J. R. Statist. Soc. B (1997)
59, *No.* 3, *pp.* 615–626

Some Properties and Generalizations of Non-negative Bayesian Time Series Models

By GARY K. GRUNWALD† and KAIS HAMZA and ROB J. HYNDMAN

University of Melbourne, Parkville, Australia — *Monash University, Clayton, Australia*

[Received January 1994. Final revision August 1996]

SUMMARY

We study the most basic Bayesian forecasting model for exponential family time series, the power steady model (PSM) of Smith, in terms of observable properties of one-step forecast distributions and sample paths. The PSM implies a constraint between location and spread of the forecast distribution. Including a scale parameter in the models does not always give an exact solution free of this problem, but it does suggest how to define related models free of the constraint. We define such a class of models which contains the PSM. We concentrate on the case where observations are non-negative. Probability theory and simulation show that under very mild conditions almost all sample paths of these models converge to some constant, making them unsuitable for modelling in many situations. The results apply more generally to non-negative models defined in terms of exponentially weighted moving averages. We use these and related results to motivate, define and apply very simple models based on directly specifying the forecast distributions.

Keywords: BAYESIAN FORECASTING; EXPONENTIAL FAMILY TIME SERIES; EXPONENTIALLY WEIGHTED MOVING AVERAGE; NON-NEGATIVE TIME SERIES; POISSON TIME SERIES; POWER STEADY MODEL

1. INTRODUCTION

Exponential family state space models have provided one of the most common approaches to modelling time series of counts, proportions, compositions and positive observations. Various researchers, including West *et al.* (1985), Harvey and Fernandes (1989), Attwell and Smith (1991) and Grunwald, Raftery and Guttorp (1993), have used state space models based on the power steady model (PSM) of Smith (1979) to describe non-stationary series. They have extended the model to include trends, covariates and time varying effects. In Section 2 we briefly summarize these Bayesian exponential family state space models and the PSM. Other forms of exponential family state space models which we do not study in this paper have been proposed by Kitagawa (1987), Fahrmeir (1992) and Kashiwagi and Yanagimoto (1992).

There has been some recent study of the PSM. Smith (1990) has described various aspects of state transition rules, and Grunwald, Guttorp and Raftery (1993) have given some general results for estimation and forecasting. In this paper we add to this discussion by studying the PSM and related models in terms of forecast distributions and properties of sample paths. These properties provide a good basis for study because they relate to identifiable properties of the data. For instance, Smith and Miller (1986) showed that the one-step forecast distributions $p(y_t|D_{t-1})$, $t = 1, \ldots, n$,

†*Address for correspondence*: 5521 East Dickenson Place, Denver, CO 80222, USA.

 0035–9246/97/59615

contain all the information that is verifiable from the data. (D_{t-1} describes the process history and is defined carefully in Section 2.1.) In the context of non-linear time series models, Ozaki (1993) has stated as a criterion that simulations produce sample paths similar to the data.

By examining the forecast distributions, it will be seen in Section 3 that the temporal structure of the PSM is very limited. The addition of a scale parameter may not provide a broader range of models using the exact Bayesian theory but does suggest larger classes of models, which we study. For these larger classes, which contain the PSM, we use stochastic processes theory to show in Section 4 that for non-negative series the process $\{Y_t\}$ describing the observed series has sample paths almost surely approaching a constant. This does not reflect the behaviour of series to which Bayesian time series models have been applied in past work. In fact, this result is more general than Bayesian time series models, and we define a class of models for non-negative series based on the exponentially weighted moving average (EWMA) that also have this property.

In Section 5, the form of the PSM forecast distribution is used to suggest how to define models with other temporal structures in terms of forecast distributions. As an example, an autoregressive model is stated and shown to be more appropriate for a series previously analysed with a Bayesian model. In Section 6 we give some discussion and conclusions about the uses of Bayesian and other related time series models.

2. BAYESIAN TIME SERIES MODELS

2.1. *Observation and Conjugate State Distributions*

For a time series y_t ($t = 1, 2, \ldots$) in sample space $\mathbf{Y}$ we consider an exponential family Bayesian time series model as follows. Assume an exponential family *observation density*

$$p(y_t|\theta_t) = \exp\{y_t\theta_t - b(\theta_t) + c(y_t)\} \tag{1}$$

where θ_t is an unobservable state variable in parameter space Θ. (We assume that y_t and θ_t are univariate, but most results hold more generally.) $b(\theta_t) = \log[\int_{\mathbf{Y}} \exp\{y_t\theta_t + c(y_t)\}\, \mathrm{d}y_t]$ normalizes density (1) to have unit probability, and $c(y_t)$ contains all terms not depending on θ_t.

