ISSN 1440-771X

MONASH UNIVERSITY
AUSTRALIA

DEPARTMENT OF ECONOMETRICS
AND BUSINESS STATISTICS

Invertibility Conditions for Exponential Smoothing Models

Rob J. Hyndman, Muhammad Akram and Blyth Archibald

Working Paper 3/2003

1.1 The modelling framework

We describe the exponential smoothing methods using a similar framework to that proposed in HKSG. Each method is denoted by two letters: the first letter denotes the type of trend (none, additive, multiplicative or damped) and the second letter denotes the type of seasonality (none, additive or multiplicative). Cell NN describes the simple exponential smoothing method, cell AN describes Holt's linear method. The additive Holt-Winters' method is given by cell AA and the multiplicative Holt-Winters' method is given by cell AM. The other cells correspond to less commonly used but analogous methods.

		Seasonal Component	
Trend Component	N (none)	A (additive)	M (multiplicative)
N (none)	NN	NA	NM
A (additive)	AN	AA	AM
M (multiplicative)	MN	MA	MM
D (damped)	DN	DA	DM

For each of these methods, HKSG proposed two state space models with a single source of error following the general approach of Ord, Koehler and Snyder (1997). The state space models enable easy calculation of the likelihood, and provide facilities to compute prediction intervals for each model. A single source of error model 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. 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. To distinguish these models, we add a third letter (A or M) before the letters denoting the type of trend and seasonality. For example, MAN refers to a model with multiplicative errors, additive trend and no seasonality. In this paper, we only consider the linear models with additive errors. Table 1 shows the equations for the models we consider in this paper.

Note that we use a slightly different parameterization from HKSG for the trend equation—we use β where HKSG used $\alpha\beta$. This change in parameters makes no difference to the models but allows us to have bounded invertibility regions. 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 $0 < \beta < \alpha$.

Trend component	Seasonal component N (none)	Seasonal component A (additive)
N (none)	$\mu_t = \ell_{t-1}$ $\ell_t = \ell_{t-1} + \alpha\varepsilon_t$ $\mu_n(h) = \ell_n$	$\mu_t = \ell_{t-1} + s_{t-m}$ $\ell_t = \ell_{t-1} + \alpha\varepsilon_t$ $s_t = s_{t-m} + \gamma\varepsilon_t$ $\mu_n(h) = \ell_n + s_{n-m+1+(h-1)*}$
A (additive)	$\mu_t = \ell_{t-1} + b_{t-1}$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t$ $b_t = b_{t-1} + \beta\varepsilon_t$ $\mu_n(h) = \ell_n + hb_n$	$\mu_t = \ell_{t-1} + b_{t-1} + s_{t-m}$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t$ $b_t = b_{t-1} + \beta\varepsilon_t$ $s_t = s_{t-m} + \gamma\varepsilon_t$ $\mu_n(h) = \ell_n + hb_n + s_{n-m+1+(h-1)*}$
D (damped)	$\mu_t = \ell_{t-1} + b_{t-1}$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t$ $b_t = \phi b_{t-1} + \beta\varepsilon_t$ $\mu_n(h) = \ell_n + \phi_h b_n$	$\mu_t = \ell_{t-1} + b_{t-1} + s_{t-m}$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t$ $b_t = \phi b_{t-1} + \beta\varepsilon_t$ $s_t = s_{t-m} + \gamma\varepsilon_t$ $\mu_n(h) = \ell_n + \phi_h b_n + s_{n-m+1+(h-1)*}$

Table 1: *State space equations for the models considered in this paper. In all cases, $Y_t = \mu_t + \varepsilon_t$. Point forecasts are given by $\mu_n(h)$. Here $\phi_j = 1 + \phi + \cdots + \phi^{j-1} = (1-\phi^j)/(1-\phi)$ and $(h-1)^* = (h-1)$ mod m.*

