^2/A_{n-1}$, of the error process v_n in (35) is the asymptotic variance of the classical prediction error $Y_n - \hat{\alpha}_{n-1}Y_{n-1} = u_n - (\hat{\alpha}_{n-1} - \alpha)Y_{n-1}$. In our Bayesian context under the measure Q_n, $\hat{\alpha}_{n-1}Y_{n-1} = E(Y_n|\mathscr{F}_{n-1})$ is the conditional expectation directly (i.e. the Wiener-Kolmogorov predictor under Q_n measure) and $f_n = \mathrm{var}(Y_n|\mathscr{F}_{n-1})$ is the conditional variance. This equivalence between the Bayes model that corresponds to the path dependent measure Q_n and the asymptotic form of the classical prediction model goes a long way towards reconciling the differences between Bayes and classical inference in simple time series models like (1).

(iii) It should be pointed out that model (35), when interpreted as a Bayes model for the data, has some advantages over the classical prediction model interpretation of (35). First, (35) is an exact finite sample result in Bayes theory whereas the classical interpretation of (35) relies on asymptotic theory, in particular, $1/A_{n-1}$ is the asymptotic variance of the estimator $\hat{\alpha}_{n-1}$ whereas it is the exact variance of the posterior distribution of α. Second, the model (35) carries with it the measure Q_n. This is especially useful when we come to compare models, e.g. a model like (35) with a similar model in which there is a unit autoregressive root (i.e. $\Delta Y_n = u_n$) for which the associated measure is P_n. Model comparisons of this type (which in the case just given amount to a test of the presence of a unit root) can then be performed by considering the likelihood ratio of the respective measures, i.e. dQ_n/dP_n in the example given. This idea forms the basis of the approach to inference that we recommend for Bayes models like (35).

(iv) It is interesting to observe that the error process v_n in the model (35) is a nonlinear ARCH process, unlike the error in the continuous time Bayes model (1^B). Note, however, that as $n \to \infty$, $Y_{n-1}^2/A_{n-1} \to_{a.s.} 0$ and the ARCH effects die out for large n.

(v) The Bayes model Q_n that corresponds to the model (35) was originally defined by the RN derivative dQ_n/dP_n given in (28) and this in turn relied on the uniform prior $\pi(h) = (2\pi)^{-1/2}$. Since Q_n is a σ-finite measure, any other choice of non zero constant for the uniform prior $\pi(h)$ would lead to a measure equivalent to Q_n (i.e. each measure would be absolutely continuous with respect to the other) and the conditional distributions dQ_n/dQ_{n-1} would be exactly the same for each measure. Thus, the Bayes model (35) is invariant to the choice of the scale in the uniform prior. The particular choice $\pi(h) = (2\pi)^{-1/2}$ is motivated by the following argument. If in place of P_n as the reference measure we chose instead the measure P_n^0 (corresponding to $\alpha = 0$ in model (22)) we would have

$$dQ_n/dP_n^0 = A_n^{-1/2} \exp\{(1/2)\hat{\alpha}_n^2 A_n\} .$$

Under P_n^0 we have the canonical model $y_t = u_t \equiv$ iid $N(0, 1)$ and then $n^{-1}A_n \rightarrow_{\text{a.s.}} 1$. Hence, twice the logarithm of the likelihood ratio dQ_n/dP_n^0 scaled by $n^{1/2}$ (to avoid a degenerate limit) is

$$2 \ln[n^{1/2}(dQ_n/dP_n^0)] = \hat{\alpha}_n^2 A_n - \ln(n^{-1}A_n) = \hat{\alpha}_n^2 A_n + o_p(1) .$$

Note that $\hat{\alpha}_n^2 A_n$ is the Wald statistic for testing the hypothesis $\alpha = 0$. Thus, with the explicit choice of the prior $\pi(\alpha) = (2\pi)^{-1/2}$ the likelihood ratio dQ_n/dP_n^0 is asymptotically equivalent to the classical Wald and likelihood ratio tests of the null hypothesis $H_0 : \alpha = 0$. □

3.3. A Bayes Model Test

Bayes methods change the frame of reference to a Bayes measure (Q_n) and Bayes model ($H_{\hat{\alpha}_{n-1}}$). It should therefore be possible to test one Bayes model against another using a likelihood ratio test. It is possible to pursue both classical (Neyman-Pearson) and Bayesian (posterior odds) approaches at this point. The former is valuable because it facilitates comparisons with other classical tests (e.g. of the unit root hypothesis). We will therefore proceed along these lines in this section and in Section 3.4. However, as will become clear in Section 3.5 our recommended procedure for practical implementation is Bayesian and is based on posterior odds.

We now apply this idea, starting with model $H_{\hat{\alpha}_{n-1}}$. From (27) we see that twice the log-likelihood ratio is

$$(38) \qquad 2 \ln(dQ_n/dP_n) = \hat{h}_n^2 A_n - \ln(A_n) .$$

Under H_1 we standardize A_n by n^{-2} to ensure a well defined limit process. This leads us to define the Bayes model likelihood ratio test statistic as

$$(39) \qquad BLR = \hat{h}_n^2 A_n - \ln(n^{-2}A_n) .$$

When the error variance σ^2 in H_1 and $H_{\hat{\alpha}_{n-1}}$ is unknown and must be estimated we employ the estimate

$$\hat{\sigma}^2 = n^{-1}\Sigma_1^n(Y_t - \hat{\alpha}_n Y_{t-1})^2$$

and then the *BLR* statistic is

$$BLR_o = \hat{h}_n^2 A_n/\hat{\sigma}^2 - \ln(n^{-1}A_n/\hat{\sigma}^2) \ . \tag{40}$$

Using standard functional limit theory we obtain:

THEOREM 3.4. *Under* H_1

$$BLR,\ BLR_o \Rightarrow \left(\int_0^1 SdS\right)^2 / \int_0^1 S^2 - \ln(\int_0^1 S^2) = g(S) \tag{41}$$

where $S(\cdot) \equiv BM(1)$ *is standard Brownian motion.* □

We may use the statistic BLR_o to conduct a classical test of H_1 against H_α ($\alpha \neq 1$). Critical values of the limit functional are readily obtained by simulation. Letting $g_{0.95}$ denote the right tail 5% critical value of $g(S)$, a 5% level test of H_1 against H_α ($\alpha \neq 1$) is provided by the criterion

$$BLR_o > g_{0.95} \ .$$

Observe that the BLR_o statistic is a nonlinear mixture of the Dickey-Fuller (squared) t-ratio statistic, $\hat{h}_n^2 A_n/\hat{\sigma}^2$, and the Anderson-Darling/Sargan-Bhargava statistic $n^{-2}A_n/\hat{\sigma}^2$. (The latter would apply precisely if σ^2 were estimated under the null by $s^2 = n^{-1}\Sigma_1^n(\Delta Y_t)^2$). Rates of divergence of the statistic BLR_o are easily seen to be $O_p(n)$ under the alternative $\alpha < 1$ and $O_p(|\alpha|^n)$ under the alternative $\alpha > 1$.

Figure 1 about here

The performance of the BLR_o test in finite samples was explored in simulations and the results are reported in Figure 1. Comparisons were made between the BLR_o test and the Dickey-Fuller t-test (DF(t)) and coefficient test (DF(a)). In each case the size of the test was set at 5% and the graphs in Figure 1 show the power functions of the three tests for the sample size $n = 100$. The results show that there is little to choose between the tests at this sample size.

3.4. Some Model Extensions

The ideas of the last subsection can be used to develop tests for a unit root that apply in models with drift, with deterministic trends and with transient dynamics. These extensions will be considered in turn to illustrate the theory. First we look at models with intercept or drift, i.e.

$$H_{\mu,\alpha} : Y_t = \mu + \alpha Y_{t-1} + u_t$$
$$H_{\mu,1} : Y_t = \mu + Y_{t-1} + u_t$$

where $u_t \equiv$ iid $N(0, 1)$ and the time series are initialized at $t = 0$ with Y_0 being $\mathcal{F}_0$-measurable. We shall proceed with the same general notation as before. The density process of $H_{\mu,1}$ with reference to H_1 (whose measure is represented by P_n) is

$$dP_n^{\mu,1}/dP_n = \exp\left\{-(1/2)\Sigma_1^n(\Delta Y_t - \mu)^2 + (1/2)\Sigma_1^n(\Delta Y_t^2)\right\} = \exp\left\{(\Sigma_1^n \Delta Y_t)\mu - (1/2)\mu^2 n\right\} .$$

Let $\pi(\mu)$ be the prior density of μ and $\hat{\mu}_n = n^{-1}\Sigma_1^n \Delta Y_t$ be the usual maximum likelihood estimate under $H_{\mu,1}$. Then the posterior process is

$$\Pi_n = \pi(\mu)(dP_n^{\mu,1}/dP_n) = \pi(\mu)\exp\{\hat{\mu}\mu n - (1/2)\mu^2 n\}$$

$$= \left[n^{-1/2}\exp\{\hat{\mu}_n^2 n/2\}\right]\left[\pi(\mu)n^{1/2}\exp\{-(1/2)(\mu - \hat{\mu}_n)^2 n\}\right] . \tag{42}$$

The Bayes model measure $Q_n^{\mu,1}$ is determined by the RN derivative

$$dQ_n^{\mu,1}/dP_n = n^{-1/2} \exp\{\hat{\mu}_n^2 n/2\} , \tag{43}$$

and associated Bayes model $B_{\mu,1}$ is

$$\Delta Y_{n+1} = \hat{\mu}_n + v_{n+1} \ ; \quad v_{n+1}|_{\mathcal{F}_n} \equiv N(0, (n+1)/n) ,$$

in place of $H_{\mu,1}$.

Following (38) and (39) the Bayes model likelihood ratio test of $H_{\mu,1}$ against the null reference model $H_{\mu=0,1}$ is just

$$BLR(\hat{\mu}_n) = 2 \ln\left(n^{1/2}dQ_n^{\mu,1}/dP_n\right) = n\hat{\mu}_n^2 ,$$

with asymptotic distribution given by

(45) $$BLR(\hat{\mu}_n) \Rightarrow \chi_1^2 \ ,$$

under P_n (i.e. $H_{\mu=0,1}$). Again, when the error variance is to be estimated we may use

$$\hat{\sigma}^2 = n^{-1}\Sigma_1^n(\Delta Y_t - \hat{\mu}_n)^2$$

and the test statistic is

$$BLR_\sigma(\hat{\mu}_n) = n\hat{\mu}_n^2/\hat{\sigma}^2 \ ,$$

with the same limit distribution as (45).

BLR and BLR_σ are Bayes model likelihood ratio tests for the presence of a drift in the model $H_{\mu,1}$ with a unit root. Our next object is to find the *BLR* test of model $H_{\mu,\alpha}$ against model $H_{\mu,1}$. The density process of $H_{\mu,\alpha}$ with reference to H_1 is

$$\begin{aligned} dP_n^{\mu,\alpha}/dP_n &= \exp\{-(1/2)\Sigma_1^n(Y_t - \mu - \alpha Y_{t-1})^2 + (1/2)\Sigma_1^n(\Delta Y_t)^2\} \\ &= \exp\{-(1/2)\Sigma_1^n(\Delta Y_t - \mu - hY_{t-1})^2 + (1/2)\Sigma_1^n(\Delta Y_t)^2\} \\ &= \exp\{-(1/2)\Sigma_1^n(\Delta Y_t - \theta' X_t)^2 + (1/2)\Sigma_1^n(\Delta Y_t)^2\} \\ &= \exp\{\theta'\Sigma_1^n X_t\Delta Y_t - (1/2)\theta'\Sigma_1^n X_t X_t'\theta\} \end{aligned}$$

where $\theta' = (\mu, \ h)$ and $X_t' = (1, Y_{t-1})$. The maximum likelihood estimator of θ is $\hat{\theta}_n = \left(\Sigma_1^n X_t X_t'\right)^{-1}(\Sigma_1^n X_t \Delta Y_t)$.

If $\pi(\theta)$ is the prior density of θ then the posterior process is

$$\Pi_n = \pi(\theta) dP_n^\theta/dP_n \ , \quad P_n^\theta = P_n^{\mu,1+h} \ .$$

Using the same approach as before we now decompose this density into two factors as

(46) $$\Pi_n = \left[|A_n|^{-1/2}\exp\{(1/2)\hat{\theta}_n' A_n \hat{\theta}_n\}\right]\left[\pi(\theta)|A_n|^{1/2}\exp\{(-1/2)(\theta - \hat{\theta}_n)' A_n(\theta - \hat{\theta}_n)\}\right] ,$$

where $A_n = \Sigma_1^n X_t X_t'$. The Bayes model measure is then obtained from the derivative

(47) $$dQ_n^\theta/dP_n = |A_n|^{-1/2}\exp\{(1/2)\hat{\theta}_n' A_n \hat{\theta}_n\} \ .$$

This is a useful general form of the Bayes model measure that will be utilized extensively in what follows. After a little calculation in the present case with $\hat{\theta}_t' = (\hat{\mu}_n, \hat{h}_n)$, we find

$$dQ_n^{\theta}/dP_n = \exp\left\{(1/2)\hat{h}_n^2\Sigma_1^n(Y_{t-1} - \bar{Y}_{-1})^2 - (1/2)\ln\left(n\Sigma_1^n(Y_{t-1} - \bar{Y}_{-1})^2\right) + (1/2)n(\hat{\mu}_n + \hat{h}_n\bar{Y}_{-1})^2\right\} .$$

Now $\hat{\mu}_n = \overline{\Delta Y} - \hat{h}_n\bar{Y}_{-1}$ so that the above expression simplifies to

$$dQ_n^{\theta}/dQ_n^{\mu,1} = \exp\left\{(1/2)\hat{h}_n\Sigma_1^n(Y_{t-1} - \bar{Y}_{-1})^2 - (1/2)\ln\left(n\Sigma_1^n(Y_{t-1} - \bar{Y}_{-1})^2\right) + (1/2)n\overline{\Delta Y}^2\right\} . \tag{48}$$

Next observe that the Bayes model measure for $B_{\mu,1}$ is, from (43),

$$dQ_n^{\mu,1}/dP_n = \exp\{(1/2)n\overline{\Delta Y}^2 - (1/2)\ln(n)\} . \tag{49}$$

Combining (48) and (49) we obtain

$$\begin{aligned} dQ_n^{\theta}/dQ_n^{\mu,1} &= (dQ_n^{\theta}/dP_n)(dP_n/dQ_n^{\mu,1}) \\ &= \exp\left\{(1/2)\hat{h}_n^2\Sigma_1^n(Y_{t-1} - \bar{Y}_{-1})^2 - (1/2)\ln\left(\Sigma_1^n(Y_{t-1} - Y_{-1})^2\right)\right\} . \end{aligned} \tag{50}$$

Factoring in the sample size n to ensure a limit distribution for $n^{-2}\Sigma_1^n(Y_{t-1} - \bar{Y}_{-1})^2$, we have

$$BLR = 2\ln\left(n(dQ_n^{\theta}/dQ_n^{\mu,1})\right) = \hat{h}_n^2\Sigma_1^n(Y_{t-1} - \bar{Y}_{-1})^2 - \ln\left(n^{-2}\Sigma_1^n(Y_{t-1} - \bar{Y}_{-1})^2\right) .$$

Finally, estimating the error variance by $\hat{\sigma}^2 = n^{-2}\Sigma_1^n(Y_t - \hat{\theta}_n'X_t)$, we have the Bayes model likelihood ratio test

$$BLR_{\sigma} = \hat{h}_n^2\Sigma_1^n(Y_{t-1} - \bar{Y}_{-1})^2/\hat{\sigma}^2 - \ln\left\{n^{-2}\Sigma_1^n(Y_{t-1} - \bar{Y}_{-1})^2/\hat{\sigma}^2\right\} .$$

THEOREM 3.5. *Under* $H_1 = H_{\mu=0,1}$

$$BLR,\ BLR_{\sigma} \Rightarrow \left[\int_0^1 \underline{S}dS\right]^2 - \ln\left(\int_0^1 \underline{S}^2\right) ,$$

where $\underline{S}(\cdot) = S(\cdot) - \int_0^1 S$ *is demeaned Brownian motion and* $S(\cdot) \equiv BM(1)$. □

Models with higher order deterministic trends can easily be accommodated in this approach. Let

$$H_{\varphi,\alpha}^k : Y_t = X_t'\varphi + \alpha Y_{t-1} + u_t\ , \quad u_t \equiv \text{iid } N(0,\ \sigma^2) \tag{51}$$

be a model with auxiliary regressors $X_t' = (1, t, t^2, \ldots, t^k)$ and parameters $\varphi' = (\varphi_0, \varphi_1, \ldots, \varphi_k)$.