The *state density* is defined as the conjugate of density (1):

$$p(\theta_t|D_s) = k(m_{t|s}, \sigma_{t|s}) \exp[\sigma_{t|s}\{m_{t|s}\theta_t - b(\theta_t)\}] \tag{2}$$

with parameters $m_{t|s}$ and $\sigma_{t|s}$. Here, D_s denotes the entire past observed history and is defined by $D_s = \{D_{s-1}, y_s\}$ with D_0 containing values of any estimated parameters, covariates or initializations. The constant $k(m_{t|s}, \sigma_{t|s})$ contains terms not involving θ_t and normalizes the density to unit probability. By the conjugate property, if the prior ($s = t - 1$) is assumed to be of this form, Bayes theorem gives a posterior (with $s = t$) of the same form, thus providing for simple recursive estimation of $\theta_t|D_t$.

2.2. *Forecast Distributions*

The one-step *forecast density* is

$$p(y_t|D_{t-1}) = \int_\Theta p(y_t|\theta_t, D_{t-1})\, p(\theta_t|D_{t-1})\, \mathrm{d}\theta_t \tag{3}$$

for each $t \geqslant 1$. (The models assume that θ_t contains all relevant information to time t, so $p(y_t|\theta_t, D_{t-1}) = p(y_t|\theta_t)$.) This density (3) has parameters $m_{t|t-1}$ and $\sigma_{t|t-1}$. It simply represents a compound distribution based on densities (1) and (2), and for many standard observation distributions gives standard forms; Poisson observation distributions give negative binomial forecast distributions, exponential observations give Pareto forecast distributions, and so on.

The forecast densities (3) play a fundamental role in time series analysis. They do not condition on the unobservable state θ_t and are therefore used for forecasting (hence the name). They are identifiable from observed data even though the joint state distributions may not be (Smith, 1990). The model is completely specified by these distributions through the likelihood

$$p(y_1, \ldots, y_n|D_0) = \prod_{t=1}^{n} p(y_t|D_{t-1}). \tag{4}$$

Smith and Miller (1986) give further discussion of these issues.

Under simple and general conditions, Grunwald, Guttorp and Raftery (1993) applied general Bayesian results of Diaconis and Ylvisaker (1985) to give the simple relationship for the mean of this compound distribution

$$\mathbb{E}(Y_t|D_{t-1}) = m_{t|t-1}. \tag{5}$$

This allows easy study and comparison of these models in terms of forecast distributions.

We denote the general form of the *compound forecast distribution* (3) parameterized in terms of $m_{t|t-1}$ and $\sigma_{t|t-1}$ by

$$Y_t|D_{t-1} \sim \mathrm{CFD}(m_{t|t-1}, \sigma_{t|t-1}). \tag{6}$$

(Throughout this paper we write all distributions in this parameterization, with the mean first, followed by any other parameters.)

2.3. *Temporal Structure*

The temporal structure of the model (the relationship between time $t-1$ and time t) enters through $m_{t|t-1}$ and $\sigma_{t|t-1}$, and for Bayesian time series models these are given by specifying $p(\theta_t|D_{t-1})$. The PSM specifies

$$p(\theta_t|D_{t-1}) \propto p(\theta_{t-1}|D_{t-1})^\gamma \tag{7}$$

with $0 < \gamma \leqslant 1$. This rule is often modified to include trends and covariates.

Applying Bayes theorem with expressions (1), (2) and (7) gives the simple recursions

$$m_{t+1|t} = \gamma m_{t|t-1} + (1-\gamma) y_t = \gamma^t m_{1|0} + (1-\gamma) \sum_{j=0}^{t-1} \gamma^j y_{t-j} \tag{8}$$

and

$$\sigma_{t+1|t} \rightarrow \gamma/(1-\gamma) \equiv \sigma(\gamma). \tag{9}$$

These are eventual forms as $t \rightarrow \infty$; see Grunwald, Guttorp and Raftery (1993) for details. We usually use these eventual forms because initialization effects and the non-stochastic evolution of $\sigma_{t+1|t}$ are unnecessary complications in studying the properties of interest.

The second equality in the recursion (8) for $m_{t+1|t}$ is found by direct expansion, and is an EWMA. Our later results will apply to models with this expression for the mean, but with other distributional assumptions.

2.4. *Examples*

Throughout the paper we study the following two examples.

2.4.1. *Example 1: Poisson observations*

Let $Y_t|\lambda_t \sim \text{Poisson}(\lambda_t)$ with $\theta_t = \log \lambda_t$ and observation and state sample spaces $\mathbf{Y} = \{0, 1, \ldots\}$ and $\mathbf{\Theta} = \mathbf{R}^1$. From densities (2) and (3), the one-step forecast distribution is negative binomial. ($Y \sim \text{NB}(m, \sigma)$ if

$$p(y) = \frac{\Gamma(\sigma m + y)}{\Gamma(y+1)\,\Gamma(\sigma m)}\, \sigma^{\sigma m}(1+\sigma)^{-(\sigma m + y)} \qquad \text{for } m > 0,\ \sigma > 0 \text{ and } y = 0, 1, \ldots.$$

As $\sigma \rightarrow \infty$, $Y \rightarrow \text{Poisson}(m)$.) For the PSM, $Y_{t+1}|D_t \sim \text{NB}\{m_{t+1|t}, \sigma(\gamma)\}$ with $\mathbb{E}(Y_{t+1}|D_t) = m_{t+1|t}$ (true in general by equation (5)) and

$$\text{var}(Y_{t+1}|D_t) = \frac{1+\sigma(\gamma)}{\sigma(\gamma)}\, m_{t+1|t}.$$

This distribution has as support the non-negative integers (as it should) and dispersion greater than the Poisson distribution.

