AISM (2008) 60:407–426
DOI 10.1007/s10463-006-0109-x

The admissible parameter space for exponential smoothing models

Rob J. Hyndman · Muhammad Akram · Blyth C. Archibald

Received: 4 November 2005 / Revised: 18 July 2006 / Published online: 14 February 2007

Abstract We discuss the admissible parameter space for some state space models, including the models that underly exponential smoothing methods. We find that the usual parameter restrictions (requiring all smoothing parameters to lie between 0 and 1) do not always lead to stable models. We also find that all seasonal exponential smoothing methods are unstable as the underlying state space models are neither reachable nor observable. This instability does not affect the forecasts, but does corrupt the state estimates. The problem can be overcome with a simple normalizing procedure. Finally we show that the admissible parameter space of a seasonal exponential smoothing model is much larger than that for a basic structural model, leading to better forecasts from the exponential smoothing model when there is a rapidly changing seasonal pattern.

Keywords Exponential smoothing · Invertibility · Observability · Parameter space · Reachability · Stability · State space models · Structural models

R. J. Hyndman · M. Akram
Department of Econometrics and Business Statistics,
Monash University, Clayton, VIC 3800, Australia
e-mail: Rob.Hyndman@buseco.monash.edu.au

M. Akram (✉)
e-mail: Muhammad.Akram@buseco.monash.edu.au

B. C. Archibald
School of Business Administration, Dalhousie University,
Halifax, Canada B3H 1Z5
e-mail: blyth.archibald@dal.ca

1 Introduction

Hyndman et al. (2002) proposed a modelling framework based on exponential smoothing methods. The framework involves 12 different methods, including the well-known simple exponential smoothing, Holt's method, and Holt–Winters additive and multiplicative methods. They demonstrated that each method in their taxonomy of exponential smoothing methods is equivalent to the forecasts obtained from a state space model.

For each of these methods, Hyndman et al. (2002) proposed two state space models with a single source of error following the general approach of Ord et al. (1997). The state space models enable easy calculation of the likelihood, and provide facilities to compute prediction intervals for each model. The two state space formulations correspond to the additive error and the multiplicative error cases. They give equivalent point forecasts although different prediction intervals and different likelihoods.

In this paper, we investigate the admissible parameter space for each of the linear state space models in the Hyndman et al. (2002) framework. We describe each of the exponential smoothing state space models using a three-letter code, following Hyndman et al. (2005). The first letter describes the error (in this paper always additive), the second letter describes the trend (none, additive or damped) and the third letter describes the seasonal component (none or additive). For example, AAN refers to a model with additive errors, additive trend and no seasonality. In this paper, we consider the six linear models: ANN, AAN, ADN, ANA, AAA and ADA. Model ANN gives forecasts equivalent to simple exponential smoothing, model AAN underlies Holt's linear method and the additive Holt–Winters' method is obtained by model AAA.

Table 1 shows the equations for the models we consider in this paper. Note that we use a slightly different parameterization from Hyndman et al. (2002) for the trend equation—we use β where Hyndman et al. (2002) used $\alpha\beta$. This change in parameters makes no difference to the models but allows us to have a bounded admissible parameter space. The usual parameter space has all parameters lie between 0 and 1. Because of our reparameterization, this means that α, γ and ϕ would lie between 0 and 1, but $0 < \beta < \alpha$.

1.1 State space models

Let $Y_1, \ldots, Y_n$ denote the time series of interest and let $\boldsymbol{x}_t = (\ell_t, b_t, s_t, s_{t-1}, \ldots, s_{t-(m-1)})$ where ℓ_t denotes the level, b_t denotes the trend and s_t denotes the seasonal component, all at time t. Then the models in Table 1 can be written as

$$Y_t = H\boldsymbol{x}_{t-1} + \varepsilon_t \tag{1}$$

$$\boldsymbol{x}_t = F\boldsymbol{x}_{t-1} + G\varepsilon_t \tag{2}$$

where $\{\varepsilon_t\}$ is a Gaussian white noise process with mean zero and variance σ^2.

Table 1 State space equations for the models considered in this paper

Model ANN	Model ANA
$Y_t = \ell_{t-1} + \varepsilon_t$	$Y_t = \ell_{t-1} + s_{t-m} + \varepsilon_t$
$\ell_t = \ell_{t-1} + \alpha\varepsilon_t$	$\ell_t = \ell_{t-1} + \alpha\varepsilon_t$
	$s_t = s_{t-m} + \gamma\varepsilon_t$
$\mu_n(h) = \ell_n$	$\mu_n(h) = \ell_n + s_{n-m+1+(h-1)*}$
Model AAN	**Model AAA**
$Y_t = \ell_{t-1} + b_{t-1} + \varepsilon_t$	$Y_t = \ell_{t-1} + b_{t-1} + s_{t-m} + \varepsilon_t$
$\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t$	$\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t$
$b_t = b_{t-1} + \beta\varepsilon_t$	$b_t = b_{t-1} + \beta\varepsilon_t$
	$s_t = s_{t-m} + \gamma\varepsilon_t$
$\mu_n(h) = \ell_n + hb_n$	$\mu_n(h) = \ell_n + hb_n + s_{n-m+1+(h-1)*}$
Model ADN	**Model ADA**
$Y_t = \ell_{t-1} + b_{t-1} + \varepsilon_t$	$Y_t = \ell_{t-1} + b_{t-1} + s_{t-m} + \varepsilon_t$
$\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t$	$\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t$
$b_t = \phi b_{t-1} + \beta\varepsilon_t$	$b_t = \phi b_{t-1} + \beta\varepsilon_t$
	$s_t = s_{t-m} + \gamma\varepsilon_t$
$\mu_n(h) = \ell_n + \phi_h b_n$	$\mu_n(h) = \ell_n + \phi_h b_n + s_{n-m+1+(h-1)*}$

Point forecasts are given by $\mu_n(h)$. Here $\phi_j = 1 + \phi + \cdots + \phi^{j-1} = (1-\phi^j)/(1-\phi)$ and $(h-1)^* = (h-1) \bmod m$

A single source of error model (with ε_t appearing in both equations) is preferable to a multiple source of error model because it allows the state space formulation of non-linear as well as linear cases, and allows the state equations to be expressed in a form which coincides with the error-correction form of the usual smoothing equations. However, we do not discuss the non-linear models in this paper.

We write $\mu_t = H\boldsymbol{x}_{t-1}$ to denote the mean of Y_t conditional on $\boldsymbol{x}_{t-1}$. The usual point forecasts are obtained as $\mu_n(h) = \mathrm{E}(Y_{n+h}|\boldsymbol{x}_n)$, so that $\mu_t = \mu_{t-1}(1)$. Hyndman et al. (2005) show that the forecast distribution of these models, defined as the distribution of Y_{n+h} conditional on $\boldsymbol{x}_n$, is normal with mean and variance given by

$$\mu_n(h) = \mathrm{E}(Y_{n+h}|\boldsymbol{x}_n) = HF^{h-1}\boldsymbol{x}_n \tag{3}$$

and

$$v_n(h) = \mathrm{Var}(Y_{n+h}|\boldsymbol{x}_n) = \sigma^2\left[1 + \sum_{j=1}^{h-1}(HF^{j-1}G)^2\right]. \tag{4}$$

The coefficient matrices F, G and H can be easily determined from Table 1 and are given below. Here I_k denotes the $k \times k$ identity matrix, $\mathbf{0}_k$ denotes a zero vector of length k and $\mathbf{0}'_k$ means transpose of $\mathbf{0}_k$.