Proceeding as before, we find the Bayes model

$$(52)\qquad B^k_{\varphi,\alpha} : Y_{n+1} = X'_{n+1}\hat{\varphi}_n + \hat{\alpha}_n Y_n + v_{n+1} = Z'_{n+1}\hat{\theta}_n + v_{n+1} ;$$

$$v_{n+1}|_{\mathscr{F}_n} \equiv N(0,\ \sigma^2(1 + Z'_{n+1}A_n^{-1}Z_{n+1}))$$

where $\hat{\theta}_n = (\hat{\varphi}'_n,\ \hat{\alpha}_n) = \left(\Sigma_1^n Z_t Z'_t\right)^{-1}(\Sigma_1^n Z_t Y_t)$, $Z'_t = (X'_t,\ Y_{t-1})$ and $A_n = \Sigma_1^n Z_t Z'_t$. The Bayes model measure for $B^k_{\varphi,\alpha}$ is then obtained from (47).

In a similar way, when we restrict the autoregressive coefficient to $\alpha = 1$ we obtain the Bayes model

$$(53)\qquad B^k_{\varphi,1} : \Delta Y_{n+1} = X'_{n+1}\tilde{\varphi}_n + v_{n+1} ;\quad v_{n+1}|_{\mathscr{F}_n} \equiv N(0,\ \sigma^2(1 + X'_{n+1}A_n^{-1}X_{n+1})) ,$$

with $\tilde{\varphi}_n = \left(\Sigma_1^n X_t X'_t\right)^{-1}(\Sigma_1^n X_t \Delta Y_t)$ and $A_n = \Sigma_1^n X_t X'_t$. The Bayes model measure for $B^k_{\varphi,1}$ is again given by expression (47). We now have

$$\begin{aligned} dQ_n^{\varphi,\alpha}/dQ_n^{\varphi,1} &= (dQ_n^{\varphi,\alpha}/dP_n)(dQ_n^{\varphi,1}/dP_n) \\ &= \exp\{(1/2)\hat{\theta}'_n(\Sigma_1^n Z_t Z'_t)\hat{\theta}_n - (1/2)\tilde{\varphi}'_n(\Sigma_1^n X_t X'_t)\tilde{\varphi}_n - (1/2)\ln(|\Sigma_1^n Z_t Z'_t|/|\Sigma_1^n X_t X'_t|)\} \\ (54)\qquad &= \exp\{(1/2)\hat{h}_n^2 Y'_{-1}Q_X Y_{-1} - (1/2)\ln(Y'_{-1}Q_X Y_{-1})\} \end{aligned}$$

where $\hat{\alpha}_n = 1 + \hat{h}_n$, $Y'_{-1} = (Y_0,\ Y_1,\ \ldots,\ Y_{n-1})$ and Q_X is the orthogonal projection matrix onto the range of $X = [X_1,\ \ldots,\ X_n]'$.

The Bayes model likelihood ratio test of $H^k_{\varphi,\alpha}$ against $H^k_{\varphi,1}$ is therefore based on the statistic

$$\begin{aligned} BLR &= 2\ \ln\{n(dQ_n^{\varphi,\alpha}/dQ_n^{\varphi,1})\} \\ &= \hat{h}_n^2 Y'_{-1}Q_X Y_{-1} - \ln(n^{-2}Y'_{-1}Q_X Y_{-1}) . \end{aligned}$$

Again, when σ^2 is estimated we have

$$(55)\qquad BLR_\sigma = \hat{h}_n^2 Y'_{-1}Q_X Y_{-1}/\hat{\sigma}^2 - \ln\{n^{-2}(Y'_{-1}Q_X Y_{-1})/\sigma^2\}$$

where $\hat{\sigma}^2 = n^{-1}\Sigma_1^n(Y_t - \hat{\varphi}'_n X_t - \hat{\alpha}_n Y_{t-1})^2$.

THEOREM 3.6. *Under the null model $H^{k-1}_{\varphi,1}$ (i.e. $H^k_{\varphi,1}$* with $\varphi_k = 0$) *we have*

$$BLR,\ BLR_\sigma \Rightarrow \left(\int_0^1 S_k dS\right)^2 / \int_0^1 S_k^2 - \ln\left(\int_0^1 S_k^2\right) \tag{56}$$

where $S_k(\cdot)$ is the detrended Brownian motion

$$S_k(r) = S(r) - \hat{\delta}_0 - \hat{\delta}_1 r - \cdots - \hat{\delta}_k r^k$$

with

$$\hat{\delta} = \left(\int_0^1 p(r)p(r)'\right)^{-1} \int_0^1 p(r)S(r)$$

and $p(r) = (1, r, \ldots, r^k)'$. □

The statistic BLR_σ in (55) may be used to test $H^{k-1}_{\varphi,1}$ (a model with a unit root and drift process of degree $k-1$) against $H^k_{\varphi,\alpha}$ (a trend stationary model with trend degree k). Both (55) and its limit distribution given in (56) are invariant to the trend coefficients φ under the maintained hypothesis that $\varphi_k = 0$, i.e. that Y_t follows a process which can be decomposed into the sum of a k^{th} order deterministic trend and a stochastic trend. The statistic (55) may therefore be used to test for the presence of a unit root in a time series model where there is a maintained deterministic trend. In this sense, the Bayes likelihood ratio test BLR_σ may be regarded as a Bayes version of the classical tests of Dickey-Fuller (1981), Phillips-Perron (1988) and Ouliaris-Park-Phillips (1989).

Figure 2 shows the power curves of the BLR_σ and Dickey-Fuller t−test (DF(t)) and coefficient test (DF(a)) when $k = 1$ (i.e. there is a fitted linear trend in (51) and (52)) and $n = 100$. The size was controlled at the 5% level for each test. The power functions show that the BLR_σ test has greater finite sample power than the DF(t) test but lower power than the DF(a) test.

Figure 2 about here

Models with trends and transient dynamics can be treated in the same way. Consider for example the commonly used augmented Dickey-Fuller model

$$H^k_{\varphi,\psi,\alpha} : Y_t = \alpha Y_{n-1} + \Sigma_{j=1}^{p-1} \psi_j \Delta Y_{t-j} + X_t'\varphi + u_t\ ,\quad u_t \equiv \text{iid } N(0,\ \sigma^2) \tag{57}$$

where X_t is the same trend polynomial as in model (51). Bayes models can now be constructed in the same way as (52) and (53) but with the additional regressors $\Delta Y_{t-1}, \ldots, \Delta Y_{t-p+1}$. The RN derivative of the respective Bayes measures of these models is then

$$(58) \qquad \frac{dQ_n^{\varphi,\psi,\alpha}}{dQ_n^{\varphi,\psi,1}} = \frac{\exp\{(1/2)\hat{h}_n^2 Y_{-1}' Q_{\underline{X}} Y_{-1}\}}{(Y_{-1}' Q_{\underline{X}} Y_{-1})^{1/2}}$$

where $\hat{h}_n = \hat{\alpha}_{n-1}$, $\hat{\alpha}_n = (Y_{-1}' Q_{\underline{X}} Y_{-1})^{-1}(Y_{-1}' Q_{\underline{X}} Y)$, $Q_{\underline{X}}$ is the projection matrix onto the orthogonal complement of the range of $\underline{X} = [\underline{X}_1, \ldots,$ (10U$\underline{X}_n$]' $\underline{X}_t' = [X_t', \Delta Y_{t-1}, \ldots, \Delta Y_{t-p}]$. Estimating σ^2 in the usual way, i.e. by the MLE $\hat{\sigma}^2$ from the more general model $H_{\varphi,\psi,\alpha}$, we obtain the likelihood ratio statistic

$$(59) \qquad \begin{aligned} BLR_o &= 2\ \ln\{n(dQ_n^{\varphi,\psi,\alpha}/dQ_n^{\varphi,\psi,1})\} \\ &= \hat{h}_n^2 Y_{-1}' Q_{\underline{X}} Y_{-1}/\hat{\sigma}^2 - \ln\{n^{-2}(Y_{-1} Q_{\underline{X}} Y_{-1})/\hat{\sigma}^2\} \ . \end{aligned}$$

Under the null hypothesis $H_{\varphi,\psi,1}^{k-1}$ (i.e. $H_{\varphi,\psi,1}^{k}$ with $\varphi_k = 0$) BLR_o in (59) has the same limit as that given in Theorem 3.6 by (56). Tests based on the BLR_o statistic (59) are therefore Bayes versions of the augmented Dickey-Fuller test (cf. Nelson-Plosser (1982) and Said-Dickey (1984)).

3.5. Bayes Model Posterior Odds

How do Bayes model likelihood ratio tests relate to conventional Bayes testing procedures like posterior odds ratios and best Bayes tests? To address this question we look at these alternatives in the context of the simple autoregressive models H_α and H_1 considered earlier (see (22)).

Let π_1 and π_α represent the prior probabilities of H_1 and H_α. The posterior odds ratio of H_α to H_1 is

$$\frac{dP_n^\alpha}{dP_n} \cdot \frac{\pi_\alpha}{\pi_1}$$

and the "Bayes factor" in favor of H_α is dP_n^α/dP_n. If we use a loss structure to penalize incorrect decisions and form a basis for action, then the Bayes solution corresponds to the choice that minimizes the Bayes risk. When the loss function is symmetrical in the sense that the losses from type I and type II errors are set to be the same, the decision rule is (cf. Zellner, 1971, pp. 295-297):

(60) *if* $dP_n^\alpha/dP_n > \pi_1/\pi_\alpha$, *then decide in favor of* H_α

i.e. decide in favor of H_α if the posterior odds > 1. The criterion (60) is sometimes called the "best Bayesian test" of H_α against H_1 (e.g. Grenander, 1981, Theorem 3, p. 111) or the Bayes solution (Hall and Heyde, 1980, p. 163).

To implement the decision rule (60) we would need to have a specific model H_α and value of α to incorporate in the criterion. As we have argued earlier in the paper, Bayes methods do not rely on the concept of a true model like H_α with a true value of the parameter α. Instead, as shown in Theorem 3.3, Bayes methods (with a uniform prior on α) imply the replacement of the model H_α by the trajectory-based version given in (35) where the parameter $\hat{\alpha}_{n-1}$ relies on the latest available data on the trajectory and is updated as soon as new data is available. As we have seen, corresponding to this model is a path dependent measure Q_n. In constructing a posterior odds test it therefore seems appropriate that we compare the prior odds with a Bayes factor that is determined by the likelihood ratio of the two competing Bayes models. In the present case, these are $H_{\hat{\alpha}_{n-1}}$ as given by (35) and the unit root model H_1. Noting that

(61) $$\frac{dQ_n}{dP_n} = \frac{\exp\{(1/2)\hat{h}_n^2\Sigma_1^n Y_{t-1}^2\}}{\left(\Sigma_1^n Y_{t-1}^2\right)^{1/2}} .$$

This leads to the following decision rule

(62) *if* $dQ_n/dP_n > \pi_1/\pi_\alpha$ *then decide in favor of the Bayes model* $H_{\hat{\alpha}_{n-1}}$ *over the model* H_1 *with a unit root.*

Since (61) relies on knowledge of σ^2, we construct the following statistic (analogous to BLR_σ in (40)) when σ^2 is unknown:

(61)′ $$\frac{dQ_n}{dP_n}(\hat{\sigma}^2) = \frac{\exp\{(1/2)\hat{h}_n^2\Sigma_1^n Y_{t-1}^2/\hat{\sigma}^2\}}{\left(\Sigma_1^n Y_{t-1}^2/\hat{\sigma}^2\right)^{1/2}} ,$$

where $\hat{\sigma}^2$ is MLE of σ^2 that is used in BLR_σ in (40). We will call (61)′ the *Bayes model posterior odds* criterion. We then have the decision rule:

(62)′ *if* $dQ_n/dP_n(\hat{\sigma}^2) > \pi_1/\pi_\alpha$, *then decide in favor of the Bayes model* $H_{\hat{\alpha}_{n-1}}$ *over the model* H_1 *with a unit root.*

Setting the prior odds to be unity, as we will often wish to in practice, this reduces to the very simple criterion

(62)″ *if* $dQ_n/dP_n(\hat{\sigma}^2) > 1$, *then decide in favor of* $H_{\hat{\alpha}_{n-1}}$ *over* H_1 .

THEOREM 3.7. *The Bayes model posterior odds test for the presence of a unit root in* (22) *is based on the criterion*

(63) *"if* $dQ_n/dP_n(\hat{\sigma}^2) < 1$, *then accept* H_1 *(i.e. a unit root)."*

This test is completely consistent in the sense that type I and type II errors both tend to zero as $n \to \infty$.

Next we consider the case of a model with trend and transient dynamics. To be explicit, we shall take the common case of a linear trend ($k = 1$) and write model (57) as

(64) $$H_{\psi,\mu,\beta,\alpha} : Y_t = \alpha Y_{t-1} + \mu + \beta t + \Sigma_{j=1}^{p-1}\psi_j \Delta Y_{t-j} + u_t \ , \quad u_t \equiv \text{iid } N(0, \sigma^2) \ .$$

Extensions to the general case follow in a straightforward manner. From (64) we construct three Bayes models corresponding to the following specifications:

(a) $\alpha = 1$ (a unit root in (64))

(b) $\alpha = 1, \beta = 0, \mu = 0$ (a unit root and no drift or trend in (64))

(c) $\alpha = 1, \beta = 0$ (a unit root and no trend in (64)).

The Bayes model posterior odds tests that correspond to these specializations are:

(a′) *if* $dQ_n^{\psi,\mu,\beta,\alpha}/dQ_n^{\psi,\mu,\beta,1}(\hat{\sigma}^2) < 1$, *accept the Bayes model with a unit root.*

(b′) *if* $dQ_n^{\psi,\mu,\beta,\alpha}/dQ_n^{\psi,0,0,1}(\hat{\sigma}^2) < 1$, *accept the Bayes model with a unit root and no drift and no trend.*

(c′) *if* $dQ_n^{\psi,\mu,\beta,\alpha}/dQ_n^{\psi,\mu,0,1}(\hat{\sigma}^2) < 1$, *accept the Bayes model with a unit root and a drift but no trend.*

The first of these tests is a Bayes model posterior odds version of the classical ADF test. The second and third tests are joint and they are the Bayes model posterior odds analogues of the classical joint F tests of Dickey-Fuller (1981), i.e. Φ_2 and Φ_3 in the Dickey-Fuller notation.

Each of the posterior odds tests in (a)-(c) above is completely consistent in the sense given in

Theorem 3.7. These tests have the additional advantage over classical tests that it is not necessary to have detailed tabulations of critical values for the multitude of tests that correspond to the different non standard distributions that apply depending on the nature of the deterministic regressors that are included in the regression model. For instance, in the "classical" version of the Bayes model likelihood ratio test given in Theorem 3.6, the limit distribution of BLR_{σ} is different and requires different tables of critical values for every value of k, the degree of the time trend in the fitted regression. For these reasons, we recommend the use of the posterior odds versions of these Bayes model tests, i.e. (65) and (a′)-(c′).

Finally, we observe that criteria such as (a′)-(c′) are really model selection criteria. In effect, the RN derivative of the measures of the respective Bayes models is an instrument for comparing models with a decision criterion that leads us to select the model with the greater density *a posteriori*. In the present case, the criterion is being used to determine the evidence in support of a unit root. But the principle has a much wider range of application and can, for instance, be used to determine lag length in an autoregression, trend degree and the presence or absence of a unit root.

