COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
AT YALE UNIVERSITY

Box 2125, Yale Station
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 980

TIME SERIES MODELING WITH A BAYESIAN FRAME OF REFERENCE: I. CONCEPTS AND ILLUSTRATIONS

Peter C. B. Phillips and Werner Ploberger

May 1991

TIME SERIES MODELING WITH A BAYESIAN FRAME OF REFERENCE: I. CONCEPTS AND ILLUSTRATIONS*

by

Peter C. B. Phillips and Werner Ploberger

Cowles Foundation for Research in Economics
Yale University

March 1991

*This 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 No. SES 8821180. Ploberger thanks the Fonds zur Förderung der wissenschaftlichen Forschung for supporting his stay at Yale with Schrödingerstipendium Nr. J0469–TEC. The authors thank Bent Sorensen for commenting on a draft version and Glena Ames for her skill and effort in keyboarding the manuscript.

0. ABSTRACT

This paper offers a general approach to time series modeling that attempts to reconcile classical and Bayesian methods. The central idea put forward to achieve this reconciliation is that the Bayesian approach relies implicitly on a frame of reference for the data generating mechanism that is quite different from the one that is employed in the classical approach. Differences in inferences from the two approaches are therefore to be expected unless the altered frame of reference is taken into account. We show that the new frame of reference in Bayesian inference is a consequence of a change of measure that arises naturally in the application of Bayes theorem. Our paper explores this change of measure and its consequences using martingale methods. Examples are given to illustrate its practical implications. No assumptions concerning stationarity or rates of convergence are required and techniques of stochastic differential geometry on manifolds are involved. Some implications for statistical testing are explored and we suggest new tests, which we call Bayes model tests, for discriminating between models.

This paper (Part I) emphasizes the new conceptual framework for thinking about Bayesian methods in time series and provides illustrations of its use in practice. A subsequent paper (Part II) develops a general and more abstract theory.

JEL Classification No. 211

Keywords: Autoregression; Bayes model; Bayes model measure; Bayes test; Bayesian inference; density process; Dolèans exponential; Girsanov theorem; likelihood; martingale; posterior; prior; quadratic variation process; semimartingale; stochastic differential equation; unit roots.

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, 1984; 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, 1989a, b, c; Schotman and Van Dijk, 1990; Phillips, 1991). Advances in simulation–based technology (Kloek and Van Dijk, 1978; Geweke, 1989) and improvements in analytic devices like the Laplace approximation method (Phillips, 1983, 1991; 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, 1991; 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 (1988/1991) have argued recently that time series models provide important examples where Bayesian and classical methods differ fundamentally. Phillips (1991) 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 merely the result of the use of uniform priors, which Phillips argues are inappropriate in a time series context (especially one that admits nonstationarity). 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 argued in Sims (1990). These additional differences between the 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 the conventional Bayesian approach implicitly involves a change in the underlying probability measure, leading to a new Bayesian frame of reference for the data generating mechanism. The measure change is accomplished by using Girsanov transformation theorems, which have been used in the mathematical finance literature recently (e.g. Duffie, 1988) but are not well known in econometrics. We explore the consequences of this change of measure 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 where no assumptions concerning stationarity or rates of convergence are required. Section 6 concludes the 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 or semimartingale, and the square bracket $[M,M]_t$ 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 $\langle M,M\rangle_n$ 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 " $<<$ " 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 in a simple way all of the main features of 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 similar to that of the simpler model.

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

$$dY_t = \beta Y_t dt + dW_t \tag{1}$$

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 $P_t = P_t^0$. The probability measure P_t^β has density with respect to P_t given by the following Radon—Nikodym (hereafter, RN) derivative

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

The form of (2) is actually well known in the literature (e.g. Ibragimov and Has'minski, 1981, p. 16). It will be justified below by indirect methods. Define

$$M_t = \beta \int_0^t Y_s dY_s = \beta \int_0^t Y_s dW_s + \beta^2 \int_0^t Y_s^2 ds = M_t^0 + \beta^2 \int_0^t Y_s^2 ds . \tag{3}$$

Note that $dM_t^0 = \beta Y_t dW_t$ satisfies $E(dM_t^0 | \mathcal{F}_t) = 0$, so that M_t^0 is a continuous martingale and M_t a continuous semimartingale under P_t. Further,

$$dM_t = \beta Y_t dY_t = \beta Y_t dW_t + \beta^2 Y_t^2 dt \tag{4}$$

and

$$d[M,M]_t = (dM_t)^2 = \beta^2 Y_t^2 dt = d[M^0, M^0]_t ,$$

so that the RN derivative (2) may be written in the more suggestive and generalizable form

(5) $$L_t = \exp\{M_t - (1/2)[M,M]_t\} .$$

The form (5) for the likelihood function is especially interesting 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 the true value of $\beta = 0$ we have $Y_t = W_t$, our original process of reference. Then $M_t = M_t^0$ is a P_t—martingale.

The process L_t given by (5) is called the Doléans exponential (cf. Meyer, 1989, p. 148). Using Ito's rule for differentiating continuous semimartingales we deduce that

(6) $$\begin{aligned} dL_t &= L_t\{dM_t - (1/2)d[M,M]_t\} + (1/2)L_t d[M,M]_t \\ &= L_t dM_t . \end{aligned}$$

Integrating over the interval $[0,t]$ we obtain

(7) $$L_t = 1 + \int_0^t L_s dM_s .$$

It follows that when $Y_t = W_t$ and M_t is a martingale, so too is L_t and, moreover, $E(L_t) = 1$. In this case L_t is known as the density martingale (again, see Meyer, 1989, p. 149).

Finally, to end this preliminary discussion, we note that the Girsanov theorem (e.g. Meyer, 1989, pp. 149—150, Protter, 1990, p. 109) gives us directly the mapping from P_t—martingales to P_t^β—martingales as

(8) $$Y_t \rightarrow Y_t - [Y,Z]_t = \bar{Y}_t$$

where $Z_t = \int_0^t (1/L_s)dL_s$ is the stochastic logarithm of L_t. Observe that the image process of the mapping (8) is

$$\bar{Y}_t = Y_t - \int_0^t (1/L_s)dY_s dL_s = Y_t - \int_0^t dY_s dM_s = Y_t - \beta\int_0^t Y_s ds .$$

Since this process is a P_t^β—martingale by the Girsanov theorem and since its quadratic variation is $[\bar{Y},\bar{Y}]_t = [Y,Y]_t = t$ it follows that $\bar{Y}_t$ is standard Brownian motion (e.g. see Protter, 1990, Theorem

38, p. 79). We deduce that Y_t is a process of the form (1) with $W_t \equiv \bar{Y}_t = BM(1)$ and with likelihood ratio given by (2). This provides an alternate justification of the formula (2).

The log–likelihood corresponding to (2) is

$$\Lambda_{\beta t} = \log(L_t) = \beta\int_0^t Y_s dY_s - (1/2)\beta^2\int_0^t Y_s^2 ds = \beta V_t - (1/2)\beta^2 A_t \text{ , say} \tag{9}$$

from which we derive the MLE

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

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 score function process is obtained by differentiating (9) and we write

$$N_t = \partial\Lambda_{\beta t}/\partial\beta = V_t - \beta A_t . \tag{10}$$

Note that at $\beta = 0$ we have $V_t = \int_0^t W_s dW_s$ and $N_t = V_t = \int_0^t W_s dW_s$, so that at this point in the parameter space N_t is a continuous L_2 martingale. We now show how to change the frame of reference so that N_t becomes a continuous martingale at each point $\beta \neq 0$ in a new frame of reference. In our approach the frame of reference is provided by the underlying probability measure. So we change the reference frame by changing the measure and this is effected by means of the Girsanov transformation (8).

It will now be convenient to introduce a path integral representation of the log–likelihood that involves the score function process (10). To do so, we consider a path $\beta = \beta(u)$ from $\beta(0) = 0$ to $\beta = \beta(r)$, say, that is continuously differentiable. In this simple scalar case we can select a linear path such as $\beta(u) = (u/r)\beta(r) = (u/r)\beta$. Now

$$\Lambda_{\beta(r),t} = \beta(r)V_t - (1/2)\beta(r)^2 A_t = f_t(r) = \int_0^r f_t'(u)du . \tag{11}$$

Since

$$\partial\Lambda_{\beta t}/\partial\beta = f_t'(u)\partial u/\partial\beta$$

an alternative way of writing (11) is as the path integral

$$(12) \qquad \Lambda_{\beta(r),t} = \int_0^r (\partial \Lambda_{\beta(u),t}/\partial\beta) d\beta(u) = \int_0^r N_t(u) d\beta(u)$$

using the score function process $N_t(u)$, which we now index according to the position in the path.

LEMMA 2.1. *Under the change of measure* $P_t \to Q_t^u = P_t^{\beta(u)}$, *the score process* $N_t(u)$ *becomes a* Q_t^u*–martingale.* □

As already noted, $N_t(0)$ is a P_t–martingale. The lemma shows that with a change of measure to Q_t^u the process $N_t(u)$ is also a martingale for $u \neq 0$. What this means is that at every point u on the path $\beta(u)$ the score function $N_t(u)$ is a martingale provided we use a new frame of reference in which the resident measure is Q_t^u. When we integrate along this path we produce the line integral (12) by which the log–likelihood is expressed as an accumulated score. According to the conventional interpretation the score vector is measured in a fixed coordinate system. But when $N_t(u)$ is reinterpreted as a Q_t^u–martingale, (12) provides a coordinate free representation of the log–likelihood in which the frame of reference changes continuously as we move along the path. The local coordinates in the tangent space to $\Lambda_{\beta t}$ where the score vector lies are, in effect, now provided by the probability measure Q_t^u.

Our next step is to show that the log–likelihood given by (12) can be decomposed into a collection of local quadratic terms. We start by decomposing the path $\beta(u)$, $u \in [0,r]$ into I segments $[u_{i-1}, u_i]$, $i = 1, ..., I$ with $u_0 = 0$, $u_I = r$. Then we have:

LEMMA 2.2. *The following two decompositions of the log–likelihood* $\Lambda_{\beta t}$ *apply:*

$$(13) \qquad \Lambda_{\beta(r),t} = \Sigma_{i=1}^I \left\{ N_t^{u_i} - (1/2)\left[N^{u_i}, N^{u_i}\right]_t \right\}$$

where

$$(14) \qquad N_t^{u_i} = N_t(u_{i-1})(\beta(u_i) - \beta(u_{i-1}))$$

is a continuous $Q_t^{u_{i-1}}$*–martingale;*

$$(15) \qquad \Lambda_{\beta(r),\tau} = \Sigma_{i=1}^I \{\xi_i h_i - (1/2) h_i^2 A\}$$

where

$$h_i = \beta(u_i) - \beta(u_{i-1})$$

and

$$\xi_i \equiv N(0,A)$$

under the random measure $Q_\tau^{u_{i-1}}$. *Here* τ *is the stopping time given by*

(16) $$\tau = \inf\{t > 0 : \int_0^t Y_s^2 ds \geq A\}$$

for $A > 0$ *constant.* □

Expression (13) of Lemma 2.2 decomposes the log–likelihood into a sum of elements, $N_t^{u_i} - (1/2)\left[N^{u_i}, N^{u_i}\right]_t$, each of which has the same local characteristics in the new frame of reference. The linear term is the continuous $Q_t^{u_{i-1}}$–martingale $N_t^{u_i}$ and $\left[N^{u_i}, N^{u_i}\right]_t$ is its quadratic variation process. The likelihood is

(17) $$L_t = \exp(\Lambda_{\beta t}) = \exp\left[\Sigma_1^I\left\{N_t^{u_i} - (1/2)\left[N^{u_i}, N^{u_i}\right]_t\right\}\right]$$

which, in form at least, is closely related to the earlier expression (5). However, in (5) the process M_t is a semimartingale not a martingale in the original coordinates, whereas the component processes $N_t^{u_i}$ $(i = 1, ..., N)$ in (13) are all martingales in the new reference frame.

