



Prediction Rules for Exponential Family State Space Models

Gary K. Grunwald; Peter Guttorp; Adrian E. Raftery

Journal of the Royal Statistical Society. Series B (Methodological), Vol. 55, No. 4
(1993), 937-943.

Stable URL:

<http://links.jstor.org/sici?sici=0035-9246%281993%2955%3A4%3C937%3APRFEFS%3E2.0.CO%3B2-E>

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

Journal of the Royal Statistical Society. Series B (Methodological) is published by Royal Statistical Society. Please contact the publisher for further permissions regarding the use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/rss.html>.

Journal of the Royal Statistical Society. Series B (Methodological)
©1993 Royal Statistical Society

JSTOR and the JSTOR logo are trademarks of JSTOR, and are Registered in the U.S. Patent and Trademark Office. For more information on JSTOR contact jstor-info@umich.edu.

©2003 JSTOR

Prediction Rules for Exponential Family State Space Models

By GARY K. GRUNWALD†,

PETER GUTTORP and ADRIAN E. RAFTERY

University of Melbourne, Parkville, Australia

University of Washington, Seattle, USA

[Received February 1991. Revised July 1992]

SUMMARY

We state and prove a simple and general expression for the marginal mean in an exponential family state space model with conjugate state density. We then discuss implications of this result on the choice of parameterization of the observation density, on the interpretation of the predictive density for steady models and on exponentially weighted moving average and trend-free forecasting.

Keywords: BAYESIAN FORECASTING; BINOMIAL TIME SERIES; POISSON TIME SERIES;
RECURSIVE UPDATING; STATE SPACE MODELS; STEADY MODEL; TIME SERIES

1. INTRODUCTION

The power steady model (PSM) of Smith (1979, 1981) has been used by several researchers (West *et al.* (1985), Harvey and Fernandes (1989), Smith (1983) or Attwell and Smith (1991), for example) as a basis for state space modelling of non-Gaussian time series. Although these models are ‘steady’ in the sense discussed later, several difficulties have also been noted. For instance, West (1986) states that the model remains ambiguous, since it can depend on the parameterization used. Smith and Miller (1986) point out that there may not be a simple formula for m -step forecasting if $m > 1$, and Key and Godolphin (1981) show that point forecasts from the model need not be either trend free or exponentially weighted moving average (EWMA). In this paper we help to clarify and resolve these difficulties.

In Section 2 we give the setting and describe the PSM. We then state and prove a simple and general expression for the marginal (hence forecast) mean in an exponential family state space model (Section 3). This result shows that if the natural exponential family parameterization is used there is no ambiguity in the definition of the model, and there is a simple formula for m -step forecasting which gives, for steady models, trend-free EWMA forecasts and also gives recursions which do not depend on the underlying distributions (Section 4).

2. POWER STEADY MODEL

A state space model for a time series $\mathbf{y}^T = \{\mathbf{y}_1, \dots, \mathbf{y}_t\}$, $\mathbf{y}_j \in \mathbf{R}^d$ for $j = 1, \dots, t$, assumes a density $p(\mathbf{y}_t | \boldsymbol{\theta}_t)$ (the *observation density*), where $\boldsymbol{\theta}_t$ is an unobservable *state*. (We attach an italicized verbal label to some equations or expressions for later reference.) Although this distribution need not belong to an exponential family, most situations of practical interest do satisfy this assumption and we consider only this case.

†Address for correspondence: Department of Statistics, University of Melbourne, Parkville, Victoria 3052, Australia.

A conjugate density is usually assumed for $p(\theta_t | \mathbf{D}_t)$, where the observed history \mathbf{D}_t is defined by $\mathbf{D}_t = \{\mathbf{D}_0, \mathbf{y}^T\}$ with \mathbf{D}_0 denoting the values of any estimated parameters and all relevant external information. Given a prior $p(\theta_{t+1} | \mathbf{D}_t)$ and a new observation \mathbf{y}_{t+1} , Bayes's theorem and the likelihood for the new observation yield a posterior $p(\theta_{t+1} | \mathbf{D}_{t+1})$ as usual.