4. TOWARD A GENERAL THEORY

The results in Sections 2 and 3 of this paper are all derived for linear models under Gaussian distributional assumptions. The purpose of this section and Section 5 is to show how the ideas we have introduced extend to a much wider class of models. We will also show that the form of the Bayes measure Q_n, as given for example by (6), (28) and (47) in linear models, is retained asymptotically even in rather general situations. Our analysis will start with continuous time processes as in Section 2 and we will later show how the theory for discrete time processes can be embodied in that of the continuous time case by a suitable embedding technique. This approach has certain advantages: (i) we can use the theory of continuous square integrable (L_2) martingales and exponential martingales, which is elegant in itself and which leads to results that are of independent interest; (ii) there is no need to distinguish stationary, nonstationary and explosive cases (as is usually the case in discrete time); and (iii) determination of the Bayes model that is implied by the use of Bayes rule in the general time series case is simpler in continuous time and can be used to infer the corresponding model in the discrete time case. The embedding theorem that enables us to achieve this (Theorem 4.7) is therefore of special interest.

4.1. General Exponential Bayes Measures

The framework we start with is fairly general but we hope the advantages of this will soon be apparent. Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $(\mathcal{F}_t)_{t\geq 0}$ be an increasing family of right continuous sub σ-fields of $\mathcal{F}$. Let $M(t)$ be a continuous local martingale with respect to $\mathcal{F}_t$ and let $A(t)$ be its quadratic variation process, i.e. $[M]_t = A(t)$.

Now suppose we have a stopping time τ. For instance in our set up τ might be a stopping time defined as

(65) $$\tau = \inf\{s : A(s) \geq c\}$$

for some constant $c > 0$. This could be interpreted as a minimal information time, where we measure information in terms of the quadratic variation $A(t)$ and prescribe a level by minimal information by the constant c. Now let $(\tau_a)_{a>0}$ be a family of monotone increasing and continuous (in a) stopping times such that $A(\tau_a)$ is a.s. (P) bounded. This could be arranged, for example, by setting

(66) $$\tau_a = \inf\{s : A(s) \geq ce^a\} \ , \quad a \geq 0$$

so that the process is, in effect, stopped before the quadratic variation gets too large. Note that with definition (66) the family $(\tau_a)_{a\geq 0}$ is, in fact, initialized at $\tau_0 = \tau$.

Consider the stopped process $M_a(t) = M(t \wedge \tau_a)$. $M_a(t)$ is a continuous local martingale with quadratic variation process $A_a(t) = [M_a]_t = A(t \wedge \tau_a)$ (e.g. Protter (1990), Theorem 22, p. 59). Moreover, since $A(\tau_a)$ is bounded we have $E\{[M_a]_\infty\} = E\{A(\tau_a)\} < \infty$ and, therefore, $M_a(t)$ is a continuous L_2 martingale (Protter (1990), Corollary 4, p. 67). This shows that we can produce martingales from the original local martingale by employing the stopping time τ_a. But, in addition, if we now index our process by a as in (66) and let $a \to \infty$, then we effect a fundamental time change in the process whereby the "new time" is measured by the information content (as measured by $A(t)$) of the original process.

Our next result shows how to construct a new measure Q_a from the original probability measure

P using exponential martingales that are based on $M(\tau_a)$ in the new time frame.

THEOREM 4.1. *The measure Q_a defined by the* RN *derivative*

$$(67)\qquad \frac{dQ_a}{dP} = \frac{\exp\{M(\tau_a) - (1/2)A(\tau_a)\}}{\exp\{M(\tau_0) - (1/2)A(\tau_0)\}}$$

is a probability measure on $(\Omega, \mathcal{F}_{\tau_a})$ *and*

$$(68)\qquad Q_b|_{\mathcal{F}_{\tau_a}} = Q_a$$

for all $\tau_b > \tau_a \geq \tau_0$, *i.e. the restriction of* Q_b *to* $\mathcal{F}_{\tau_a}$ *is given by* Q_a. □

REMARKS

(i) Theorem 4.1 shows how to construct probability measures that correspond to exponential families directly from given local martingales. This is done even without assuming the existence of the expectation of the exponentials by allowing for a suitable time change in the process. In effect, given any continuous local martingale we can find a suitable initialization and time change in the process so that a new probability measure belonging to an exponential family is constructed from the given local martingale.

(ii) The process of construction described in Remark (i) is precisely what was done in the explicit case of the martingales $V_t = \int_0^t W_s dW_s$ $(t \geq 0)$ and $G_t = \int_{t_0}^t \hat{\beta}_s dV_s$ $(t \geq t_0)$ in Section 2, see Lemma 2.3 in particular.

(iii) To illustrate, let V_t $(t \geq 0)$ be a continuous martingale with $V_0 = 0$ and quadratic variation process $A_t = [V]_t$. In this case we can set the initialization at $t = 0$ and for $\theta \in \mathbb{R}$ we have the exponential family

$$(69)\qquad dQ_t^\theta/dP = \exp\{\theta V_t - (1/2)\theta^2 A_t\} .$$

With a suitable sequence of stopping times $\tau_a \geq 0$ this now defines a parameterized sequence of new probability measures $Q_{\tau_a}^\theta$ as in (67). When $V_t = \int_0^t W_s dW_s$ and $\theta = \beta$, (69) reduces to the likelihood process L_t given earlier in (2).

(iv) Next consider the process $K(t)$ constructed from the martingale $V(t)$ and its quadratic variation $A(t)$ as follows

$$K(t) = V(t)^2/2A(t) - (1/2)\ln(A(t)) \ .$$

(We assume that $A(t) > 0$ a.s. (P) for $t > 0$.) Note that

$$\text{(70)} \qquad \exp(K(t)) = A(t)^{-1/2}\exp\{(1/2)V_t^2A_t^{-1}\}$$

corresponding to the form of the process Z_t given by (7), where $V_t = \int_0^t W_s dW_s$ and W_s is Brownian motion. We use the following lemma which gives the Ito differential of $K(t)$.

LEMMA 4.2

$$\text{(71)} \qquad dK(t) = [V(t)/A(t)]dV(t) - (1/2)[V(t)/A(t)]^2 dA(t) \ . \ \square$$

Now define, as in (67),

$$R_a = \exp\{K(\tau_a) - K(\tau_0)\} = \exp\left\{ \int_{\tau_0}^{\tau_a} K(t)dt \right\} = \exp\{G_a - (1/2)[G]_a\} \ ,$$

where $G_a = G(\tau_a) = \int_{\tau_0}^{\tau_a}[V(t)/A(t)]dV(t)$. Note that since $V(t)$ is a martingale $G(t) = \int_{\tau_0}^{t}[V(s)/A(s)]dV(s)$ is a martingale also and its quadratic variation process is $[G](t) = \int_{\tau_0}^{t}[V(s)/A(t)]^2 dA(s)$. This gives us the exponential process

$$R(t) = \exp\{G(t) - (1/2)[G](t)\} \ ,$$

and using the stopping time τ_a we have

$$\text{(72)} \qquad R_a = \exp\{G_a - (1/2)[G]_a\} = \exp\{K(\tau_a) - K(\tau_0)\} \ ,$$

as required, where $[G]_a = [G](\tau_a)$. For a suitable choice of stopping times τ_a (such that $E(\exp\{(1/2)[G]_a\}) < \infty$), the process R_a given in (72) is a martingale with $E[R_a] = 1$ and thus, as in Theorem 4.1, R_a effects a change in probability measure. Note that in this case the new measure is a path-dependent measure that depends only on the history of the original martingale $V(t)$, i.e. it is not a parameterized measure like that induced by (69). Given the form of (70) and the role that we have seen specialized versions of this process play in the transition to the posterior density using Bayes rule, it will be useful to describe the path dependent measures that are induced by R_a in (72) *exponential Bayes measures*. In the following we will see how they can be constructed in a fairly general case.

(v) Observe that (67) and (72) are defined for the "time changed" σ-fields $\mathcal{F}_{\tau_a}$. If $\sup_{t\geq 0}A(t) = \infty$ a.s. (P) then $\tau_a \to \infty$ a.s. (P) as $a \to \infty$ and then $\mathcal{F}_t = \vee_{a\geq 0}\, \mathcal{F}_{t\wedge\tau_a}$. Since (67) and (72) also apply for the stopped process $M(t\wedge\tau_a)$ and $G(t\wedge\tau_a)$ and therefore for the stopped fields $\mathcal{F}_{t\wedge\tau_a}$, we can see that they apply approximately for $\mathcal{F}_t$ to the extent that $\mathcal{F}_{t\wedge\tau_a}$ approximates $\mathcal{F}_t$ for large a (and hence large τ_a). If, on the other hand, $\sup_{t\geq 0}A(t) < \infty$ a.s. (P), then $A(t)$ and $\exp\{(1/2)A(t)\}$ are bounded a.s. (P) and $\mathrm{E}[A(t)] < \infty$ and $E(\exp\{(1/2)\mathrm{A}(t)\}) < \infty$. In this case, $M(t)$ is a martingale with $E(\exp\{(1/2)[M]_t\}) < \infty$ and (67) defines a new probability measure without the time change to τ_a. □

4.2. Approximating the Bayesian Data Measure by an Exponential Bayes Measure

Let us start by assuming that we have given a parameterized family of probability measures P_t^θ on the sequence of filtered spaces $(\Omega, \mathcal{F}_t)$. Suppose $\theta \in \mathbf{R}$ and $P_t^\theta \ll \nu$, some σ-finite measure on $(\Omega, \mathcal{F}_t)$. If $\pi(\theta)$ is a prior density on θ then the mixture

$$\mathcal{P}_t = \int_{\mathbf{R}}\pi(\theta)P_t^\theta d\theta$$

is the Bayesian data measure. Note that if $\pi(\theta)$ is improper then $\mathcal{P}_t$ will be σ-finite and we will accommodate this possibility in what follows. Let θ^0 be the "true value" of θ and set $P_t^0 = P_t^{\theta^0}$. We will write the likelihood function in terms of the density $L_t(\theta) = dP_t^\theta/dP_t^0$ and then the density of $\mathcal{P}_t$ is

$$d\mathcal{P}_t/dP_t^0 = \int_{\mathbf{R}}\pi(\theta)(dP_t^\theta/dP_t^0)d\theta = \int_{\mathbf{R}}\pi(\theta)L_t(\theta)d\theta\ .$$

Our object is now to show that $\mathcal{P}_t$ can be approximated asymptotically by an exponential Bayes measure Q_t and to find its general form.

THEOREM 4.3. *Assume the following conditions hold:*

(C1) $\ell_t(\theta) = \ln(L_t(\mathrm{t}))$ *is twice continuously differentiable with derivatives* $\ell_t^{(1)}(\theta)$ *and* $\ell_t^{(2)}(\theta)$.

(C2) *Under* P_t^θ, $\ell_t^{(1)}(\theta)$ *is a continuous local martingale with quadratic variation process* $A_t(\theta)$ *and* $A_t(\theta) \to \infty$ *a.s.* (P^θ) *as* $t \to \infty$.

(C3) $(\ell_t^{(2)}(\theta) + A_t(\theta))/A_t(\theta) \to 0$ a.s. (P^θ) *as* $t \to \infty$.

(C4) *There exist continuous functions* $w_t(\theta, \theta')$ *such that* $w_t(\theta, \theta) = 0$ *and such that for all* θ, θ' *in some neighborhood* $N_\delta(\theta^0) = \{\theta : |\theta - \theta^0| < \delta\}$ *of* θ^0 *we have*

$$\{\ell_t^{(2)}(\theta) - \ell_t^{(2)}(\theta')\}/A_t(\theta) \leq w_t(\theta, \theta') \quad \text{a.s. } (P^0)$$

for each t, *and* $w_t(\theta, \theta') \to w_\infty(\theta, \theta)$ a.s. (P^0) *uniformly for* $\theta, \theta' \in N_\delta(\theta^0)$ *and* $w_\infty(0, 0) = 0$.

(C5) *The maximum likelihood estimate* $\hat{\theta}_t$ *for* θ^0 *is consistent, i.e.* $\hat{\theta}_t \to \theta^0$ a.s. (P^0).

(C6) *For any* $\delta > 0$ *and* $\omega_\delta = \{\theta : |\theta - \theta^0| \geq \delta\}$ *we have*

$$A_t^{1/2} \int_{\omega_\delta} (dP_t^\theta / dP_t^0) d\theta \to 0 \quad \text{a.s } (P^0) \ .$$

where $A_t = A_t(\theta^0)$.

(C7) *The prior* $\pi(\theta) = (2\pi)^{-1/2}$ *is uniform on* $\mathbf{R}$.

Then

$$\frac{d\mathcal{P}_t}{dP_t^0} / \frac{dQ_t}{dP_t^0} \to 1 \quad \text{a.s. } (P^0) \tag{74}$$

where Q_t *is the exponential Bayes measure defined by the following* RN *derivative with respect to* P_t^0

$$\frac{dQ_t}{dP_t^0} = \frac{\exp\{(1/2)V_t^2 A_t^{-1}\}}{A_t^{1/2}} \ , \tag{75}$$

where $V_t = \ell_t^{(1)}(\theta^0)$ *and* $A_t = A_t(\theta^0)$. *The derivative* (75) *may also be written in the following asymptotically equivalent forms:*

$$dQ_t/dP_t^0 = \exp\left\{(1/2)\left(\hat{\theta}_t - \theta^0\right)^2 A_t\right\}/A_t^{1/2} \tag{76}$$

and

$$dQ_t/dP_t^0 = \exp\{\ell_t(\hat{\theta}_t)\}/A_t^{1/2} \ . \quad \square \tag{76}$$

REMARKS ON (C1)-(C7)

(i) Condition (C1) is standard in the asymptotic theory of regular estimators. So is the first part of (C2) -- in the usual maximum likelihood theory (e.g. Hall and Heyde (1980), p. 157) $\ell_t^{(1)}(\theta)$ is a continuous L_2 martingale under P_t^θ. The conditional variance process of $\ell_t^{(1)}(\theta)$ (the time clock of the martingale) is the quadratic variation $A_t(\theta)$ since $\ell_t^{(1)}(\theta)$ is continuous and the square bracket process $A_t(\theta)$ is therefore identical to the angle bracket or conditional variance process (e.g. Protter (1990), p. 63). The requirement that $A_t(\theta) \to \infty$ a.s. (P^θ) ensures that there is eventually an infinite amount of information about the process in the likelihood function. It corresponds to the usual persistent excitation condition in regression theory.

(ii) Condition (C3) says that $\ell_t^{(2)}(\theta) + A_t(\mathrm{t})$ must be small relative to $A_t(\theta)$ as $t \to \infty$. This is a version of the usual requirement that $J_n/I_n \to -1$ as $n \to \infty$ in the theory of maximum likelihood where I_n is the conditional variance of the score and J_n is the Hessian of the likelihood (see, for instance, Hall-Heyde, 1980, p. 160). In the present case note that $\ell_t^{(2)}(\theta) + (\ell_t^{(1)}(\theta))^2$ is a local martingale (as in the standard ML theory) and, moreover, since $\ell_t^{(1)}(\theta)$ is a local P^θ martingale, so also is $(\ell_t^{(1)}(\theta))^2 - A_t(\theta)$. Hence, we would expect the sum of these local martingales

$$\ell_t^2(\theta) + A_t(\theta) = \left\{\ell_t^2(\theta) + \left(\ell_t^{(1)}(\theta)\right)^2\right\} + \left\{A_t(\theta) - \left(\ell_t^{(1)}(\theta)\right)^2\right\}$$

to be small relative to the quadratic variation process $A_t(\theta)$, thereby giving intuitive support to (C3).

(iii) Condition (C4) is a smoothness condition. It requires, in effect, that relative differences in $\ell_t^{(2)}(\theta)$ and $\ell_t^{(2)}(\theta')$ be bounded above by an equicontinuous family of functions $\omega_t(\theta, \theta')$ in some neighborhood of θ^0 with the property that when $\theta = \theta'$ the limit function $\omega_\infty(\theta, \theta) = 0$.