Expression (15) of Lemma 2.2 shows that when the "information content of the data," viz. $[N,N]_t = [V,V]_t = \int_0^t Y_s^2 ds$, is preset and fixed at some constant level (here, given by A), then the log-likelihood takes the usual form of a local quadratic in which the linear term is Gaussian and the quadratic term is fixed, representing the nonrandom information content A . Further, observe that with the stopping rule (16) imposed, the estimator $\hat{\beta}_\tau$ is itself Gaussian, i.e.

(18) $$\hat{\beta}_\tau = \int_0^\tau Y_s dY_s / \int_0^\tau Y_s^2 \equiv N(\beta, 1/A) ,$$

just as it would be in the case of a fixed regressor model without the use of a stopping time. For instance, if the model (1) were replaced by

(1)′ $$dY_t = \beta S_t dt + dW_t$$

where S_t is a smooth nonrandom signal in continuous time then $\hat{\beta}_t = \int_0^t S_s dY_s / \int_0^t S_s^2 dS \equiv N(\beta,\ 1/\int_0^t S_s^2 ds)$, which is entirely analogous to (18).

Next, we define a grid that leads to a very important and useful decomposition of the likelihood. Set $I = 2$ and define $\beta(u_0) = 0$, $\beta(u_1) = \hat{\beta}_t$ and $\beta(u_2) = \beta$, the last being an arbitrary value of the parameter. Then

$$N_t^{u_1} = N_t(u_0)\hat{\beta}_t = V_t\hat{\beta}_t\,,\quad \left[N^{u_1}, N^{u_1}\right]_t = \hat{\beta}_t^2 A_t\,,$$

$$N_t^{u_2} = N_t(u_1)(\beta - \hat{\beta}) = 0\,,\quad \left[N^{u_2}, N^{u_2}\right]_t = (\beta - \hat{\beta}_t)^2 A_t\,.$$

Using this decomposition in (13) the likelihood (17) becomes

$$\begin{aligned} L_t &= \exp\left\{N_t^{u_1} - (1/2)\left[N^{u_1}, N^{u_1}\right]_t\right\}\exp\left\{-(1/2)\left[N^{u_2}, N^{u_2}\right]_t\right\} \\ &= \exp\{V_t\hat{\beta}_t - (1/2)\hat{\beta}_t^2 A_t\}\exp\{-(1/2)(\beta - \hat{\beta}_t)^2 A_t\} \\ &= \exp\{(1/2)\hat{\beta}_t^2 A_t\}\exp\{-(1/2)(\beta - \hat{\beta}_t)^2 A_t\}\,. \end{aligned} \tag{19}$$

Note that only the second exponential factor of (19) depends explicitly on β and is, moreover, proportional to a $N(\hat{\beta}_t,\ A_t^{-1})$ density. In conventional Bayesian inference it is this latter factor that plays a 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 a uniform density. Combining the prior $\pi(\beta)$ with the likelihood as given in (19) we have the posterior process

$$\begin{aligned} \Pi_t &= \pi(\beta)(dP_t^{\beta}/dP_t) = \pi(\beta)L_t = \pi(\beta)\exp\{(1/2)\hat{\beta}_t^2 A_t\}\exp\{-(1/2)(\beta - \hat{\beta}_t)^2 A_t\} \\ &= \left[A_t^{-1/2}\exp\{(1/2)\hat{\beta}_t^2 A_t\}\right]\left[\pi(\beta)A_t^{1/2}\exp\{-(1/2)(\beta - \hat{\beta}_t)^2 A_t\}\right] \\ &= \left[\exp\{(1/2)\hat{\beta}_t^2 A_t - (1/2)\ell n(A_t)\}\right]\left[\pi(\beta)A_t^{1/2}\exp\{-(1/2)(\beta - \hat{\beta}_t)^2 A_t\}\right]\,. \end{aligned} \tag{20}$$

The decomposition of the posterior process Π_t into the two factors in square brackets in (20) is very important in what follows. The first factor is a local martingale and, as we shall see, produces the density process that changes the measure to a Bayesian frame of reference. Observe, however, that the first factor does not explicitly involve the parameter β, so that the Bayesian posterior is, in effect, proportional to the second factor in square brackets in (20). 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.

In order to avoid integrability problems and the need to work with local martingales we let t_0 be a stopping time for which a "minimal amount" of information in the data has accumulated. Specifically, we set

$$t_0 = \inf\{s > 0 : A_s \geq c\} \tag{21}$$

where $c > 0$ is a preassigned (possibly small) constant. Let

$$c_0 = A_{t_0}^{-1/2} \exp\left\{(1/2)\hat{\beta}_{t_0}^2 A_{t_0}\right\} .$$

Then

$$A_t^{-1/2} \exp\{(1/2)\hat{\beta}_t^2 A_t\} = c_0 \exp\left\{(1/2)\int_{t_0}^t d[\hat{\beta}_s^2 A_s - \ln(A_s)]\right\} \tag{22}$$

and t_0 becomes, in effect, the new initialization of the process (22). Define

$$R_t = \exp\left\{(1/2)\int_{t_0}^t d[\hat{\beta}_s^2 A_s - \ln(A_s)]\right\} . \tag{23}$$

The following lemmas show that we can write the factor (23) in a much more revealing form

LEMMA 2.3

$$d[\hat{\beta}_t^2 A_t - \ln(A_t)] = \hat{\beta}_t dV_t - (1/2)\hat{\beta}_t^2 dA_t . \quad \square \tag{24}$$

LEMMA 2.4

$$R_t = \exp\{G_t - (1/2)[G,G]_t\} \tag{25}$$

where $G_t = \int_{t_0}^{t} \hat{\beta}_s dV_s$.

$$R_t = 1 + \int_{t_0}^{t} R_s dG_s \tag{26}$$

is the Doléans exponential of G_t . *If* V_t *and, hence,* G_t *are a martingales and if* $E[\exp\{(1/2)[G,G]_t\}] < \infty$ *then*

$$E(R_t) = 1 \tag{27}$$

and R_t *is a density process.* □

Our next step is to make a change of probability measure from $P_t \to Q_t^B$ according to the following RN derivative which defines Q_t^B , viz.

$$dQ_t^B/dP_t = R_t = \exp\left\{\int_{t_0}^{t} [\hat{\beta}_s dV_s - (1/2)\hat{\beta}_s^2 dA_s]\right\} . \tag{28}$$

We shall call Q_t^B the *Bayes model measure.* Using (20), (22), (24) and (28) we now obtain the following expression for the posterior density process. Specifically,

$$\Pi_t = \pi(\beta)(dP_t^{\beta}/dP_t) = c_0(dQ_t^B/dP_t)[\pi(\beta)A_t^{1/2} \exp\{-(1/2)(\beta - \hat{\beta}_t)^2 A_t\}] . \tag{29}$$

We next define the posterior density process with respect to Q_t^B measure as

$$\Pi_t^B = \pi(\beta)(dP_t^{\beta}/dQ_t^B) = c_0\pi(\beta)A_t^{1/2} \exp\{-(1/2)(\beta - \hat{\beta}_t)^2 A_t\} . \tag{30}$$

Note that it is this posterior density that is used in practice in traditional Bayesian inference. This is because when we condition on the data the factor (dQ_t^B/dP_t) that appears in (29) is absorbed into the constant of proportionality. Hence, the implicit change of measure in Bayesian inference is lost in the passage from the likelihood to the posterior. Under data conditioning and from a Bayesian perspective it is equivalent to work with either Π_t or Π_t^B . However, only Π_t^B as given by (30) makes explicit the underlying probability measure that is implicit in Bayesian inference.

<u>THEOREM 2.5.</u> (a) *Under a uniform prior for* β *the posterior density process* Π_t^B *is a Gaussian process whose finite dimensional distribution at time* t *is* $N(\hat{\beta}_t, A_t^{-1})$, *i.e. normal with mean* $\hat{\beta}_t$ *and variance* A_t^{-1} .

(b) *Irrespective of the prior distribution that is used for* β , *Bayes methods imply a replacement of the underlying probability measure* P_t *with the Bayes model measure* Q_t^B , *i.e. the likelihood function on which Bayes inference is based is given by* dP_t^{β}/dQ_t^B *not* dP_t^{β}/dP_t *as in* (2).

(c) *The Bayes model measure* Q_t^B *is the probability measure of the output process* Y_t *of the nonlinear stochastic differential equation*

$$(1)^B \qquad dY_t = \hat{\beta}_t Y_t dt + dW_t$$

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

This theorem tells us that in a pure Bayesian analysis there is no concept of a true dynamic model (1) with a true value of β . Instead, in 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 model measure* Q_t^B . Thus, to a Bayesian the reference model evolves according to the recorded history of the process on which all inference is conditioned. That model is the nonlinear stochastic differential equation $(1)^B$ and is an evolving parameter model. We conclude that Bayesian inference based on the posterior density process Π_t^B relates in effect to a trajectory based version of the original model, i.e. $(1)^B$, rather than the true original model.

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

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

Our model is the Gaussian AR(1)

$$H_\alpha : Y_t = \alpha Y_{t-1} + u_t \,, \quad u_t \equiv \text{iid } N(0,1) \tag{31}$$

initialized at $t = 0$ with Y_0 any $\mathcal{F}_0$–measurable variable. We use P_n^α to represent the probability measure of $\{Y_t\}_1^n$. 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, is

$$\begin{aligned}\Lambda_{hn} &= \ln(dP_n^\alpha/dP_n) = \ln[(dP_n^\alpha/dP_n^0)(dP_n^0/dP_n)] \\ &= -(1/2)\Sigma_1^n(Y_t - \alpha Y_{t-1})^2 + (1/2)\Sigma_1^n(Y_t - Y_{t-1})^2 \\ &= h\Sigma_1^n Y_{t-1}\Delta Y_t - (1/2)h^2\Sigma_1^n Y_{t-1}^2 \end{aligned} \tag{32}$$

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 (32) above.

The likelihood function is given by

$$L_n = dP_n^\alpha/dP_n = \exp(\Lambda_{hn}) \,.$$

Our next lemma shows that the discrete likelihood has a form analogous to that of (5) in the continuous case.

<u>LEMMA 3.1</u>

$$L_n = \exp\{M_n - (1/2)\langle M,M\rangle_n\} \tag{33}$$

where $M_n = h\Sigma_1^n Y_{t-1}\Delta Y_t$ *is a* P_n*–martingale and* $\langle M,M\rangle_n$ *is its conditional quadratic variation;*

$$L_n = 1 + \Sigma_{j=1}^n L_{j-1}\{L_j/L_{j-1} - 1\} = 1 + \Sigma_{j=1}^n L_{j-1} r_j \tag{34}$$

where $E(r_j | \mathcal{F}_{j-1}\} = 0$ *; and*

(35) $$E(L_n) = 1 . \ \square$$

Observe that the Dolèans exponential formula (33) for the likelihood L_n involves the conditional (or predictable) quadratic variation represented by the sharp bracket formula $<M,M>_n$. In the continuous time case we used the square bracket process and in that case, since the martingales were continuous, the distinction is unimportant because the square bracket process and sharp bracket process are the same. Here, we must use $<M,M>_n$ not $[M,M]_n$ for the validity of (33).

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

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

At $h = 0$, N_n is a P_n–martingale. When $h \neq 0$ we may change the measure to ensure that N_n becomes a martingale under the new measure. Let $h = h(u)$ be a continuously differentiable path from $h(0) = 0$ to $h(r) = h$. Then, as in (12), we have a path integral representation of Λ_{hn}, viz.

(36) $$\Lambda_{hn} = \int_0^r(\partial\Lambda_{hn}/\partial h)dh(u) = \int_0^r N_n(u)dh(u)$$

where

(37) $$N_n(u) = V_n - h(u)A_n .$$

Let $Q_n^u = P_n^{1+h(u)}$. The following lemmas mirror earlier results for the continuous time case.