ANN: $H = F = 1, \quad G = \alpha$

ADN: $H = [1 \quad 1], \quad F = \begin{bmatrix} 1 & 1 \\ 0 & \phi \end{bmatrix}$ and $G = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$

ANA: $H = [1\ \mathbf{0}'_{m-1}\ 1]$, $F = \begin{bmatrix} 1 & \mathbf{0}'_{m-1} & 0 \\ 0 & \mathbf{0}'_{m-1} & 1 \\ \mathbf{0}_{m-1} & I_{m-1} & \mathbf{0}_{m-1} \end{bmatrix}$ and $G = \begin{bmatrix} \alpha \\ \gamma \\ \mathbf{0}_{m-1} \end{bmatrix}$

ADA: $H = [1\ 1\ \mathbf{0}'_{m-1}\ 1]$, $F = \begin{bmatrix} 1 & 1 & \mathbf{0}'_{m-1} & 0 \\ 0 & \phi & \mathbf{0}'_{m-1} & 0 \\ 0 & 0 & \mathbf{0}'_{m-1} & 1 \\ \mathbf{0}_{m-1} & \mathbf{0}_{m-1} & I_{m-1} & \mathbf{0}_{m-1} \end{bmatrix}$ and

$$G = \begin{bmatrix} \alpha \\ \beta \\ \gamma \\ \mathbf{0}_{m-1} \end{bmatrix}$$

The matrices for AAN and AAA are the same as for ADN and ADA respectively, but with $\phi = 1$.

In this paper, we study some of the properties of these models; in particular, reachability, observability and stability. The parameters are deemed 'admissible' if the model is stable. We derive the admissible parameter space for each of the models. Traditionally, the parameters α, β/α and γ are allowed to lie between 0 and 1. We compare the admissible parameter space with this traditional space.

In Sect. 2, we define various model properties for these state space models. In Sect. 3, we note that the seasonal models (ANA, AAA and ADA) have some undesirable properties, and so we introduce some alternative models designed to overcome these problems. The admissible parameter spaces for these alternative seasonal models are discussed in Sect. 4. A competitor model closely related to AAA is Harvey (1989) basic structural model. In Sect. 5, we compare the admissible parameter space for the basic structural model with the exponential smoothing models, showing that the exponential smoothing models are able to cater for a larger range of seasonal patterns than the basic structural model. Finally, in Sect. 6, we compare the various seasonal models on a real data set, and Sect. 7 illustrate a Monte Carlo simulation study.

2 Model properties

This section examines the various properties of time-invariant state space models. The following definitions are given by Hannan and Deistler (1988, p. 44–45):

Definition 1 The model (1) and (2) is said to be *observable* if

$$\mathrm{Rank}[H', F'H', (F')^2H', \ldots, (F')^{p-1}H'] = p$$

where p is the length of the state vector $\boldsymbol{x}_t$.

Definition 2 The model (1) and (2) is said to be *reachable* if

$$\mathrm{Rank}[G, FG, F^2G, \ldots, F^{p-1}G] = p$$

where p is the length of the state vector $\boldsymbol{x}_t$.

Reachability and observability are desirable properties of a state space model because they guarantee a minimal state dimension (see Hannan and Deistler, 1988, p. 48), and consequently, that the state elements will be identifiable. It is easily seen that the non-seasonal models ANN, AAN and ADN are both reachable and observable, and therefore of minimal dimension.

To motivate the next property, note that we can write the state vector as

$$\boldsymbol{x}_t = M^t \boldsymbol{x}_0 + \sum_{j=0}^{t-1} M^j G Y_{t-j}$$

where $M = F - GH$. So for initial conditions to be negligible, we need M^t to converge to zero. Therefore, we require M to have all eigenvalues inside the unit circle. We call this condition *stability* (following Hannan and Deistler, 1988, p. 48).

Definition 3 The model (1) and (2) is said to be *stable* if all eigenvalues of $M = F - GH$ lie inside the unit circle.

Snyder et al. (2001) show that this property is equivalent to invertibility of the underlying ARIMA model. Stability is a desirable property of a time series model because we want models where the distant past has a negligible effect on the present state.

State space models that are not stable can still produce stable point forecasts. Therefore, it is sometimes useful to have a weaker notion of stability which we shall call *forecastability*.

Definition 4 Let $(\lambda_i, \boldsymbol{v}_i)$ denote an eigenvalue-eigenvector pair of $M = F - GH$. Then the model (1) and (2) is said to be *forecastable* if, for all i, either $|\lambda_i| < 1$ or $HF^j \boldsymbol{v}_i = 0$ for $j = 0, 1, \ldots$.

Obviously, any model that is stable is also forecastable. The notion of forecastability is motivated by the idea that an unstable model can still produce stable forecasts provided the eigenvalues which cause the instability have no effect on the forecasts. This arises because M may have unit eigenvalues where H is orthogonal to the eigenvectors corresponding to the unit eigenvalues. Under these conditions, the forecast function is unique and asymptotically independent of the initial state $\boldsymbol{x}_0$.

If a model is forecastable, the forecast mean and variance given by (3) and (4) are unaffected by the eigenvalues on or outside the unit circle. The concept of forecastability was noted by Sweet (1985) and Lawton (1998) for AAA (additive Holt–Winters) forecasts, although neither author used a stochastic model as we do here. The phenomenon was also observed by Snyder and Forbes (2003) in connection with the AAA model. To our knowledge, ours is the first general definition of this property. Burridge and Wallis (1998, Theorem 3.1) give a weaker condition than forecastability which guarantees convergence of the covariance matrix. Therefore, forecastability also guarantees this convergence.

Table 2 Stability conditions for models without seasonality

ANN:	$0 < \alpha < 2$
AAN:	$0 < \alpha < 2$ $0 < \beta < 4 - 2\alpha$
ADN:	$1 - 1/\phi < \alpha < 1 + 1/\phi$ $\alpha(\phi - 1) < \beta < (1 + \phi)(2 - \alpha)$ $0 < \phi \le 1$

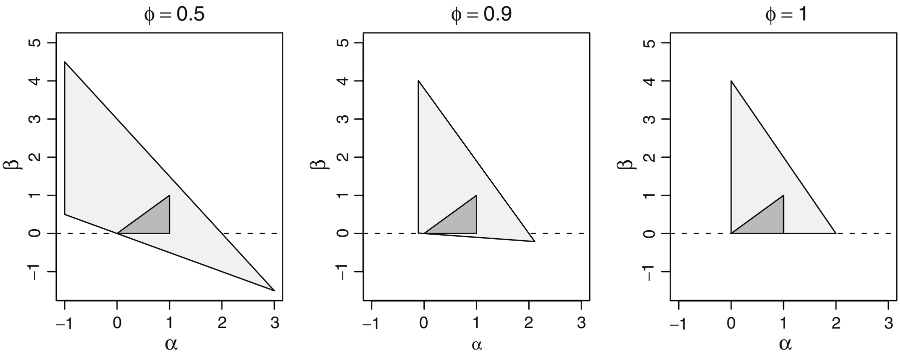

Fig. 1 Parameter spaces for model ADN. The *right hand* graph shows the region for model AAN (when $\phi = 1$). In each case, the *light-shaded regions* represent the admissible regions; the *dark-shaded regions* are the usual regions constructed by restricting each parameter in the conventional parameterization to lie between 0 and 1

The value of M for each model is given below.