(iv) Condition (C5) is standard. It could be replaced by an explicit condition on the behavior of the likelihood ratio dP_t^θ/dP_t^0 as $t \to \infty$ in closed sets like $\omega_\delta = \{\theta : |\theta - \theta^0| \geq \delta\}$ that do not contain θ^0. For instance, one commonly occurring condition (e.g. Walker, 1969, p. 83 and Hall-Heyde, 1980, p. 158) would, in the present case, take the form that for every $\delta > 0$ there is a $k(\delta) > 0$ such that

$$P\left[\sup_{\theta \in \omega_\delta} \frac{dP_t^\theta}{dP_t^0} < \exp\{-A_t(\theta^0)k(\delta)\}\right] \to 1 \ .$$

A somewhat stronger version of this condition is that for $\delta > 0$ there exists a $k(\delta)$ such that

$$(C5') \qquad \exp\{A_t(\theta^0)k(\delta)\} \sup_{\theta \in \omega_\delta} \frac{dP_t^\theta}{dP_t^0} < 1 \quad \text{a.s. } (P^0)$$

as $t \to \infty$. Then, if the prior density $\pi(\theta)$ were proper we would have

$$A_t(\theta^0)^{1/2} \int_{\omega_\delta} \pi(\theta)(dP_t^\theta/dP_t^0)d\theta \le A_t(\theta^0)^{1/2} \sup_{\theta \in \omega_\delta} (dP_t^\theta/dP_t^0) \int_{\omega_\delta} \pi(\theta)d\theta$$

$$(C6') \qquad \le A_t(\theta^0)^{1/2} \exp\{-A_t(\theta^0)k(\delta)\} \to 0 \text{ a.s. } (P^0), \text{ as } t \to \infty$$

in view of (C5′). Result (C6′) is the natural alternative to condition (C6) when the prior density $\pi(\theta)$ is proper. As it stands, (C6) simply requires that the average density dP_t^θ/dP_t^0 over a closed set like ω_δ that does not contain θ^0 is small relative to $A_t^{1/2}$ as $t \to \infty$. When $\pi(\theta)$ is proper (C6′) shows that the average density, which in this case is $\int_{\omega_\delta} \pi(\theta)(dP_t^\theta/dP_t^0)d\theta$, is exponentially small in $A_t(\theta^0)$ as $t \to \infty$. The explicit condition (C6) does not therefore seem to be overly strong and allows us the extra convenience of working with improper priors.

(v) Condition (C7) sets the prior $\pi(\theta)$ to be uniform at the constant level $(2\pi)^{-1/2}$. This corresponds to the specification used earlier in Sections 2 and 3. We may in the present case replace this condition by

(C7′) *The prior density $\pi(\theta)$ is proper and continuous at $\theta = \theta^0$ with $\pi(\theta_0) > 0$*

(c.f. Walker, 1969, p. 81 and Hartigan, 1983, p. 109). In view of result (C6′) we could now eliminate (C6), replace (C5) by (C5′), (C7) by (C7′) and we would obtain

$$(75') \qquad \frac{d\mathcal{P}_t}{dP_t^0} / \frac{dQ_t}{dP_t^0} \to (2\pi)^{1/2}\pi(\theta^0) \quad \text{a.s. } (P^0)$$

in place of the stated result (75). In this case the exponential Bayes measure defined in (76) is asymptotically proportional (up to the constant scalar $(2\pi)^{1/2}\pi(\theta^0)$) to the Bayesian data measure $\mathcal{P}_t$. □

REMARKS ON THEOREM 4.3

(i) We can write result (74) in the form $d\mathcal{P}_t/dQ_t \to 1$ a.s. (P^0), which tells us that the exponential Bayes measure Q_t is identical to the Bayes data measure $\mathcal{P}_t$ as $t \to \infty$. In its present form (74) tells us that the data density $d\mathcal{P}_t/dP_t^0$ is asymptotically equivalent to the exponential local martingale given in (75). This local martingale defines the exponential Bayes measure Q_t.

(ii) Using (75) we have

$$2\ln(dQ_t/dP_t^0) = V_t^2 A_t^{-1} - \ln(A_t)$$

the first term of which is a quadratic form in the score $V_t = \ell_t^{(1)}(\theta^0)$ that corresponds to the classical score statistic. The statistic dQ_t/dP_t^0, which as we have seen in Section 3.5 is our posterior odds test statistic for testing $\theta = \theta^0$, can therefore be interpreted as a form of penalized score test in which the size of the penalty (for estimating θ) is determined by the quadratic variation A_t. Note that (75) can, in fact, be computed under the null hypothesis $\theta = \theta^0$, just as the classical score or LM test, simply by using the value of θ under the null.

(iii) Formulae (76) and (77) give alternative (asymptotically equivalent) expressions for dQ_t/dP_t^0. Of particular interest is (77) because we can rewrite this in the form

$$\ln(dQ_t/dP_t^0) = \ell_t(\hat{\theta}_t) - (1/2)\ln A_t ,$$

which is related to the BIC criterion of Schwarz (1978). Here $\ell_t(\hat{\theta}_t)$ is the maximized value of the likelihood (ratio) functional (compare the error sum of squares minimand that appears in the usual form of the BIC criterion) and $\ln(A_t)$ is the penalty for including θ as a free parameter (i.e. free rather than set equal to the fixed value $\theta = \theta^0$, as in the competing model for which P_t^0 is the probability measure).

(iv) Theorem 4.3 remains true under random time changes. Thus, if $(\tau_a)_{a>0}$ is a family of monotone increasing and continuous stopping times for which $\tau_a \to \infty$ as $a \to \infty$, then in place of (74) we have for the time changed measures $\mathcal{P}_{\tau_a}$ and Q_{τ_a} a similar convergence as $a \to \infty$, i.e.

$$\frac{d\mathcal{P}_{\tau_a}}{dQ_{\tau_a}} = \frac{d\mathcal{P}_{\tau_a}}{dP^0_{\tau_a}} / \frac{dQ_{\tau_a}}{dP^0_{\tau_a}} \to 0 \text{ a.s. } (P^0), \text{ as } a \to \infty .$$

The time change could be selected, as in (66), to ensure that the quadratic variation $A(\tau_a)$ is a.s. (P^0) bounded, and then $V_a = V_{t\wedge\tau_a}$ is a continuous L_2 martingale. If we also condition on a minimal information time such as $\tau_0 = \tau$ in (65), then the exponential Bayes measure given in (75) is a proper probability measure in the new time frame and with the new initialization, just as in Theorem 4.1.

(v) Define the posterior density process for θ by the ratio (using Bayes rule)

$$\text{(78)} \qquad \begin{aligned} \Pi^B_t(\theta) &= \pi(\theta)(dP^\theta_t/dP^0_t)/\textstyle\int_{\mathbf{R}}\pi(\theta)(dP^\theta_t/dP^0_t)d\theta \\ &= \pi(\theta)(dP^\theta_t/dP^0_t)/(d\mathcal{P}_t/dP^0_t) \\ &= \pi(\theta)dP^\theta_t/d\mathcal{P}_t . \end{aligned}$$

Applying Theorem 4.3 we see that the Bayes data measure $\mathcal{P}_t$ in this expression for $\Pi^B_t(\theta)$ can be replaced by the exponential Bayes measure Q_t with a relative error that tends to zero as $t \to \infty$. Thus, the asymptotic form of the posterior density process is simply

$$\text{(79)} \qquad \Pi^B_t(\theta) \sim \pi(\theta)(dP^\theta_t/dQ_t) , \quad \text{as} \quad t \to \infty ,$$

which is obtained by multiplying the prior by the likelihood (ratio) function taken with respect to Q_t as the reference measure. As the following Corollary to Theorem 4.3 shows, the density $\Pi^B_t(\theta)$ is, in fact, asymptotically Gaussian in form with a $N(\hat\theta_t, A_t^{-1})$ density. The equivalence in (79) shows that this result on the asymptotic Gaussianity of $\Pi^B_t(\theta)$ should be interpreted in the light of the reference measure Q_t with respect to which the likelihood (viz. dP^θ_t/dQ_t) is implicitly being computed. As in the case of the linear models discussed in Sections 2 and 3 this change of reference measure from P^0_t to Q_t alters the frame of reference with respect to the interpretation of the posterior density for θ.

COROLLARY 4.4. *Suppose the conditions of Theorem* 4.3 *hold. Given* $M > 0$, *let* $N^M_t = \{\theta : (\theta - \hat\theta_t)^2 A_t \le M\}$ *and define*

$$\text{(80)} \qquad \varphi(\theta;\, \hat\theta_t,\, A_t^{-1}) = (2\pi)^{-1/2}A_t^{1/2}\exp\{-(1/2)(\theta - \hat\theta_t)^2 A_t\} .$$

The posterior density process (78), *i.e.* $\Pi^B_t = \pi(\theta)dP^\theta_t/d\mathcal{P}_t$, *is asymptotically Gaussian* $N(\hat\theta_t, \hat A_t^{-1})$

with density (80) *in the sense that*

(81) $$\sup_{\theta \in N_t^M} \left| \frac{\Pi_t^B(\theta)}{\varphi(\theta;\ \hat{\theta}_t,\ A_t^{-1})} - 1 \right| \to 0 \quad \text{a.s. } (P^0)$$

as $t \to \infty$. □

(vi) The density ratio $\Re(t) = dQ_t/dP_t^0$ which defines the measure Q_t in (75) can be written as in (70) in the form $\Re(t) = \exp\{K(t)\}$ with $K(t) = V(t)^2/2A(t) - (1/2)\ln(A(t))$. Using the stopping time sequence $(\tau_a)_{a\geq 0}$ and initialization $\tau_0 = \tau$ given in (65) and (66), we can define the time changed density process

$$\Re_a = \Re(\tau_a)/\Re(\tau_0) - \exp\{K(\tau_a) - K(\tau_0)\}$$
$$= \exp\{G_a - (1/2)[G]_a\} ,$$

where $G_a = G(\tau_a) = \int_{\tau_a}^{\tau_0}(V(t)/A(t))dV(t)$ just as in (72). Now let $\hat{h}(t) = V(t)/A(t)$, which is, in effect, a linearized MLE (i.e., $\hat{\theta}_t - \theta^0 = V_t/A_t + o(\hat{\theta}_t - \theta^0)$ as $t \to \infty$ -- see equation A(24) in the appendix). Then, $G_a = \int_{\tau_a}^{\tau_0}\hat{h}(t)dV(t)$. The process $\Re(t)$ and stopped process $\Re_a$ are analogous to the density ratio R_t of the linear model studied in Section 2. As in Lemma 2.3, we find that $\Re_a$ satisfies the integral equation

$$\Re_a = 1 + \int_{\tau_0}^{\tau_a}\Re(s)dG(s) ,$$

and

(82) $$d\Re_a = \Re(\tau_a)dG(\tau_a) = \Re(\tau_a)\hat{h}(\tau_a)dV(\tau_a) .$$

This is a nonlinear stochastic differential equation for $\Re_a = \Re(\tau_a)$, showing how the density process $\Re(\tau_a)$ is updated using the latest available (in $\mathcal{F}_{\tau_a}$) value of the linearized MLE $\hat{h}(\tau_a)$ and the increment in the score process $dV(\tau_a)$ at time τ_a. The model to which the exponential Bayes measure Q_t relates is therefore determined by the nonlinear stochastic differential equation (82), which prescribes the evolution of the path dependent density $\Re_a = \Re(\tau_a)$ from a given initialization at $a = 0$ (i.e. τ_0) in terms of the linearized MLE $\hat{h}(\tau_a)$ and the increment in the score $dV(\tau_a)$, both of which are continuously updated as a increases.

(vii) Equation (82) is the updating equation for the likelihood ratio density process $\mathfrak{R}_a$. In the linear model (1) we were able to obtain an explicit Bayes model for the data, viz. $(1)^B$, corresponding to the path dependent Bayes measure Q_t. Interestingly, with a further time change in the process it is possible to construct a similar Bayes model for the data in the general case. To do this we use the following lemma

LEMMA 4.5. *Suppose V_t is a continuous local martingale with $V_0 = 0$ and quadratic variation process A_t for which $A_t \to \infty$ a.s. (P^0). Then there exists a Brownian motion X_t and a family of stopping times $(\sigma_t)_{t\geq 0}$ with $\sigma_t \to \infty$ as $t \to \infty$ such that V_t is indistinguishable from $\int_0^{\sigma_t} X dX$.* □

The time change σ_t in the theorem is constructed using the rule

$$\sigma_t = \inf\left\{p : \int_0^p X_s^2 ds \geq A_t\right\} . \tag{83}$$

Then

$$\hat{h}_t = V_t / A_t = \int_0^{\sigma_t} X dX \Big/ \int_0^{\sigma_t} X_s^2 ds , \tag{84}$$

$$G_a = G(\tau_a) = \int_{\tau_0}^{\tau_a} \hat{h}_t dV_t = \int_{\tau_0}^{\tau_a} \hat{h}_t X_{\sigma_t} dX_{\sigma_t} ,$$

and

$$[G]_a = \int_{\tau_0}^{\tau_a} \hat{h}_t^2 dA_t = \int_{\tau_0}^{\tau_a} \hat{h}_t^2 X_{\sigma_t}^2 dt .$$

The time changed density process $\mathfrak{R}_a$ now has the form of the exponential martingale

$$\mathfrak{R}_a = \exp\{G_a - (1/2)[G]_a\} = \exp\left\{ \int_{\tau_0}^{\tau_a} \hat{h}_t X_{\sigma_t} dX_{\sigma_t} - (1/2) \int_{\tau_0}^{\tau_a} \hat{h}_t^2 X_{\sigma_t}^2 dt \right\} . \tag{85}$$

As in the proof of Theorem 2.4 we observe that (84) is the likelihood ratio density process for the model

$$dX_{\sigma_t} = \hat{h}_t X_{\sigma_t} dt + dW_t , \quad t \geq \tau_0 , \tag{86}$$

where W_t is a Brownian motion and $\hat{h}_t$ is given in (85). The nonlinear stochastic differential equation (86) is an explicit form of the Bayes model to which the path dependent exponential Bayes measure Q_t relates. The model (86), like $(1)^B$, is trajectory dependent and here relies on the path of the Brownian motion X, which is time changed so that the martingale $\int_0^{a_t} XdX$ is indistinguishable from the score process $V_t = \ell_t^{(1)}(\theta^0)$. We can think of (86) as being the model for the score process V_t under the Bayes measure Q_t.

(viii) One other way of thinking about Theorem 4.3 is that the measure Q_t as determined by (75) is the "Bayes measure" of the continuous time linear model whose probability measure $\bar{P}_t^h$, say, for $h = \theta - \theta^0$ belongs to the exponential family of densities (with respect to P_t^0) given by

$$d\bar{P}_t^h / dP_t^0 = \exp\{hV_t - (h^2/2)A_t\} \ .$$

When $V_t = \int_0^t Y_s dY_s$ and $A_t = \int_0^t Y_s^2 ds$, this corresponds with the density (2) and the continuous time model is then the diffusion equation (1).

4.3. The Discrete Time Case

Let $\{Y_t\}_1^n$ be a discrete time series defined on the filtered sequence of measurable spaces $(\Omega, \mathcal{F}_t\}$. Let P_n^θ be a parameterized probability measure of $\{Y_t\}_1^n$ with $\theta \in \mathbf{R}$. Suppose θ^0 is the true value of θ and that $P_n^\theta \ll \nu_n$, some σ-finite measure on $(\Omega, \mathcal{F}_n)$. We write the RN derivative of P_n^θ with respect to $P_n^0 = P_n^{\theta^0}$ as

$$L_n(\theta) = dP_n^\theta / dP_n^0 = (dP_n^\theta / d\nu_n)/(dP_n^0 / d\nu_n) \ . \tag{87}$$

If $\pi(\theta)$ is a prior density on θ then the Bayesian data measure is given by the mixture $\mathcal{P}_n = \int_{\mathbf{R}} P_n^\theta d\theta$, as in the continuous time case. Suppose θ^0 is the "true value" of θ and let $P_n^0 = P_n^{\theta^0}$. The following result is the analogue of Theorem 4.3, showing that $\mathcal{P}_n$ can be asymptotically approximated by an exponential Bayes measure in discrete time.