1.2 State space models

Let $Y_1, \dots, Y_n$ denote the time series of interest and let $\boldsymbol{x}_t = (\ell_t, b_t, s_t, s_{t-1}, \dots, s_{t-(m-1)})$ 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

$$Y_t = H\boldsymbol{x}_{t-1} + \varepsilon_t \tag{1.1}$$

$$\boldsymbol{x}_t = F\boldsymbol{x}_{t-1} + G\varepsilon_t \tag{1.2}$$

where $\{\varepsilon_t\}$ is a Gaussian white noise process with mean zero and variance σ^2. We write $\mu_t = H\boldsymbol{x}_{t-1}$ to denote the mean of Y_t. The usual point forecasts are obtained as $\mu_n(h) = \mathrm{E}(Y_{n+h} \mid \boldsymbol{x}_n) = \boldsymbol{f}_h'\boldsymbol{x}_n$, so that $\mu_t = \mu_{t-1}(1)$. The expressions for $\mu_n(h)$ given in Table 1 are derived in Hyndman, Koehler, Ord and Snyder (2001) who also derive forecast variances for these models (this is "Class 1" of the models they consider); Snyder, Koehler, Hyndman and Ord (2001) provide lead-time variances. The forecast distributions are all normal, so this allows easy computation of prediction intervals.

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 and $\mathbf{0}_k$ denotes a zero vector of length 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: $H = [1\ \mathbf{0}'_{m-1}\ 1], \quad F = \begin{bmatrix} 1 & \mathbf{0}'_{m-1} & 0 \\ 0 & \mathbf{0}'_{m-1} & 1 \\ \mathbf{0}_{m-1} & I_{m-1} & \mathbf{0}_{m-1} \end{bmatrix}$ and $G = \begin{bmatrix} \alpha \\ \gamma \\ \mathbf{0}_{m-1} \end{bmatrix}$

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

2 Invertibility conditions

Invertibility is a desirable property of a time series model because we are interested in associating present events with past and present happenings in a sensible manner. Specifically, we want to avoid models where the distant past has a non-negligible effect on the present. More precisely, we define invertibility as follows.

Definition 1 *The model (1.1) and (1.2) is said to be* invertible *if there exists a sequence of constants* $\{\pi_j\}$ *such that* $\sum_{j=0}^{\infty} |\pi_j| < \infty$ *and*

$$\varepsilon_t = \sum_{j=0}^{\infty} \pi_j Y_{t-j}.$$

This is analogous to the definition of invertibility for an ARMA process. See, for example, Brockwell and Davis (1991).

Theorem 1 *Let* Y_t *be defined by the state space model (1.1) and (1.2), and let* $M = F - GH$*. Then the model is invertible if and only if all eigenvalues of* M *lie inside the unit circle (Snyder, Ord and Koehler, 2001).*

Under some circumstances, it is useful to have a weaker notion of invertibility which we shall call *forecast invertibility*.

Definition 2 *Let $(\lambda_i, \boldsymbol{v}_i)$ denote an eigenvalue-eigenvector pair of M. Then the model (1.1) and (1.2) is said to be* forecast invertible *if, for all i, either $|\lambda_i| < 1$ or $\boldsymbol{f}_h'\boldsymbol{v}_i = 0$ where $\mu_n(h) = \mathrm{E}(Y_{n+h} \mid \boldsymbol{x}_n) = \boldsymbol{f}_h'\boldsymbol{x}_n$.*

The notion of forecast invertibility is motivated by the idea that a non-invertible model can still produce stable point forecasts provided the eigenvalues which cause the non-invertibility have no effect on the point forecasts. The concept was introduced by Lawton (1998) for AAA (additive Holt-Winters) forecasts, although he did not have a stochastic state space model as we do here. Note that forecast invertibility is only useful if we want point forecasts but require no other information about the forecast distributions. If prediction intervals are required, or some features of the forecast distribution other than the mean, than full invertibility is necessary. Obviously, any model that is invertible is also forecast invertible.

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 now establish invertibility 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 an invertible model.