ANN: $M = 1 - \alpha$

ADN: $M = \begin{bmatrix} 1-\alpha & 1-\alpha \\ -\beta & \phi-\beta \end{bmatrix}$

ANA: $M = \begin{bmatrix} 1-\alpha & \mathbf{0}'_{m-1} & -\alpha \\ -\gamma & \mathbf{0}'_{m-1} & 1-\gamma \\ \mathbf{0}_{m-1} & I_{m-1} & \mathbf{0}_{m-1} \end{bmatrix}$

ADA: $M = \begin{bmatrix} 1-\alpha & 1-\alpha & \mathbf{0}'_{m-1} & -\alpha \\ -\beta & \phi-\beta & \mathbf{0}'_{m-1} & -\beta \\ -\gamma & -\gamma & \mathbf{0}'_{m-1} & 1-\gamma \\ \mathbf{0}_{m-1} & \mathbf{0}_{m-1} & I_{m-1} & \mathbf{0}_{m-1} \end{bmatrix}$

Again, for AAN and AAA, the analogous result is obtained from ADN and ADA by setting $\phi = 1$.

We establish stability and forecastability conditions for each of the linear models. For the damped models, we assume ϕ is a fixed damping parameter between 0 and 1, and we consider the values of the other parameters that would lead to a stable model.

The stability conditions for models without seasonality (i.e., ANN, AAN and ADN) are summarized in Table 2. These are given in McClain and Thomas (1973) for the AAN model; results for the ADN and ANN models are obtained in a similar way. To visualize these regions, we have plotted them in Fig. 1. The light-shaded regions represent the stability regions; the dark-shaded regions are the usual regions defined by $0 < \beta < \alpha < 1$. Note that the usual parameter region is entirely within the stability region in each case. Therefore non-seasonal models obtained using the usual constraints are always stable (and always forecastable).

3 Two seasonal models

Consider the ANA model, for which the rank of $(F')^{p-1}H' < p$ and the rank of $F^{p-1}G < p$. This is because, for the ANA model, $(F')^{p-1} = F^{p-1} = I_{p+1}$ where I_{p+1} denotes the $(p+1) \times (p+1)$ identity matrix (see Definition 1 and Definition 2 of Sect. 2). Therefore, model ANA is neither reachable nor observable. Further, its characteristic equation for matrix M is $f(\lambda) = (1-\lambda)P(\lambda) = 0$ where

$$P(\lambda) = \lambda^m + \alpha\lambda^{m-1} + \alpha\lambda^{m-2} + \cdots + \alpha\lambda^2 + \alpha\lambda + (\alpha + \gamma - 1). \tag{5}$$

Thus, M has a unit eigenvalue regardless of the values of the model parameters, and so the model is always unstable. A similar argument shows that models AAA and ADA are also neither reachable nor observable and always unstable. These problems arise because of a redundancy in the model. For example, the ANA model is given by $y_t = \ell_{t-1} + s_{t-m} + \varepsilon_t$ where the level and seasonal components are given by

$$\ell_t = \ell_{t-1} + \alpha\varepsilon_t \quad \text{and} \quad s_t = s_{t-m} + \gamma\varepsilon_t.$$

So both level and seasonal components have long run features due to unit roots. In other words, both can model the level of the series and the seasonal component is not constrained to lie anywhere near zero.

In fact, by expanding $s_t = e_t/(1 - B^m)$ where $e_t = \gamma\varepsilon_t$ and B is the backshift operator, it can be seen that s_t can be decomposed into two processes, a level displaying a unit root at the zero frequency and a purely seasonal process, having unit roots at the seasonal frequency:

$$\begin{aligned} & s_t = \ell_t^* + s_t^* \\ \text{where} \qquad & \ell_t^* = \ell_{t-1}^* + \tfrac{1}{m} e_t, \\ & S(B) s_t^* = \theta(B) e_t, \end{aligned}$$

$S(B) = 1 + B + \cdots + B^{m-1}$ representing the seasonal summation operator and

$$\theta(B) = \frac{1}{m}\left[(m-1) + (m-2)B + \cdots + 2B^{m-3} + B^{m-2}\right].$$

The long run component ℓ_t^* should be part of the level term.

This leads to an alternative model specification where the seasonal equation for models ANA, AAA and ADA is replaced by

$$S(B)s_t = \theta(B)\gamma\varepsilon_t. \tag{6}$$

The other equations remain the same as the additional level term can be absorbed into the original level equation by a simple change of parameters. Noting that $\theta(B)/S(B) = [1 - \frac{1}{m}S(B)]/(1 - B^m)$, we see that (6) can be written as

$$s_t = s_{t-m} + \gamma\varepsilon_t - \frac{\gamma}{m}\left[\varepsilon_t + \varepsilon_{t-1} + \cdots + \varepsilon_{t-m+1}\right].$$

In other words the seasonal term is calculated as in the original models, but then adjusted by subtracting the average of the last m shocks. The effect of this adjustment is equivalent to the normalized updating proposal of Roberts (1982) in which the seasonal terms $s_t, \ldots, s_{t-m+1}$ are adjusted every time period to ensure they sum to zero. Models using the seasonal component (6) will be referred to as "normalized" versions of ANA, AAA and ADA.

4 Stability of seasonal models

4.1 Standard models

As noted in the previous section, M has a unit eigenvalue in each of the seasonal models ANA, AAA and ADA. In fact, the characteristic equation of M for model ADA is $f(\lambda) = (1 - \lambda)P(\lambda) = 0$ where

$$\begin{aligned} P(\lambda) = \lambda^{m+1} &+ (\alpha + \beta - \phi)\lambda^m + (\alpha + \beta - \alpha\phi)\lambda^{m-1} + \cdots + (\alpha + \beta - \alpha\phi)\lambda^2 \\ &+ (\alpha + \beta - \alpha\phi + \gamma - 1)\lambda + \phi(1 - \alpha - \gamma). \end{aligned} \tag{7}$$

However, it is easy to see that the eigenvector associated with $\lambda = 1$ is orthogonal to $\boldsymbol{f}_h = HF^{h-1}$. For example, with ADA the eigenvector is $\boldsymbol{v}_1 = [-1, 0, 1, \ldots, 1]'$ and $\boldsymbol{f}_h = [1, \phi_h, k_{1,h}, \ldots, k_{m,h}]$ where $k_{i,h} = 1$ if $i + h = 1$ (mod m) and $k_{i,h} = 0$ otherwise. Thus $\boldsymbol{f}'_h\boldsymbol{v}_1 = 0$. Therefore, the models can still be forecastable, even though they are not stable. No other eigenvectors are orthogonal to $\boldsymbol{f}_h$. Forecastability requires the roots of $P(\lambda)$ to lie inside the unit circle. The conditions for forecastability are derived in the Appendix and summarized in Table 3.

The inequalities involving only α and γ provide necessary conditions for forecastability that are easily implemented. The final condition (giving a range for β) is more complicated to use in practice than finding the numerical roots of (7). Therefore, we suggest that in practice the conditions on α and γ be checked first, and if satisfied, then the roots of (7) be calculated and tested.