THEOREM 4.6. *Assume the following conditions hold:*

(D1) $\ell_n(\theta) = \ln(L_n(\theta))$ *is twice continuously differentiable with derivatives* $\ell_n^{(1)}(\theta)$ *and* $\ell_n^{(2)}(\theta)$.

(D2) *Under* P_n^θ, $\ell_n^{(1)}(\theta)$ *is a zero mean* L_2 *martingale with conditional quadratic variation process* $B_n(\theta)$ *and* $B_n(\theta) \to \infty$ a.s. (P^θ) *as* $n \to \infty$.

(D3) $(\ell_n^{(2)}(\theta) + B_n(\theta))/B_n \to 0$ a.s. (P^θ) *as* $n \to \infty$.

(D4) *There exist continuous functions* $w_n(\theta, \theta')$ *such that* $w_n(\theta, \theta) = 0$ *and such that for all* $\theta, \theta' \in N_\delta(\theta^0) = \{\theta : |\theta - \theta^0| < \delta\}$ *of* θ^0 *we have*

$$\{\ell_n^{(2)}(\theta) - \ell_n^{(2)}(\theta')\}/B_n(\theta) \le w_n(\theta, \theta') \text{ a.s. } (P^0)$$

for each n and $w_n(\theta, \theta') \to w_\infty(\theta, \theta')$ a.s. (P^0) *uniformly for* $\theta, \theta' \in N_\delta(\theta^0)$.

(D5) *The maximum likelihood estimate* $\hat{\theta}_n \to \theta^0$ a.s. (P^0).

(D6) *For any* $\delta > 0$ *and* $\omega_\delta = \{\theta : |\theta - \theta^0| \ge \delta\}$ *we have*

$$B_n^{1/2} \int_{\omega_\delta} (dP_n^\theta / dP_n^0) d\theta \to 0 \quad \text{a.s. } (P^0)$$

where $B_n = B_n(\theta^0)$.

(D7) *The prior* $\pi(\theta) = (2\pi)^{-1/2}$ *is uniform on* $\mathbf{R}$.

Then

$$(88) \qquad \frac{d\mathcal{P}_n}{dP_n^0} / \frac{dQ_n}{dP_n^0} \to 1 \text{ a.s. } (P^0)$$

where Q_n *is the exponential Bayes measure defined by the following* RN *derivative with respect to* P_n^0

$$(89) \qquad \frac{dQ_n}{dP_n^0} = \frac{\exp\{(1/2)V_n^2 B_n^{-1}\}}{B_n^{1/2}},$$

where $V_n = \ell_n^{(1)}(\theta^0)$. *The derivative* (89) *has the following asymptotically equivalent forms*

$$(90) \qquad dQ_n/dP_n^0 = \exp\left\{(1/2)\left(\hat{\theta}_n - \theta^0\right)^2 B_n\right\}/B_n^{1/2}$$

and

$$(91) \qquad dQ_n/dP_n^0 = \exp\{\ell_n(\hat{\theta}_n)\}/B_n^{1/2}. \quad \square$$

REMARKS

(i) Conditions (D1)-(D7) mirror (C1)-(C7) used in the continuous time case. The only difference is that we now use the conditional quadratic variance process $B_n(\theta) = \langle \ell_n^{(1)}(\theta) \rangle$ in place of $A_t(\theta) = [\ell_n^{(1)}(\theta)]$. Writing the log likelihood as

(92) $\ell_n^{(1)}(\theta) = \Sigma_{k=1}^n (\partial/\partial\theta)[\ln(L_k(\theta)) - \ln(L_{k-1}(\theta))] = \Sigma_{k=1}^n \varepsilon_k(\theta)$, say.

Then

$$B_n(t) = \Sigma_{k=1}^n E(\varepsilon_k(\theta)^2 | \mathcal{F}_{k-1}) = < \ell_n^{(1)}(\theta) > ,$$

which is the conditional variance of the martingale $\ell_n^{(1)}(\theta)$ under P_n^θ (c.f. Hall-Heyde, 1980, p. 157).

(ii) Writing the discrete time posterior density as

$$\Pi_n^B(\theta) = \pi(\theta)(dP_n^\theta/dP_n^0)/\int_{\mathbf{R}}\pi(\theta)(dP_n^\theta/dP_n^0)d\theta = \pi(\theta)dP_n^\theta/d\mathcal{P}_n$$

we deduce in the same way as Corollary 4.4 that $\Pi_n^B(\theta)$ is asymptotically Gaussian $N(\hat{\theta}_n, B_n^{-1})$. Moreover, in view of the equivalence (88) we have

(94) $\Pi_n^B \sim \pi(\theta)(dP_n^\theta/dQ_n)$, as $n \to \infty$,

so that the asymptotic Gaussian posterior $N(\hat{\theta}_n, B_n^{-1})$ is to be interpreted with respect to the path dependent Bayes measure Q_n. □

It is rather more difficult than in the continuous time case to determine the form of the implied Bayes model from the form of the discrete time local exponential martingale (89). We can however use our theory for the continuous time case to analyze the discrete time case by an embedding technique. We will show that we can embed the local martingale (89) into a corresponding continuous time process whose Bayes model we have already studied in Section 4.2. The discrete time Bayes model can then be regarded as simply the model of the discrete observations from the continuous process. An advantage of this embedding is that we can analyze the model without making a special cut in the asymptotic theory for nonstationary time series (i.e. in the case of a unit root). This is because in the continuous time case there is no difference in treatment between the stationary and nonstationary cases.

To begin, we continue to assume conditions (D1)-(D7) hold and then the asymptotic approximation (88) applies. Our objective is to find an alternative representation of (89) in terms of a continuous process. It will be convenient for us to write the increments in the score process given in (92) at $\theta = \theta^0$ as $\varepsilon_k = \varepsilon_k(\theta^0)$. Then we have

(95) $$V_n = \ell_n^{(1)}(\theta^0) = \Sigma_{k=1}^n \varepsilon_k ,$$

which is a P_n^0-martingale with conditional variance process B_n. Let $\mathcal{F}_n$ be the σ-field generated by $(\varepsilon_j)_1^k$.

THEOREM 4.7. *Assume* (D1), (D2) *and the following conditions hold:*

(D8) $$\sup_{k\geq 1} E(\varepsilon_k^4) < \infty .$$

(D9) $$\sup_{k\geq 1} \frac{E(\varepsilon_k^4 | \mathcal{F}_{k-1})}{\{E(\varepsilon_k^2 | \mathcal{F}_{k-1})\}^2} \leq C_a \text{ a.s. } (P^0)$$

for some constant $C_a > 0$.

(D10) *There exists some* γ *with* $0 < \gamma < 1$ *such that*

$$\frac{E(\varepsilon_n^2 | \mathcal{F}_{n-1})}{B_n^\gamma} \to 0 \text{ a.s. } (P^0) .$$

Then there exists a probability space (Ω, G, P) *supporting* $(V_n, B_n)_{n\geq 1}$, *a standard Brownian motion* W, *and stopping times* $(\tau_n)_{n\geq 1}$ *such that*

(96) $$\frac{\exp\{(1/2)V_n^2 B_n^{-1}\}}{B_n^{1/2}} / \frac{\exp\{(1/2)W(\tau_n)^2/\tau_n\}}{\tau_n^{1/2}} \to 1 \text{ a.s. } (P^0) . \quad \square$$

REMARKS ON (D8)-(D10)

(i) Condition (D8) requires that fourth moments of the martingale differences ε_k in (95) exist and are uniformly (in k) bounded above. It could be relaxed to a weaker $(2+r)$-moment requirement on ε_k for some r with $0 < r < 2$, at the expense of making the proof (and some of the other conditions) of Theorem 4.7 a little more complicated.

(ii) Condition (D9) imposes a bound on the relative conditional fourth moments of ε_k. (D9) requires that the ratio of the conditional fourth moment to the square of the conditional second moment of ε_k be uniformly bounded above. This means that the kurtosis of the conditional distribution of ε_k cannot be too large relative to the square of the variance. For a stochastic linear regression model $y_t = \theta' x_t + u_t$ with $u_t \equiv$ iid $N(0, 1)$ and $\mathcal{F}_{t-1}$-measurable regressors, the score process increments are $\varepsilon_k = x_k u_k$ and then

$$\sup_{k\geq 1} \frac{E(\varepsilon_k^4|\mathcal{F}_{k-1})}{\{E(\varepsilon_k^2|\mathcal{F}_{k-1})\}^2} = \sup_{k\geq 1} \frac{2\sigma^4 x_k^4}{\sigma^4 x_k^4} = 2 \ .$$

In this case the condition (D9) is fulfilled regardless of the structure of the regressor x_t.

(iii) The conditional variance process B_n is often interpreted as the time clock of the martingale V_n in the sense that it records the information content of the process up to time period n. The increment in the information content from period $n-1$ to period n is

$$d_n = B_n - B_{n-1} = E(\varepsilon_n^2|\mathcal{F}_{n-1}) \ .$$

Condition (D10) requires that the incremental information d_n be small (by an order of magnitude or power of B_n) relative to the total information content B_n. We can explore the implications of this requirement in the linear AR(1) model (22). In this case we have $\varepsilon_k = y_{k-1}u_k$ and $E(\varepsilon_k^2|\mathcal{F}_{k-1}) = y_{k-1}^2\sigma^2$. (D10) requires that

$$\text{(97)} \qquad \frac{y_{n-1}^2\sigma^2}{\left(\Sigma_1^n y_{k-1}^2\sigma^2\right)^\gamma} \to 0 \text{ a.s. } (P) \ ,$$

for some γ in the interval $0 < \gamma < 1$. Take the stationary case first. Here $|\alpha| < 1$ and we have $n^{-1}\Sigma_1^n y_{k-1}^2 = O_{\text{a.s.}}(1)$ and (97) holds if

$$y_{n-1}^2/n^\gamma \to 0 \text{ a.s. } (P) \ ,$$

which holds by the Borel Cantelli lemma if $\sup_n E(y_n^4) < \infty$ and $\gamma > 1/2$. In the unit root case where $\alpha = 1$ we rescale the numerator and denominator of (97) as follows:

$$\text{(98)} \qquad \frac{n\ \ln(\ln(n))\{y_{n-1}^2\sigma^2/n\ \ln(\ln(n))\}}{n^{2\gamma}/(\ln(\ln(n)))^\gamma\left\{\Sigma_1^n y_{k-1}^2\sigma^2/[n^2/\ln(\ln(n))]\right\}^\gamma} \ .$$

By the law of the iterated logarithm we have

$$\limsup_{n\to\infty} \frac{y_{n-1}^2\sigma^2}{n\ \ln(\ln(n))} = 2\sigma^2 \text{ a.s. } (P) \ ,$$

and by a result of Donsker-Varadhan (1977, p. 751) that is used in Lai-Wei (1982, p. 364) we have

$$\liminf_{n\to\infty} \frac{\Sigma_1^n y_{n-1}^2 \sigma^2}{n^2/\ln(\ln(n))} = \sigma^4/4 \text{ a.s. } (P) ,$$

so that (98) is of order $O(n^{-2\gamma+1}(\ln(\ln(n))^{1+\gamma})$ as $\to 0$ a.s. (P) provided $\gamma > 1/2$. Hence, (97) and thus (D10) hold in the stationary and nonstationary AR(1) model for $\gamma > 1/2$. □

REMARKS ON THEOREM 4.7

(i) In the proof of Theorem 4.7 we use the fact that the discrete time martingale V_n can be embedded in a Brownian motion so that, by changing the probability space if necessary, we can write $V_n = W(\tau_n)$ a.s. (P) for some stopping time τ_n. This is simply an application of the conventional Skorokhod embedding of a martingale, as discussed in detail by Hall-Heyde (1980, Appendix 1). What Theorem 4.7 shows in addition is that it is also possible to approximate the conditional variance process B_n by τ_n asymptotically. This means that the discrete exponential process

(99) $$M_n = B^{-1/2}\exp\{(1/2)V_n^2/B_n\}$$

can be embedded asymptotically in the local exponential martingale

(100) $$Z_{\tau_n} = \tau_n^{-1/2}\exp\{(1/2)W(\tau_n)^2/\tau_n\} .$$

(ii) What is the model corresponding to the local exponential martingale Z_{τ_n}? As discussed in the argument following equation (8) where we first encountered this exponential process, Z_{τ_n} is a time changed version of the process Z_t given in (7) and this process defines the Bayes measure for the model $(1)^B$. thus, under the conditions of the theorem, the original discrete time process can be embedded into a continuous process whose score function evolves according to a nonlinear stochastic differential equation of the form $(1)^B$.

(iii) One consequence of the proof of Theorem 4.7 is that (on a new probability space if necessary) $V_n = W(\tau_n)$ and

$$\frac{V_n}{B_n} / \frac{W(\tau_n)}{\tau_n} \to 1 \text{ a.s. } (P) ,$$

as $n \to \infty$. Using Lemma 4.5 there exists a Brownian motion X and stopping times ρ_n such that

$$V_n = W(\tau_n) = \int_0^{\rho_n} XdX$$

and

$$W(\tau_n)/\tau_n = \int_0^{\rho_n} XdX \,/\, \int_0^{\rho_n} X_s^2 ds \ .$$

With this equivalence we can interpret the martingale V_n as a sequence of discrete observations taken from the stochastic differential equation (86) at the discrete times $\sigma_t = \rho_n$, $n = 1, 2, \ldots$. The Bayes model corresponding to the discrete exponential Bayes measure Q_n defined by (89) can therefore be regarded as a model for discrete data embedded in the continuous system (86). □

5. CONCLUSION

This paper puts forward the idea that Bayesian modeling of time series involves a special frame of reference, one that seems very different from classical modeling. In classical models, the starting point is a model or likelihood in which a hypothetical true value of the parameter is postulated. By contrast, the conventional Bayes treatment of the same problem involves the replacement of the classical model with one where the parameter is updated each period according to the latest observation. Conceptually, the Bayesian frame of reference, which eschews the notion of a true parameter value, is a time varying parameter model in which the parameter value is determined by the penultimate value of the MLE, i.e. by recursive maximum likelihood. We call the new model the Bayes model and its associated probability measure the Bayes model measure.

The new frame of reference in Bayesian modeling arises incidentally in the passage from prior to posterior density and results from the data conditioning that is explicit in the likelihood principle. One consequence that has important practical consequences is that the Bayes model inevitably inherits the statistical properties of the recursive MLE on which it is based. In time series models, this includes the bias and skewness of the finite sample distribution of the MLE. Whereas classical methods compensate for these properties by taking the sampling properties of the estimator into account, conventional Bayes methods that are based on the use of the posterior distribution do not.

This helps to explain the poor sampling properties of such Bayes methods in autoregressions that were reported in the simulation exercises in Phillips (1991).

In spite of the above mentioned difficulties, we have shown that it is possible to mount meaningful Bayes model tests by taking into account the correct Bayes model measure that underlies conventional Bayesian inference that relies on the posterior distribution. When applied to autoregressions, it turns out that this principle can be used to derive classical Dickey-Fuller and augmented Dickey-Fuller tests. Alternative Bayes model likelihood ratio tests and posterior odds criteria are also suggested. In the case of posterior odds we find that correct use of the Bayes model measure in computation of the Bayes factor leads to a scaling that is equivalent to the use of a conditional type of Jeffreys prior and leads to a criterion that is, in fact, a generalization of the Schwarz (1978) BIC criterion. This new criterion can be used for model selection and for hypothesis testing and we have shown in the paper how to apply it in the context of unit root tests in autoregressions with trends.

This paper is a beginning. Our main concerns have been: (i) the conceptual framework that the Bayes frame of reference implies; (ii) the practical import of the new frame of reference in modeling and in statistical tests; and (iii) the development of an asymptotic theory that justifies our procedures in non-Gaussian models. Later work will extend our treatment to more general multivariate models, explore the properties of our posterior odds model selection criterion and illustrate the use of our methods in empirical work.