2.4.2. *Example 2: exponential observations*

Let $Y_t|\lambda_t \sim \text{Exp}(1/\lambda_t)$ (exponentially distributed with mean λ_t) with $\theta_t = -\lambda_t$ and observation and state sample spaces $\mathbf{Y} = (0, \infty)$ and $\mathbf{\Theta} = (-\infty, 0)$. From densities (2) and (3), the one-step forecast distribution is of the Pareto form. ($Y \sim \text{Pareto}(m, \sigma)$ if

$$p(y) = \frac{(\sigma+1)(\sigma m)^{\sigma+1}}{(y+\sigma m)^{\sigma+2}} \qquad \text{for } m > 0,\ \sigma > 0 \text{ and } y > 0.$$

As $\sigma \rightarrow \infty$, $Y \rightarrow \text{Exp}(1/m)$.) For the PSM, $Y_{t+1}|D_t \sim \text{Pareto}\{m_{t+1|t}, \sigma(\gamma)\}$ with $\mathbb{E}(Y_{t+1}|D_t) = m_{t+1|t}$ and

$$\text{var}(Y_{t+1}|D_t) = \frac{\sigma(\gamma)+1}{\sigma(\gamma)-1}\, m_{t+1|t}^2 \qquad \text{for } \sigma(\gamma) > 1$$

and is infinite for $0 < \sigma(\gamma) \leqslant 1$. This distribution has as support $(0, \infty)$ (as it should) and dispersion greater than the exponential distribution.

3. SCALED POWER STEADY MODELS

We first show that the PSM as it stands has a constraint between the location and dispersion of the forecast distribution that can greatly limit the range of models. We examine the idea of adding a scale parameter to eliminate this restriction, and we define a more general class of models containing the PSM but free of the location–dispersion constraint. We illustrate with the Poisson and exponential cases in examples 1 and 2, but the results are more general.

3.1. *Class of Power Steady Models*

For given observation and state distributions, the eventual PSM one-step forecast distribution (6) is specified by $m_{t+1|t}$ and $\sigma(\gamma)$. Equations (8) and (9) show that there is a single model parameter, γ. As a result, the range of possible models is quite limited. Both the temporal characteristics of the model (how past information is used in adjusting the forecast mean through the recursion (8)) and the dispersion of the forecast distribution are controlled by γ. This has interesting consequences that can be seen most easily as $\gamma \rightarrow 1$ or 0.

(a) $\gamma \rightarrow 1$ gives $\sigma(\gamma) \rightarrow \infty$, $m_{t+1|t} \rightarrow m_{1|0}$ and $p(y_{t+1}|D_t)$ approaches the form of the observation density (1) with mean $m_{1|0}$. This can best be seen by noting that the state density (2) is unimodal for Poisson and exponential observations and so approaches a point mass as $\sigma(\gamma) \rightarrow \infty$, thus giving no compounding in density (3). This model describes a sequence of independent and identically distributed (IID) observations from density (1) with mean $m_{1|0}$.

(b) $\gamma \rightarrow 0$ gives $m_{t+1|t} \rightarrow y_t$. This could be described as a generalization of a Gaussian random walk. However, $\sigma(\gamma) \rightarrow 0$ so that $\mathrm{var}(Y_{t+1}|D_t) \rightarrow \infty$. Again, this can be seen most easily by noting that the state density approaches a constant, and for Θ unbounded this is not a proper distribution.

In these models the mean and variance are tied too closely to allow specification of models like a finite variance 'random walk' or a sequence of IID observations with greater variance than density (1).

3.2. *Scale Parameters*

West (1985) and West *et al.* (1985) have suggested the addition of a scale parameter $\phi > 0$ to the observation density (1), giving

$$p(y_t|\theta_t, \phi) = \exp[\phi\{y_t\theta_t - b(\theta_t)\} + c(y_t, \phi)]. \tag{10}$$

Formally applying Bayes theorem gives mean recursion identical with equation (8) but with the second recursion (9) replaced by

$$\sigma_{t+1|t} \rightarrow \phi\gamma/(1-\gamma) \equiv \sigma(\gamma, \phi). \tag{11}$$

Since varying ϕ over $(0, \infty)$ allows $\sigma(\gamma, \phi)$ to vary over $(0, \infty)$ for any given γ, this construction gives a wider range of models not subject to the constraint noted in Section 3.1. However, the distributions (10) may not exist. For instance, Jørgensen (1986) showed that no discrete distributions of the form (10) with support on the non-negative integers exist for $\phi \neq 1$. For discrete distributions, he gave the form

$$p(y_t|\theta_t, \phi) = \exp\{y_t\theta_t - \phi\, b(\theta_t) + c(y_t, \phi)\}. \tag{12}$$

Using this form may not give the broader class desired, as follows.