<u>LEMMA 3.2.</u> *Under the change of measure* $P_n \to Q_n^u$, *the score process* $N_n(u)$ *is a discrete* Q_n^u*–martingale. The conditional quadratic variation of* $N_n(u)$ *under* Q_n^u *is*

$$<N(u), N(u)>_n = A_n . \ \square$$

LEMMA 3.3

(38) $$\Lambda_{hn} = \Sigma_{i=1}^{I}\left\{N_n^{u_i} - (1/2)<N^{u_i}, N^{u_i}>_n\right\}$$

where

$$N_n^{u_i} = N_n(u_{i-1})(h(u_i) - h(u_{i-1}))$$

is a discrete $Q_n^{u_{i-1}}$*–martingale and*

$$\{[u_{i-1}, u_i] : i = 1, ..., I;\ u_0 = 0,\ u_I = u;\ h(0) = 0,\ h(u) = h\}$$

is a partition that decomposes the path $h(u)$ *into* I *segments.* □

As in the continuous case, now take a grid with $I = 2$ such that $h(0) = 0$, $h(u_1) = \hat{h}_n$ and $h(u_2) = h$. Then

$$N_n^{u_1} = N_n(u_0)\hat{h}_n = V_h\hat{h}_n\ ,\quad <N^{u_1}, N^{u_1}>_n = \hat{h}_n^2 A_n$$

$$N_n^{u_2} = N_n(u_1)(h - \hat{h}_n) = 0\ ,\quad <N^{u_2}, N^{u_2}>_n = (h - \hat{h}_n)^2 A_n\ .$$

From (38) we find the likelihood is

$$L_n = \exp\left\{N_n^{u_1} - (1/2)<N^{u_1}, N^{u_1}>_n\right\}\exp\left\{-(1/2)<N^{u_2}, N^{u_2}>_n\right\}$$

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

entirely analogous to the continuous case (19).

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

Suppose we have given a prior density $\pi(h)$ on $h = \alpha-1$. The posterior density process is then

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

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

LEMMA 3.4

$$(41)\qquad (1/2)\{\hat{h}^2_{n+1}A_{n+1} - \ln(A_{n+1})\} = (1/2)\{\hat{h}^2_n A_n - \ln(A_n)\} + \{\hat{h}_n Y_n \Delta Y_{n+1} - (1/2)\hat{h}^2_n Y^2_n\} + O_p(Y^2_n/A_n)$$

$$(42)\qquad (1/2)\{\hat{h}^2_{n+1}A_{n+1} - \ln(A_{n+1})\} = (1/2)\{\hat{h}^2_{n_0} A_{n_0} - \ln(A_{n_0})\} + \Sigma^n_{s=n_0}\{\hat{h}_s Y_s \Delta Y_{s+1} - (1/2)\hat{h}^2_s Y^2_s\} + O_p(\Sigma^n_{n_0+1} Y^2_s/A_{n_0}) . \ \square$$

We now set n_0 to be a stopping time for which a minimal amount of information about the process has accumulated. Specifically, for some preassigned $c > 0$ set

$$n_0 = \min\{k : A_k \geq c\}$$

and let

$$c_0 = A^{-1/2}_{n_0} \exp\{(1/2)\hat{h}^2_{n_0} A_{n_0}\}$$

be the initialization of the process in the first square bracket of (40). The approximation suggested by (42) will be adequate if there is sufficient initial information A_{n_0} . The posterior density process is approximately

$$(43)\qquad \Pi_n \sim c_0 \exp\left\{\Sigma^{n-1}_{n_0}[\hat{h}_s Y_s \Delta Y_{s+1} - (1/2)\hat{h}^2_s Y^2_s]\right\}\left[\pi(h) A^{1/2}_n \exp\{-(1/2)(h - \hat{h}_n)^2 A_n\}\right] .$$

LEMMA 3.5. *Under* H_1

$$(44)\qquad R_n = \exp\left\{\Sigma^{n-1}_{n_0}[\hat{h}_s Y_s \Delta Y_{s+1} - (1/2)\hat{h}^2_s Y^2_s]\right\} = \exp\{G_n - (1/2)<G,G>_n\}$$

where $G_n = \Sigma^{n-1}_{n_0}\hat{h}_s Y_s \Delta Y_{s+1}$ *is a discrete martingale under* H_1 ;

$$(45)\qquad E(R_{n+1}|\mathcal{F}_n) = R_n$$

$$(46)\qquad E(R_n) = 1 . \ \square$$

It follows from Lemma 3.5 that R_n is the discrete Doléans exponential of G_n and is a martingale under H_1 with $E(R_n) = 1$. R_n therefore represents a discrete density process. We use it to define the new probability measure Q_n^B by setting

(47) $$dQ_n^B/dP_n = R_n = \exp\left\{\Sigma_{n_0}^{n-1}[\hat{h}_s Y_s \Delta Y_{s+1} - (1/2)\hat{h}_s^2 Y_s^2]\right\}.$$

Following (28) we call Q_n^B the *discrete Bayes model measure.* Using (47) in (43) we have

$$\Pi_n = \pi(h)(dP_n^\alpha/dP_n) - c_0(dQ_n^B/dP_n)\left[\pi(h)A_n^{1/2}\exp\{-(1/2)(h-\hat{h}_n)^2 A_n\}\right].$$

The discrete posterior density process with respect to Q_n^B measure is given by

(48) $$\begin{aligned}\Pi_n^B &= \pi(h)(dP_n^\alpha/dQ_n^B) \\ &\quad - c_0\pi(h)A_n^{1/2}\exp\{-(1/2)(h-\hat{h}_n)^2 A_n\}.\end{aligned}$$

Similar comments apply to the *discrete* process Π_n^B as those made in connection with the continuous process Π_t^B following (30). In particular, we have:

<u>THEOREM 3.6.</u>

(a) *Under a uniform prior* $\pi(h)$, Π_n^B *is approximately Gaussian with distribution* $N(\hat{h}_n, A_n^{-1})$ *at time* n.

(b) *Bayes methods imply the use of the discrete Bayes model measure* Q_n^B *as the reference measure in constructing the likelihood.*

(c) *The model to which* Q_n^B *applies is the time varying parameter model*

(49) $$H_{\hat{\alpha}_{n-1}} : Y_n = \hat{\alpha}_{n-1}Y_{n-1} + u_n$$

where the evolving parameter $\hat{\alpha}_{n-1}$ *is given by*

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

and is trajectory dependent. □

Thus, 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^B) in which the parameters evolve according to the observed trajectory of the process.

3.3. A Bayes Model Test

Bayes methods change the frame of reference to a Bayes measure (Q_n^B) 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. We now apply this idea, starting with model $H_{\hat{\alpha}_{n-1}}$.

From (44) and (42) we deduce that twice the log–likelihood ratio is approximately

$$2\ \ln(dQ_n^B/dP_n) \sim \hat{h}_n^2 A_n - \ln(A_n) - \ln c_0 . \tag{50}$$

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

$$BLR = \hat{h}_n^2 A_n - \ln(n^{-2} A_n) . \tag{51}$$

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

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

and then the BLR statistic is

$$BLR_\sigma = \hat{h}_n^2 A_n / \hat{\sigma}^2 - \ln(n^{-2} A_n / \hat{\sigma}^2) . \tag{52}$$

Using standard functional limit theory we obtain:

<u>THEOREM 3.7.</u> *Under* H_1

$$BLR,\ BLR_\sigma \Rightarrow \left[\int_0^1 S dS\right]^2 / \int_0^1 S^2 - \ln\left(\int_0^1 S^2\right) = g(S) \tag{53}$$

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

We may use the statistic BLR_σ to test H_1 against H_α $(\alpha < 1)$. 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 one sided 5% level test of H_1 against H_α $(\alpha < 1)$ is provided by the criterion

$$BLR_\sigma > g_{0.95} .$$

Likewise, a one–sided 5% level test of H_1 against H_α $(\alpha > 1)$ is provided by

$$\mathrm{BLR}_\sigma < g_{0.05}$$

where $g_{0.05}$ is the left tail 5% critical value of $g(S)$.

Observe that the BLR_σ 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_σ under both alternatives (viz. $\alpha < 1$, $\alpha > 1$) are easily seen to be $O_p(n)$.

3.4. Some Model Extensions

The ideas of the last subsection can be used to develop tests that apply in models with drift and deterministic trends. We start by considering the models

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

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

where $u_t \equiv$ iid $N(0,1)$ and the time series are initialized at $t = 0$ with Y_0 $\mathcal{F}_0$–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

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

$$= \left[n^{-1/2}\exp\{\hat{\mu}_n^2 n/2\}\right]\left[\pi(\mu)n^{1/2}\exp\{-(1/2)(\mu - \hat{\mu}_n)^2 n\}\right]. \tag{54}$$

The Bayes model measure is

$$dQ_n^{\mu,1}/dP_n = \exp\left\{\Sigma_{n_0+1}^{n-1}[\hat{\mu}_s \Delta Y_{s+1} - (1/2)\hat{\mu}_s^2]\right\} \tag{55}$$

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

(56) $$\Delta Y_{n+1} = \hat{\mu}_n + u_n$$

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

Following (50) and (51) the Bayes model likelihood ratio test of $H_{\mu,1}$ against the null reference model $H_{\mu=0,1}$ is just

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

with asymptotic distribution given by

(57) $$BLR(\hat{\mu}_n) \Rightarrow N(0,1)\ ,$$

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 (57).

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

$$\Pi_n = [|A_n|^{-1/2} \exp\{(1/2)\hat{\theta}_n' A_n \hat{\theta}_n\}][\pi(\theta)|A_n|^{1/2} \exp\{-(1/2)(\theta - \hat{\theta}_n)' A_n (\theta - \hat{\theta}_n)\}] \tag{58}$$

where $A_n = \Sigma_1^n X_t X_t'$. Ignoring the step of initializing on minimal information, we deduce that the nonintegrable version of the Bayes model measure is

$$dQ_n^\theta / dP_n = |A_n|^{-1/2} \exp\{(1/2)\hat{\theta}_n' A_n \hat{\theta}_n\} \,. \tag{59}$$

This is a very useful general form of the Bayes model measure that will be utilized extensively in what follows and in Part II. We shall call this version of the measure the *unconditional Bayes model measure* since it is not conditional on an initialization in which there is minimal information.

After a little calculation in the present case with $\hat{\theta}_n' = (\hat{\mu}_n, \hat{h}_n)$, we find

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

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

$$dQ_n^\theta / dP_n = \exp\{(1/2)\hat{h}_n \Sigma_1^n (Y_{t-1} - \bar{Y}_{-1})^2 - (1/2)\ln(n\Sigma_1^n (Y_{t-1} - \bar{Y}_{-1})^2) + (1/2)n\overline{\Delta Y}^2\} \,. \tag{60}$$

Next observe that the unconditional version of the Bayes model measure for $B_{\mu,1}$ is, from (54),

$$dQ_n^{\mu,1} / dP_n = \exp\{(1/2)n\overline{\Delta Y}^2 - (1/2)\ln(n)\} \,. \tag{61}$$

Combining (60) and (61) we obtain

$$\begin{aligned} dQ_n^\theta / dQ_n^{\mu,1} &= (dQ_n^\theta / dP_n)(dP_n / dQ_n^{\mu,1}) \\ &= \exp\{(1/2)\hat{h}_n^2 \Sigma_1^n (Y_{t-1} - \bar{Y}_{-1})^2 - (1/2)\ln(\Sigma_1^n (Y_{t-1} - \bar{Y}_{-1})^2)\} \,. \end{aligned} \tag{62}$$

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

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

Finally, estimating the error variance by $\hat{\sigma}^2 = n^{-1}\Sigma_1^n(Y_t - \hat{\theta}_n' X_t)$, we have the Bayes model likelihood ratio test

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

<u>THEOREM 3.8.</u> *Under* $H_1 = H_{\mu=0,1}$

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

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

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

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

$$B^k_{\varphi,\alpha} : Y_{n+1} = X_{n+1}'\hat{\varphi}_n + \hat{\alpha}_n Y_n + u_{n+1}$$

where $\hat{\theta}_n = (\hat{\varphi}_n',\ \hat{\alpha}_n) = \left[\Sigma_1^n Z_t Z_t'\right]^{-1}(\Sigma_1^n Z_t Y_t)$, $Z_t' = (X_t',\ Y_{t-1})$. The unconditional Bayes model measure for $B^k_{\varphi,\alpha}$ is given by (59) with $A_n = \Sigma_1^n Z_t Z_t'$.