Smith (1979) proposed the *PSM prediction rule*

$$p(\theta_{t+1} | \mathbf{D}_t) \propto p(\theta_t | \mathbf{D}_t)^\gamma \quad (0 < \gamma < 1) \tag{2.1}$$

as a non-Gaussian analogue of the state prediction in the normal steady model. Models obtained by using the PSM rule are steady in the sense that $\theta_{t+1} | \mathbf{D}_t$ has the same mode as $\theta_t | \mathbf{D}_t$ but greater dispersion. (The exact sense in which these remarks hold is discussed at length by Smith (1990).)

The *marginal data distribution*

$$p(\mathbf{y}_s | \mathbf{D}_t) = \int_{\Theta} p(\mathbf{y}_s | \mathbf{D}_t, \theta_s) p(\theta_s | \mathbf{D}_t) d\theta_s \quad (s = t \text{ or } s = t + 1) \tag{2.2}$$

is fundamental since it does not condition on the unobservable state θ_s . It is used, for example, to calculate the likelihood, and it is the *forecast distribution* when $s = t + 1$. The *marginal mean* $E[\mathbf{y}_s | \mathbf{D}_t]$ is defined as usual using equation (2.2) (with $s = t$ or $s = t + 1$) and is the *forecast mean* when $s = t + 1$.

2.1. *Example 1: Binomial Observation Distribution, n, Known*

For $y_t | \pi_t \sim \text{bin}(n, \pi_t)$ the density in the 'usual' parameterization is

$$p(y_t | \pi_t) = \binom{n}{y_t} \pi_t^{y_t} (1 - \pi_t)^{n - y_t}, \quad y_t \in \{0, 1, \dots, n\}, 0 < \pi_t < 1.$$

We assume for simplicity that $n_{t+1} = n_t = n$; this sequence is contained in \mathbf{D}_0 . The conjugate density for $\pi_t | \mathbf{D}_t$ is beta with parameters, say, $a_{t|t}$ and $b_{t|t}$, where the subscript notation $s|t$ indicates that the parameter is estimated at time s given \mathbf{D}_t ($s \geq t$). ($\pi \sim \text{beta}(a, b)$ if $p(\pi) = B(a, b)^{-1} \pi^{a-1} (1 - \pi)^{b-1}$ where $B(a, b)$ is the beta function.)

Application of the PSM rule (2.1) gives prior $\pi_{t+1} | \mathbf{D}_t \sim \text{beta}(a_{t+1|t}, b_{t+1|t})$, and the one-step forecast distribution computed from equation (2.2) is beta-binomial (see Mosimann (1962)) with mean

$$E[y_{t+1} | \mathbf{D}_t] = \frac{na_{t+1|t}}{a_{t+1|t} + b_{t+1|t}} = n \frac{\gamma a_{t|t} - \gamma + 1}{\gamma(a_{t|t} + b_{t|t}) - 2\gamma + 2} \neq E[y_t | \mathbf{D}_t] = n \frac{a_{t|t}}{a_{t|t} + b_{t|t}}.$$

Thus, the forecasts are not trend free, and therefore they are not EWMA. In later sections, we explain the source of these difficulties and show how to formulate a binomial model free from these problems.

3. MARGINAL MEAN

We adopt the multivariate exponential family setting of Diaconis and Ylvisaker (1979, 1985). Let \mathbf{y}_t follow an *exponential family observation distribution*

$$p(\mathbf{y}_t | \theta_t) = \exp\{\mathbf{y}_t^T \theta_t - M(\theta_t) + S(\mathbf{y}_t)\}, \quad \mathbf{y}_t \in \mathbf{Y} \tag{3.1}$$

and let Ψ be the interior of the convex hull of the support \mathbf{Y} of this distribution, with Ψ assumed to be non-empty and open in \mathbf{R}^d . Further assume that the natural parameter space $\Theta = \{\boldsymbol{\theta}_t \in \mathbf{R}^d: M(\boldsymbol{\theta}_t) < \infty\}$ is non-empty and open, where $M(\boldsymbol{\theta}_t) \equiv \log[\int \exp\{\mathbf{y}_t^\top \boldsymbol{\theta}_t + S(\mathbf{y}_t)\} d\mathbf{y}_t]$.