Table 3 Forecastability conditions for models ANA and ADA

ANA:	$\max(-m\alpha, 0) < \gamma < 2-\alpha$ and $\frac{-2}{m-1} < \alpha < 2-\gamma$
ADA:	$0 < \phi \le 1$ $\max(1-1/\phi-\alpha, 0) < \gamma < 1+1/\phi-\alpha$ $1-1/\phi-\gamma(1-m+\phi+\phi m)/(2\phi m) < \alpha < (B+C)/(4\phi)$ $-(1-\phi)(\gamma/m+\alpha) < \beta < D+(\phi-1)\alpha$
where	$B = \phi(4-3\gamma)+\gamma(1-\phi)/m$ $C = \sqrt{B^2-8\left[\phi^2(1-\gamma)^2+2(\phi-1)(1-\gamma)-1\right]+8\gamma^2(1-\phi)/m}$ $D = \min_\theta \left\{(\phi-\phi\alpha+1)(1-\cos\theta)-\gamma\left[\frac{(1+\phi)(1-\cos\theta-\cos m\theta)+\cos(m-1)\theta+\phi\cos(m+1)\theta}{2(1-\cos m\theta)}\right]\right\}$ and θ is a solution to $\frac{\phi\alpha-\phi+1}{\gamma} + \frac{(\phi-1)(1+\cos\theta-\cos m\theta)+\cos(m-1)\theta-\phi\cos(m+1)\theta}{2(1+\cos\theta)(1-\cos m\theta)} = 0.$

Conditions for AAA can be obtained from ADA by setting $\phi = 1$

To visualize these regions, we have plotted them in Figs. 2 and 3. The light-shaded regions represent the forecastability regions; the dark-shaded regions are the usual regions where each parameter (in the Hyndman et al. (2002) parameterization) lies in [0,1].

The forecastable region for α and γ is illustrated in Fig. 2. The upper limit of γ is obtained when the upper limit of α equals the lower limit of α. For $\phi = 1$ this simplifies to $\gamma < 2m/(m-1)$ as given by Archibald (1991), but for smaller values of ϕ the upper limit of γ is slightly smaller than this.

The right hand column of Fig. 2 shows that the usual parameter region of an ANA model is entirely within the forecastability region. Therefore ANA models obtained using the usual constraints are always forecastable.

The forecastable region for α and β is depicted in Fig. 3 for $m = 4$. From Fig. 3, it can be seen that the usual parameter region and the forecastability region intersect for model ADA but neither is contained within the other. Therefore, models obtained using the usual constraints may not be forecastable. This problem is greatest when the seasonal smoothing parameter γ is large.

Note that Sweet (1985) and Lawton (1998) in discussing this problem used a different parameterization where they required $\gamma/(1-\alpha) < 1$. Under this parameterization, the model is always forecastable for $m \le 4$ when the parameters lie in $[0, 1]$. However, under our parameterization, the model may not be forecastable, even when $m \le 4$.

4.2 Normalized models

Archibald (1984, 1990) discussed the stable region for the normalized version of AAA and Archibald (1991) provides some preliminary steps towards the stable region for the normalized version of ADA.

To write the normalized model in state space form, we need to use a different state vector given by $\boldsymbol{x}_t = (\ell_t, b_t, s_{1,t}, \ldots, s_{m-1,t})'$. Here, $s_{i,t}$ denotes the

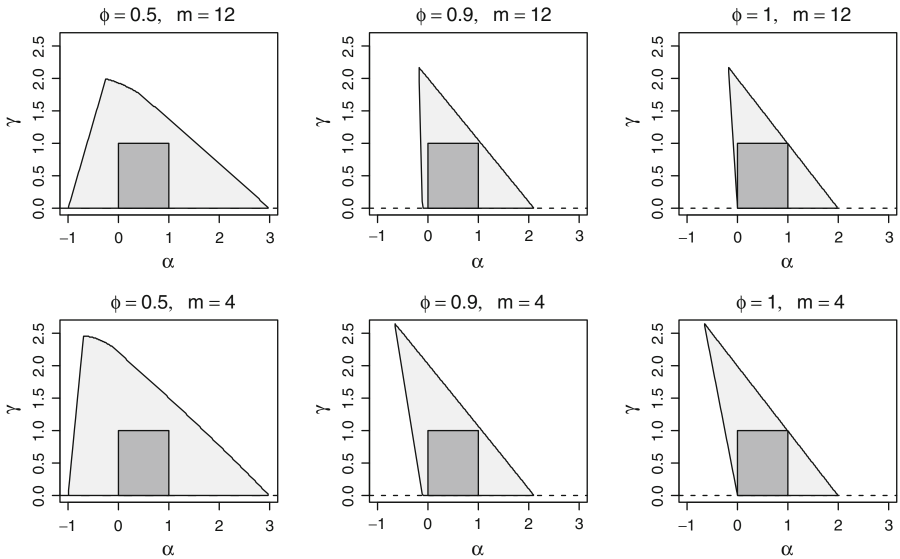

Fig. 2 *Light shaded region* the forecastable region of α and γ for model ADA. *Dark shaded region* usual region where both parameters are bounded by 0 and 1. The *right column* shows the regions for model AAA (when $\phi = 1$). These are also the regions for model ANA as they are independent of β

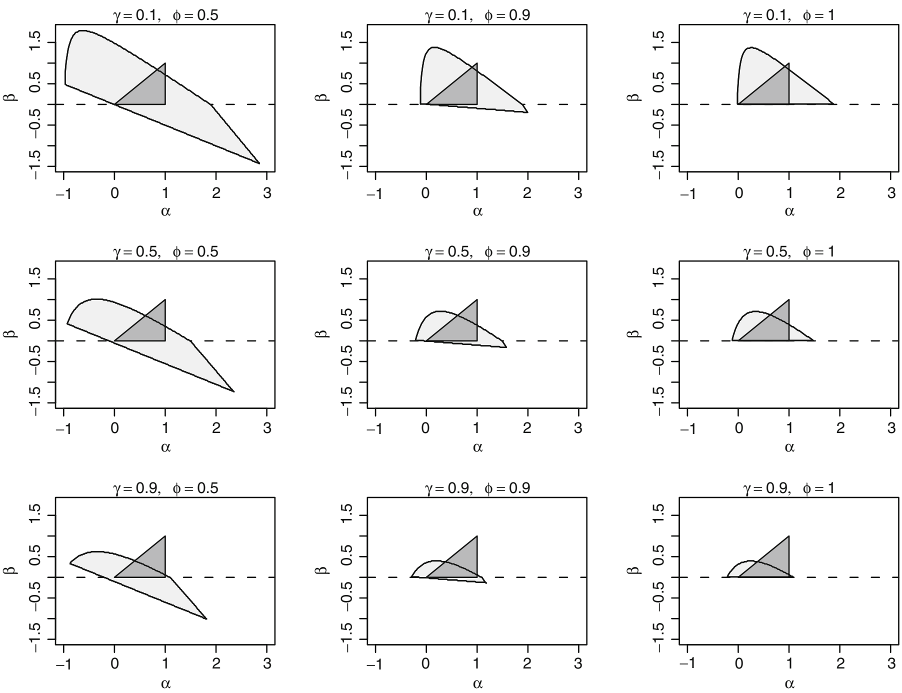

Fig. 3 *Light shaded region* the forecastable region of α and β for model ADA with $m = 4$. *Dark shaded region* usual region where all parameters in the Hyndman et al. (2002) parameterization are bounded by 0 and 1 ($0 < \beta < \alpha$). The right column shows the region for model AAA (when $\phi = 1$)

estimate of the seasonal factor for the ith month ahead made at time t. Note that $s_{m,t} \equiv s_{0,t} = 1 - s_{1,t} - \cdots - s_{m-1,t}$. Following (Roberts, 1982, Sect. 3), the seasonal updating is defined as follows.

$$\begin{aligned} s_{0,t} &= s_{1,t-1} + \gamma(1 - \tfrac{1}{m})e_t \\ s_{i,t} &= s_{i+1,t-1} - \tfrac{\gamma}{m}e_t. \end{aligned}$$