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

$$B^k_{\varphi,1} : \Delta Y_{n+1} = X_{n+1}'\tilde{\varphi}_n + u_n$$

with $\tilde{\varphi}_n = \left[\Sigma_1^n X_t X_t'\right]^{-1}(\Sigma_1^n X_t \Delta Y_t)$. The unconditional Bayes model measure for $B^k_{\gamma,1}$ is again given by (59) but with $A_n = \Sigma_1^n X_t X_t'$. We now have

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

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

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

Again, when σ^2 is estimated we have

$$\mathrm{BLR}_\sigma = \hat{h}^2 Y_{-1}' Q_X Y_{-1}/\hat{\sigma}^2 - \ln\{n^{-2}(Y_{-1}' Q_X Y_{-1})/\hat{\sigma}^2\} \tag{63}$$

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

<u>THEOREM 3.9.</u> *Under* $H^{k-1}_{\varphi,1} = H^k_{\varphi,1}$ *with* $\varphi_k = 0$ *we have*

$$\mathrm{BLR},\ \mathrm{BLR}_\sigma \Rightarrow \left[\int_0^1 S_k dS\right]^2 / \int_0^1 S_k^2 - \ln\left(\int_0^1 S_k^2\right) \tag{64}$$

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 (63) may be used to test $H^{k-1}_{\varphi,1}$ against $H^{k-1}_{\varphi,\alpha}$. Both (63) and its limit distribution given in (64) 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 (63) 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).

3.5. Posterior Odds, Best Bayes Tests and Bayes Model Tests

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 models H_α and H_1 considered earlier (see (31)).

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):

$$\text{(65)} \qquad \textit{if } dP_n^\alpha/dP_n > \pi_1/\pi_\alpha, \textit{ then decide in favor of } H_\alpha$$

i.e. decide in favor of H_α if the posterior odds > 1. The criterion (65) 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).

From (32) we have the density process

$$\text{(66)} \qquad dP_n^\alpha/dP_n = \exp\{h\Sigma_1^n Y_{t-1}\Delta Y_t - (1/2)h^2\Sigma_1^n Y_{t-1}^2\}$$

where $h = \alpha-1$. Theorem 3.6 tells us that use of Bayes methods implies the existence of the Bayes model $H_{\hat{\alpha}_{n-1}}$ given by (49) with an evolving, data dependent parameter $\hat{\alpha}_{n-1}$. Substituting the latest estimate $\hat{\alpha}_n = 1 + \hat{h}_n$ in (66), we get

$$\text{(67)} \quad dP_n^{\hat{\alpha}_n}/dP_n = \exp\{\hat{h}_n\Sigma_1^n Y_{t-1}\Delta Y_t - (1/2)\hat{h}_n^2\Sigma_1^n Y_{t-1}^2\} = \exp\{(1/2)\hat{h}_n^2\Sigma_1^n Y_{t-1}^2\}.$$

Twice the logarithm of (67) gives us

$$\text{(68)} \qquad 2\ln(dP_n^{\hat{\alpha}_n}/dP_n) = \hat{h}_n^2(\Sigma_1^n Y_{t-1}^2) \Rightarrow \left[\int_0^1 S dS\right]^2 / \int_0^1 S^2,$$

which is the square of the Dickey–Fuller t–ratio statistic and its limit process. That is, (67) leads to the usual classical test if conventional critical values are used.

Let us suppose, however, that the posterior odds criterion is to be employed rather than the classical test and that we set $\pi_1 = \pi_\alpha$ so that the prior odds are equal. In that case since

$$dP_n^{\hat{\alpha}_n}/dP_n > 1 = \pi_1/\pi_\alpha \tag{69}$$

we would always accept $H_{\hat{\alpha}_n}$ using the posterior odds criterion (65). Clearly, this is not a fair test and we may well ask why not. The reason is that $dP_n^{\hat{\alpha}_n}/dP_n$ is not a density and (69) is therefore not a proper Bayes test in the sense of (65). This raises the next question, which is whether we can modify (69) so that it is a proper Bayes test. The answer is yes and, furthermore, the resulting test relies on the Bayes model likelihood ratio introduced earlier.

Recall from (68) that

$$dP_n^{\hat{\alpha}_n}/dP_n = \exp\{(1/2)\hat{h}_n^2\Sigma_1^n Y_{t-1}^2\} .$$

To transform this expression into a density we need to standardize it by the square root of the conditional quadratic variation $\Sigma_1^n Y_{t-1}^2$, giving the Bayes model density process

$$\frac{dQ_n^B}{dP_n} = \frac{dP_n^{\hat{\alpha}_n}/dP_n}{\left[\Sigma_1^n Y_{t-1}^2\right]^{1/2}} = \frac{\exp\{(1/2)\hat{h}_n^2\Sigma_1^n Y_{t-1}^2\}}{\left[\Sigma_1^n Y_{t-1}^2\right]^{1/2}} . \tag{70}$$

If we now employ (70), which is a proper density upon suitable initialization, in the posterior odds criterion we have the decision rule

$$\textit{if } dQ_n^B/dP_n > \pi_1/\pi_\alpha, \textit{ then decide in favor of } H_\alpha . \tag{71}$$

This decision rule can be translated into a criterion for $dP_n^{\hat{\alpha}_n}/dP_n$. Indeed, from (70) and (71) we have the equivalent rule

$$(72) \qquad \frac{dP_n^{\hat{\alpha}_n}}{dP_n} > \frac{\pi_1 \left[\Sigma_1^n Y_{t-1}^2\right]^{1/2}}{\pi_\alpha} .$$

Thus, our "Bayes model likelihood ratio posterior odds test" is equivalent to a "best Bayes test" when the prior for H_1 is $\pi_1 \left[\Sigma_1^n Y_{t-1}^2\right]^{1/2}$. In other words, to make the Bayes posterior odds test that is based on the "likelihood ratio" $dP_n^{\hat{\alpha}_n}/dP_n$ valid, we need to weight the prior odds on H_1 with a Jeffreys–type prior. (See Phillips, 1991, for a development of Jeffreys priors in models where no stationarity assumption is made.) In the present case, $\Sigma_1^n Y_{t-1}^2$ is the conditional variance of the martingale $\Sigma_1^n Y_{t-1} \Delta Y_t$ under H_1 and is a form of conditional information measure that measures the amount of information there is in the data about α.

4. TOWARD A GENERAL THEORY

4.1. The Likelihood

Let $\{Y_t\}_1^n$ be a time series defined on the filtered sequence of measurable spaces $(\Omega, \mathcal{F}_t)$. Let P_n^θ be a parameterised probability measure of $\{Y_t\}_1^n$ in which $\theta \in \Theta$, an open subset of $\mathbb{R}$. Suppose θ^0 is the true value of θ and that $P_n^\theta << \nu_n$, 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

$$(73) \qquad L_n(\theta) = dP_n^\theta/dP_n^0 = (dP_n^\theta/d\nu_n)/(dP_n^0/d\nu_n)$$

and set $L_0(\theta) = 1$. Finally, we assume that $L_n(\theta)$ is twice continuously differentiable and that these derivatives are dominated by absolutely integrable functions, so that passage under the integral of differentiation with respect to θ is permissible.

The log–likelihood is

$$(74) \qquad \Lambda_{\theta n} = \ln(L_n(\theta)) = \Sigma_{k=1}^n [\ln(L_k(\theta)) - \ln(L_{k-1}(\theta))]$$

and the score function process is

$$(75) \qquad N_n = \partial \Lambda_{\theta n}/\partial\theta = \Sigma_{k=1}^n (\partial/\partial\theta)[\ln(L_n(\theta) - \ln(L_{k-1}(\theta))] = \Sigma_{k=1}^n \epsilon_k(\theta) \text{ , say.}$$

Set $\epsilon_k^0 = \epsilon_k(\theta^0)$, $N^0 = \Sigma_{k=1}^n \epsilon_k^0$ and define

$$I_n(\theta) = \Sigma_{k=1}^n E(\epsilon_k(\theta)^2 | \mathcal{F}_{k-1}) = \langle N,N \rangle_n \; , \; I_n^0 = I_n(\theta^0) \; .$$

We also define the second derivatives

$$\eta_k(\theta) = \partial \epsilon_k(\theta)/\partial\theta \; , \; \eta_k^0 = \eta_k(\theta^0) \; ,$$

and the accumulated derivative process

$$J_n(\theta) = \Sigma_{k=1}^n \eta_k(\theta) \; , \; J_n^0 = J_n(\theta^0) \; .$$

Following conventional theory for maximum likelihood estimation of stochastic processes, we assume that

(76) $$I_n(\theta) \to \infty \text{ a.s. } (P_n^0)$$

and that under P_n^0

(77) $$J_n(\theta)/I_n(\theta) \to_p -1$$

uniformly in θ (cf. Hall and Heyde, 1980, p. 160).

Under P_n^0, $E(\epsilon_k^0 | \mathcal{F}_{k-1}) = 0$ and so N_n^0 is a P_n^0–martingale, a well known feature of maximum likelihood theory for stochastic processes (e.g. Hall and Heyde, 1980, p. 157). Similarly, $E((\epsilon_k^0)^2 + \eta_k^0 | \mathcal{F}_{k-1}) = 0$ and $J_n^0 + I_n^0$ is also a P_n^0–martingale. Suitable transforms of measures preserve these martingale properties for $\theta \neq \theta^0$.

Let $\theta = \theta(u)$ be a continuously differentiable path from $\theta(0) = \theta^0$ to $\theta(r) = \theta$. As in (36), we have

(78) $$\Lambda_{\theta n} = \int_0^r (\partial \Lambda_{\theta n}/\partial\theta) d\theta(u) = \int_0^r N_n(u) d\theta(u) \; .$$

We write $Q_n^u = P_n^{\theta(u)}$ and

(79) $$N_n(u) = \Sigma_{k=1}^n \epsilon_k(\theta(u)) = \Sigma_{k=1}^n \epsilon_k^u \; , \text{ say.}$$

We now assume that for some $a > 0$ the moment generating function of ϵ_k^u is bounded above for $0 \le u \le r$, i.e.

(80) $$\sup_u \{E(\exp\{a\epsilon_k^u\})\} < \infty \; .$$

This assumption, which is restrictive, will facilitate some of the derivations below. It will not be needed in the alternative approach given in Section 5. Let $E_u(\cdot)$ signify expectation without respect to Q_n^u measure. Then

$$E_u(\epsilon_k^u | \mathcal{F}_{k-1}) = 0$$

and so $N_n(u)$ is a Q_n^u–martingale. Note also that, by the discrete time Girsanov transformation theorem, the P_n^0–martingale N_n^0 is transformed into a Q_n^u–martingale by the mapping

$$N_n^0 \to N_n^0 - \Sigma_{k=1}^n E(\epsilon_k^0 L_k^u / L_{k-1}^u | \mathcal{F}_{k-1}) = \bar{N}_n(u)\ , \text{ say.}$$

Using (74) and (76) we write

$$\begin{aligned} L_k^u &= L_k(\theta(u)) = \exp\{\Lambda_{\theta(u)k}\} = \exp\{\textstyle\int_0^u N_k(v) d\theta(v)\} \\ &= L_{k-1}^u \exp\{\textstyle\int_0^u \epsilon_k^v d\theta(v)\}\ . \end{aligned}$$

Hence,

$$\begin{aligned} \bar{N}_n(u) &= N_n^0 - \Sigma_{k=1}^n E[\epsilon_k^0 \exp\{\textstyle\int_0^u \epsilon_k^v d\theta(v)\} | \mathcal{F}_{k-1}] \\ &= \Sigma_{k=1}^n \left\{ \epsilon_k^0 - E[\epsilon_k^0 \exp\{\textstyle\int_0^u \epsilon_k^v d\theta(v)\} | \mathcal{F}_{k-1}] \right\} \\ &= \Sigma_{k=1}^n \bar{\epsilon}_k^u\ , \text{ say.} \end{aligned}$$

Now $E_u(\bar{\epsilon}_k^u | \mathcal{F}_{k-1}) = 0$ and $\bar{N}_n(u)$ is a Q_n^u–martingale. It is, in fact, an approximation to the Q_n^u–martingale $N_n(u)$ given in (79).