3 Invertibility of non-seasonal models

The invertibility conditions for models without seasonality (i.e., ANN, AAN and ADN) are derived in Section A of the Appendix and summarized in Table 2. To visualize these regions, we have plotted them in Figure 1. The light-shaded regions represent the invertibility regions; the dark-shaded regions are the usual regions constructed by restricting each parameter to lie between 0 and 1 and $0 < \beta < \alpha$. Note that the usual parameter region is entirely within the invert-

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

Table 2: *Invertibility conditions for models without seasonality.*

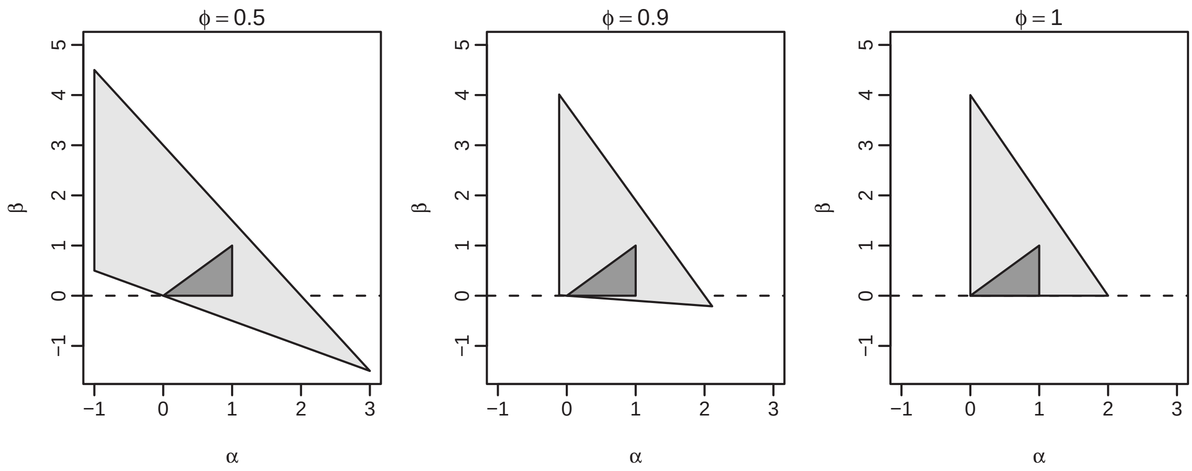

Figure 1: *Invertible region of model ADN. The right hand graph shows the region for model AAN (when $\phi = 1$). In each case, the light-shaded regions represent the invertibility regions; the dark-shaded regions are the usual regions constructed by restricting each parameter in the conventional parameterization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

ibility region in each case. Therefore non-seasonal models obtained using the usual constraints are always invertible (and always forecast invertible).

4 Three seasonal models

For the seasonal models ANA, AAA and ADA, the matrix M has a unit eigenvalue regardless of the values of the model parameters. Therefore the models are always non-invertible. This problem arises because of a redundancy in the model. For example, the ANA model has level and seasonal components given by

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

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

$$S(B)s_t = \theta(B)\gamma\varepsilon_t. \tag{4.1}$$

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 (4.1) 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 (4.1) will be referred to as "normalized" versions of ANA, AAA and ADA.

A third, and simpler, specification arises by dropping $\theta(B)$ in the above model giving

$$S(B)s_t = \gamma\varepsilon_t. \tag{4.2}$$

As with model (4.1), this ensures the seasonal component s_t does not wander too far from zero. Models using the seasonal component (4.2) will be referred to as "modified" versions of ANA, AAA and ADA.

In the following section, we examine the invertibility conditions for each of these seasonal models.

5 Invertibility of seasonal models

5.1 Standard models

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

$$\begin{aligned} P(\lambda) = \lambda^{m+1} + (\alpha+\beta-\phi)\lambda^m + (\alpha+\beta-\alpha\phi)\lambda^{m-1} + \cdots + (\alpha+\beta-\alpha\phi)\lambda^2 \\ + (\alpha+\beta-\alpha\phi+\gamma-1)\lambda + \phi(1-\alpha-\gamma). \end{aligned} \tag{5.1}$$

However, it is easy to see that the eigenvector associated with $\lambda = 1$ is orthogonal to $\boldsymbol{f}_h$. 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 forecast invertible, even though they are not strictly invertible. No other eigenvectors are orthogonal to $\boldsymbol{f}_h$. Forecast invertibility requires the roots of $P(\lambda)$ to lie inside the unit circle. The conditions for forecast invertibility are derived in Section B of the Appendix and summarized in Table 3.