3.2.1. *Example 1 (continued)*

For Poisson observations, equation (12) reduces to $Y_t|\lambda_t, \phi \sim \text{Poisson}(\phi\lambda_t)$, and direct computation of forecast densities shows that any form of conjugate state density gives the PSM.

3.3. *Scaled Power Steady Models*

Even though the addition of a scale parameter ϕ as in the previous subsection may mean that some distributions do not exist, the recursions (8) and (11) suggest how to form one-step forecast distributions (and thus fully define the model via likelihood (4)) for a class of models containing the PSM but allowing separate control over the forecast mean and dispersion. For a given observation distribution (1), we simply define the forecast distribution to be

$$Y_{t+1}|D_t \sim \text{CFD}\{m_{t+1|t}, \sigma(\gamma, \phi)\} \tag{13}$$

where $m_{t+1|t}$ is given by recursion (8) and $\sigma(\gamma, \phi)$ is given by expression (11). Note that this definition does not involve the state θ_t.

An alternative parameterization is obtained by noting that the scale parameter ϕ can be viewed as a device allowing separate control over the mean and dispersion. Recognizing this, we could define the forecast distribution by ignoring ϕ and simply specifying $\sigma > 0$ in distribution (13) unrelated to γ and ϕ, giving

$$Y_{t+1}|D_t \sim \text{CFD}(m_{t+1|t}, \sigma) \tag{14}$$

where $m_{t+1|t}$ is again given by recursion (8) and $\sigma > 0$. We refer to either distribution (13) or (14) as a *scaled power steady model* (SPSM). This class contains the eventual PSM: let $\phi = 1$ in distribution (13) or $\sigma = \gamma/(1-\gamma)$ in distribution (14).

Diaconis and Ylvisaker (1985) showed that, in the conjugate exponential family setting described in Section 2, the forecast distributions $\text{CFD}(m, \sigma)$ exist and can be normalized to proper probability distributions whenever $\sigma > 0$ and $m \in \Psi$, the interior of the convex hull of the sample space of the observation distribution (1). Thus, the SPSM is well defined provided that $m_{1|0} \in \Psi$.

4. LIMITING MODEL PROPERTIES

We now turn to the asymptotic behaviour of the PSM and SPSM. (In fact the results hold much more generally, as we show.) As far as we know there has not been a study of the long-term behaviour of these models. This is an important issue because the Bayesian models were developed as generalizations of the Gaussian random walk observed with error, which is non-stationary. As a result, various work (West *et al.* (1985), Harvey and Fernandes (1989) and Grunwald, Raftery and Guttorp (1993) for instance) used non-Gaussian versions to describe non-stationary time series. However, as we show in this section, for many non-Gaussian models the behaviour of sample paths may not look very much like the data being modelled.

General results are not immediate because the process $\{Y_t\}$ is neither Markov nor a martingale when $0 < \gamma < 1$. Furthermore, as Harvey and Fernandes (1989) and Smith (1990) have pointed out, forecast distributions for more than one step ahead are not of the form (6) with parameters obtained by iterating recursions (8) and (9) but rather, for $h > 1$, $p(y_{t+h}|D_t)$ are obtained by integrating over $y_{t+1}, \ldots, y_{t+h-1}$ in expressions of the form (4). In general this is not a standard distribution.

4.1. *Non-negative Exponentially Weighted Moving Average Models*

Because the results of this section are of wider interest, we consider general assumptions on the distributional form of the forecast distribution. Let FD(m) denote a distribution defined on a subset of $[0, \infty)$ with mean m. FD(m) may depend on other parameters that do not vary in time (such as σ in the SPSM), but to simplify the notation we suppress this. Such parameters are contained in D_0, and we assume conditioning on D_0 throughout. We have now replaced the notation 'CFD' by 'FD' to indicate that the forecast distribution need not be obtained by compounding density (1) with density (2) but can be specified by any probability distribution on a subset of $[0, \infty)$ with mean m.

Let $\Phi(\lambda, m)$ be the generating function (or Laplace transform) of the forecast distribution FD(m) with mean m, i.e.