To see this let $\eta_k^u = \eta_k(\theta(u))$, $\delta = \theta(u) - \theta^0$ and note that

$$\epsilon_k^v = \epsilon_k^0 + \eta_k^0(\theta(v) - \theta^0) + o_p(\delta)\ .$$

Then, we have to the given order of approximation

$$\begin{aligned} \bar{N}_n(u) &= N_n^0 - \Sigma_{k=1}^n E[\epsilon_k^0\{1 + \epsilon_k^0(\theta(u) - \theta^0) + o_p(\delta)\} | \mathcal{F}_{k-1}] \\ &= N_n^0 - \Sigma_{k=1}^n E((\epsilon_k^0)^2 | \mathcal{F}_{k-1})(\theta(u) - \theta^0) + o_p(\delta)\} \\ &= N_n^0 - I_n(\theta^0)(\theta(u) - \theta^0) + o_p(\delta)\} \\ &= N_n^0 - \langle N^0, N^0 \rangle_n (\theta(u) - \theta^0) + o_p(\delta)\ . \end{aligned} \tag{81}$$

Turning to (79) and expanding about $u = 0$ we have

$$N_n(u) = N_n^0 + J_n^0(\theta(u) - \theta^0) + o_p(\delta)\ , \tag{82}$$

which differs from (81) by a term that tends in probability to zero in view of (77). Thus, $\bar{N}_n(u)$ may be taken as a local approximation to $N_n(u)$ and both are Q_n^u—martingales.

Now let $\{[u_{i-1}, u_i] : i = 1, ..., I;\ u_0 = 0,\ u_I = u\}$ be a partition that decomposes the path $\theta(u)$ into I segments with maximum mesh size δ. In a similar fashion to (82) and (81), we have

$$\begin{aligned} N_n(u_i) &= N_n(u_{i-1}) + J_n(\theta(u_{i-1}))(\theta(u_i) - \theta(u_{i-1})) + o_p(\delta) \\ &\sim N_n(u_{i-1}) - \langle N(u_{i-1}), N(u_{i-1})\rangle_n(\theta(u_i) - \theta(u_{i-1})) + o_p(\delta) \end{aligned} \tag{83}$$

where $N_n(u_{i-1}) = \Sigma_{k=1}^n \epsilon_k^{u_{i-1}}$ is a $Q_n^{u_{i-1}}$—martingale and $\langle N(u_{i-1}), N(u_{i-1})\rangle_n = \Sigma_{k=1}^n E((\epsilon_k^{u_{i-1}})^2 \mid \mathcal{F}_{k-1})$ is its conditional quadratic variation process. Expression (83) leads directly to the rolling quadratic approximation to the log—likelihood $\Lambda_{\theta n}$ given in the following result.

THEOREM 4.1

$$\Lambda_{\theta n} \sim \Sigma_{i=1}^I \left\{ N_n^{u_i} - (1/2)\langle N^{u_i}, N^{u_i}\rangle_n \right\} \tag{84}$$

where

$$N_n^{u_i} = N_n(u_{i-1})(\theta(u_i) - \theta(u_{i-1}))$$

is a discrete $Q_n^{u_{i-1}}$*—martingale for each* $i = 1, ..., I$. □

Now consider a grid with $I = 2$. Set $\theta = \theta(u_2)$ and select u_0 and u_1 so that $\hat{\theta}_n = \theta(u_1)$ is the maximum likelihood estimate and $\tilde{\theta} = \theta(u_0)$ is some preliminary consistent estimate. Note that with the initialization $\tilde{\theta}$ (in place of $\theta^0 = \theta(u_0)$), the reference measure is $\tilde{P}_n = P_n^{\tilde{\theta}}$ and the likelihood function is

$$L_n(\theta) = dP_n^\theta / d\tilde{P}_n\ . \tag{85}$$

Since the score at $\hat{\theta}_n$ is zero, we have from (83)

$$0 = N_n(u_1) = N_n(u_0) - <N(u_0), N(u_0)>_n(\hat{\theta}_n - \tilde{\theta}) + o_p(\delta) .$$

This leads to the following approximate representation of the MLE

$$\hat{\theta}_n = \tilde{\theta} + <\tilde{N},\tilde{N}>_n^{-1}\tilde{N}_n ,$$

where $\tilde{N}_n = N(u_0)$. In this form, $\hat{\theta}_n$ is a linearised MLE constructed from the initial estimate $\tilde{\theta}$.

Now apply Theorem 4.1 with the same grid and set $\hat{h}_n = \hat{\theta}_n - \tilde{\theta}$, $h = \theta - \tilde{\theta}$ and rename $\Lambda_{\theta n}$ as Λ_{hn} in this new parametrization. We obtain from (84)

$$\Lambda_{hn} = \tilde{N}_n\hat{h}_n - (1/2)<\tilde{N},\tilde{N}>_n\hat{h}_n^2 - (1/2)<\hat{N},\hat{N}>_n(h - \hat{h}_n)^2 , \tag{86}$$

where

$$\hat{N}_n = \Sigma_{k=1}^n \hat{\epsilon}_k = \Sigma_{k=1}^n \epsilon_k(\hat{\theta}_n) = 0 ,$$

and

$$<\hat{N},\hat{N}>_n = \Sigma_{k=1}^n E(\hat{\epsilon}_k^2 | \mathcal{F}_{k-1}) .$$

Note that, in view of (83), we have

$$<\hat{N},\hat{N}>_n = <\tilde{N},\tilde{N}>_n + O_p(\delta^2) . \tag{87}$$

We deduce the following approximate form of the likelihood function (85) when it is reparameterized as $L_n(h) = dP_n^h/d\tilde{P}_n$ with $h = \theta - \tilde{\theta}$, viz.

$$\begin{aligned} L_n(h) &= \exp\{\tilde{N}_n\hat{h}_n - (1/2)\hat{h}_n^2<\tilde{N},\tilde{N}>_n\}\exp\{-(1/2)(h - \hat{h}_n)^2<\hat{N},\hat{N}>_n \} \\ &= \exp\{(1/2)\hat{h}_n^2\tilde{A}_n\}\exp\{-(1/2)(h - \hat{h}_n)^2\hat{A}_n\} , \end{aligned} \tag{88}$$

where

$$\tilde{A}_n = <\tilde{N},\tilde{N}>_n , \quad \hat{A}_n = <\hat{N},\hat{N}>_n .$$

The approximation (88) is a local approximation to the likelihood in the vicinity of the MLE $\hat{h}_n$.

4.2. Bayesian Inference

Let $\pi(h)$ be a prior density for h, which may be implied by a corresponding prior for θ. The posterior density process is obtained by combining this prior with the approximate likelihood function (88) giving

$$\Pi_n = \pi(h)L_n(h) = \pi(h)\exp\{(1/2)\hat{h}_n^2\tilde{A}_n\}\exp\{-(1/2)(h-\hat{h}_n)^2\hat{A}_n\} .$$

Using (87) we may write this process as

$$\Pi_n \sim [\tilde{A}_n^{-1/2}\exp\{(1/2)\hat{h}_n^2\tilde{A}_n\}][\pi(h)\hat{A}_n^{1/2}\exp\{-(1/2)(h-\hat{h}_n)^2\hat{A}_n\}] \tag{89}$$

$$\sim [\hat{A}_n^{-1/2}\exp\{(1/2)\hat{h}_n^2\hat{A}_n\}][\pi(h)\hat{A}_n^{1/2}\exp\{-(1/2)(h-\hat{h}_n)^2\hat{A}_n\}] . \tag{90}$$

<u>LEMMA 4.2</u>

$$\begin{aligned}(1/2)\{\hat{h}_{n+1}^2\tilde{A}_{n+1} - \ln(\tilde{A}_{n+1})\} &= (1/2)\{\hat{h}_n^2\tilde{A}_n - \ln(\tilde{A}_n)\} \\ &+ \{\hat{h}_n\tilde{\epsilon}_{n+1} - (1/2)\hat{h}_n^2 E(\tilde{\epsilon}_{n+1}^2|\mathcal{F}_{n+1})\} + O_p([\tilde{\epsilon}_{n+1}^2 + E(\tilde{\epsilon}_{n+1}^2|\mathcal{F}_n)]/\tilde{A}_n) ,\end{aligned} \tag{91}$$

$$\begin{aligned}(1/2)\{\hat{h}_{n+1}^2\tilde{A}_{n+1} - \ln(\tilde{A}_{n+1})\} &= (1/2)\{\hat{h}_{n_0}^2\tilde{A}_{n_0} - \ln(\tilde{A}_{n_0})\} \\ &+ \Sigma_{s=n_0}^{n}\{\hat{h}_s\tilde{\epsilon}_{s+1} - (1/2)\hat{h}_s^2 E(\tilde{\epsilon}_{s+1}^2|\mathcal{F}_s)\} + O_p(\Sigma_{n_0+1}^{n+1}[\tilde{\epsilon}_s^2 + E(\tilde{\epsilon}_s^2|\mathcal{F}_s)]/\tilde{A}_{n_0}) .\end{aligned} \tag{92}$$

As in the discrete AR(1) case we now let

$$c_0 = \tilde{A}_{n_0}^{-1/2}\exp\{(1/2)\hat{h}_{n_0}^2\tilde{A}_{n_0}\} \tag{93}$$

be an initialization of the process in the first square bracket factor of (89). We may determine n_0 as the stopping time for which there is minimal information in $\tilde{A}_k$, i.e.

$$n_0 = \min\{k : \tilde{A}_k \geq c\}$$

where $c > 0$ is a preassigned constant.

Using (89), (92) and (93) we find that the posterior density process is approximately

$$\Pi_n \sim \left[c_0 \exp\{\Sigma_{n_0}^{n-1}[\hat{h}_s\tilde{\epsilon}_{s+1} - (1/2)\hat{h}_s^2 E(\tilde{\epsilon}_{s+1}^2|\mathcal{F}_s)]\}\right]\left[\pi(h)\hat{A}_n^{1/2}\exp\{-(1/2)(h-\hat{h}_n)^2\hat{A}_n\}\right] . \tag{94}$$

LEMMA 4.3

$$R_n = \exp\left\{\Sigma_{n_0}^{n-1}[\hat{h}_s\tilde{\epsilon}_{s+1} + (1/2)\hat{h}_s^2\tilde{E}(\tilde{\epsilon}_{s+1}^2|\mathcal{F}_s)]\right\}$$

$$=\exp\{G_n - (1/2)<G,G>_n\} \tag{95}$$

where $G_n = \Sigma_{n_0}^{n-1}\hat{h}_s\tilde{\epsilon}_{s+1}$ *is a discrete* $\tilde{P}_n$*—martingale and where* $\tilde{E}(\cdot)$ *is expectation taken with respect to* $\tilde{P}_n$ *measure.* □

The process R_n defines an approximate change of measure from the reference measure $\tilde{P}_n$ to the new measure Q_n^B given by

$$dQ_n^B/d\tilde{P}_n = R_n = \exp\{G_n - (1/2)<G,G>_n\} . \tag{96}$$

Using (95) and (96) in (94) we have

$$\Pi_n = \pi(h)(dP_n^h/d\tilde{P}_n) \sim c_0(dQ_n^B/d\tilde{P}_n)[\pi(h)\tilde{A}_n^{1/2}\exp\{-(1/2)(h-\hat{h}_n)^2\hat{A}_n\}] .$$

The posterior density process with respect to Q_n^B—measure is then

$$\Pi_n^B = \pi(h)(dP_n^h/dQ_n^B)$$

$$\sim c_0\pi(h)\hat{A}_n^{1/2}\exp\{-(1/2)(h-\hat{h}_n)^2\hat{A}_n\} . \tag{97}$$

We now give the main result on Bayesian inference for the case of a general likelihood.

THEOREM 4.4

(a) *Under a uniform prior* $\pi(h)$, Π_n^B *is approximately Gaussian with distribution* $N(\hat{h}_n, \hat{A}_n^{-1})$ *at time* n .