The inequalities involving only α and γ provide necessary conditions for invertibility 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 (5.1). Therefore, we suggest that in practice the conditions on α and γ be checked first, and if satisfied, then the roots of (5.1) be calculated and tested.

To visualize these regions, we have plotted them in Figures 2–3. The light-shaded regions represent the forecast invertibility regions; the dark-shaded regions are the usual regions where each parameter (in the HKSG parameterization) lies in [0,1].

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	$B = \phi(4-3\gamma)+\gamma(1-\phi)/m$
	$C = \sqrt{B^2-8\big[\phi^2(1-\gamma)^2+2(\phi-1)(1-\gamma)-1\big]+8\gamma^2(1-\phi)/m}$
	$D = \min_\theta \Big\{(\phi-\phi\alpha+1)(1-\cos\theta)-\gamma\Big[\frac{(1+\phi)(1-\cos\theta-\cos m\theta)+\cos(m-1)\theta+\phi\cos(m+1)\theta}{2(1-\cos m\theta)}\Big]\Big\}$
	and θ is a solution to
	$\frac{\phi\alpha-\phi+1}{\gamma}+\frac{(\phi-1)(1+\cos\theta-\cos m\theta)+\cos(m-1)\theta-\phi\cos(m+1)\theta}{2(1+\cos\theta)(1-\cos m\theta)} = 0.$

Table 3: *Forecast invertibility conditions for models ANA and ADA. Conditions for AAA can be obtained from ADA by setting* $\phi = 1$.

The invertible region for α and γ is illustrated in Figure 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 Figure 2 shows that the usual parameter region of an ANA model is entirely within the forecast invertibility region. Therefore ANA models obtained using the usual constraints are always forecast invertible.

The invertible region for α and β is depicted in Figure 3 for $m = 4$.

From Figure 3, it can be seen that the usual parameter region and the forecast invertibility region intersect for model ADA but neither is contained within the other. Therefore, models obtained using the usual constraints may not be forecast invertible. This problem is greatest when the seasonal smoothing parameter γ is large which, fortunately, does not happen often in practice.

5.2 Normalized models

Archibald (1984, 1990) discussed the invertible region for the normalized version of AAA and Archibald (1991) provides some preliminary steps towards the invertible region for the normalized version of ADA. However, the damping used in the latter paper is slightly different from

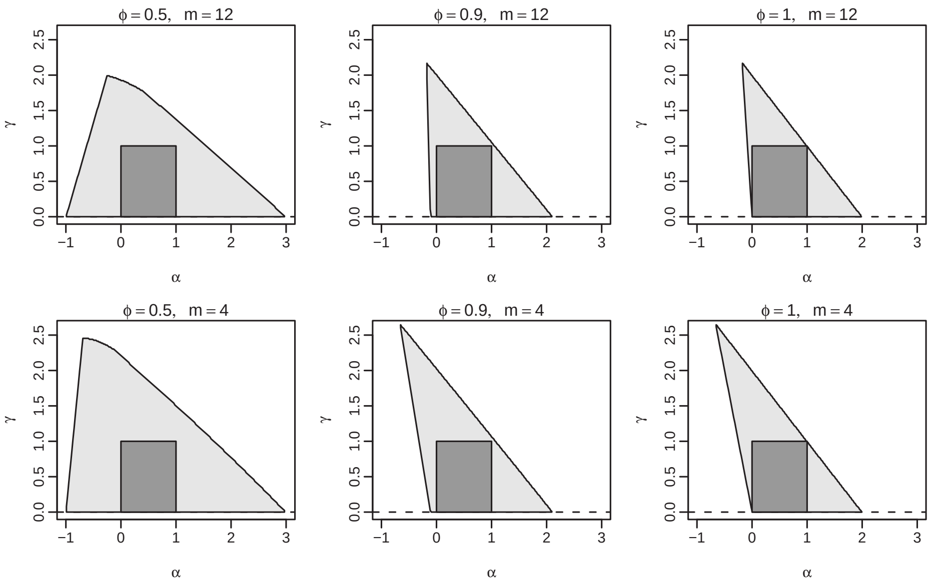

Figure 2: *Light shaded region: the forecast invertible region of α and γ for model ADA. Dark shaded region: usual region where both parameters are bounded by 0 and 1. The right column shows the regions for model AAA (when $\phi = 1$). These are also the regions for model ANA as they are independent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