The level and trend equations are updated as with the standard model. Then $H = [1, 1, 1, \mathbf{0}'_{m-2}]$,

$$F = \begin{bmatrix} 1 & 1 & 0 & \mathbf{0}'_{m-2} \\ 0 & \phi & 0 & \mathbf{0}'_{m-2} \\ \mathbf{0}_{m-2} & \mathbf{0}_{m-2} & \mathbf{0}_{m-2} & I_{m-2} \\ 0 & 0 & -1 & -\mathbf{1}'_{m-2} \end{bmatrix}, \qquad G = \begin{bmatrix} \alpha \\ \beta \\ -(\gamma/m)\mathbf{1}_{m-1} \end{bmatrix}$$

and

$$M = \begin{bmatrix} 1-\alpha & 1-\alpha & -\alpha & \mathbf{0}'_{m-2} \\ -\beta & \phi-\beta & -\beta & \mathbf{0}'_{m-2} \\ (\gamma/m)\mathbf{1}_{m-2} & (\gamma/m)\mathbf{1}_{m-2} & (\gamma/m)\mathbf{1}_{m-2} & I_{m-2} \\ \gamma/m & \gamma/m & \gamma/m - 1 & -\mathbf{1}'_{m-2} \end{bmatrix},$$

where $\mathbf{1}_k$ denotes a k-vector of ones. The characteristic equation for M is given by

$$f(\lambda) = \sum_{i=0}^{m+1} \theta_i \lambda^{m+1-i} \tag{8}$$

$$\begin{aligned} \text{where} \quad \theta_0 &= 1 \\ \theta_1 &= \alpha + \beta - \gamma/m - \phi \\ \theta_i &= \alpha(1-\phi) + \beta - (1-\phi)\gamma/m, \qquad i = 2, \ldots, m-1 \\ \theta_m &= \alpha(1-\phi) + \beta + \gamma[1 - (1-\phi)/m] - 1 \\ \text{and} \quad \theta_{m+1} &= \phi[1 - \gamma(1 - 1/m) - \alpha]. \end{aligned}$$

Note that this is equivalent to (7) if we reparamaterize the model, replacing α in (7) by $\alpha - \gamma/m$. Therefore the forecastability conditions for the standard ADA model are the same as the stability conditions for the normalized ADA model, apart from this minor reparameterization.

5 Comparison with the basic structural model

An alternative seasonal model is Harvey's (1989) basic structural model (BSM), which is similar to the normalized AAA model except $\theta(B)$ is dropped in (6) and it uses multiple disturbance terms that are independent of each other. That is,

$$
\begin{aligned}
Y_t &= \ell_t + s_t + \varepsilon_t \\
\ell_t &= \ell_{t-1} + b_{t-1} + \eta_t \\
b_t &= b_{t-1} + \zeta_t \\
S(B)s_t &= \sum_{j=0}^{m-1} s_{t-j} = \omega_t
\end{aligned}
$$

where η_t, ζ_t and ω_t are mutually uncorrelated white-noise disturbances with zero mean and variances $\sigma_\eta^2, \sigma_\zeta^2$ and σ_ω^2, respectively.

The disturbances of the BSM can be combined to give a model with a single disturbance, known as the "reduced form" which is equivalent to an ARIMA$(0,1,m+1)(0,1,0)_m$ model with some parameter constraints (Harvey, 1989, p.69). Stability of the reduced form requires a positive value of the seasonal disturbance variance, σ_ω^2, irrespective of the values of σ_η^2 and σ_ζ^2.

Note that the normalized AAA model can also be written as an ARIMA $(0,1,m+1)(0,1,0)_m$ model with some parameter constraints (Roberts, 1982), and the characteristic equation (8) shows that the ADA model is also of this form. So we can compare the BSM with the ADA and AAA models by writing all models in reduced ARIMA form.

Figure 4 shows some projections of the parameter spaces of the two model classes onto the two-dimensional space spanned by the ARIMA parameters θ_1 and θ_m. Each graph of the AAA model is for different values of the parameter γ. For this model, $\gamma = (1 - \theta_1 + \theta_m - \theta_{m+1})/2$. To enable comparisons with the BSM, we have computed the same quantity for the BSM. In both models, high values of γ correspond to a rapidly changing seasonal pattern. In the BSM, this corresponds to a large value of σ_ω^2 relative to σ_ε^2. In Fig. 4, the dark shaded region corresponds to the admissible parameter space of the BSM and the light shaded region corresponds to the admissible parameter space for the exponential smoothing model. Note that, the BSM parameter space is very small for large γ which will make it too restrictive for use with some data sets.

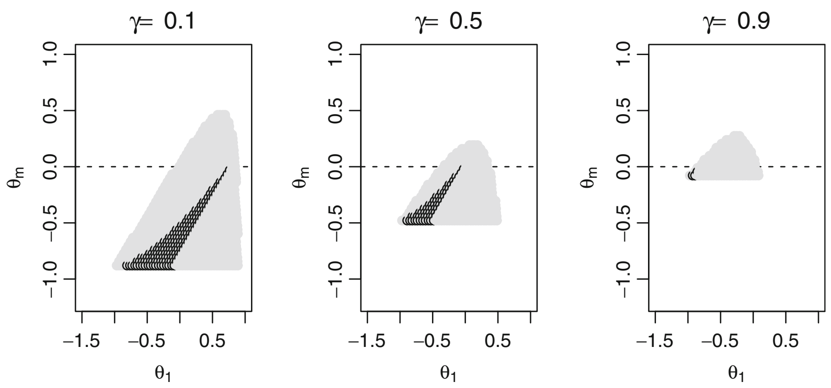

Fig. 4 Admissible parameter space of BSM (*dark shaded region*) and for the normalized AAA (*light shaded region*)

6 Examples

We apply the normalized AAA model and the BSM model to some real data to demonstrate the differences between the models in application. The data used are Australian electricity production data (taken from Makridakis et al., 1998), sales of Australian wine (fortified wine and sparkling bulk), Australian and Turkish data on construction permits issued, Germany data on sales of intermediate goods, and sales and retail trade of car registration data from the UK and Ireland. The sales of Australian wine data are taken from Australian Bureau of Statistics web site (http://www.abs.gov.au (ABS Cat. No. 8504.0)), and the remaining data sets (AUS.ODCNPI03.ML, TUR.ODCNPI02.ML, DEU.SLMNIG01.IXOB, IRL.SLRTCR01.ML, GBR.SLRTCR03.ML) are taken from DX database (Australia). The Australian electricity production (AEP) data along with the Turkish construction permits (TCP) data are shown in Fig. 5. Note that all series have changing seasonal patterns over time. Thus it is expected that γ will be relatively high in each case.

For each logged data series, we hold out the last four years of data and fit the normalized AAA and basic structural models to the rest of the data using maximum likelihood estimation. The estimates are constrained to give a stable model, and are given in Table 4. Note that the parameter estimates of σ^2_ζ for the BSM are very small. This is a result of the small admissible parameter space for this model.

To compare the performance of the models, we have plotted forecasts from AEP and TCP in Fig. 6. The actual values in the forecast period are also shown. The normalized AAA model gives reasonable forecasts, while those from the BSM have not reacting sufficiently to recent changes in the seasonal pattern seen in the data. This is a direct result of the small estimated parameters, which occur because of the restrictive admissible parameter spaces. Similar performance of forecasts were obtained in other data sets. For all data sets we have calculated the root mean squared error (RMSE) over the forecast period for each model. These measures of accuracy are given in Table 5. While these are only examples, and so general conclusions may not be drawn, the evidence is consistent with Fig. 4 in suggesting that the small parameter space of the BSM can harm forecast accuracy, especially when seasonal patterns are changing.