The *state prior* (if $s = t + 1$) or *state posterior* (if $s = t$) is assumed conjugate to the exponential family observation distribution (3.1) and is of the form

$$p(\boldsymbol{\theta}_s | \mathbf{D}_t) \propto \exp[\sigma_{s|t} \{\boldsymbol{\kappa}_{s|t}^\top \boldsymbol{\theta}_s - M(\boldsymbol{\theta}_s)\}], \quad \boldsymbol{\theta}_s \in \Theta. \tag{3.2}$$

Diaconis and Ylvisaker (1979), theorem 1, show that if $\sigma_{s|t} > 0$ and $\boldsymbol{\kappa}_{s|t} \in \Psi$ this family can be normalized to a family of probability distributions. Here we assume arbitrary forms for the prior parameters $\boldsymbol{\kappa}_{t+1|t}$ and $\sigma_{t+1|t}$ since the next result does not depend on the PSM rule.

Theorem. Assuming observation and state distributions of the forms (3.1) and (3.2) with Ψ and Θ non-empty and open, and $\sigma_{s|t} > 0$ and $\boldsymbol{\kappa}_{s|t} \in \Psi$ with $s = t$ or $s = t + 1$, then the marginal mean satisfies

$$E[\mathbf{y}_s | \mathbf{D}_t] = \boldsymbol{\kappa}_{s|t}. \tag{3.3}$$

Proof. Computing directly gives

$$\begin{aligned} E[\mathbf{y}_s | \mathbf{D}_t] &= \int_{\Theta} E[\mathbf{y}_s | \boldsymbol{\theta}_s, \mathbf{D}_t] p(\boldsymbol{\theta}_s | \mathbf{D}_t) d\boldsymbol{\theta}_s = \int_{\Theta} \nabla M(\boldsymbol{\theta}_s) p(\boldsymbol{\theta}_s | \mathbf{D}_t) d\boldsymbol{\theta}_s \\ &= E[\nabla M(\boldsymbol{\theta}_s) | \mathbf{D}_t] = \boldsymbol{\kappa}_{s|t}, \end{aligned}$$

where

$$\nabla M(\boldsymbol{\theta}) = \left(\frac{\partial M(\boldsymbol{\theta})}{\partial \theta_1}, \dots, \frac{\partial M(\boldsymbol{\theta})}{\partial \theta_d} \right)^\top.$$

The last equality follows from theorem 2 of Diaconis and Ylvisaker (1979), and the equality $E[\mathbf{y}_s | \boldsymbol{\theta}_s, \mathbf{D}_t] = \nabla M(\boldsymbol{\theta}_s)$ is standard exponential family theory (e.g. Bickel and Doksum (1977)). □

3.1. Remarks

- (a) The conditions of the theorem are easily checked and hold in most cases of interest, including observation distributions that are multinomial, Poisson, normal, exponential, negative binomial, Dirichlet and Dirichlet with fixed dispersion (on suitable reparameterization; see Grunwald *et al.* (1993)).
- (b) Guttorp and Lockhart (1988) consider, in a non-time series setting, a von Mises observation density with known concentration parameter and von Mises conjugate. In that situation, the conditions do not hold (Θ is not open).
- (c) In the context of modelling time series of compositions by using a Dirichlet observation distribution with fixed dispersion (Grunwald *et al.*, 1993), the state and marginal distributions are not of known form and the moments cannot be computed directly, but the theorem gives a simple and easily computed expression for the forecast mean (and hence residuals).
- (d) There does not appear to be a result like equation (3.3) for the marginal variance; examples show that it depends on both $\boldsymbol{\kappa}_{s|t}$ and $\sigma_{s|t}$ but not according to any general rule that we can discover.

- (e) If we transform the *observation* \mathbf{x}_t to obtain an exponential family form in \mathbf{y}_t (for example, when $\mathbf{x}_t \sim \text{log-normal}$ then $\mathbf{y}_t = \text{log } \mathbf{x}_t$), none of the problems associated with non-exponential family forms arise. However, all results such as equation (3.3) pertain to \mathbf{y}_t and not to \mathbf{x}_t .