6. APPENDIX

PROOF OF LEMMA 2.1. In view of the equivalence (10) under the time change (9) we need only prove the result for the process Z_t'. We can proceed to compute the conditional expectation $E(Z_t'|\mathcal{F}_s)$, $s < t$, directly as follows.

Let $X = W(t) - W(s) \equiv N(0, t-s)$ and write $Z_t' = t^{-1/2}\exp\{(1/2t)(W(s) + X)^2\}$. Then

$$\begin{aligned}
E(Z_t'|\mathcal{F}_s) &= t^{-1/2}\exp\{(1/2t)W(s)^2\}[2\pi(t-s)]^{-1/2}\int_{\mathbf{R}}\exp\{(1/2)W(s)X-(1/2t)X^2-(1/2(t-s))X^2\}dX \\
&= t^{-1/2}\exp\{(1/2t)W(s)^2\}\exp\{((t-s)/2ts)W(s)^2\} \\
&\quad \cdot[2\pi(t-s)]^{-1/2}\int_{\mathbf{R}}\exp\left\{-\frac{s}{2t(t-s)}\left[X^2 - 2\frac{t-s}{s}XW(s) + \left[\frac{t-s}{s}\right]^2 W(s)^2\right]dX\right\} \\
&= t^{-1/2}\exp\{(1/2t)W(s)^2\}\exp\{((t-s)/2ts)W(s)^2\} \\
&\quad .[2\pi(t-s)]^{-1/2}\int_{\mathbf{R}}\left\{-\frac{s}{2t(t-s)}\left[X - \frac{t-s}{s}W(s)\right]^2 dX\right\} \\
&= t^{-1/2}\exp\{(1/2s)W(s)^2\}(t/s)^{1/2} \\
&= s^{-1/2}\exp\{(1/2s)W(s)^2\} \\
&= Z_s' \,, \qquad \text{(A1)}
\end{aligned}$$

as required. Note, however, that Z_t' is not integrable even though the conditional expectation exists. To show the local L_2 property of Z_t' we introduce the stopping time sequence

$$\sigma_n = \inf\{t > \tau : W(t)^2 \geq nt\} \,, \quad n \in \mathbf{N}$$

and note that $\sigma_n \to \infty$ a.s. as $n \to \infty$ by the law of the iterated logarithm for Brownian motion. Define the new process $Z_t^n = Z'_{t\wedge\sigma_n}$ and note that Z_t^n is bounded and therefore square integrable for all $t > \tau$. Now, using the fact that Z_t' satisfies the martingale property, we have for $\tau \leq s \leq t$

$$\begin{aligned}
E(Z_t^n|\mathcal{F}_s) &= E(Z'_{t\wedge\sigma_n}|\mathcal{F}_s) = E(E(Z'_{\sigma_n}|\mathcal{F}_t)|\mathcal{F}_s) \\
&= E(Z'_{\sigma_n}|\mathcal{F}_s) = Z'_{\sigma_n\wedge s} = Z_s^n \,.
\end{aligned}$$

Thus, Z_t^n satisfies the martingale property and is in L_2 for $t > \tau$. The process Z_t^n is initialized at $t = \tau$ where the stopping time $\tau > 0$ a.s. (P), because $Z_0' = \infty$ a.s. (P) and thus the process is not properly defined with an initialization at $t = 0$. The process $Z_t^n = Z'_{t \wedge \sigma_n}$ is a uniformly integrable martingale for each n and, consequently, Z_t' and Z_t are continuous local L_2 martingales, giving the required result.

PROOF OF LEMMA 2.2. Taking stochastic differentials and employing Ito's rule, we have

$$d[(1/2)\hat{\beta}_t^2 A_t - (1/2)\ln(A_t)] = \hat{\beta}_t d\hat{\beta}_t A_t + (1/2)d[\hat{\beta}, \hat{\beta}]_t A_t + (1/2)\hat{\beta}_t^2 dA_t - (1/2)A_t^{-1}dA_t \;. \tag{A2}$$

Recall that $\hat{\beta}_t = A_t^{-1}V_t$ so that

$$d[\hat{\beta}, \hat{\beta}]_t = -A_t^{-2}dA_t V_t + A_t^{-1}dV_t \;. \tag{A3}$$

and

$$d[\hat{\beta}, \hat{\beta}]_t = A_t^{-2}dA_t \;. \tag{A4}$$

Using (A3)-(A4) in (A2), we have

$$\hat{\beta}_t dV_t - \hat{\beta}_t^2 dA_t + (1/2)A_t^{-1}dA_t + (1/2)\hat{\beta}_t^2 dA_t - (1/2)A_t^{-1}dA_t = \hat{\beta}_t dV_t - (1/2)\hat{\beta}_t^2 dA_t \;,$$

giving the required result (14).

PROOF OF LEMMA 2.3. From (13) and (14), we deduce that

$$\begin{aligned} R_t &= \exp\left\{ \textstyle\int_{t_0}^t [\hat{\beta}_s dV_s - (1/2)\hat{\beta}_s^2 dA_s] \right\} \\ &= \exp\left\{ \textstyle\int_{t_0}^t \hat{\beta}_s dV_s - (1/2) \int_{t_0}^t \hat{\beta}_s^2 dA_s \right\} \\ &= \exp\left\{ G_t - (1/2) \textstyle\int_{t_0}^t d[G, G]_s \right\} \\ &= \exp\{G_t - (1/2)[G, G]_t\} \;, \qquad \text{(A5)} \end{aligned}$$

as required for (15). By stochastic differentiation

$$dR_t = R_t\{dG_t - (1/2)d[G, G]_t\} + \frac{1}{2}R_t d[G, G]_t = R_t dG_t \;,$$

and integrating we have

$$\int_{t_0}^t dR_s = \int_{t_0}^t R_s dG_s$$

or

$$R_t = 1 + \int_{t_0}^{t} R_s dG_s ,$$

since $R_{t_0} = 1$. This proves (16). Finally, note that since V_t is a P_t-martingale (recall that when $\beta = 0$, $V_t = \int_0^t W_s dW_s$) and $dG_t = \hat{\beta}_t dV_t$ we have

$$E(dG_t|\mathscr{F}_t) = \hat{\beta}_t E(dV_t|\mathscr{F}_t) = 0 ,$$

so that G_t is also a P_t-martingale. It follows from (16) and Theorem 5.3 of Ikeda and Watanabe (1981, p. 142) that $E(R_t) = 1$, as required.

PROOF OF THEOREM 2.4

(a) Under a uniform prior for β the posterior density process is

$$\Pi_t^B \propto a_t^{1/2}\exp\{-(1/2)(\beta - \hat{\beta}_t)^2 A_t\} \propto N(\hat{\beta}_t, A_t^{-1}) .$$

When the prior $\pi(\beta) = (2\pi)^{-1/2}$, we have $\Pi_t^B = N(\hat{\beta}_t, A_t^{-1})$ exactly as in (12). When the initialization is at $t_0 > 0$ and $\pi(\beta) = \Pi_{t_0}^B$ we get the posterior $\Pi_{t,t_0}^B = N(\hat{\beta}_t, A_t^{-1})$ again as shown in the analysis leading up to equation (21) of the text.

(b) As explained in the argument that follows equations (11) and (12) of the text the posterior density is

$$\Pi_t^B = \pi(\beta)(dP_t^\beta/dQ_t) .$$

Under Bayes methods of inference, working with Π_t^B is equivalent to working with

$$\Pi_t = \pi(\beta)(dP_t^\beta/dP_t)$$

since the factor dQ_t^B/dP_t by which they differ is only dependent on the data (i.e. it does not depend on β) and, hence, becomes constant upon data conditioning. The absorption of this factor in the constant of proportionality ensures that the effective likelihood function for Bayes inference is the RN derivative dP_t^β/dQ_t. Thus, the new frame of reference for interpreting the posterior Π_t^B is the reference measure Q_t not P_t.

(c) From (13) and (14) we have the density process

$$R_t = \exp\left\{ \int_{t_0}^{t} [\hat{\beta}_s dV_s - (1/2)\hat{\beta}_s^2 dA_s] \right\}$$

$$= \exp\left\{ \int_{t_0}^{t} [\hat{\beta}_s Y_s dY_s - (1/2)\hat{\beta}_s^2 Y_s^2 ds] \right\}$$

which is the likelihood ratio density process for the model

$$dY_t = \hat{\beta}_t Y_t dt + dW_t \,, \quad t \geq t_0 \,.$$

This holds for all $t \geq t_0$ and any $t_0 > 0$. It therefore holds for $t > 0$. Thus, R_t is the likelihood ratio of the output process of the nonlinear stochastic differential equation $(1)^B$.

PROOF OF LEMMA 3.1. Note that $A_n = A_{n-1} + Y_{n-1}^2 = A_{n-1}(1 + Y_{n-1}^2/A_{n-1})$ and thus by recursion we have:

(A6) $$A_n = A_{n_0} \prod_{i=0}^{n-n_0-1} (1 + Y_{n_0+i}^2/A_{n_0+i}) = A_{n_0} \prod_{t=n_0+1}^{n} f_t \,.$$

Next

$$V_n^2 A_n^{-1} - V_{n-1}^2 A_{n-1}^{-1} = A_n^{-1}\{(V_{n-1} + Y_{n-1}\Delta Y_n)^2 - V_{n-1}^2(1 + Y_{n-1}^2/A_{n-1})\}$$

$$= A_n^{-1}\{2\hat{h}_{n-1}A_{n-1}Y_{n-1}\Delta Y_n + (\Delta Y_n Y_{n-1})^2 - \hat{h}_{n-1}^2 A_{n-1} Y_{n-1}^2\}$$

$$= A_n^{-1}\{-(\Delta Y_n - \hat{h}_{n-1}Y_{n-1})^2 A_{n-1} + (\Delta Y_n)^2 A_n\}$$

$$= -(\Delta Y_n - \hat{h}_{n-1}Y_{n-1})^2(A_{n-1}/A_n) + (\Delta Y_n)^2$$

$$= -(\Delta Y_n - \hat{h}_{n-1}Y_{n-1})^2/f_n + (\Delta Y_n)^2$$

and by recursion we have

(A7) $$V_n^2 A_n^{-1} - V_{n_0}^2 A_{n_0}^{-1} = -\sum_{t=n_0+1}^{n} (\Delta Y_t - \hat{h}_{t-1}Y_{t-1})^2/f_t + \sum_{t=n_0+1}^{n} (\Delta Y_t)^2 \,.$$

Combining (A6) and (A7) in (31) we get

$$R_n = (A_n/A_{n_0})^{-1/2}\exp\{(1/2)V_n^2 A_n^{-1} - (1/2)V_{n_0}^2 A_{n_0}^{-1}\}$$

$$= \prod_{t=n_0+1}^{n} \frac{(2\pi f_t)^{-1/2}\exp\{-(1/2f_t)(\Delta Y_t - \hat{h}_{t-1}Y_{t-1})^2\}}{(1/2\pi)^{1/2}\exp\{-(1/2)(\Delta Y_t)^2\}}$$

as required.

PROOF OF THEOREM 3.2. Rewrite the ratio R_n in (31) as $R_n = (dQ_n/dP_n)/(dQ_{n_0}/dP_{n_0}) = (dQ_n/dQ_{n_0})/(dP_n/dP_{n_0})$. Observe that

$$dP_n/dP_{n_0} = \prod_{t=n_0+1}^{n} (1/2\pi)^{1/2}\exp\{-(1/2)(\Delta Y_t)^2\}$$

being the conditional density of $\{Y_t\}_{n_0+1}^{n}$ given $\mathscr{F}_{n_0}$ under H_1 (i.e. when $h = 0$). The required result follows directly.

PROOF OF THEOREM 3.3. Parts (a) and (b) follow directly from the form of (30). To prove (c) note that from (32) that R_n is the likelihood ratio and Q_n is the measure for the model which conditioned on $\mathscr{F}_{n-1}$ is

$$\Delta Y_n = \hat{h}_{n-1}Y_{n-1} + u_n \,, \quad u_n|_{\mathscr{F}_{n-1}} \equiv N(0, f_n)$$

where f_n is given by (36). This model is the same as

$$Y_n = (1 + \hat{h}_{n-1})Y_{n-1} + u_n \,, \quad u_n|_{\mathscr{F}_n} \equiv N(0, f_n)$$

and the stated result follows.

PROOF OF THEOREM 3.4. Under H_1 we have

$$\begin{aligned} BLR &= \left(\Sigma_1^n Y_{t-1}u_t\right)^2/(\Sigma_1^n Y_{t-1}^2) - \ln(n^{-2}\Sigma_1^n Y_{t-1}^2) \\ &\Rightarrow \left(\int_0^1 SdS\right)^2/\int_0^1 S^2 - \ln(\int_0^1 S^2) \,. \end{aligned}$$

Since $\hat{\sigma}^2 \to \sigma^2$ a.s. under H_1, the same result applies to BLR_σ.

PROOF OF THEOREMS 3.5 AND 3.6. These proofs follow the same line and involve a routine application of functional limit theory and the L_2 projection geometry given in Park and Phillips (1988, 1989) for sample moments of residuals from regressions of integrated processes on deterministic trends.

PROOF OF THEOREM 3.7. Note that

$$dQ_n/dP_n(\hat{\sigma}^2) = \left(\Sigma_1^n Y_{t-1}^2/\hat{\sigma}^2\right)^{-1/2} \exp\{(1/2)\hat{h}_n^2 \Sigma_1^n Y_{t-1}^2/\hat{\sigma}^2\}$$

where $\hat{\sigma}^2 = n^{-1}\Sigma_1^n(Y_t - \hat{\alpha}_n Y_{t-1})^2$. When H_1 is true (i.e. $\alpha = 1$ in (22)) we have

$$\hat{h}_n^2 \Sigma_1^n Y_{t-1}^2/\hat{\sigma}^2 = [n(\hat{\alpha}_n - 1)]^2[n^{-2}\Sigma_1^n Y_{t-1}^2/\hat{\sigma}^2] = O_p(1)$$

so that

$$dQ_n/dP_n(\hat{\sigma}^2) \to_p 0 \ .$$

Hence, $P(dQ_n/dP_n(\hat{\sigma}^2) < 1) \to 0$ as $n \to \infty$ and the type I error tends to zero as $n \to \infty$.

When H_1 is false and $|\alpha| < 1$, say, then

$$\hat{h}_n^2 \Sigma_1^n Y_{t-1}^2/\hat{\sigma}^2 = [\sqrt{n}(\hat{\alpha}_n - 1)]^2[n^{-1}\Sigma_1^n Y_{t-1}^2/\hat{\sigma}^2] = O_p(n)$$

and $\Sigma_1^n Y_{t-1}^2/\hat{\sigma}^2 = O_p(n)$, so that

$$\ln[dQ_n/dP_n(\hat{\sigma}^2)] = O_p(n)$$

as $n \to \infty$. It follows that $dQ_n/dP_n(\hat{\sigma}^2)$ diverges as $n \to \infty$ and $P(dQ_n/dP_n(\hat{\sigma}^2) > 1) \to 1$ as $n \to \infty$. Thus, the power of the test tends to unity and the type II error tends to zero as $n \to \infty$. By a similar argument the same behavior obtains when $\alpha > 1$.

PROOF OF THEOREM 4.1. Let $X_a = M(\tau_a) - M(\tau_0)$ and note that X_a is a bounded continuous local martingale and hence a continuous L_2 martingale. Then

$$dQ_a/dP = \exp\{X_a - (1/2)[X_a]\}$$

is the stochastic exponential of the martingale X_a. But since X_a is bounded for all $a > 0$, we have

$$E(\exp\{(1/2)[X]_a\}) < \infty \ , \quad \text{for all} \quad a > 0 \ ,$$

and thus by Theorem 5.3, p. 152 of Ikeda-Watanabe (1989) $R_a = dQ_a/dP$ is a continuous martingale with $E[R_a] = 1$, for all $a > 0$. The measure Q_a is defined by the integral of R_a, viz

$$Q_a(B) = \int_B R_a dP \ , \quad \forall B \in \mathscr{F}_{\tau_a} \ ,$$

and $Q_b|\mathscr{F}_a = Q_a$ for all $\tau_b > \tau_a \geq 0$ as in Ikeda-Watanabe, p. 191. Since $E(R_a) = 1$, Q_a defines a probability measure on $(\Omega, \mathscr{F}_{\tau_a})$.