$$\Phi(\lambda, m) = \mathbb{E}[\exp\{-\lambda\xi(m)\}],$$

where $\xi(m) \sim \mathrm{FD}(m)$. We make the following assumptions on FD(m). There is a closed subinterval Δ of $[0, \infty)$, or $\Delta = [0, \infty)$, such that, for any $m \in \Delta$, $\xi(m) \in \Delta$ (almost surely), $\mathbb{E}\{\xi(m)\}$ exists and equals m, and $\Phi(\lambda, m)$ is continuous in m within Δ. The set $\Lambda \equiv \{m \in \Delta : \xi(m) = m \text{ (almost surely)}\}$ will also be important below. Since $\mathbb{E}\{\xi(m)\} = m$, this is the set of values m for which the distribution of $\xi(m)$ is degenerate and thus $\xi(m) = m$ (almost surely). Define the processes $\{M_t\}$ and $\{Y_t\}$ by

$$\begin{aligned} Y_{t+1}|D_t &\sim \mathrm{FD}(M_{t+1}), \\ M_{t+1} &= \gamma M_t + (1-\gamma) Y_t \end{aligned} \tag{15}$$

with $0 \leqslant \gamma \leqslant 1$. We now use M_t in place of $m_{t|t-1}$ to denote the forecast mean, emphasizing that it is a stochastic process related to recursion (8). Here, we assume that $M_0 \in \Delta$ (almost surely) and is integrable. This class clearly contains the non-negative PSM and SPSM. Since the recursion in expression (15) for M_t expands as in recursion (8) to an EWMA, we refer to these as *non-negative exponentially weighted moving average models*.

4.2. *Limiting Model Properties*

Proposition 1. Under the assumptions in Section 4.1, the process Y_t has constant mean: $\mathbb{E}(Y_t) = \mathbb{E}(M_0)$ for all $t \geqslant 1$.

Proof. Direct calculation using expression (15) shows that M_t is a martingale, and the constant mean property gives $\mathbb{E}(M_t) = \mathbb{E}(M_0)$ for $t \geqslant 1$. Expanding M_{t+1} in $\mathbb{E}(M_{t+1}) = \mathbb{E}(M_0)$ and solving gives $\mathbb{E}(Y_t) = \mathbb{E}(M_0)$. □

For $\gamma = 1$, it is easily seen that $M_t = M_0$ for all $t \geqslant 0$ so $\{Y_t\}$ is a sequence of IID random variables with distribution FD(M_0). Other values of γ are more interesting.

Theorem 1. For $0 \leqslant \gamma < 1$, both Y_t and M_t converge almost surely to an integrable random variable, say M_∞, and $M_\infty \in \Lambda$ (almost surely).

Proof. As in the proof of proposition 1, M_t is a non-negative martingale, so by the martingale limit theorem (see Shiryayev (1984), p. 476) it must converge almost surely to an integrable random variable, say M_∞. Since M_t is non-negative, so is M_∞. It follows from expression (15) that Y_t converges almost surely to the same limit M_∞.

Let Ψ_t denote the generating function of Y_t, so

$$\Psi_t(\lambda) = \mathbb{E}\{\exp(-\lambda Y_t)\} = \mathbb{E}[\mathbb{E}\{\exp(-\lambda Y_t)|D_{t-1}\}] = \mathbb{E}\{\Phi(\lambda, M_t)\}.$$

By the dominated convergence theorem and since $\exp(-\lambda m)$ and $\Phi(\lambda, m)$ are bounded by 1, $\Psi_t(\lambda) \rightarrow \mathbb{E}\{\exp(-\lambda M_\infty)\}$ and $\mathbb{E}\{\Phi(\lambda, M_t)\} \rightarrow \mathbb{E}\{\Phi(\lambda, M_\infty)\}$. We must then have

$$\mathbb{E}\{\exp(-\lambda M_\infty)\} = \mathbb{E}\{\Phi(\lambda, M_\infty)\}. \tag{16}$$

Now, by the convexity of $\exp(-\lambda m)$, $\Phi(\lambda, m) \geqslant \exp(-\lambda m)$, which combined with equation (16) gives

$$\exp(-\lambda M_\infty) = \Phi(\lambda, M_\infty) \qquad \text{(almost surely)},$$

i.e. $M_\infty \in \{m \in \Delta: \exp(-\lambda m) = \Phi(\lambda, m)\}$ (almost surely) which, again because of the convexity of $\exp(-\lambda m)$, means that $M_\infty \in \Lambda$ (almost surely). □

This result applies to the PSM and SPSM (as we illustrate in the next section) and also generalizes several known special cases. For instance, the Poisson and negative binomial SPSMs can be written as branching processes with time varying rate of immigration. When $\gamma = 0$ we have the critical case and no immigration, and then the result is well known (Guttorp (1991), proposition 1.2 for instance).

Theorem 1 applies more generally than Bayesian time series models. In Gaussian time series modelling, the EWMA forecasting rule (8) for $m_{t+1|t}$ is obtained when the forecast mean is calculated for an ARIMA(0, 1, 1) model (Box and Jenkins, 1976) or, equivalently, for a random walk observed with error (Harvey, 1990). So, we could be led naturally to define non-Gaussian models with $m_{t+1|t}$ given by recursion (8) and $Y_{t+1}|D_t \sim \mathrm{FD}(m_{t+1|t})$ for some FD appropriate to the problem of interest (Poisson or negative binomial for a series of counts, for instance), in an attempt to mimic the non-stationary behaviour of the corresponding Gaussian models. Theorem 1 shows that for non-negative series such an approach would be ill fated.