4. APPLICATIONS

In this section we use the theorem to clarify and resolve the difficulties discussed in Sections 1 and 2, and to unify some results concerning exponential family state space models.

4.1. Model Ambiguity

As mentioned earlier the PSM rule (2.1) does not fully define the time series model, as we first illustrate in the binomial case and then explain generally.

4.1.1. Example 1 (continued)

We change to binomial proportions $w_t = y_t/n_t$ so that Ψ does not depend on time, and we assume that the sequence $\{n_t\}$ is contained in \mathbf{D}_0 . $n_t w_t | \pi_t, n_t \sim \text{bin}(n_t, \pi_t)$ leads to $\theta_t = n_t \log\{\pi_t/(1 - \pi_t)\}$, $\mathbf{Y}_t = \{0, 1/n_t, \dots, 1\}$, $\Psi = (0, 1)$ and $\Theta = \mathbf{R}^1$ in distribution (3.1). The last two sets are both non-empty and open, and the dependence of S_t, M_t and \mathbf{Y}_t on t causes no difficulties.

The PSM rule applied in the exponential family parameterization and in the usual parameterization gives prior state densities respectively

$$\begin{aligned} \pi_{t+1} | \mathbf{D}_t &\sim \text{beta}\{\gamma n_{t+1} \sigma_{t|t} \kappa_{t|t}, \gamma n_{t+1} \sigma_{t|t} (1 - \kappa_{t|t})\}, \\ \pi_{t+1} | \mathbf{D}_t &\sim \text{beta}\{\gamma n_{t+1} \sigma_{t|t} \kappa_{t|t} - \gamma + 1, \gamma n_{t+1} \sigma_{t|t} (1 - \kappa_{t|t}) - \gamma + 1\}. \end{aligned}$$

The different priors give different time series models.

Both parameterizations have appeared in the literature. For example, Smith (1979), Key and Godolphin (1981) and West (1986) use the usual parameterization whereas West *et al.* (1985) use the exponential family forms. Harvey and Fernandes (1989) quote the usual form but, by requiring that $E[w_{t+1} | \mathbf{D}_t] = E[w_t | \mathbf{D}_t]$ is to hold for the steady model, derive the prior for the exponential family form.

The results of Diaconis and Ylvisaker (1985) cover the general situation. If the exponential family observation density (3.1) is written in a parameterization that is different from the natural exponential family form, say $\xi_t = \psi(\theta_t)$ where $\psi: \Theta \rightarrow \Theta_\psi \subseteq \mathbf{R}^d$ is a diffeomorphism (one to one and continuously differentiable with continuously differentiable inverse ψ^{-1}), the observation density is

$$p(\mathbf{y}_t | \xi_t) = \exp[\mathbf{y}_t^\top \psi^{-1}(\xi_t) - M\{\psi^{-1}(\xi_t)\} + S(\mathbf{y}_t)]$$

with conjugate

$$p(\xi_t | \mathbf{D}_t) \propto \exp(\sigma_{t|t} [\boldsymbol{\kappa}_{t|t}^\top \psi^{-1}(\xi_t) - M\{\psi^{-1}(\xi_t)\}]) \frac{1}{\psi'\{\psi^{-1}(\xi_t)\}} \tag{4.1}$$

where ψ' is the Jacobian. The prior given by the PSM rule in this parameterization now contains the factor $\psi'\{\psi^{-1}(\xi_t)\}^{-\gamma}$, and it is this term which causes the ambiguity in the prior (unless ψ is linear, since then ψ' is constant).

4.1.2. *Example 1 (continued)*

In the binomial case, $\xi_t = \pi_t = \psi(\theta_t) = \exp(\theta_t/n_t)/\{1 + \exp(\theta_t/n_t)\}$ is not linear, and $\psi'\{\psi^{-1}(\xi_t)\}^{-1} = n_t/\{\xi_t(1 - \xi_t)\}$, so the two parameterizations (ξ_t and θ_t) give different priors under the PSM rule.