(b) *Bayes methods imply the use of a new reference measure* Q_n^B *in constructing the likelihood.*

(c) *The model for the data that is implied by* Q_n^B *has a likelihood function (with respect to the reference measure* $\tilde{P}_n$ *) that is obtained by the recursive updating formula*

$$R_n = R_{n-1}\exp\{\hat{h}_{n-1}\tilde{\epsilon}_n - (1/2)\hat{h}_n^2E(\tilde{\epsilon}_n^2|\mathcal{F}_{n-1})\} \tag{98}$$

that relies on the MLE, $\hat{h}_{n-1}$ *, from the first* $n-1$ *observations and the increment in the score* $\tilde{\epsilon}_n$ *arising from the latest observation.* □

Following the terminology of the previous two sections, Q_n^B may be called the *Bayes model measure*. The *Bayes model* here is again a trajectory dependent one and is defined recursively through the formula (98) that updates the likelihood as new observations arrive. In (98) $\hat{h}_{n-1}$ provides the present estimate of the parameter change (viz. $\hat{\theta}_{n-1} - \tilde{\theta}$) while $\tilde{\epsilon}_n$ gives the change in the score with the latest data, so that the product $\hat{h}_{n-1}\tilde{\epsilon}_n$ provides the increment in the likelihood. This is compensated by the quadratic term in the exponent (viz. $-(1/2)\hat{h}_n^2 E(\tilde{\epsilon}_n^2 | \mathcal{F}_{n-1})$) which keeps the result close to a density process. The resulting process, R_n, is here only approximately a $\tilde{P}_n$—martingale, in contrast to the corresponding process in the Gaussian AR(1) model of Sections 3.1—3.2 where R_n is precisely a P_n—martingale (cf. Lemma 3.5). Nevertheless, the implications of the replacement of the measure remain the same: viz. (i) that Bayesian inference relates to a different model concept from classical inference; and (ii) that the Bayesian model has a time varying parameter that evolves according to the path of the recursive MLE computed from the given data trajectory.

5. AN ALTERNATIVE APPROACH TO A GENERAL THEORY

In dealing with the case of a general likelihood function, the approach adopted in the previous section relies on discrete time martingales. Use of the discrete Girsanov transformation necessarily involves the conditional quadratic variation process, whose existence puts moment restrictions on the score function process. In an asymptotic theory we would expect that strong moment restrictions such as (80) are unnecessary. It is therefore of interest to explore ways in which the likelihood approximations derived in Section 4.1 can be developed under less restrictive conditions.

This section will look at a different approach to a general theory that presents large sample approximations in terms of continuous L_2 martingales. Our starting point is related to recent work by Strasser (1986) and Jacob and Shiryaev (1987, Ch. X) on the limiting form of likelihood ratios. We combine the limiting likelihood density process with a prior to induce the posterior density process. Then, following the analysis of earlier sections, we derive the implied limiting Bayes model and limiting Bayes model measure under which Bayesian inference takes place.

5.1. The Limiting Likelihood Density Process

We start with the following representation for the likelihood

$$L_n(\theta) = \exp\{\Lambda_{n\theta}\} = \exp\{\int_0^r N_n(u)d\theta(u)\} \tag{99}$$

based on (74) and (78). We shall now consider a local quadratic approximation to the log–likelihood in the neighborhood of $\theta = \theta^0$. This will lead to a simple and recognizable form of the limiting density process.

Let θ_n be a sequence of parameter values for θ with $\theta_n \to \theta^0$ and define the deviation $h_n = \theta_n - \theta^0$. Recall that at $u = 0$ the score process given in (99), viz.

$$N_n^0 = N_n(0) = \Sigma_{k=1}^n \epsilon_k^0 ,$$

is a P_n^0–martingale. We define

$$a_n = E(I_n^0) = E\{\Sigma_{k=1}^n E((\epsilon_k^0)^2 | \mathcal{F}_{n-1})\} .$$

In the usual asymptotic theory for stochastic processes, a_n serves as a normalization factor for the martingale N^0 (e.g. Hall and Heyde, 1980, p. 160, Proposition 6.1). Following this approach, we may let

$$h_n = h/a_n^{1/2} , \quad h > 0 \tag{100}$$

so that θ_n is, in effect, a sequence in Θ that is local to θ^0 in the traditional sense. Corresponding to the path $\{\theta_n(u) : u \in [0,r]\}$ from $\theta_n(0) = \theta^0$ to $\theta_n(r) = \theta_n$, we now have the induced path $\{h(u) : u \in [0,r]\}$ with $h(0) = 0$, and $h = h(r) = a_n^{1/2}(\theta_n - \theta^0)$.

In addition to (76) and (77) we shall assume that

$$I_n(\theta_n)/I_n^0 \to 1 \text{ a.s. } (P^0) \tag{101}$$

(cf. Hall and Heyde, 1980, p. 160).

Next let s, $0 \leq s \leq 1$, denote an arbitrary fraction of the total sample n. Then, working from (78), we can write the likelihood process as

$$L_{[ns]}^h = L_{[ns]}(\theta_n) = \exp\{\int_0^r N_{[ns]}(u)d\theta_n(u)\} = \exp\{\int_0^r a_n^{-1/2} N_{[ns]}(u)dh(u)\} \tag{102}$$

in terms of the random element $N_{[ns]}(u)$. To proceed with the asymptotics we assume that $N_{[ns]}$ converges weakly in the space $D[0,1]$ after suitable standardization. In this latter respect, we may construct the localizing sequence h_n in (100) as is required for this purpose. Specifically, we shall assume that we have joint weak convergence, as $n \to \infty$, of the process $N^0_{[ns]} = N_{[ns]}(0)$ and its conditional quadratic variation, i.e.

$$(103) \qquad (a_n^{-1/2} N^0_{[ns]}, a_n^{-1} \langle N^0, N^0 \rangle_{[ns]}) \Rightarrow (T^0_s, \langle T^0, T^0 \rangle_s)$$

In (103), $N^0_{[ns]}$ and the limit process T^0_s are defined on a probability space $(\Omega, \mathcal{F}, P^0)$ where $\cup_1^\infty \mathcal{F}_n \subset \mathcal{F}$ and $P^0_n = E(P^0 | \mathcal{F}_n)$. We shall assume that the limit process T^0_s is a continuous $L_2(P^0)$ integrable process and a P^0—martingale, just as its finite sample counterpart N^0_n is a P^0_n—martingale. We call T^0_s the *limiting score process.* Since T^0_s is a continuous martingale (by assumption) we have $\langle T,T \rangle_s = [T,T]_s$.

The following theorem describes the limiting behavior of the likelihood function $L^h_{[ns]}$ given in (102).

<u>THEOREM 5.1</u>

$$(104) \qquad a_n^{-1/2} N_{[ns]}(u) \Rightarrow T_s(u) = T^0_s - [T^0, T^0]_s h(u) ,$$

$$(105) \qquad L^h_{[ns]} \Rightarrow \mathcal{L}^h_s = \exp\left\{\int_0^r T_s(u) dh(u)\right\} = \exp\left\{h T^0_s - (1/2) h^2 [T^0, T^0]_s\right\} .$$

$\mathcal{L}^h_s$ is a P^0—martingale with $E(\mathcal{L}^h_s) = 1$. □

<u>COROLLARY 5.2</u>. *Let P^h_s be the measure defined by the limiting density process $\mathcal{L}^h_s$, i.e. define P^h_s by the RN derivative*

$$dP^h_s / dP_s = \mathcal{L}^h_s$$

Under the change of measure $P^0_s \to P^h_s$, the limit process $T_s(u)$ in (104) becomes a P^h_s—martingale. □

Suppose we now select $u = \hat{u}$ such that

$$\hat{h}_n = h_n(\hat{u}) = a_n^{1/2} (\hat{\theta}_n - \theta^0) ,$$

where $\hat{\theta}_n$ is the MLE of θ^0. Setting $s = 1$ so that the full sample is used, we have

$$\hat{h}_n \Rightarrow \hat{h} = \operatorname{argmax}(\mathcal{L}_1^h) = T_1^0/[T^0, T^0]_1 .$$

As an illustration, take the case of the AR(1) (31) with $\alpha^0 = 1$. Then

$$T_s^0 = \int_0^s W dW , \ [T^0, T^0]_s = \int_0^s W^2 ,$$

and

$$\hat{h}_n = n(\hat{\alpha}_n - \alpha^0) \Rightarrow \int_0^1 W dW / \int_0^1 W^2 ,$$

where $W(\cdot) \equiv BM(1)$.

As a further illustration, suppose we have a sequence of models like (31) with $\alpha_n^0 = \exp(-c/n)$ for some constant c, so that the autoregressive root is local to unity. Then,

$$T_s^0 = \int_0^s J_c dW , \ [T^0, T^0]_s = \int_0^s J_c^2 ,$$

where $J_c(\cdot)$ is an Ornstein—Uhlenbeck process satisfying the linear stochastic differential equation

$$dJ_c(r) = cJ_c(r)dr + dW(r) .$$

We then have the limit theory

$$\hat{h}_n = n(\hat{\alpha}_n - \alpha^0) \Rightarrow \int_0^1 J_c dW / \int_0^1 J_c^2$$

(cf. Phillips, 1988).

We now return to the general form of the limit density process $\mathcal{L}_s^h$ given in (105). Define

$$\hat{h}(s) = h(\hat{u}) = \operatorname{argmax}(\mathcal{L}_s^h) = T_s^0/[T^0, T^0]_s . \tag{106}$$

Note that at the point $\hat{u}$ we have

$$\hat{T}_s = T_s(\hat{u}) = T_s^0 - \hat{h}(s)[T^0, T^0]_s = 0 , \tag{107}$$

i.e. a zero value for the limiting score. We decompose the path integral of the score process in (105) as

$$\int_0^r T_s(u)dh(u) = \int_0^{\hat{u}} T_s(u)dh(u) + \int_{\hat{u}}^r T_s(u)dh(u) .$$

Now

$$\int_0^{\hat{u}} T_s(u)dh(u) = \hat{h}(s)T_s^0 - (1/2)\hat{h}(s)^2[T^0, T^0]_s , \tag{108}$$

and writing, for $u \in [\hat{u}, r]$,

$$T_s(u) = \hat{T}_s - [\hat{T},\hat{T}]_s(h(u) - \hat{h}(s)) = -[\hat{T},\hat{T}]_s(h(u) - \hat{h}(s)) ,$$

we have

$$(109) \quad \int_{\hat{u}}^{r} T_s(u)dh(u) = -[\hat{T},\hat{T}]_s \int_{\hat{u}}^{r}(h(u) - \hat{h}(s))dh(u) = -[\hat{T},\hat{T}]_s\{(1/2)(h^2 - \hat{h}(s)^2) - \hat{h}(s)(h - \hat{h}(s))\} ,$$
$$= -[\hat{T},\hat{T}]_s(1/2)(h - \hat{h}(s))^2 .$$

Combining (106), (108) and (109) we obtain

$$(110) \qquad \int_0^r T_s(u)dh(u) = (1/2)\hat{h}(s)^2[T^0, T^0]_s - (1/2)(h - \hat{h}(s))^2[\hat{T},\hat{T}]_s$$

which leads directly to the following result.

<u>THEOREM 5.3</u>. *The limiting process* $\mathcal{L}_s^h$ *may be expressed in the form*

$$(111) \qquad \mathcal{L}_s^h = \exp\{(1/2)\hat{h}(s)^2[T^0, T^0]_s\}\exp\{-(1/2)(h - \hat{h}(s))^2[\hat{T},\hat{T}]_s\}$$

where $\hat{h}(s)$ *is the process given in* (106) *that represents the limiting form of the standardized and centered* MLE. □

5.2. The Posterior Density Process

Based on (102) and constructed from the fraction s , $0 < s \leq 1$, of the total sample n , the posterior density process is

$$\Pi_{[ns]} = \pi(h)L_{[ns]}^h .$$

Using (105) we have

$$\Pi_{[ns]} \Rightarrow \Pi(s) = \pi(h)\mathcal{L}_s^h .$$

Next define

$$\mathcal{A}_s = [T^0, T^0]_s ,$$

and then, from (111), we have the following alternate form for $\Pi(s)$, viz.

$$\Pi(s) = [\mathcal{I}_s^{-1/2}\exp\{(1/2)\hat{h}(s)^2\mathcal{I}_s\}][\Pi(h)\mathcal{I}_s^{1/2}\exp\{-(1/2)(h-\hat{h}(s))\mathcal{I}_s\}] \;. \tag{112}$$

Let

$$\mathcal{R}(s) = \mathcal{I}_s^{-1/2}\exp\{(1/2)\hat{h}(s)^2\mathcal{I}_s\} \;, \text{ and } r_s = \mathcal{R}(s)/\mathcal{R}(s_0)$$

where $s_0 > 0$ is some given initialization of the process $R(s)$. We note the properties of these processes in the following lemma.