4.3. *Examples*

4.3.1. *Example 1 (continued): scaled power steady model with Poisson observations*

Let $\mathrm{FD}(m) \sim \mathrm{NB}(m, \sigma)$ and $\Delta = [0, \infty)$. $\xi(m) = m$ (almost surely) if and only if $m = 0$ so $\Lambda = \{0\}$. Thus $M_\infty = 0$ (almost surely) and $Y_t \rightarrow 0$ (almost surely).

4.3.2. *Example 2 (continued): scaled power steady model with exponential observations*

$\mathrm{FD}(m) \sim \mathrm{Pareto}(m, \sigma)$ and $\Delta = [0, \infty)$. Again $\xi(m) = m$ (almost surely) if and only if $m = 0$ so $\Lambda = \{0\}$. Thus $M_\infty = 0$ (almost surely), and again $Y_t \rightarrow 0$ (almost surely).

4.3.3. *Example 3: scaled power steady model with binomial observations*

A direct calculation of density (3) shows that for $Y_t|p_t \sim \text{binomial}(p_t, n)$ (for fixed n) the conjugate state distribution is a transformation of a beta distribution and the forecast distribution is beta binomial with support on $\{0, 1, \ldots, n\}$. (We omit the details because they are not important to the point of this example.) We thus have $\Delta = [0, n]$, and $\xi(m) = m$ (almost surely) if and only if $m = 0$ or $m = n$ so $\Lambda = \{0, n\}$. M_∞ takes two values, 0 and n, and $Y_t \rightarrow 0$ or n (almost surely).

In this case we can say more about the limit distribution. As Δ is a finite interval, M_t is bounded and by the dominated convergence theorem we have $\mathbb{E}(M_\infty) = \mathbb{E}(M_0)$. The distribution of M_∞ is then

$$P(M_\infty = 0) = 1 - \mathbb{E}(M_0)/n$$

and

$$P(M_\infty = n) = \mathbb{E}(M_0)/n.$$

Note that the limit distribution depends only on M_0 and n but not on γ. However, simulations indicate that the rate of convergence is faster for smaller γ.

4.3.4. *Example 4: case where M_∞ takes more than two values*

Let $u(m)$ be a continuous function such that $0 \leqslant u(m) \leqslant m$ and $\mathrm{U} = \{m\text{: } u(m) = 0\}$ contains at least three elements (e.g. $u(m) = m(1 + \sin m)/2$). Here, $\Delta = [0, \infty)$. Now, if $\mathrm{FD}(m) = R\{m - u(m), m + u(m)\}$ where $R(a, b)$ is a uniform distribution on (a, b), then $\xi(m) = m$ (almost surely) means that m satisfies $u(m) = 0$, i.e. $M_\infty \in \mathrm{U} = \Lambda$ (almost surely).

4.4. *Example 1 (Continued)*

The theorem is also important in fitting models to real data. Harvey and Fernandes (1989) used a Poisson PSM to model the number of goals scored by England against Scotland in Scotland from 1872 to 1987. Assuming an equally spaced series of length 53, maximum likelihood gives $\hat{\gamma} = 0.844$. Even with this very slight amount of correlation ($\gamma = 1$ gives an IID series), the convergence to 0 can be seen. For instance, of 100 sample paths of length 100 simulated from this model with the initialization $M_1 = m_{1|0} = 1$ (the first non-zero value) and $\sigma_{1|0} = 0$ used in fitting the model, 45 showed that England scored 0 goals in the last 47 of 100 matches, a result that is not likely to be well received by English fans. For these 45 games, $M_{101} = m_{101|100}$ was between 9.4×10^{-9} and 1.4×10^{-4}, indicating little better chance for England in the future. With stronger correlation (smaller γ), the result is more dramatic.

Fig. 1 shows a time series plot of the series along with several sample paths simulated from the fitted model. Some of the simulations appear to have similar stochastic behaviour to the goals series (e.g. simulation 1 and perhaps simulation 4), but the others do not. (We do not carefully define the term 'similar stochastic behaviour' but leave this use of simulations as an informal model diagnostic in the spirit of Ozaki (1993).)

Other temporal structures can be considered by using different rules for $m_{t+1|t}$ in place of recursion (8). For instance, linear first-order autoregressive structure is specified by $m_{t+1|t} = \rho y_{t-1} + (1 - \rho)\mu$ where $0 \leqslant \rho < 1$ and $\mu \geqslant 0$. This, along with