7 Results from Monte Carlo simulation

We have also done a simulation study in order to gain further insight into the findings of the previous section. First, we simulated data from the AAA model by keeping all parameters and the initial state vector constant. Than we fitted the normalized AAA and the BSM to the simulated data. Second, we simulated data from the BSM and repeated the model fitting. To ensure realistic models, we use estimates of the model applied to real data as an initial state vector to start the simulations.

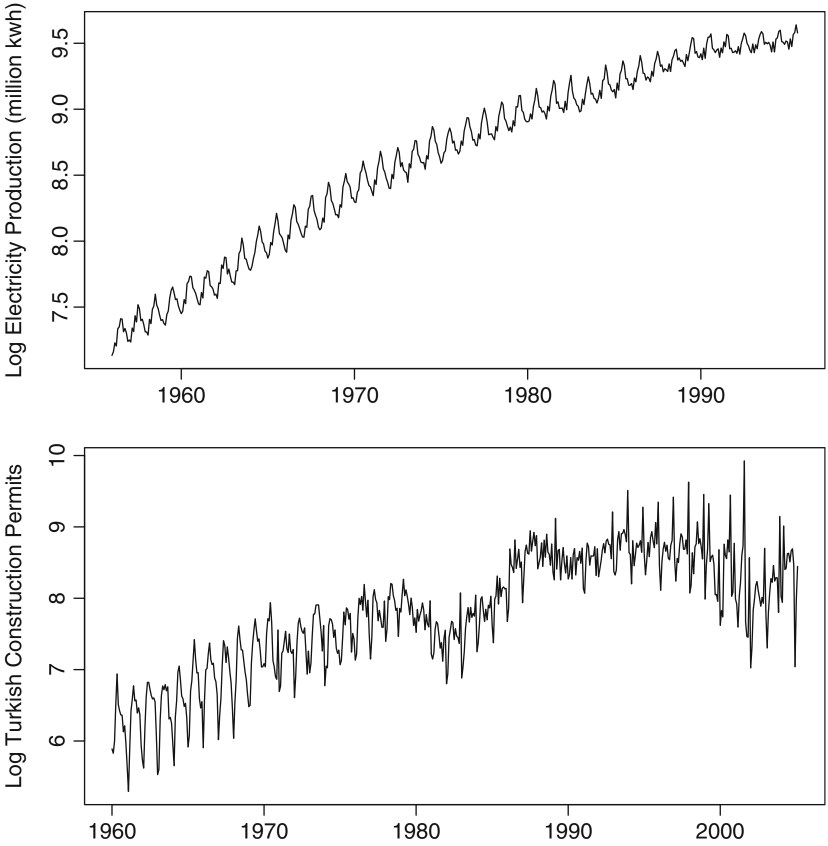

Fig. 5 *Top panel* Australian electricity production, January 1956–August 1995. *Bottom panel* Construction permits issued for buildings in Turkey, January 1960–March 2005

Table 4 Maximum likelihood estimates of the parameters for each seasonal model applied to each seasonal data

	AAA			***BSM*** **($\sigma^2 \times 10^{-4}$)**				
	α	β	γ	σ^2_ε	σ^2_η	σ^2_ζ	σ^2_ω	γ
AEP	0.2705	0.0047	0.4872	1.28	0.71	0.0	0.57	0.3578
Sparkling bulk	0.2752	0.0025	0.5158	143.97	75.83	0.0	7.55	0.1329
Fortified wine	0.0508	0.0030	0.7258	51.42	0.06	0.0	5.92	0.2458
Aus C.P.	0.5198	0.0016	0.2538	33.86	24.58	0.0	0.06	0.0430
Turkey C.P.	0.3219	0.0081	0.3490	133.20	59.17	0.0	42.62	0.3256
Germany	0.3092	0.0016	0.3089	10.43	2.03	0.0	0.27	0.1228
Ireland	0.2203	0.0003	0.4852	164.94	47.40	0.0	32.06	0.3030
UK	0.0.1963	0.0003	0.7809	37.88	18.86	0.0	89.23	0.5738

Last column shows the value of γ computed for the BSM

We simulated 1,000 monthly series of size 500 each from both models. For the model AAA, we use $\alpha = 0.27$, $\beta = 0.005$, $\gamma = 0.49$. The initial level $\ell_0 = 7$, initial trend $b_0 = 0.007$, and initial seasonal estimates are

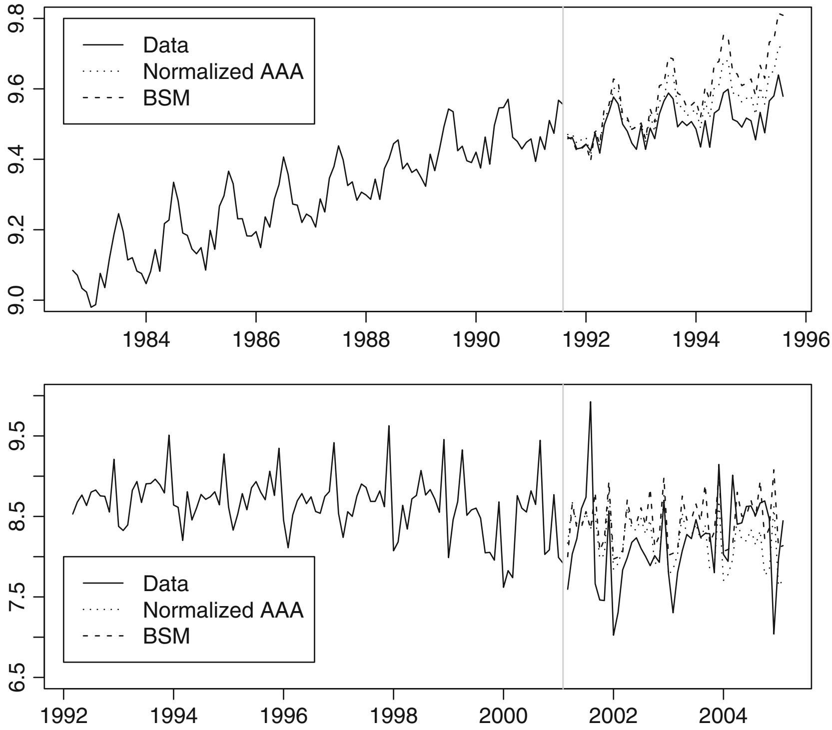

Fig. 6 *Top panel* Historical data and forecasts from Australian electricity production data, forecasts are based on models fitted to the data up to August 1991. *Bottom panel* Historical data and forecasts from construction permits issued for buildings in Turkey, forecasts are based on models fitted to the data up to March 2001

Table 5 Root mean squared error for each seasonal model

	AEP	Sparkling bulk	Fortified wine	Aus CP	Turkey CP	Germany	Ireland	UK
AAA	0.3246	1.4653	0.1644	0.9502	0.2474	0.3063	1.2431	0.2690
BSM	0.5520	1.7442	0.1825	0.9922	1.8062	0.3940	1.6595	0.2926

$s_{-11} = -0.0947, s_{-10} = -0.1045, s_{-9} = -0.0179, s_{-8} = -0.0473, s_{-7} = 0.0679$, $s_{-6} = 0.0614$, $s_{-5} = 0.1309$, $s_{-4} = 0.0988$, $s_{-3} = 0.0097$, $s_{-2} = 0.01$, $s_{-1} = -0.0304, s_0 = -0.0839$. For the BSM, we use $\sigma_\eta = 0.00843, \sigma_\varepsilon = 0.01133, \sigma_\zeta = 0$, $\sigma_\omega = 0.00755$, initial level (ℓ_0)= 7.5, initial trend (b_0)= 0.0052 and initial seasonal estimates are $s_{-11} = -0.04074, s_{-10} = -0.092847, s_{-9} = -0.01227, s_{-8} = -0.054205$, $s_{-7} = 0.043273$, $s_{-6} = 0.045063$, $s_{-5} = 0.11442$, $s_{-4} = 0.10393$, $s_{-3} = 0.0014834, s_{-2} = -0.011222, s_{-1} = -0.04927, s_0 = -0.047618$.