<u>LEMMA 5.4</u>

$$E(\mathcal{R}(s)\,|\,\mathcal{F}_t) = \mathcal{R}(t) \;, \quad s > t > 0 \tag{113}$$

$$r_s = 1 + \int_{s_0}^{s} r_t dg_t = \exp\{g_s - (1/2)[g,g]_s\} \tag{114}$$

where $g_s = \int_{s_0}^{s} \hat{h}(t)dT_t^0$ *and* r_s *are* P^0*—martingales and*

$$E(r_s) = 1 \;. \;\square \tag{115}$$

The limit process r_s is a density process. It defines a change in measure from P_s^0 to a measure Q_s^B given by the RN derivative

$$dQ_s^B/dP_s^0 = r_s = \exp\{g_s - (1/2)[g,g]_s\} \;. \tag{116}$$

The limiting posterior density process is then

$$\Pi(s) = c_0(dQ_s^B/dP_s^0)[\pi(h)\mathcal{I}_s^{1/2}\exp\{-(1/2)(h-\hat{h}(s))\mathcal{I}_s\}] \;,$$

where $c_0 = \mathcal{R}(s_0)$. Changing the measure by which $\Pi(s)$ is defined we have

$$\begin{aligned}\Pi^B(s) &= \pi(h)(dP_s^h/dQ_s^B) = \pi(h)(dP_s^h/dP_s^0)(dP_s^0/dQ_s^B)\\ &= c_0\pi(h)\mathcal{I}_s^{1/2}\exp\{-(1/2)(h-\hat{h}(s))\mathcal{I}_s\} \;. \end{aligned} \tag{117}$$

We deduce the following analogue of Theorem 2.5 for the limiting posterior density $\Pi^B(s)$.

THEOREM 5.5

(a) *Under a uniform prior the limiting posterior density process* $\Pi^B(s)$ *is Gaussian with marginal distribution* $N(\hat{h}(s), \mathcal{I}_s^{-1})$ *centered at the limiting* MLE $\hat{h}(s)$.

(b) *Bayes methods involve a reference measure given by the limiting Bayes model measure* Q_s^B *defined by* (116).

(c) *The model for the data implied by* Q_s^B *has limiting likelihood density process* (114) *which satisfies the stochastic differential equation*

$$dr_s = r_s \hat{h}(s) dT_s^0 , \tag{118}$$

according to which the likelihood is updated from the limiting MLE $\hat{h}(s)$ *and the increment in the score process* T_s^0 . □

Note that the major difference between this case and that of the continuous time Gaussian AR(1) considered in Section 2 is that the updating equation (118) above relates to the density process r_s that defines the limiting Bayes model measure. (In this sense, Theorem 5.4 is a limiting process version of Theorem 4.4.) However, for the Gaussian AR(1) in Theorem 2.5 we were able to deduce the nonlinear stochastic differential equation, viz. equation $(1)^B$, to which the Bayes model measure directly relates. In other words, an explicit Bayes model was available in that special case. In the general case, we are forced to deal with models in terms of their associated likelihood functions or density processes. We cannot, therefore, be any more explicit about the implied Bayes model than to give the differential equation (118) by which the density process that defines the Bayes model measure is updated.

6. CONCLUSION

This paper puts forward the idea that Bayesian modeling of time series involves a special frame of reference, 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 do not. This helps to explain the poor sampling properties of 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. 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 Jeffreys—type prior.

This paper is a beginning. Our main concerns have been: (i) the conceptual framework that the Bayes frame of reference implies; and (ii) the practical import of the new frame of reference in modeling and in statistical tests. Later work will attempt to extend our treatment to more general models and to illustrate the use of our methods in empirical work.

7. APPENDIX

PROOF OF LEMMA 2.1. Write

$$L_s = dQ_s^u/dP_s = dP_s^{\beta(u)}/dP_s = \exp\{\beta(u)V_s - (1/2)\beta(u)^2 A_s\} .$$

Taking stochastic differentials and employing Ito's rule, we have

$$\begin{aligned} dL_s &= L_s\{dV_s\beta(u) - (1/2)\beta(u)^2 dA_s\} + (1/2)L_s\beta(u)^2 dA_s \\ &= L_s dV_s \beta(u) . \end{aligned}$$

The quadratic covariation differential is now

$$d[L,V]_s = L_s dA_s \beta(u) .$$

Under the change of measure $P_t \to Q_t^u$ and using the Girsanov theorem we have the mapping

$$\text{(A1)} \qquad V_t \to V_t - \int_0^t (1/L_s) d[L,V]_s$$

from the P_t—martingale V_t to the Q_t^u—martingale

$$V_t - \int_0^t \beta(u) dA_s = V_t - \beta(u)A_t = N_t(u) .$$

Hence, $N_t(u)$ is a Q_t^u—martingale as required.

PROOF OF LEMMA 2.2. Since $N_t(u) = V_t - \beta(u)A_t$ we have

$$\begin{aligned} \Lambda_{\beta(r)} &= \int_0^r N_t(u) d\beta(u) = \Sigma_{i=1}^I \int_{u_{i-1}}^{u_i} N_t(u) d\beta(u) \\ &= \Sigma_{i=1}^I \{V_t(\beta(u_i) - \beta(u_{i-1})) - (1/2)A_t(\beta(u_i)^2 - \beta(u_{i-1})^2)\} \\ &= \Sigma_{i=1}^I \{N_t(u_{i-1})(\beta(u_i) - \beta(u_{i-1})) + A_t\beta(u_{i-1})(\beta(u_i) - \beta(u_{i-1})) - (1/2)A_t(\beta(u_i)^2 - \beta(u_{i-1})^2)\} \\ &= \Sigma_{i=1}^I \{N_t(u_{i-1})(\beta(u_i) - \beta(u_{i-1})) - (1/2)A_t(\beta(u_i) - \beta(u_{i-1}))^2\} \\ &= \Sigma_{i=1}^I \left\{N_t^{u_i} - (1/2)\left[N^{u_i}, N^{u_i}\right]_t\right\} \end{aligned}$$

as stated in (13), since

$$\left[N^{u_i}, N^{u_i}\right]_t = (\beta(u_i) - \beta(u_{i-1}))^2[N,N]_t = (\beta(u_i) - \beta(u_{i-1}))^2[V,V]_t = (\beta(u_i) - \beta(u_{i-1}))^2 A_t \, .$$

To prove (15) we first note that the continuous $Q_t^{u_{i-1}}$–martingale $N_t(u_{i-1})$ can be represented under an appropriate time change as Brownian motion, viz.

$$N_t(u_{i-1}) = W_{[N,N]_t} \quad \text{a.s.}$$

(e.g. Protter, 1990, Theorem 41, p. 81).

Now

$$[N,N]_t = [V,V]_t = \int_0^t Y_s^2 ds$$

and the stopping time τ is defined so that

$$[N,N]_\tau = \int_0^\tau Y_s^2 ds = A \, ,$$

where $A > 0$ is fixed. It follows that under the stated time change

$$\xi_i = N_\tau(u_{i-1}) = W_A \equiv N(0,A) \, .$$

Moreover,

$$\left[N^{u_i}, N^{u_i}\right]_\tau = h_i^2[N,N]_\tau = h_i^2 A$$

and the required result now follows.

PROOF OF LEMMA 2.3. Taking stochastic differentials and employing Ito's rule, we have

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

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

(A3) $$d\hat{\beta}_t = -A_t^{-2} dA_t V_t + A_t^{-1} dV_t$$

and

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

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 (24).

PROOF OF LEMMA 2.4. From (23) and (24), we deduce that

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

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

$$= \exp\left\{G_t - (1/2)\int_{t_0}^t [G,G]_s\right\}$$

$$= \exp\left\{G_t - (1/2)[G,G]_t\right\} \ , \tag{A5}$$

as required for (25). By stochastic differentiation

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

$$= R_t dG_t \ ,$$

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 $G_{t_0} = 1$. This proves (26). Finally, note that when V_t is a martingale (i.e. when $\beta = 0$) we have $dG_t = \hat{\beta}_t dV_t$ and

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

so that G_t is also a martingale. It follows from (26) and Theorem 5.3 of Ikeda and Watanabe (1981) that

$$E(R_t) = 1 \ ,$$

as required.

PROOF OF THEOREM 2.5

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

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

(b) As explained in the argument that follows equations (30) of the text the posterior density is

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

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 data dependent and 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^B .

(c) From (28) we have the RN derivative

$$\begin{aligned} dQ_t^B/dP_t &= \exp\left\{\int_{t_0}^t [\hat{\beta}_s dV_s - (1/2)\hat{\beta}_s^2 dA_s]\right\} \\ &= \exp\left\{\int_{t_0}^t [\hat{\beta}_s Y_s dY_s - (1/2)\hat{\beta}_s^2 Y_s^2 ds]\right\} \end{aligned}$$

which is the likelihood ratio density process for the model

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

In other words, Q_t^B is the probability measure of the output process of this nonlinear stochastic differential equation in place of the linear model (1).

PROOF OF LEMMA 3.1

$$L_n = \exp(\Lambda_{hn}) = \exp\{h\Sigma_1^n Y_{t-1}\Delta Y_t - (1/2)h^2\Sigma_1^n Y_{t-1}^2\} .$$

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

$$<M,M>_n = h^2\Sigma_1^n Y_{t-1}^2 E((\Delta Y_t)^2 | \mathcal{F}_{t-1}) = h^2\Sigma_1^n Y_{t-1}^2$$

and expression (33) for L_n follows directly.

Setting $M_0 = 0$, (34) holds identically and

$$E(L_j/L_{j-1} | \mathcal{F}_{j-1}) = E\left[\exp\{hY_{j-1}u_j - (1/2)h^2Y^2_{j-1}\} | \mathcal{F}_{j-1}\right] = 1$$

so that $E(r_j | \mathcal{F}_{j-1}) = 0$. Thus, r_j is a martingale difference, L_n is a martingale and $E(L_n) = 1$.

PROOF OF LEMMA 3.2. Observe that

$$L_n = dQ^u_n/dP_n = \exp\{h(u)V_n - (1/2)h(u)^2A_n\}$$
$$= \exp\{U^u_n - (1/2)<U^u, U^u>_n\}$$

where $U^u_n = h(u)V_n$. By the Girsanov theorem for discrete time processes (e.g. Jacod and Shiryaev, 1987, Theorem 3.46, p. 165), the mapping

$$\text{(A6)} \qquad V_n \to V_n - <V, U^u>_n$$

takes the P_n–martingale V_n into the Q^u_n–martingale

$$V_n - <V,U>_n = V_n - h(u)<V,V>_n = V_n - h(u)A_n = N_n(u) .$$

Hence, $N_n(u)$ is a Q^u_n–martingale as required. Next note that

$$N_n(u) = \Sigma^n_1 Y_{t-1}(\Delta Y_t - h(u)Y_{t-1}) = \Sigma^n_1 Y_{t-1}(Y_t - (1 + h(u))Y_{t-1}) = \Sigma^n_1 Y_{t-1}v_t(u) , \text{ say.}$$

Under Q^u_n, $E(v_t(u) | \mathcal{F}_{t-1}) = 0$, so that $v_t(u)$ is a Q^u_n–martingale difference and thus, by definition

$$<N(u), N(u)>_n = \Sigma^n_1 Y^2_{t-1} E(v_t(u)^2 | \mathcal{F}_{t-1}) = \Sigma^n_1 Y^2_{t-1} = A_n .$$

PROOF OF LEMMA 3.3. The proof is identical to the proof of the first part of Lemma 2.2 upon replacement of square bracket processes with sharp bracket processes.