PROOF OF LEMMA 4.2. By Ito differentiation we have

$$dK(t) = \{V(t)dV(t) + (1/2)(dV(t))^2\}/A(t) - (1/2)V(t)^2A(t)^{-2}dA(t) - (1/2)A(t)^{-1}dA(t)$$
$$= [V(t)/A(t)]dV(t) - (1/2)[V(t)/A(t)]^2dA_t$$

since $(dV(t))^2 = dA(t)$. This gives the required result.

PROOF OF THEOREM 4.3. The proof follows the general idea given in Walker (1969) and Hartigan (1983, Sec. 11.2), but does not rely on a specific rate of convergence for the MLE $\hat{\theta}_t$, nor on asymptotic normality of $\hat{\theta}_t$, nor on any ergodic properties for the Fisher information.

As in (C6) define $\omega_\delta = \{\theta : |\theta - \theta^0| \geq \delta > 0\}$ and let $N_\delta = \mathbf{R} - \omega_\delta$. We can choose $\delta > 0$ such that N_δ corresponds to the neighborhood of θ^0 in (C4) i.e. $N_\delta(\theta^0)$. Then,

(A8) $$d\mathcal{P}_t/dP_t^0 = (\int_{N_\delta} + \int_{\omega_\delta})\pi(\theta)(dP_t^\theta/dP_t^0)d\theta = I_\delta + I_\delta^c\ , \quad \text{say}$$

Under (C7), $I_\delta^c = (2\pi)^{-1/2}\int_{\omega_\delta}(dP_t^\theta/dP_t^0)d\theta$ and

(A9) $$A_t^{1/2}I_\delta^c \to 0 \quad \text{a.s. } (P^0)$$

by (C6).

Next using (C7), we write I_δ as

$$I_\delta = (2\pi)^{-1/2}\int_{N_\delta}(dP_t^\theta/dP_t^0)d\theta = (2\pi)^{-1/2}\int_{N_\delta}\exp\{\ell_t(\theta)\}d\theta$$

and define for some large $M > 0$ the neighborhood of $\hat{\theta}_t$

(A10) $$N_t = \{\theta : (\theta - \hat{\theta}_t)^2A_t < M\}$$

with $N_t^c = \mathbf{R} - N_t$. Then

(A11) $$I_\delta = (2\pi)^{-1/2}[\int_{N_\delta\cap N_t} + \int_{N_\delta\cap N_t^c} = (2\pi)^{-1/2}[I_1 + I_2]\ , \quad \text{say.}$$

Consider I_1 first. Taking a second order Taylor expansion of $\ell_t(\theta)$ we have

(A12) $$\ell_t(\theta) = \ell_t(\hat{\theta}_t) + (1/2)\ell_t^{(2)}(\theta_m)(\theta - \hat{\theta}_t)^2$$

where θ_m lies on the line segment between $\hat{\theta}_t$ and θ. Now

(A13) $$\ell_t^{(2)}(\theta_m)(\theta-\hat{\theta}_t)^2 = -A_t(\theta-\hat{\theta}_t)^2 + \{[\ell_t^{(2)}(\theta_m) - \ell_t^{(2)}(\theta^0)]/A_t + [\ell_t^{(2)}(\theta^0)+A_t]/A_t\}(\theta-\hat{\theta}_t)^2A_t\ .$$

Under (C3)

(A14) $\quad [\ell_t^{(2)}(\theta_0) + A_t]/A_t \to 0 \quad \text{a.s. } (P^0)$

and under (C4)

(A15) $\quad |\ell_t^{(2)}(\theta_m) - \ell_t^{(2)}(\theta^0)|/A_t \le w_t(\theta_m, \theta^0) \to 0 \quad \text{a.s. } (P^0)$

uniformly for $\theta \in N_\delta$. Hence combining (A12)-(A15) we have

$$\ell_t(\theta) = \ell_t(\hat{\theta}_t) - (1/2)A_t(\theta - \hat{\theta}_t)^2[1 + \varepsilon_t(\theta)]$$

where $\varepsilon_t(\theta) \to 0$ a.s. (P^0) uniformly for $\theta \in N_\delta$. Using this expansion we have

$$I_1 = \exp\{\ell_t(\hat{\theta}_t)\} \int_{N_\delta \cap N_t} \exp\{-(1/2)A_t(\theta - \hat{\theta}_t)^2[1 + \varepsilon_t(t)]\}d\theta$$

(A16) $$= A_t^{-1/2}\exp\{\ell_t(\hat{\theta}_t)\}(2\pi)^{1/2}[1 + O(\exp(-M^2/2)) + O(\eta_t)]$$

where for $\theta \in N_\delta$ we have $|\varepsilon_t(\mathrm{t})| \le \eta_t \to 0$ a.s. (P^0). It is in fact possible to choose M in (A10) in such a way that $M \to \infty$ as $t \to \infty$. We may, for instance choose $M = M_t = (1/2)\delta^2 A_t$ and then

(A17) $\quad M_t \to \infty \quad \text{a.s. } (P^0) \text{ as } t \to \infty$.

Now consider I_2 in (A11). Using (A12) again we have

$$I_2 = \int_{N_\delta \cap N_t^c} \exp\{\ell_t(\theta)\}d\theta = \exp\{\ell_t(\hat{\theta}_t)\} \int_{N_\delta \cap N_t^c} \exp\{(1/2)\ell_t^{(2)}(\theta_m)(\theta - \hat{\theta}_t)^2\}d\theta \ .$$

Now

$$\ell_t^{(2)}(\theta_m) = -A_t\{1 - [A_t + \ell_t(\theta^0)]/A_t - [\ell_t^{(2)}(\theta_m) - \ell_t(\theta^0)]/A_t\}$$

and in view of (A14) and (A15) we find that for large enough t

$$8\ell_t^{(2)}(\theta_m) < -(1/2)A_t \quad \text{a.s. } (P^0)$$

for $\theta \in N_\delta$. It follows that we may bound I_2 by the expression

$$I_2 \le \exp\{\ell_t(\hat{\theta}_t)\} \int_{N_\delta \cap N_t^c} \exp\{-(1/4)A_t(\theta - \hat{\theta}_t)^2\}d\theta$$

$$\le \exp\{\ell_t(\hat{\theta}_t)\} \int_{N_t^c} \exp\{-(1/4)A_t(\theta - \hat{\theta}_t)^2\}d\theta$$

(A18) $$= A_t^{-1/2}\exp\{\ell_t(\hat{\theta}_t)\}(2\pi)^{1/2}O(\exp\{-(1/4)M^2\}) \ .$$

Combining (A16) and (A18) we have

(A19) $$I_\delta = A_t^{-1/2}\exp\{\ell_t(\hat\theta_t)\}[1 + o_{as}(1)]$$

and then, using (A8), (A9) and (A19) we obtain

(A20) $$d\mathcal{P}_t/dP_t^0 = I_\delta + I_\delta^c = A_t^{-1/2}\exp\{\ell_t(\hat\theta_t)\}[1 + o_{as}(1)] \ .$$

To complete the proof of the theorem we find an alternative representation of the factor $\exp\{\ell_t(\hat\theta_t)\}$ in (A20). Noting that $\ell_t(\theta^0) = 0$ we have the two Taylor expansions

(A21) $$\ell_t(\hat\theta_t) = \ell_t^{(1)}(\theta^0)(\hat\theta_t - \theta^0) + (1/2)\ell_t^{(2)}(\theta_{m_1})\left(\hat\theta_t - \theta^0\right)^2$$

and

(A22) $$0 = \ell_t^{(1)}(\hat\theta_t) = \ell_t^{(1)}(\theta^0) + \ell_t^{(1)}(\theta_{m_2})(\hat\theta_t - \theta^0)$$

with θ_{m_1} and θ_{m_2} lying on the line segment joining $\hat\theta_t$ and θ^0. Combining (A21) and (A22) we have

$$\begin{aligned}\ell_t(\hat\theta_t) &= (1/2)\left(\hat\theta_t - \theta^0\right)^2\{\ell_t^{(2)}(\theta_{m_1}) - 2\ell_t^{(2)}(\theta_{m_2})\} \\ &= (1/2)\left(\hat\theta_t - \theta^0\right)^2 A_t\Big\{[\ell_t^{(2)}(\theta_{m_1}) - \ell_t^{(2)}(\theta^0) + \ell_t^{(2)}(\theta^0) + A_t]/A_t \\ &\quad - 2[\ell_t^{(2)}(\theta_{m_2}) - \ell_t^{(2)}(\theta^0) + \ell_t^{(2)}(\theta^0) + A_t]/A_t + 1\Big\} \\ &= (1/2)\left(\hat\theta_t - \theta^0\right)^2 A_t[1 + o_{as}(1)] \ ,\end{aligned}$$

using (C3) and (C4). It follows that (A20) may also be written as

(A23) $$d\mathcal{P}_t/dP_t^0 = A_t^{-1/2}\exp\Big\{(1/2)\left(\hat\theta_t - \theta^0\right)^2 A_t\Big\}[1 + o_{as}(1)]$$

giving the stated result (74) using the form (76) for the derivative dQ_t/dP_t.

Finally, we can use (A22) again, giving

$$0 = \ell_t^{(1)}(\theta^0) + \{\ell_t^{(2)}(\theta_{m_2}) - \ell_t^{(2)}(\theta^0) + \ell_t^{(2)}(\theta^0) + A_t - A_t\}(\hat\theta_t - \theta^0)$$

(A24) $$= \ell_t^{(1)}(\theta^0) - A_t(\hat\theta_t - \theta^0)[1 + o_{as}(1)]$$

in view of (C3) and (C4). Noting that $\ell_t^{(1)}(\theta^0) = V_t$ is a P_t^0 martingale, we can combine (A24) and (A23) to give

(A25) $$d\mathcal{P}_t/dP_t^0 = A_t^{-1/2}\exp\{(1/2)V_t^2 A_t^{-1}\}[1 + o_{as}(1)]$$

as required by expression (75) of the theorem for the exponential martingale

$$dQ_t/dP_t^0 = A_t^{-1/2}\exp\{(1/2)V_t^2A_t^{-1}\} .$$

Thus, using all three asymptotically equivalent forms of dQ_t/dP_t given by (75), (76) and (77) we have

$$\frac{d\mathcal{P}_t}{dP_t^0}/\frac{dQ_t}{dP_t^0} \to 1 \quad \text{a.s. } (P^0)$$

and the theorem is proved. □

PROOF OF COROLLARY 4.4. Using the same line of argument as that leading up to (A16) in the proof of Theorem 4.3 we have

$$\begin{aligned}\Pi_t^B(\theta) &= \pi(\theta)(dP_t^\theta/d\mathcal{P}_t) = \pi(\theta)(dP_t^\theta/dP_t^0)/(d\mathcal{P}_t/dP_t^0)\\ &= (2\pi)^{-1/2}(dP_t^\theta/dP_t^0)/(dQ_t/dP_t^0)[1 + o_{as}(1)]\\ &= (2\pi)^{-1/2}A_t^{1/2}\exp\{\ell_t(\theta) - \ell_t(\hat\theta_t)\}[1 + o_{as}(1)]\\ &= (2\pi)^{-1/2}A_t^{1/2}\exp\{-(1/2)(\theta - \hat\theta_t)^2A_t[1 + \varepsilon_t(\theta)][1 + o_{as}(1)]\end{aligned}$$

where $\varepsilon_t(\theta) \to$ a.s. (P^0) uniformly in $N_\delta \cap N_t^M$. Since $\hat\theta_t \to_{\text{a.s.}} \theta^0$ and $A_t = A_t(\theta^0) \to \infty$ a.s. (P^0) we have $N_t^M \subset N_\delta$ a.s. (P^0) for large enough t and fixed $M > 0$. Then

$$\sup_{\theta\in N_t^M}\left|\frac{\Pi_t^b(\theta)}{\varphi(\theta;\ \hat\theta_t,\ A_t^{-1})} - 1\right| \to 0 \quad \text{a.s. } (P^0)$$

giving (81) as required.

PROOF OF LEMMA 4.5. Under the stated conditions it is well known that there is a stopping time

$$\nu_{\underline{t}} = \inf\{s : A_{\underline{s}} \geq t\}$$

such that V_{ν_t} is indistinguishable from a Brownian motion W_t (e.g. Protter, 1990, Theorem 41, p. 81). We can write this equivalence as

$$V_{\underline{t}} = W_{\underline{A}_t} \quad \text{a.s. } 0 \leq t < \infty .$$

Now suppose we let X_t be another Brownian motion on the same space and construct the family of stopping times $(\sigma_t)_{t\geq 0}$ as

$$\sigma_t = \inf\left\{p : \int_0^p X_s^2 ds \geq A_t\right\} .$$

Then $\int_0^{\sigma_t} XdX$ is a martingale with quadratic variation process $\int_0^{\sigma_t} X_s^2 ds = A_t$ a.s. $0 \leq t < \infty$. Like V_t, the process $\int_0^{\sigma_t} XdX$ is equivalent to the time changed Brownian motion W_{A_t}. Hence, we have

$$V_{\nu_t} = \int_0^{\sigma_{\nu_t}} XdX = W_t \quad \text{a.s.} \quad 0 \leq t < \infty$$

and

$$V_t = \int_0^{\sigma_t} XdX = W_{A_t} \quad \text{a.s.} \quad 0 \leq t < \infty$$

giving the required result.

PROOF OF THEOREM 4.6. The proof is virtually identical to the proof of Theorem 4.3 but uses the conditional variance process $B_n = \langle \ell_n^{(1)}(\theta^0) \rangle$ in place of the quadratic variation $A_t = [\ell_n^{(1)}(\theta^0)]$.

PROOF OF THEOREM 4.7. Since $\{V_n, \mathcal{F}_n, n \geq 1\}$ is a zero mean L_2 martingale we can embed this process in a standard Brownian motion. By Theorem A1, p. 269 of Hall-Heyde (1980) there exists a probability space $(\Omega, \mathcal{E}, P)$ supporting $(V_n = \Sigma_1^n \varepsilon_k)_{n\geq 1}$, a standard Brownian motion W and stopping times $(\tau_n)_{n\geq 1}$ such that $V_n = W(\tau_n)$ and, if $\mathcal{E}_n \subset \mathcal{E}$ is the σ-field generated by $(V_k)_1^n$ and $W(t)$ for $0 \leq t \leq \tau_n$, then

H^2(i) $\quad \tau_n$ is $\mathcal{E}_n$-measurable,

H^2(ii) $\quad E\{(\tau_n - \tau_{n-1})^2 | \mathcal{E}_{n-1}\} \leq C_2 E(\varepsilon_n^4 | \mathcal{E}_{n-1})$ a.s. (P) where $C_2 = 32/\pi^2$,

and

H^2(iii) $\quad E\{(\tau_n - \tau_{n-1}) | \mathcal{E}_{n-1}\} \leq E(\varepsilon_n^2 | \mathcal{E}_{n-1})$ a.s. (P).

To prove (96) we need to show that

(A26) $\quad [V_n^2/B_n - W(\tau_n)^2/\tau_n] - \ln(B_n/\tau_n) \to 0 \quad$ a.s. (P) .