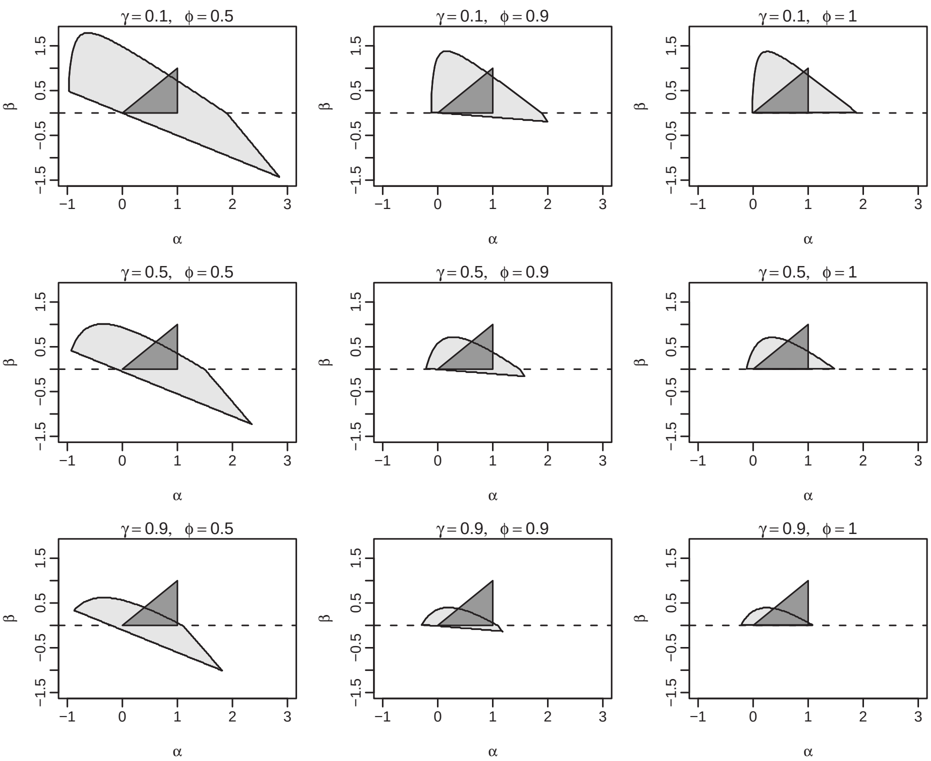

Figure 3: *Light shaded region: the forecast invertible region of α and β for model ADA with $m = 4$. Dark shaded region: usual region where all parameters in the HKSG parameterization are bounded by 0 and 1. The right column shows the region for model AAA (when $\phi = 1$).*

that described here; the damping in our models only takes effect for forecasts two or more periods ahead, whereas Archibald (and Gardner, 1985) use a damping term which applies immediately from the forecast one period ahead.

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

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

Note that this is equivalent to (5.1) if we reparamaterize the model, replacing α in (5.1) by $\alpha - \gamma/m$. Therefore the forecast invertibility conditions for the standard ADA model are the same as the full invertibility conditions for the normalized ADA model, apart from this minor reparameterization.

5.3 Modified models

For the modified models, the matrix M is given below. (Again, the results for AAA are obtained from ADA by setting $\phi = 1$.)

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

The characteristic equation for the modified ADA is

$$f(\lambda) = \lambda^{m+2} + (\alpha + \beta - \phi)\lambda^{m+1} + \sum_{i=2}^{m}(\alpha + \beta - \alpha\phi)\lambda^i + (\gamma - 1)\lambda^2 + [\phi(1 - \alpha - \gamma) - \gamma]\lambda + \phi\gamma.$$

For invertibility, we require the roots of $f(\lambda)$ to lie inside the unit circle. Derivations of these conditions follow a similar approach to those given in the Appendix for the standard model, and lead to conditions analogous to those given in Table 3.