4.1.3. *Example 2: exponential observation distribution*

For $y_t|\lambda_t \sim \text{exponential}(\lambda_t)$, $\theta_t = -\lambda_t$ and ψ is linear, so the same (gamma) conjugate prior is obtained by applying the PSM rule in either parameterization. This explains why Key and Godolphin (1981) did not encounter the same difficulties with exponential observations as with binomial observations.

4.2. *Trend-free Forecasts*

The theorem of Section 3 shows that if the PSM rule is applied with distributions in exponential family form the one-step forecasts are trend free: $E[y_{t+1}|\mathbf{D}_t] = E[y_t|\mathbf{D}_t]$. In a different (ξ_t) parameterization, computing $E[y_{t+1}|\mathbf{D}_t]$ directly from expressions (2.2) and (4.1) shows that, for a steady model obtained from the PSM rule applied in the ξ_t -parameterization, the one-step forecasts are trend free if and only if ψ is linear.

4.2.1. *Example 1 (continued)*

By the theorem, $E[w_{t+1}|\mathbf{D}_t] = \kappa_{t|t} = E[w_t|\mathbf{D}_t]$ if the PSM rule is applied in the exponential family parameterization, whereas in the usual parameterization

$$E[w_{t+1}|\mathbf{D}_t] = \frac{\gamma n_{t+1} \sigma_{t|t} \kappa_{t|t} - \gamma + 1}{\gamma n_{t+1} \sigma_{t|t} - 2\gamma + 2}.$$

In the latter case we have $E[w_{t+1}|\mathbf{D}_t] \rightarrow \frac{1}{2}$ as $\gamma \rightarrow 0$ which, although possibly sometimes giving a useful model, does not seem to give a steady model in this sense.

4.3. *Unified Recursions*

Here, we consider a conjugate state prior of the form (3.2) defined for any general state transition rules by $\kappa_{t+1|t}$ and $\sigma_{t+1|t}$. Regardless of the parameterization used, if the prior can again be written in the exponential family form (3.2) (as it always can if the PSM rule is applied to the exponential family form) then Bayes's theorem yields *recursions*

$$\kappa_{t+1|t+1} = (1 - g_{t+1})\kappa_{t+1|t} + g_{t+1} \mathbf{y}_{t+1}, \tag{4.2a}$$

$$\sigma_{t+1|t+1} = \sigma_{t+1|t} + 1, \tag{4.2b}$$

$$g_{t+1} = 1/\sigma_{t+1|t+1} \tag{4.2c}$$

regardless of the particular observation distribution assumed. Although this is not a consequence of the theorem of Section 3 or of the parameterization used, it appears naturally if exponential family forms are used throughout and is also very convenient for computer implementation.

If the exponential family parameterization of the observation density (3.1) is used, the theorem and the common recursions (4.2) clarify the computation of the marginal posterior mean as a convex combination of the forecast mean and the new observation

(for $0 < \sigma_{t+1|t} < \infty, 0 < g_{t+1} < 1$). Since Ψ is convex, recursion (4.2a) shows that $\kappa_{t+1|t+1} \in \Psi$ whenever $\kappa_{t+1|t} \in \Psi$, ensuring a proper posterior distribution (by Diaconis and Ylvisaker (1979), theorem 1).

4.4. *Exponentially Weighted Moving Average Forecasting*

For the steady model, where $\sigma_{t+1|t} = \gamma\sigma_{t|t}$, recursions (4.2b) and (4.2c) show that $\sigma_{t|t} \rightarrow 1/(1-\gamma)$ as $t \rightarrow \infty$ so $g_t \rightarrow 1-\gamma$ as $t \rightarrow \infty$. If these limiting values are used initially ($\sigma_{0|0} = 1/(1-\gamma)$) then $\sigma_{t|t} = 1/(1-\gamma)$ and $g_t = 1-\gamma$ for all $t \geq 0$. The theorem and recursion (4.2a) then show that the forecasts are of EWMA form:

$$E[\mathbf{y}_{t+1} | \mathbf{D}_t] = \kappa_{t+1|t} = \gamma^t \kappa_{1|0} + (1-\gamma) \sum_{j=0}^{t-1} \gamma^j \mathbf{y}_{t-j} \tag{4.3}$$

regardless of the underlying observation distribution.