Take some positive constant $\beta < 1$. (Later on in the proof we will require that β lie in the interval $(1+\gamma)/2 < \beta < 1$.) Then

$$\text{(A27)} \quad \frac{V_n^2}{B_n} - \frac{W(\tau_n)^2}{\tau_n} = \frac{W(\tau_n)^2}{\tau_n^{2-\beta}} \frac{\tau_n - \beta_n}{\tau_n^{\beta}} \frac{\tau_n}{B_n} .$$

By the law of the iterated logarithm for Brownian motion (e.g. Shorack and Wellner, 1986, p. 27)

$$\limsup_{n\to\infty} \frac{W(\tau_n)}{\{2\tau_n \ln(\ln(\tau_n))\}^{1/2}} = 1 ,$$

so that

$$\text{(A28)} \quad W(\tau_n)^2/\tau_n^{2-\beta} \to 0 \quad \text{a.s. } (P) ,$$

since $2-\beta > 1$. Next observe that

$$\text{(A29)} \quad \frac{\tau_n - B_n}{\tau_n^{\beta}} = \frac{\tau_n(1 - B_n/\tau_n)}{\tau_n^{\beta}} \to 0 \quad \text{a.s. } (P) ,$$

and $\beta < 1$ imply that $B_n/\tau_n \to 1$ a.s. (P), and hence, in view of (A27) and (A28), it is sufficient for (A26) to prove that (A29) holds. This is easily seen to be equivalent to proving

$$\text{(A30)} \quad (\tau_n - B_n)/B_n^{\beta} \to 0 \quad \text{a.s. } (P)$$

for $\beta < 1$, which we now set out to do.

Set $\tau_0 = 0$ and $B_0 = 0$ and define

$$d_j = B_j - B_{j-1} = E(\varepsilon_j^2 | \mathscr{E}_{j-1}) ,$$

and

$$\Delta_j = \tau_j - \tau_{j-1} .$$

Then

$$\tau_n - B_n = \Sigma_1^n\{(\tau_j - \tau_{j-1}) - B_j - B_{j-1})\} = \Sigma_1^n(\Delta_j - d_j)$$

and so rewriting (A30) we need to prove

$$\text{(A31)} \quad B_n^{-\beta}\Sigma_1^n(\Delta_j - d_j) \to 0 \quad \text{a.s. } (P) .$$

By Kronecker's lemma, (A31) holds if $\Sigma_1^\infty(\Delta_j - d_j)/B_j^\beta < \infty$, a.s (P), which holds by Chow's theorem (Hall-Heyde, 1980, p. 35, Theorem 2.17) if

$$\text{(A32)} \quad \Sigma_1^\infty E\left\{\left[(\Delta_j - d_j)/B_j^\beta\right]^2 | \mathscr{E}_{j-1}\right\} < \infty \ , \quad \text{a.s. } (P)$$

since $E(\Sigma_j - d_j)/B_j^\beta$ is a martingale.

Now $E\{(\Delta_j - d_j)^2 | \mathscr{E}_{j-1}\} = E(\Delta_j^2 | \mathscr{E}_{j-1}) - d_j^2 \leq E(\Delta_j^2 | \mathscr{E}_{j-1})$, so that it is sufficient for (A32) to prove that

$$\text{(A33)} \quad \Sigma_1^\infty E(\Sigma_j^2 | \mathscr{E}_{j-1})/B_j^{2\beta} < \infty \ , \quad \text{a.s. } (P) \ .$$

Using H^2(ii) we have

$$E(\Sigma_j^2 | \mathscr{E}_{j-1}) \leq C_2 E(\varepsilon_j^4 | \mathscr{E}_{j-1}) \quad \text{a.s. } (P)$$

$$\leq C_2 C_a \left\{E(\varepsilon_j^2 | \mathscr{E}_{j-1})\right\}^2 \quad \text{a.s. } (P)$$

because of (D9). Therefore, (A33) holds if

$$\text{(A34)} \quad \Sigma_1^\infty \left(E(\varepsilon_j^2 | \mathscr{E}_{j-1})\right)^2 / B_j^{2\beta} < \infty \ , \quad \text{a.s. } (P) \ .$$

holds. Since $d_j = E(\Delta_j^2 | \mathscr{E}_{j-1}) = B_j - B_{j-1}$, we may write the left side of (A34) as

$$\text{(A35)} \quad \Sigma_1^\infty \left[\frac{B_j - B_{j-1}}{B_j^{2\beta}}\right] E(\varepsilon_j^2 | \mathscr{E}_{j-1}) \ .$$

Now take some $M > 0$, possibly large. Then by (D10) we have

$$\text{(A36)} \quad P[E(\varepsilon_n^2 | \mathscr{E}_{n-1})/B_n^\gamma > M \text{ at most finitely often}] = 1 \ .$$

The event

$$\text{(A37)} \quad [E(\varepsilon_n^2 | \mathscr{E}_{n-1})/B_n^\gamma > M \text{ at most finitely often}]$$

implies the event

$$\left[\Sigma_1^\infty \left[\frac{B_j - B_{j-1}}{B_j^{2\beta}}\right] E(\varepsilon_j^2 | \mathscr{E}_{j-1}) \leq \Sigma_{N+1}^\infty \left[\frac{B_j - B_{j-1}}{B_j^{2\beta}}\right] M B_j^\gamma + \Sigma_1^N \left[\frac{B_j - B_{j-1}}{B_j^{2\beta}}\right] E(\varepsilon_j^2 | \mathscr{E}_{j-1}) \text{ for some finite } N\right]$$

which implies

(A38) $\left[\Sigma_1^\infty\left(\frac{B_j-B_{j-1}}{B_j^{2\beta}}\right)E(\varepsilon_j^2|\mathscr{E}_{j-1}) \le M\Sigma_1^\infty\left(\frac{B_j-B_{j-1}}{B_j^{2\beta-\gamma}}\right) + \Sigma_1^N\left(\frac{B_j-B_{j-1}}{B_j^{2\beta}}\right)E(\varepsilon_j^2|\mathscr{E}_{j-1}) \text{ for some finite } N\right]$.

Let $p = 2\beta-\gamma$ and since (D10) holds for γ with $0 < \gamma < 1$ we may choose β in the interval $(1+\gamma)/2 < \beta < 1$ and then $p = 2\beta-\gamma > 1$. We have

$$\Sigma_1^\infty(B_j - B_{j-1})/B_j^p = \Sigma_1^\infty d_j/B_j^p$$

where $B_j = B_{j-1} + d_j = \Sigma_1^j d_k$. Since $d_k \ge 0$ a.s. (P) for all k and $B_j \to \infty$ a.s. (P) as $j \to \infty$ by (D2) it follows by Dini's theorem (e.g. Knopp, 1956, Theorem 1, p. 125) that

(A39) $\Sigma_1^\infty d_j/B_j^p < \infty$ a.s. (P)

because $p > 1$.

Event (A37) implies (A38) which in view of (A39) implies

(A40) $\left[\Sigma_1^\infty\left(\frac{B_j - B_{j-1}}{B_j^{2\beta}}\right)E(\varepsilon_j^2|\mathscr{E}_{j-1}) < \infty\right]$

In view of (A36) we deduce that

$$P\left[\Sigma_1^\infty\left(\frac{B_j - B_{j-1}}{B_j^{2\beta}}\right)E(\varepsilon_j^2|\mathscr{E}_{j-1}) < \infty\right] \ge P\left[E(\varepsilon_n^2|\mathscr{E}_{n-1})B_n^\gamma > M \text{ at most finitely often}\right] = 1$$

thereby proving (A34). This in turn establishes (A31), (A30) and thus (A26), which gives the stated result (96).

7. REFERENCES

Barnard, G. A., G. M. Jenkins, and C. B. Winsten (1962). "Likelihood inference and time series," *Journal of the Royal Statistical Society*, A, 125, 321-372.

Basu, D. (1973/1975). "Statistical information and likelihood," *Sankhya*, A, 37, 1-71.

Chung, K. L. and R. J. Williams (1990). *Introduction to Stochastic Integration.* Boston: Birkhauser.

DeJong, D. N. and C. H. Whiteman (1991). "Trends and random walks in macroeconomic time series: A reconsideration based on the likelihood principle," *Journal of Monetary Economics*.

Dickey, D. A. and W. A. Fuller (1981). "Likelihood ratio statistics for autoregressive time series with a unit root," *Econometrica*, 49, 1057-1072.

Doan, T., R. B. Litterman and C. Sims (1984). "Forecasting and conditional projections using realistic prior distributions," *Econometric Reviews*, 3, 1-100.

Donsker, M. D. and S. R. S. Varadhan (1977). "On laws of the iterated logarithm for local times," *Communications in Pure and Applied Mathematics*, 30, 707-753.

Duffie, D. (1988). *Security Markets: Stochastic Models.* San Diego: Academic Press.

Emery, M. (1989). *Stochastic Calculus in Manifolds.* New York: Springer Verlag.

Geweke, J. (1988). "The secular and cyclical behavior of real GDP in nineteen OECD countries, 1957-1983," *Journal of Business and Economic Statistics*, 6, 479-486.

Geweke, J. (1989). "Bayesian inference in econometric models using Monte Carlo integration," *Econometrica*, 57, 1317-1340.

Grenander, U. (1981). *Abstract Inference.* New York: Wiley.

Hall, P. and C. C. Heyde (1980). *Martingale Limit Theory and its Application.* New York: Academic Press.

Hartigan, J. A. (1964). "Invariant prior distributions," *Annals of Mathematical Statistics*, 35, 836-845.

Hartigan, J. A. (1983). *Bayes Theory.* New York: Springer Verlag.

Ibragimov, I. A. and R. Z. Has'minskii (1981). *Statistical Estimation: Asymptotic Theory.* New York: Springer Verlag.

Ikeda, N. and S. Watanabe (1989). *Stochastic Differential Equations and Diffusion Processes* (Second Edition). Amsterdam: North Holland.

Jacod, J. and A. N. Shiryaev (1987). *Limit Theorems for Stochastic Processes.* New York: Springer Verlag.

Jeffreys, H. (1946). "An invariant form for the prior probability in estimation problems," *Proceedings of the Royal Society of London*, Series A, 186, 453-461.

Jeffreys, H. (1961). *Theory of Probability, 3rd Edition.* London: Oxford University Press.

Kim, J-Y (1992). "Large sample properties of posterior densities in a time series model with a unit root and an optimal test of the unit root hypothesis," Yale University, mimeographed.

Kloek, T. and H. K. van Dijk (1978). "Bayesian estimates of equation system parameters: An application of integration by Monte Carlo," *Econometrica*, 46, 1-20.

Lai, T. L. and C. Z. Wei (1983). "Asymptotic prperties of general autoregressive models and strong consistency of least squares estimates to their parameters," *Journal of Multivariate Analysis*, 12, 346-370.

LeCam, L. (1986). *Asymptotic Methods in Statistical Decision Theory.* New York: Springer.

Litterman, R. B. (1986). "Forecasting with Bayesian vector autoregressions: Five years of experience," *Journal of Business and Economic Statistics*, 4, 25-38.

Meyer, P. A. (1989). "A short presentation of stochastic calculus," in M. Emery, *Stochastic Calculus in Manifolds*. New York: Springer Verlag.

Ouliaris, S., J. Y. Park and P. C. B. Phillips (1989). "Testing for a unit root in the presence of a maintained trend," in B. Raj (ed.), *Advances in Econometrics and Modeling*. Amsterdam: Kluwer Academic.

Park, J. Y. and P. C. B. Phillips (1988). "Statistic inference in regressions with integrated processes: Part 1," *Econometric Theory*, 4, 468-497.

Park, J. Y. and P. C. B. Phillips (1989). "Statistical inference in regressions with integrated processes: Part 2," *Econometric Theory*, 5, 95-131.

Phillips, P. C. B. (1983). "Marginal densities of instrumental variables estimators in the general single equation case," *Advances in Econometrics*, 2, 1-24.

Phillips, P. C. B. (1987). "Time series regression with a unit root," *Econometrica*, 55, 277-301.

Phillips, P. C. B. (1991). "Optimal inference in cointegrated systems," *Econometrica*, 59, 283-306.

Phillips, P. C. B. (1991a). "To criticize the critics: An objective Bayesian analysis of stochastic trends," *Journal of Applied Econometrics*, 6, 333-364.

Phillips, P. C. B. (1991b). "Bayesian routes and unit roots: de rebus prioribus semper est disputandum," *Journal of Applied Econometrics*, 6(4), 435-474.

Phillips, P. C. B. and P. Perron (1988). "Testing for a unit root in time series regression," *Biometrika*, 75, 335-346.

Poirier, D. J. (1988). "Frequentist and subjectivist perspectives on the problems of model building in economics" (with discussion), *Journal of Economic Perspectives*, 2, 121-144.

Protter, P. (1990). *Stochastic Integration and Differential Equations: A New Approach.* New York: Springer Verlag.

Said, S. E. and D. A. Dickey (1984). "Testing for unit roots in autoregressive-moving average models of unknown order," *Biometrika*, 71, 599-607.

Schotman, P. and H. K. van Dijk (1991). "A Bayesian analysis of the unit root in real exchange rates," *Journal of Econometrics*, 49, 195-238.

Sims, C. A. (1988). "Bayesian skepticism on unit root econometrics," *Journal of Economic Dynamics and Control*, 12, 463-474.

Sims, C. A. (1990). "Asymptotic behavior of the likelihood function in an autoregression with a unit root," preliminary draft, Yale University.

Sims, C. A. and H. Uhlig (1991). "Understanding unit rooters: A helicopter tour," Federal Reserve Bank of Minneapolis Institute for Empirical Macroeconomics," *Econometrica*,

Strasser, H. (1986). "Martingale difference arrays and stochastic integrals," *Probability Theory and Related Fields*, 72, 83-98.

Tierney, L. and J. B. Kadane (1986). "Accurate approximations for posterior moments and marginal densities," *Journal of the American Statistical Association*, 81, 82-86.

Tierney, L., R. E. Kass and J. B. Kadane (1989). "Approximate marginal densities of nonlinear functions," *Biometrika*, 76, 425-433.

Walker, A. M. (1969). "Asymptotic behavior of posterior distributions," *Journal of the Royal Statistical Society*, Series B, 31, 80-88.

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

Zellner, A. (1984). "Maximal data information prior distributions," in A. Zellner, *Basic Issues in Econometrics*. Chicago: University of Chicago Press.

Zellner, A (1990). "Bayesian methods and entropy in economics and econometrics." Invited Paper to 10'th International Max Ent Workship, University of Wyoming.

Power functions for Unit Root Tests

Figure 1

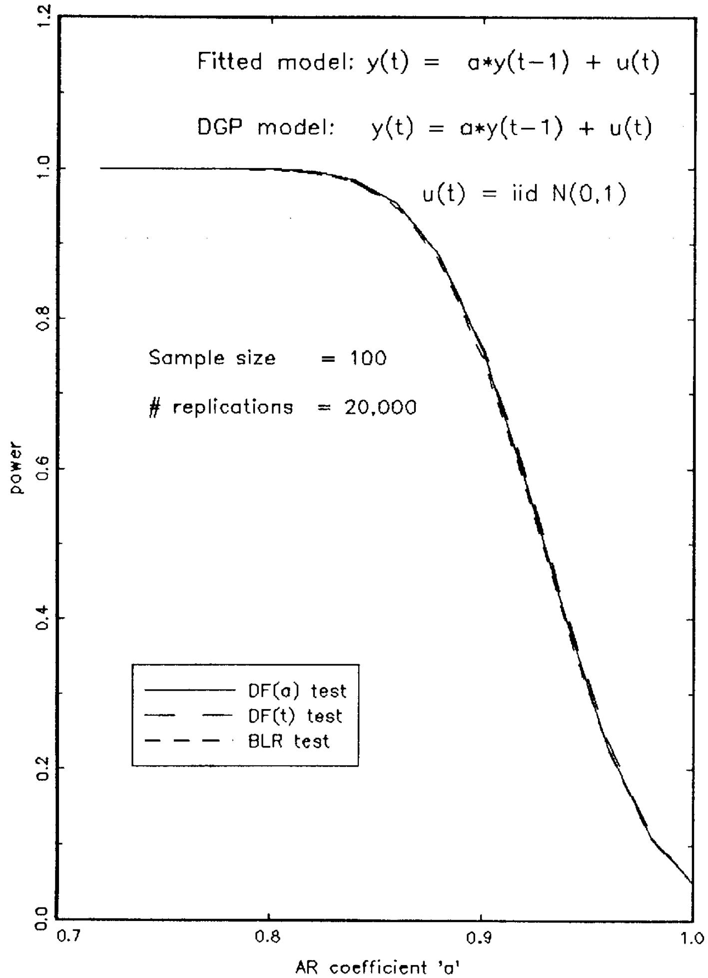

Figure 2