We have plotted the invertibility regions obtained in this manner in Figure 4 for fixed values of γ and ϕ. The light-shaded regions represent the invertibility regions; the dark-shaded regions are the usual (0,1) regions. Note that for the usual parameter region includes non-invertible parameters in all cases, especially for large γ. A striking feature of Figure 4 is that when γ is large and ϕ is close to 1, the invertible region becomes very small. These features may make the modified models too restrictive for use with some data sets.

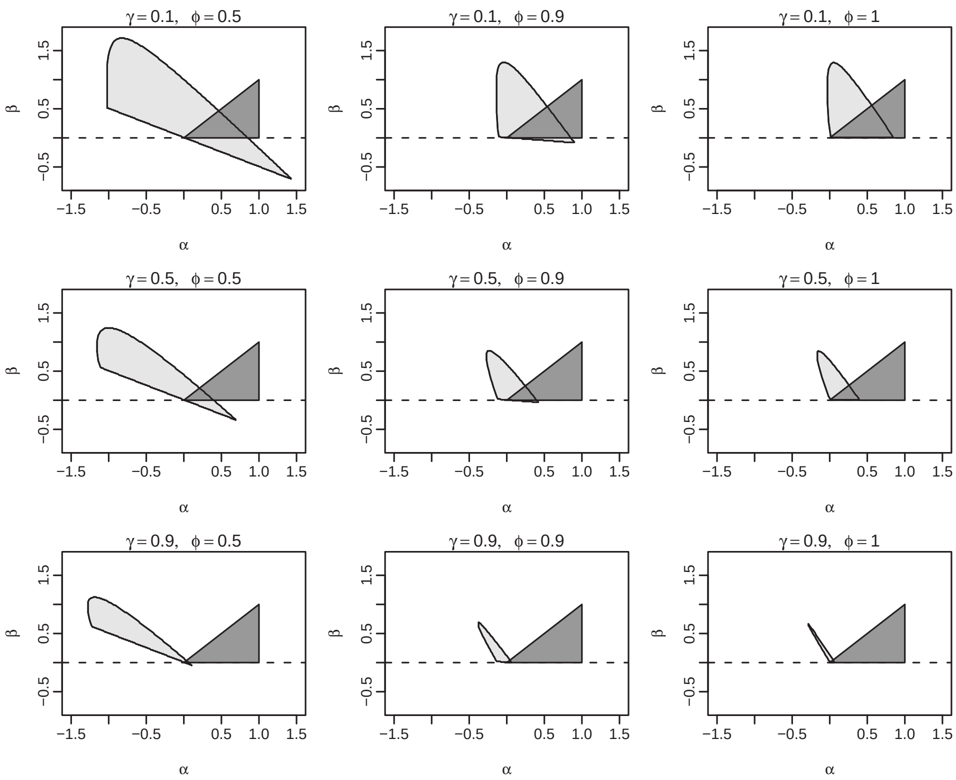

Figure 4: *Invertible region of modified model ADA. The right hand graphs show the region for the modified model AAA with* $\phi = 1$.

6 Conclusions

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

However, our empirical experience suggests that the increased parameter space will not necessarily lead to better forecast performance. Restricting the parameter space makes the forecasts more robust to unusual observations. The [0,1] space has the added advantage that it makes the model equations more interpretable as weighted averages.

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 non-invertible 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. We have shown that the model can be made "forecast invertible", so that the forecast means are unaffected by the non-invertibility, but this does not fix the problem for other attributes of the forecast distribution.

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

The modified model (introduced here) achieves a similar result by requiring the seasonal states to have mean zero. Of these two models, we prefer the normalized model because its parameter space is bounded, its invertible parameter space is larger, and because it has the property that the seasonal components always sum to zero.

For the same reasons as given above for the non-seasonal models, we have found that that the intersection of the invertible region with the usual [0,1] region provides good results in practice. This provides more robust forecasts, allows the model to remain easily interpretable, and gives invertible forecasts.

Acknowledgements

We would like to thank Richard Lawton for useful comments on an earlier draft of this work which motivated our definition of "forecast invertibility". Keith Ord, Andrew Harvey and Simone Grose also provided helpful comments which led to several improvements of the paper.