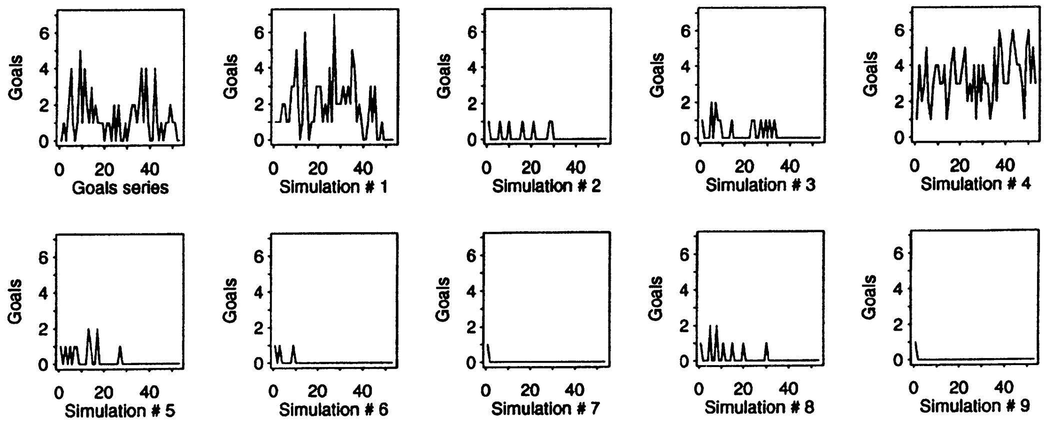

Fig. 1. Goals scored by England against Scotland in Scotland, 1872–1987, and simulations from a fitted Poisson PSM: many of these sample paths have stochastic behaviour quite different from that of the data series

$Y_t|D_{t-1} \sim \mathrm{NB}(m_{t+1|t}, \sigma)$, is easily fitted to the goals series by using likelihood (4), and gives $\hat{\rho} = 0.017$, $\hat{\mu} = 1.31$ and $\hat{\sigma} = 4.76$. Interpretation of ρ in this model is very straightforward and indicates virtually no correlation from one game to the next, which is not surprising since these matches are at least two years apart. The apparent correlation in the previously estimated PSM was induced by the EWMA relation for $m_{t+1|t}$. $\hat{\sigma} > 1$ indicates slightly greater variation than Poisson. Simulated sample paths from this model are given in Fig. 2 and indicate similar stochastic behaviour to the series.

The negative binomial autoregressive model above can be written as a branching process with immigration, so many properties are well known (Guttorp (1991), for instance). More general formulations of models specified directly through forecast

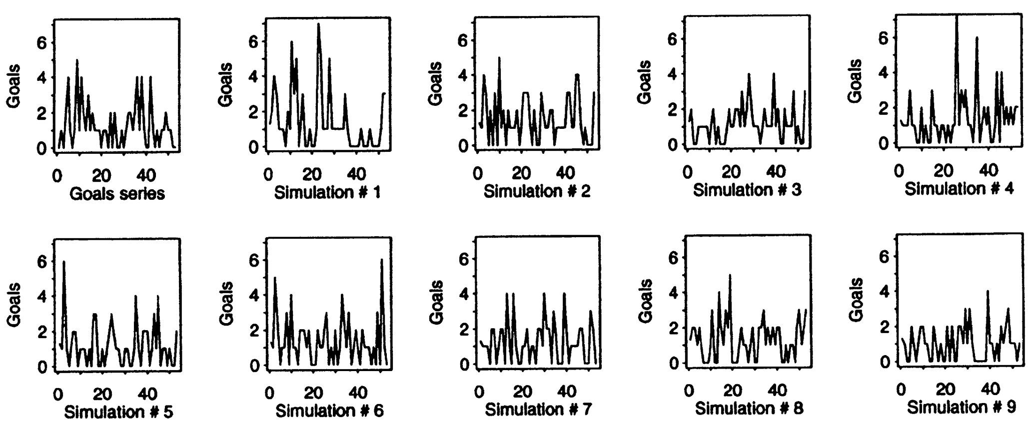

Fig. 2. Goals scored by England against Scotland in Scotland, 1872–1987, and simulations from a fitted negative binomial AR(1) model: all these sample paths appear to have stochastic behaviour similar to the data series

distributions are given by Diggle *et al.* (1994) and Lindsey (1993), but little seems to be known in general about such processes, even in the simple AR(1) case.

5. DISCUSSION

We now give some comments concerning the use of the PSM and related Bayesian forecasting models with exponential family time series.

If the goal is filtering (estimating an unobservable signal $p(\theta_t|D_t)$ as in Smith and West (1983) for instance), Bayesian forecasting models as described in Section 2 may be appropriate tools since they postulate an unobserved state θ_t. There, interest is in the short term so issues relating to limiting model behaviour may not arise.

If interest is in introducing subjective information or modelling time varying effects as in West *et al.* (1985), Bayesian models are also useful. However, Diggle *et al.* (1994) have given an approach to introducing regression information into models specified directly through the forecast distribution.

If smoothing is of interest, Bayesian models might offer help, though as Smith (1990) has pointed out the models do not completely specify the joint state distributions $p(\theta_1, \ldots, \theta_t|D_0)$, and thus exact smoothing (estimation of $p(\theta_s|D_t)$ for $1 < s < t$) is not available. The methods of Kitagawa (1987), Fahrmeir (1992) or Kashiwagi and Yanagimoto (1992) provide alternative approaches. However, the behaviour of these models does not seem to have been studied from the point of view of observable quantities such as forecast distributions and sample paths.