To compare the performance of the models through simulation, we have provided the forecast accuracy measure, root mean squared error (RMSE), over the forecast period for each simulated model. For the simulated data from the model AAA, RMSE = 5.7310 for the normalized AAA model and RMSE = 6.3859 for the BSM. While for the simulated data from BSM, RMSE = 0.2269 for the normalized AAA model and RMSE = 0.2576 for the BSM. These figures confirm the results of the previous section and are consistent with our finding that the larger parameter space of the AAA model allows greater forecast accuracy. Surprisingly, the AAA model performs (slightly) better then the BSM, even when the BSM is the true data generating process. This shows that the AAA is able to closely mimic the BSM for forecasting purposes.

8 Conclusions

With the non-seasonal exponential smoothing models, our results are clear—the models are stable using the usual constraints. In fact, it is possible to allow parameters to take values in a larger space, and still retain a stable model. The stability region is identical to that for the equivalent invertible ARIMA model. This is in contrast to the stability region for the analogous structural models of Harvey (1989) which require a reduced parameter space.

With the seasonal exponential smoothing methods, the situation is more complicated. The most striking results derived here show that the usual Holt–Winters' equations are fundamentally flawed, being unstable for any values of the model parameters. The problem arises because of the unit root in the seasonal component, which occurs because the seasonal states are not constrained. However, we have shown that the model can be made "forecastable", so that the forecast distributions are unaffected by the instability.

The normalized model (introduced by Roberts, 1982) circumvents this problem by requiring the seasonal states to sum to zero. Thus, stability in a seasonal model can be achieved via the simple step of removing the inherent redundancy in the seasonal terms.

The BSM achieves a similar result by requiring the seasonal states to have mean zero. But the resulting parameter space is much smaller than for the normalized model, and our examples and simulations show that this can lead to poorer forecasts.

Acknowledgements We would like to thank Richard Lawton for useful comments on an earlier draft of this work which motivated our definition of "forecastability". Keith Ord, Andrew Harvey, Simone Grose and two referees also provided helpful comments which led to several improvements of the paper. This research was supported by the Business and Economic Forecasting Unit, Monash University.

Appendix: Proof of results in Table 3

The Schur Method may be used to determine whether any zero of a polynomial lies within the unit circle.

Definition Let $f(z) = a_0 + a_1 z + a_2 z^2 + \cdots + a_n z^n$ be a polynomial of degree n with real coefficients. Then the Schur Transformation of $f(z)$ is

$$T[f(z)] = a_0 f(z) - a_n z^n f(z^{-1}).$$

We shall denote multiple transformations using a superscript notation: $T^j[f(z)] = T[T^{j-1}f(z)]$.

The following lemma is a corollary of Theorem 8.4 of Ralston (1965).

Lemma 1 (Schur Method) *Let $f(z) = a_0+a_1z+a_2z^2+\cdots+a_nz^n$ be a polynomial of degree n with real coefficients where $a_0 \neq 0$ and define*

$$g(z) = a_0z^n + a_1z^{n-1} + \cdots + a_{n-1}z + a_n.$$

Then all roots of $f(z)$ have modulus less than 1 if and only if

$$T^j[g(0)] > 0 \quad \textit{for } j = 1, 2, \ldots, k$$

where $k \leq n$, $T^k[g(0)] = 0$ and $T^{k-1}[g(z)]$ is constant.

The characteristic equation of model ADA is $f(\lambda) = (1-\lambda)P(\lambda) = 0$ where $P(\lambda)$ is given by (7). Our approach will be to consider λ with moduli 1, and then determine what values of the smoothing parameters lead to a solution to the characteristic equation. This gives us the boundary of the region: when the parameters are inside all these bounds the moduli of all roots are less than 1 and the model is forecastable. For a few λ values we can examine the equation $P(\lambda) = 0$ and easily obtain a boundary. For general λ, we will have to examine $\|P(\lambda)\| = 0$ which involves a lot of algebraic manipulation, for which we only present an outline.

Now $P(1) = m(\alpha+\beta-\alpha\phi) + \gamma(1-\phi)$. So $P(\lambda)$ has a unit root if and only if $(\alpha+\beta-\alpha\phi) = \gamma(\phi-1)/m$. Setting $P(1) = 0$, we get what turn out to be the lower limit on β for fixed α and γ. Therefore to ensure the roots are within the unit circle we require

$$\beta > -(1-\phi)(\alpha+\gamma/m). \tag{9}$$

Another simple bound is obtained by noting that if $\lambda \neq 1$ then $P(\lambda)$ can be written as

$$P(\lambda) = (\lambda^m - 1)(1+\alpha\phi-\phi) + \frac{(\alpha+\beta-\alpha\phi)\lambda(1-\lambda^m)}{1-\lambda} + \gamma(\lambda-\phi)$$

If we consider any λ that is a solution to $\lambda^m = 1$ and $P(\lambda) = 0$ (other than $\lambda = 1$) we have $\gamma(\lambda-\phi) = 0$ which gives $\gamma = 0$. So a lower bound is

$$\gamma > 0. \tag{10}$$

Now setting $(\alpha+\beta-\alpha\phi) = \gamma(\phi-1)/m$ in $P(\lambda)$, and dividing the resultant equation by $(\lambda-1)$, we get

$$f^*(\lambda) = \frac{P(\lambda)}{\lambda-1} = \lambda^m + (b+c)\lambda^{m-1} + (2b+c)\lambda^{m-2} + \cdots + [(m-1)b+c]\lambda - \phi(1-\alpha-\gamma)$$

where $b = \alpha + \beta - \alpha\phi$ and $c = \phi(\alpha - 1) + 1$. Then applying Lemma 1 to $f^*(\lambda)$ we get the additional following conditions for forecastability:

$$1 - 1/\phi < \alpha + \gamma < 1 + 1/\phi \qquad \text{and} \qquad B - C < 4\phi\alpha < B + C \tag{11}$$

$$\begin{aligned}\text{where}\qquad & C = \sqrt{B^2 - 8\left[\phi^2(1-\gamma)^2 - 2(1-\phi)(1-\gamma) - 1\right] + 8\gamma^2(1-\phi)/m} \\ \text{and}\qquad & B = \phi(4 - 3\gamma) + \gamma(1-\phi)/m.\end{aligned}$$