References

ARCHIBALD, B.C. (1984) Invertible region of Holt-Winters' model, *Working paper 31, March (School of Business Administration, Dalhousie University, Halifax, NS. Canada).*

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, (School of Business Administration, Dalhousie University, Halifax, NS. Canada).*

BROCKWELL AND DAVIS (1991) *Introduction to time-series and forecasting*, John Wiley & Sons: New York.

GARDNER, E.S. (1985) Exponential smoothing: the state of the art, *Journal of Forecasting*, **4**, 1–28.

HARVEY, A.C. (1989) *Forecasting, structural time series models and the Kalman filter*, Cambridge University Press: Cambridge.

HYNDMAN, R.J, A.B. KOEHLER, R.D. SNYDER and S. GROSE (2002) A state space framework for automatic forecasting using exponential smoothing methods, *International Journal of Forecasting*, **18**(3), 439–454.

HYNDMAN, R.J, A.B. KOEHLER, J.K. ORD and R.D. SNYDER (2001) Prediction intervals for exponential smoothing state space models. Working paper 11/01, Department of Econometrics & Business Statistics, Monash University.

LAWTON, R. (1998) How should additive Holt-Winters estimates be corrected? *International Journal of Forecasting*, **14**, 393–403.

MAKRIDAKIS, S., S.C. WHEELWRIGHT and R.J. HYNDMAN (1998) *Forecasting: methods and applications*, 3rd ed., John Wiley & Sons: New York.

ORD, J.K., A.B. KOEHLER and R.D. SNYDER (1997) Estimation and prediction for a class of dynamic nonlinear statistical models, *J. Amer. Statist. Assoc.*, **92**, 1621–1629.

PEGELS, C.C. (1969) Exponential smoothing: some new variations, *Management Science*, **12**, 311–315.

RALSTON, A. (1965) *A first course in numerical analysis*, McGraw-Hill : New York.

ROBERTS, S.A. (1982) A general class of Holt-Winters type forecasting models, *Management Science*, **28**(8), 808–820.

SNYDER, R.D., A.B. KOEHLER, R.J. HYNDMAN, and J.K. ORD (2001) Exponential smoothing for inventory control: mean and variances of lead-time demand. Working paper 3/02, Department of Econometrics & Business Statistics, Monash University.

SNYDER, R.D., J.K. ORD, and A.B. KOEHLER (2001) Prediction intervals for ARIMA models, *J. Bus. Econ. Statist.*, **19**(2), 217–225.

Appendix: Proofs

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 $f(z) = a_0 + a_1 z + a_2 z^2 + \cdots + a_n z^n$ be a polynomial of degree n with real coefficients where $a_0 \neq 0$ and define*

$$g(z) = a_0 z^n + a_1 z^{n-1} + \cdots + a_{n-1} z + a_n.$$

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

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

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

For polynomials of order 1, this obviously requires $|a_0| < |a_1|$. For polynomials of order 2, it leads to the following well-known corollary.

Corollary 1 *Let $f(z) = a_0 + a_1 z + a_2 z^2$. Then all roots of $f(z)$ have modulus less than 1 if and only if*

$$|a_0| < |a_2| \qquad \text{and} \qquad |a_1| < |a_0 + a_2|.$$

A: Non-seasonal models

For the ANN model, the eigenvalue of M is $1 - \alpha$. So the model is invertible if $0 < \alpha < 2$ (e.g., Harvey, 1989).

For the ADN model, the eigenvalues of M satisfy $\lambda^2 + b\lambda + c = 0$ where $b = \alpha + \beta - 1 - \phi$ and $c = (1 - \alpha)\phi$. We apply the first condition of Corollary 1 to obtain $-1 < \phi(\alpha - 1) < 1$ and so

$1-(1/\phi) < \alpha < 1+(1/\phi)$ since we assume $\phi > 0$. The second condition of Corollary 1 gives $(\phi-1)\alpha < \beta < (1+\phi)(2-\alpha)$.

For AAN, we set $\phi = 1$ to obtain the required result.