If the object is forecasting in a simple situation such as the example discussed in Section 4.4, directly defining the forecast distribution provides a very simple and flexible alternative to the more complex and perhaps not appropriate Bayesian models. Little use is made of the state in these applications of Bayesian models. If trends, covariates or interventions with effects that do not change over time are to be introduced in a linked linear way as in Harvey and Fernandes (1989) or Grunwald, Raftery and Guttorp (1993), making an appropriate adjustment to the forecast distribution may again provide a simpler and more easily understood alternative to the Bayesian models. Such adjustments are discussed by Diggle *et al.* (1994) in the context of generalized linear models.

Finally, we note the important roles that simulation has played in this paper in studying time series models for which properties may not be known. First, the sample path behaviour proven in theorem 1 was first noted in simulations by Allen (1990), and this suggested the theoretical work in Section 4. Second, simulations from the fitted models provide a very simple and general model diagnostic which can be used as in Section 4.4 with models for which properties are not known. This use of simulation has been suggested by Ozaki (1993) and developed further as a parametric bootstrap method by Tsay (1992).

ACKNOWLEDGEMENTS

The authors are grateful to Fima Klebaner and Paul Feigin for several discussions leading to the proof of theorem 1, and to the referees and Associate Editor for helpful comments which greatly improved the paper. The authors were supported in part by Australian Research Council grants SG6935488 and SG6935619.

REFERENCES

Allen, W. (1990) Poisson time series. *BSc Thesis*. Melbourne University, Parkville.

Attwell, D. N. and Smith, J. Q. (1991) A Bayesian forecasting model for sequential bidding. *J. Forecast.*, **10**, 565–577.

Box, G. E. P. and Jenkins, G. M. (1976) *Time Series Analysis: Forecasting and Control*, 2nd edn. Oakland: Holden-Day.

Diaconis, P. and Ylvisaker, D. (1985) Quantifying prior opinion. In *Bayesian Statistics 2* (eds J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith). Amsterdam: North-Holland.

Diggle, P. J., Liang, K. Y. and Zeger, S. L. (1994) *Analysis of Longitudinal Data*. Oxford: Oxford University Press.

Fahrmeir, L. (1992) State space modeling and conditional mode estimation for categorical time series. In *New Directions in Time Series Analysis*, part 1 (eds D. Brillinger, P. Caines, J. Geweke, E. Parzen, M. Rosenblatt and M. S. Taqqu). New York: Springer.

Grunwald, G. K., Guttorp, P. and Raftery, A. E. (1993) Prediction rules for exponential family state space models. *J. R. Statist. Soc.* B, **55**, 937–943.

Grunwald, G. K., Raftery, A. E. and Guttorp, P. (1993) Time series of continuous proportions. *J. R. Statist. Soc.* B, **55**, 103–116.

Guttorp, P. (1991) *Statistical Inference for Branching Processes*. New York: Wiley.

Harvey, A. C. (1990) *Forecasting, Structural Time Series Models and the Kalman Filter*. New York: Cambridge University Press.

Harvey, A. C. and Fernandes, C. (1989) Time series models for count or qualitative observations. *J. Bus. Econ. Statist.*, **7**, 407–422.

Jørgensen, B. (1986) Some properties of exponential dispersion models. *Scand. J. Statist.*, **13**, 187–198.

Kashiwagi, N. and Yanagimoto, T. (1992) Smoothing serial count data through a state space model. *Biometrics*, **48**, 1187–1194.

Kitagawa, G. (1987) Non-Gaussian state-space modeling of nonstationary time series (with discussion). *J. Am. Statist. Ass.*, **82**, 1032–1063.

Lindsey, J. K. (1993) *Models for Repeated Measurements*. Oxford: Oxford University Press.

Ozaki, T. (1993) Non-Gaussian characteristics of exponential autoregressive processes. In *Developments in Time Series Analysis* (ed. T. S. Rao). London: Chapman and Hall.

Shiryayev, A. N. (1984) *Probability*. New York: Springer.

Smith, A. F. M. and West, M. (1983) Monitoring renal transplants: an application of the multiprocess Kalman filter. *Biometrics*, **39**, 867–878.

Smith, J. Q. (1979) A generalization of the Bayesian steady forecasting model. *J. R. Statist. Soc.* B, **41**, 375–387.

——— (1990) Nonlinear state space models with partially specified distributions on states. *J. Forecast.*, **9**, 137–150.

Smith, R. L. and Miller, J. E. (1986) A non-Gaussian state space model and application to prediction of records. *J. R. Statist. Soc.* B, **48**, 79–88.

Tsay, R. S. (1992) Model checking via parametric bootstraps in time series analysis. *Appl. Statist.*, **41**, 1–15.

West, M. (1985) Generalized linear models: scale parameters, outlier accommodation and prior distributions. In *Bayesian Statistics 2* (eds J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith). Amsterdam: North-Holland.

West, M., Harrison, P. J. and Migon, H. S. (1985) Dynamic generalized linear models and Bayesian forecasting (with discussion). *J. Am. Statist. Ass.*, **80**, 73–96.