The upper bound on β is much more difficult to obtain, and we give only an outline of the procedure here. A more detailed version can be obtained from the authors. Using the polar coordinate system, we define $\lambda = \cos\theta + i\sin\theta$ so that we can write

$$\begin{aligned}P(\lambda) = {} & a + (b + \gamma - 1)\cos\theta + b\cos 2\theta + b\cos 3\theta + \cdots + b\cos(m-1)\theta \\ & + (b + \alpha\phi - \phi)\cos m\theta + \cos(m+1)\theta + i\Big[(b + \gamma - 1)\sin\theta + b\sin 2\theta \\ & + \cdots + b\sin(m-1)\theta + (b + \alpha\phi - \phi)\sin m\theta + \sin(m+1)\theta\Big]\end{aligned}$$

where $a = \phi(1 - \alpha - \gamma)$ and $b = \alpha + \beta - \alpha\phi$. Then

$$\begin{aligned}& |P(\cos\theta + i\sin\theta)|^2 \\ & = 2\Big[1 + \phi^2 - 2\phi\cos\theta + \phi\cos(m-1)\theta - \phi^2\cos m\theta - \cos m\theta \\ & \quad + \phi\cos(m+1)\theta\Big] + 2\phi^2\alpha^2(1 - \cos m\theta) + b^2(1 - \cos m\theta)/(1 - \cos\theta) \\ & \quad + \gamma^2(1 + \phi^2 - 2\phi\cos\theta) + 2b\Big[\gamma\big\{(1-\phi)(\cos\theta + \cdots + \cos(m-1)\theta) \\ & \quad - \phi\cos m\theta + 1\big\} - \big\{\phi(1-\alpha) + 1\big\}(1 - \cos m\theta)\Big] \\ & \quad + 2\gamma\Big[2\phi\cos\theta + (1 + \phi^2)(\cos m\theta - 1) - \phi\cos(m-1)\theta - \phi\cos(m+1)\theta\Big] \\ & \quad - 2\phi\alpha\gamma\Big[\cos\theta - \cos(m-1)\theta - \phi(1 - \cos m\theta)\Big] \\ & \quad + 2\phi\alpha\Big[-2\phi + 2\cos\theta - \cos(m-1)\theta + 2\phi\cos m\theta - \cos(m+1)\theta\Big].\end{aligned} \tag{12}$$

Since the above function is positive by definition and quadratic in α, b, and γ, we have to determine the minimum value of b for which (12) is equal to zero. Differentiating (12) with respect to b and setting the result to zero gives the upper bound on b for fixed α and γ: $b < D$ where $D = [\phi(1-\alpha) + 1](1 - \cos\theta) - \gamma\psi(\theta, \phi)$ and

$$\psi(\theta, \phi) = \frac{(1 + \phi)(1 - \cos\theta - \cos m\theta) + \cos(m-1)\theta + \phi\cos(m+1)\theta}{2(1 - \cos m\theta)}.$$

Equivalently

$$\beta < D - \alpha(1-\phi). \tag{13}$$

In expression (13), only θ is unknown while α, γ and ϕ are fixed. Now we have to find the value of θ, for which b is minimum. We substitute (13) in (12) and simplify using the trigonometric identity

$$1+\cos\theta+\cos 2\theta+\cdots+\cos(n-1)\theta = \frac{(\cos n\theta - 1)(\cos\theta - 1) + \sin\theta \sin n\theta}{(\cos\theta-1)^2+\sin^2\theta}$$

to obtain

$$\begin{aligned}
&|P(\cos\theta + i\sin\theta)|^2 \\
&\quad = 2\Big[1+\phi^2 - 2\phi\cos\theta + \phi\cos(m-1)\theta - \phi^2\cos m\theta - \cos m\theta + \phi\cos(m+1)\theta\Big] \\
&\qquad + 2\phi^2\alpha^2(1-\cos m\theta) - \frac{(1-\cos m\theta)}{(1-\cos\theta)}\Big[\big\{\phi(1-\alpha)+1\big\}(1-\cos\theta) - \gamma A(\theta,\phi)\Big]^2 \\
&\qquad + 2\gamma\Big[2\phi\cos\theta + (1+\phi^2)(\cos m\theta - 1) - \phi\cos(m-1)\theta - \phi\cos(m+1)\theta\Big] \\
&\qquad - 2\phi\alpha\gamma\Big[\cos\theta - \cos(m-1)\theta - \phi(1-\cos m\theta)\Big] + \gamma^2(1+\phi^2-2\phi\cos\theta) \\
&\qquad + 2\phi\alpha\Big[2\cos\theta - 2\phi - \cos(m-1)\theta + 2\phi\cos m\theta - \cos(m+1)\theta\Big].
\end{aligned} \tag{14}$$

Then partially differentiating (14) with respect to α and equating the result to zero gives

$$\frac{\phi\alpha - \phi + 1}{\gamma} + \frac{(\phi-1)(1+\cos\theta-\cos m\theta) + \cos(m-1)\theta - \phi\cos(m+1)\theta}{2(1+\cos\theta)(1-\cos m\theta)} = 0 \tag{15}$$

Then θ will be a solution to (15). We solve this equation numerically for given α, γ and ϕ. We consider only $\theta \in (0,\pi)$ as outside this range gives identical results.

Combining results (9), (10), (11) and (13) gives the required parameter space for model ADA. Forecastability conditions for AAA are obtained by setting $\phi = 1$. Forecastability conditions for ANA are obtained from (10) and (11) by setting $\phi = 1$.

References

Archibald, B. C. (1984). Invertible region of Holt–Winters' model, Working paper 31/1984, School of Business Administration, Halifax: Dalhousie University.

Archibald, B. C. (1990). Parameter space of the Holt-Winters' model. *International Journal of Forecasting*, *6*, 199–229.

Archibald, B. C. (1991). Invertible region of damped trend, seasonal, exponential smoothing model, Working paper 10/1991, School of Business Administration, Halifax: Dalhousie University.
Hannan, E. J., Deistler, M. (1988). The statistical theory of linear systems. New York: Wiley.
Harvey, A. C. (1989). Forecasting, structural time series models and the Kalman filter. Cambridge: Cambridge University Press.
Hyndman, R. J., Koehler, A. B., Ord, J. K., Snyder, R. D. (2005). Prediction intervals for exponential smoothing state space models. *Journal of Forecasting*, *24*, 17–37.
Hyndman, R. J., Koehler, A. B., Snyder, R. D., Grose, S. (2002). A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of Forecasting*, *18*(3), 439–454.
Lawton, R. (1998). How should additive Holt-Winters' estimates be corrected? *International Journal of Forecasting*, *14*, 393–403.
Makridakis, S., Wheelwright, S. C., Hyndman, R. J. (1998). Forecasting: methods and applications (3rd ed.) New York: Wiley.
McClain, J. O., Thomas, L. J. (1973). Response–variance tradeoffs in adaptive forecasting. *Operations Research*, *21*, 554–568.
Ord, J. K., Koehler, A. B., Snyder, R. D. (1997). Estimation and prediction for a class of dynamic nonlinear statistical models. *Journal of American Statistical Association*, *92*, 1621–1629.
Ralston, A. (1965). A first course in numerical analysis. New York: McGraw-Hill.
Roberts, S. A. (1982). A general class of Holt-Winters type forecasting models. *Management Science*, *28*(8), 808–820.
Snyder, R. D., Forbes, C. S. (2003). Reconstructing the Kalman filter for stationary and non-stationary time series. *Studies in nonlinear dynamics and econometrics*, *7*(2), 1–18.
Snyder, R. D., Ord, J. K., Koehler, A. B. (2001). Prediction intervals for ARIMA models. *Journal of Business and Economics Statists*, *19*(2), 217–225.
Sweet, A. L. (1985). Computing the variance of the forecast error for the Holt-Winters seasonal models. *Journal of Forecasting*, *4*, 235–243.