B: Seasonal models

The characteristic equation of model ADA is $f(\lambda) = (1-\lambda)P(\lambda) = 0$ where $P(\lambda)$ is given by (5.1). 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 forecast invertible. 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 $(\alpha+\beta-\alpha\phi) = \gamma(\phi-1)/m$. If $(\alpha+\beta-\alpha\phi) < \gamma(\phi-1)/m$, then the roots are outside the unit circle by the mean value theorem. Therefore to ensure the roots are within the unit circle we require

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

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

$$\gamma > 0. \tag{A.2}$$

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 forecast invertibility:

$$1 - 1/\phi < \alpha + \gamma < 1 + 1/\phi \qquad \text{and} \qquad B - C < 4\phi\alpha < B + C \tag{A.3}$$

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

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

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

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

$$\begin{aligned} &|P(\cos\theta + i\sin\theta)|^2 \\ &\quad = 2\Big[1 + \phi^2 - 2\phi\cos\theta + \phi\cos(m-1)\theta - \phi^2\cos m\theta - \cos m\theta + \phi\cos(m+1)\theta\Big] \\ &\qquad + 2\phi^2\alpha^2(1-\cos m\theta) + b^2(1-\cos m\theta)/(1-\cos\theta) + \gamma^2(1+\phi^2 - 2\phi\cos\theta) \\ &\qquad + 2b\Big[\gamma\big\{(1-\phi)(\cos\theta + \cdots + \cos(m-1)\theta) - \phi\cos m\theta + 1\big\} - \big\{\phi(1-\alpha)+1\big\}(1-\cos m\theta)\Big] \\ &\qquad + 2\gamma\Big[2\phi\cos\theta + (1+\phi^2)(\cos m\theta - 1) - \phi\cos(m-1)\theta - \phi\cos(m+1)\theta\Big] \\ &\qquad - 2\phi\alpha\gamma\Big[\cos\theta - \cos(m-1)\theta - \phi(1-\cos m\theta)\Big] \\ &\qquad + 2\phi\alpha\Big[-2\phi + 2\cos\theta - \cos(m-1)\theta + 2\phi\cos m\theta - \cos(m+1)\theta\Big]. \end{aligned} \tag{A.4}$$

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 (A.4) is equal to zero. Differentiating (A.4) 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

$$\beta < D - \alpha(1-\phi). \tag{A.5}$$

In expression (A.5), 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 (A.5) in (A.4) 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

$$\begin{aligned}
&|P(\cos\theta + i\sin\theta)|^2 \\
&\quad = 2\Big[1 + \phi^2 - 2\phi\cos\theta + \phi\cos(m-1)\theta - \phi^2\cos m\theta - \cos m\theta + \phi\cos(m+1)\theta\Big] \\
&\qquad + 2\phi^2\alpha^2(1-\cos m\theta) - \frac{(1-\cos m\theta)}{(1-\cos\theta)}\Big[\big\{\phi(1-\alpha)+1\big\}(1-\cos\theta) - \gamma A(\theta,\phi)\Big]^2 \\
&\qquad + 2\gamma\Big[2\phi\cos\theta + (1+\phi^2)(\cos m\theta - 1) - \phi\cos(m-1)\theta - \phi\cos(m+1)\theta\Big] \\
&\qquad - 2\phi\alpha\gamma\Big[\cos\theta - \cos(m-1)\theta - \phi(1-\cos m\theta)\Big] + \gamma^2(1+\phi^2-2\phi\cos\theta) \\
&\qquad + 2\phi\alpha\Big[2\cos\theta - 2\phi - \cos(m-1)\theta + 2\phi\cos m\theta - \cos(m+1)\theta\Big].
\end{aligned} \tag{A.6}$$

Then partially differentiating (A.6) with respect to α and equating the result to zero gives

$$\frac{\phi\alpha - \phi + 1}{\gamma} + \frac{(\phi-1)(1+\cos\theta-\cos m\theta) + \cos(m-1)\theta - \phi\cos(m+1)\theta}{2(1+\cos\theta)(1-\cos m\theta)} = 0 \tag{A.7}$$

Then θ will be a solution to (A.7). 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 (A.1), (A.2), (A.3) and (A.5) gives the required parameter space for model ADA. Forecast invertibility conditions for AAA are obtained by setting $\phi = 1$. Forecast invertibility conditions for ANA are obtained from (A.2) and (A.3) by setting $\phi = 1$.