A more common initialization uses $\sigma_{1|0} = 0$ in the state prior (3.2), giving a flat, possibly improper, prior for θ_1 , and $\kappa_{1|0}$ is ignored. If this is done, $\kappa_{1|1} \in \Psi$ whenever $\mathbf{y}_1 \in \Psi$ (\mathbf{y}_t can be on the boundary of Ψ , as when $w_t = 0$ or $w_t = 1$ in the binomial example) and the forecasts are of EWMA form except for these initial effects. These results generalize similar results derived by Harvey and Fernandes (1989) under particular distributional assumptions.

4.5. *Trend-free m-step Forecasts*

As Harvey and Fernandes (1989) point out, m -step forecasts are not made by simply iterating the PSM rule, but rather by iterating the marginal data distributions (2.2). The resulting m -step forecast distribution may be quite complex. However, the theorem of Section 3 can be used to show that for the steady model the process $\mathbf{K}(t) = \kappa_{t|t}$ defined recursively by equations (4.2) ($\mathbf{K}(t+1) = (1-g_{t+1})\mathbf{K}(t) + g_{t+1}\mathbf{y}_{t+1}$) is a martingale: $E[\mathbf{K}(t+1) | \mathbf{K}(t), \mathbf{D}_{t-1}] = \mathbf{K}(t)$. It then follows by direct calculation and the constant mean property of martingales that $E[\mathbf{y}_{t+m} | \mathbf{D}_t] = \mathbf{K}(t)$ for ($m \geq 1$). Thus, m -step point forecasts remain trend free if the exponential family forms are used.

ACKNOWLEDGEMENTS

Guttorp’s research was partially supported by a co-operative agreement between the United States Environmental Protection Agency and the Societal Institute for Mathematical Sciences. Raftery’s research was supported by the Office of Naval Research under contract N-00014-88-k-0265. The authors are grateful to Andrew Bruce and William Allen for many helpful discussions, and to Dermot Duggan and a referee for important suggestions which improved the paper.

REFERENCES

Attwell, D. N. and Smith, J. Q. (1991) A Bayesian forecasting model for sequential bidding. *J. Forecast.*, **10**, 565-577.
 Bickel, P. J. and Doksum, K. A. (1977) *Mathematical Statistics: Basic Ideas and Selected Topics*, p. 71. San Francisco: Holden-Day.

- Diaconis, P. and Ylvisaker, D. (1979) Conjugate priors for exponential families. *Ann. Statist.*, **7**, 269–281.
- (1985) Quantifying prior opinion. In *Bayesian Statistics 2* (eds J. M. Bernardo, M. DeGroot, D. V. Lindley and A. F. M. Smith). Amsterdam: North-Holland.
- 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. and Lockhart, R. A. (1988) On finding the location of a signal: a Bayesian analysis. *J. Am. Statist. Ass.*, **83**, 322–330.
- Harvey, A. C. and Fernandes, C. (1989) Time series models for count or qualitative observations. *J. Bus. Econ. Statist.*, **7**, 407–422.
- Key, P. B. and Godolphin, E. J. (1981) On the Bayesian steady forecasting model. *J. R. Statist. Soc. B*, **43**, 92–96.
- Mosimann, J. E. (1962) On the compound multinomial distribution, the multivariate β -distribution, and correlations among proportions. *Biometrika*, **49**, 65–82.
- Smith, J. Q. (1979) A generalization of the Bayesian steady forecasting model. *J. R. Statist. Soc. B*, **41**, 375–387.
- (1981) The multiparameter steady model. *J. R. Statist. Soc. B*, **43**, 256–260.
- (1983) Forecasting accident claims for an assurance company. *Statistician*, **32**, 109–115.
- (1990) Non linear state space models with partially specified distributions on states. *J. Forecast.*, **9**, 137–149.
- 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.
- West, M. (1986) Bayesian model monitoring. *J. R. Statist. Soc. B*, **48**, 70–78.
- West, M., Harrison, P. J. and Migon, H. S. (1985) Dynamic generalized linear models and Bayesian forecasting (with discussion). *J. Am. Statist. Ass.*, **80**, 73–97.