PROOF OF LEMMA 3.4

$$(1/2)A_{n+1}\hat{h}^2_{n+1} = (1/2)[A_n\hat{h}_n + Y_n\Delta Y_{n+1}]^2/(A_n + Y_n^2)$$
$$= (1/2)[A_n\hat{h}_n^2 + 2\hat{h}_n Y_n \Delta Y_{n+1} + Y_n^2(\Delta Y_{n+1})^2/A_n]\left[1 + Y_n^2/A_n\right]^{-1}$$
$$= (1/2)\hat{h}_n^2 A_n + \{\hat{h}_n Y_n \Delta Y_{n+1} - (1/2)\hat{h}_n^2 Y_n^2\} + O_p(Y_n^2/A_n)\ .$$

Also

$$(1/2)\ln(A_{n+1}) = (1/2)\ln[A_n(1 + Y_n^2/A_n)]$$
$$= (1/2)\ln A_n + O_p(Y_n^2/A_n)$$

and (41) follows. A similar decomposition yields (42).

PROOF OF LEMMA 3.5. Under H_1 we have

$$E(\hat{h}_s Y_s \Delta Y_{s+1} | \mathcal{F}_s) = E(\hat{h}_s Y_s u_{s+1} | \mathcal{F}_s) = 0$$

and G_n is a discrete martingale. Its conditional quadratic variation is

$$\langle G,G\rangle_n = \Sigma_{n_0}^{n-1} \hat{h}_s^2 Y_s^2 E(u_{s+1}^2 | \mathcal{F}_s) = \Sigma_{n_0}^{n} \hat{h}_s^2 Y_s^2$$

proving (44). Next

$$R_n = 1 + \Sigma_{j=n_0+1}^{n} R_{j-1}(R_j/R_{j-1} - 1)$$

and $E(R_j/R_{j-1} | \mathcal{F}_{j-1}) = 1$, so that R is a martingale with $E(R_n) = 1$.

PROOF OF THEOREM 3.6. Parts (a) and (b) follow directly from the forms of (48). To prove (c) note that from (44) we see that R_n is the likelihood function and Q_n^B is the probability measure for the model

$$\Delta Y_{s+1} = \hat{h}_s Y_s + u_{s+1}$$

i.e.

$$Y_{s+1} = (1 + \hat{h}_s)Y_s + u_{s+1}\ .$$

The stated result now follows.

PROOF OF THEOREM 3.7. Under H_1 we have

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

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

PROOF OF THEOREMS 3.8 AND 3.9. 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 4.1. From (78) and (83) we have

$$\begin{aligned} \Lambda_{\theta n} &= \int_0^u N_n(v) d\theta(v) \\ &= \Sigma_{i=1}^I \int_{u_{i-1}}^{u_i} N_n(v) d\theta(v) \\ &\quad - \Sigma_{i=1}^I \int_{u_{i-1}}^{u_i} [N_n(u_{i-1}) - \langle N(u_{i-1}), N(u_{i-1})\rangle_n (\theta(v) - \theta(u_{i-1}))] d\theta(v) \\ &= \Sigma_{i=1}^I \left[N_n(u_{i-1}) \int_{u_{i-1}}^{u_i} d\theta(v) - \langle N(u_{i-1}), N(u_{i-1})\rangle_n \left\{\int_{u_{i-1}}^{u_i} \theta(v) d\theta(v) - \theta(u_{i-1}) \int_{u_{i-1}}^{u_i} d\theta(v)\right\}\right] \\ &= \Sigma_{i=1}^I [N_n(u_{i-1})(\theta(u_i) - \theta(u_{i-1})) - (1/2)\langle N(u_{i-1}), N(u_{i-1})\rangle_n (\theta(u_i) - \theta(u_{i-1}))^2] \\ &= \Sigma_{i=1}^I \left[N_n^{u_i} - (1/2)\langle N^{u_i}, N^{u_i}\rangle_n\right] \end{aligned}$$

as required for (84).

PROOF OF LEMMA 4.2.. First,

$$\begin{aligned}\hat{h}^2_{n+1}\tilde{A}_{n+1} &= \tilde{N}^2_{n+1}/<\tilde{N},\tilde{N}>_{n+1} = \left[\Sigma_1^{n+1}\tilde{\epsilon}_k\right]^2/(\Sigma_1^{n+1}E(\tilde{\epsilon}_k^2|\mathcal{F}_{k-1}))\\ &= \left[\Sigma_1^n\tilde{\epsilon}_k + \tilde{\epsilon}_{n+1}\right]^2/\left[\Sigma_1^n E(\tilde{\epsilon}_k^2|\mathcal{F}_{k-1}) + E(\tilde{\epsilon}^2_{n+1}|\mathcal{F}_n)\right]\\ &= [\tilde{N}_n^2 + 2\tilde{N}_n\tilde{\epsilon}_{n+1} + \tilde{\epsilon}^2_{n+1}]/[\tilde{A}_n + E(\tilde{\epsilon}^2_{n+1}|\mathcal{F}_n)]\\ &= \hat{h}_n^2\tilde{A}_n + 2\hat{h}_n\tilde{\epsilon}_{n+1} - \hat{h}_n^2E(\tilde{\epsilon}^2_{n+1}|\mathcal{F}_n)\ .\end{aligned}$$

Next,

$$\begin{aligned}\ln(\tilde{A}_{n+1}) &= \ln[\tilde{A}_n(1 + E(\tilde{\epsilon}^2_{n+1}|\mathcal{F}_n)/\tilde{A}_n)]\\ &= \ln(\tilde{A}_n) + O_p(E(\tilde{\epsilon}^2_{n+1}|\mathcal{F}_n)/\tilde{A}_n)\ .\end{aligned}$$

Combining these expansions we have (91). Recursive calculations then lead to the decomposition (92).

PROOF OF LEMMA 4.3. Under $\tilde{P}_n$ measure we have

$$E(\tilde{\epsilon}_{s+1}|\mathcal{F}_s) = 0$$

so that $G_n = \Sigma_{n_0}^{n-1}\hat{h}_s\hat{\epsilon}_{s+1}$ is a $\tilde{P}_n$—martingale. Its conditional quadratic variation is

$$<G,G>_n = \Sigma_{n_0}^{n-1}\hat{h}_s^2E(\tilde{\epsilon}^2_{s+1}|\mathcal{F}_s)$$

giving (95).

PROOF OF THEOREM 5.1. First note that

$$\begin{aligned}a_n^{-1/2}N_{[ns]}(u) &= a_n^{-1/2}\Sigma_1^{[ns]}\epsilon_k^u\\ &= a_n^{-1/2}N^0_{[ns]} + a_n^{-1/2}J^*_{[ns]}[a_n^{-1/2}h(u)]\ ,\end{aligned}$$

where $J^*_{[ns]} = J_{[ns]}(\theta^*)$ for some θ^* on the line segment connecting θ_n and θ^0. Since $\theta_n \to \theta^0$ and in view of (77) we have

$$J^*_{[ns]}/I^*_{[ns]} \xrightarrow[p]{} -1\ ,$$

where $I^*_{[ns]} = I_{[ns]}(\theta^*)$. However, from (101) we also have

$$I^*_{[ns]}/I^0_{[ns]} \to 1 \text{ a.s. } (P^0) .$$

Hence,

$$\begin{aligned} a_n^{-1} J^*_{[ns]} &= a_n^{-1} I^*_{[ns]} (J^*_{[ns]}/I^*_{[ns]}) , \\ &= -a_n^{-1} I^0_{[ns]} + o_p(1) \\ &= -\langle a_n^{-1/2} N^0, a_n^{-1/2} N^0 \rangle_{[ns]} + o_p(1) \\ &\Rightarrow -\langle T^0, T^0 \rangle_s \end{aligned}$$

using (103). It follows that

$$a_n^{-1/2} N_{[ns]}(u) \Rightarrow T^0_s - \langle T^0, T^0 \rangle_s h(u)$$

and since T^0_s is continuous we have $\langle T^0, T^0 \rangle_s = [T^0, T^0]_s$, leading to the stated result (104).

Next from (102) and (104) we have

$$L^h_{[ns]} \Rightarrow \exp\left\{ \int_0^r T_s(u) dh(u) \right\}$$

where

$$T_s(u) = T^0_s - [T^0, T^0]_s h(u) .$$

Thus

$$\begin{aligned} \int_0^r T_s(u) dh(u) &= T^0_s h - [T^0, T^0]_s \int_0^r h(u) dh(u) \\ &= T^0_s h - (1/2)[T^0, T^0]_s h^2 \end{aligned}$$

where $h = h(r) - h(0) = h(r)$, as required for (105). As in the proof of Lemma 2.4 we have the stochastic differential

$$d\mathcal{L}^h_s = h\mathcal{L}^h_s dT^0_s$$

so that $E(d\mathcal{L}^h_s | \mathcal{F}_s) = h\mathcal{L}^h_s E(dT^0_s | \mathcal{F}_s) = 0$ since T^0_s is a P^0–martingale. Integrating and using the initialization $\mathcal{L}^h_0 = 1$ we have

$$\mathcal{L}^h_t - 1 = \int_0^t d\mathcal{L}^h_s = h \int_0^t \mathcal{L}^h_s dT^0_s ,$$

so that

$$E(\mathcal{L}_t^h) = 1 ,$$

and $\mathcal{L}_t^h$ is a limit density process as required.

PROOF OF COROLLARY 5.2. By the Girsanov theorem we have the mapping

$$T_s^0 \to T_s^0 - \int_0^s \left[\mathcal{L}_t^{h(u)}\right]^{-1} d[\mathcal{L}^{h(u)}, T^0]_t = T_s^u$$

from the P^0–martingale T_s^0 to a P_s^h–martingale T_s^u. But $d\mathcal{L}_t^{h(u)} = h(u)\mathcal{L}_t^{h(u)} dT_t^0$, so that

$$d[\mathcal{L}^{h(u)}, T^0]_t = h(u)\mathcal{L}_t^{h(u)} d[T^0, T^0]_t$$

and then

$$\begin{aligned} T_s^u &= T_s^0 - h(u)\int_0^s d[T^0, T^0]_t \\ &= T_s^0 - h(u)[T^0, T^0]_s = T_s(u) , \end{aligned}$$

giving the required result.

PROOF OF THEOREM 5.3. The stated result (111) follows immediately from (105) and (110).

PROOF OF LEMMA 5.4. It will be sufficient to show (114) and (115) since these latter results imply (113). However, the proofs of (114) and (115) are identical (after suitable translation of notation) to the proofs of Lemmas 2.3 and 2.4 above.

8. 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.

DeJong, D. N. and C. H. Whiteman (1989a). "Trends and cycles as unobserved components in US real GNP: A Bayesian perspective," *Proceedings of the American Statistical Association.*

DeJong, D. N. and C. H. Whiteman (1989b). "Trends and random walks in macroeconomic time series: A reconsideration based on the likelihood principle," Working Paper No. 89—4, Department of Economics, University of Iowa. To appear in *Journal of Monetary Economics*

DeJong, D. N. and C. H. Whiteman (1989c). "Temporal stability of dividends and stock prices: Evidence from the likelihood function," Working Paper No. 89—3, Department of Economics, University of Iowa.

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.

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

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 (1981). *Stochastic Differential Equations and Diffusion Processes.* 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.

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.

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

Litterman, R. B. (1984). "Forecasting with Bayesian vector autoregressions: Four years of experience," Federal Reserve Bank of Minneapolis.

Meyer, P. A. (1989). "A short presentation of stochastic calculus," in M. Emory, *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. (1988). "Optimal inference in cointegrated systems," Cowles Foundation Discussion Paper No. 866R, Yale University. To appear in *Econometrica*, 1991.

Phillips, P. C. B. (1991). "To criticize the critics: An objective Bayesian analysis of stochastic trends," Cowles Foundation Discussion Paper No. 950 (1990), to appear in *Journal of Applied Econometrics*.

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 Perspective*, 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* (forthcoming).

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 (1988). "Understanding unit rooters: A helicopter tour," Federal Reserve Bank of Minneapolis Institute for Empirical Macroeconomics, Discussion Paper No. 4. To appear in *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.

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.