A nonlinear dynamic model for multiplicative seasonal–trend decomposition*

Tohru Ozaki
Institute of Statistical Mathematics
Tokyo, Japan

Peter Thomson†
Statistics Research Associates Ltd
Wellington, New Zealand

11 February 2000

Abstract

A nonlinear dynamic model is introduced for multiplicative seasonal time series that follows and extends the X-11 paradigm where the observed time series is a product of trend, seasonal and irregular factors. A selection of standard seasonal and trend component models used in additive dynamic time series models are adapted for the multiplicative framework and a nonlinear filtering procedure is proposed. The results are illustrated and compared to X-11 and log-additive models using real data. In particular it is shown that the new procedures do not suffer from the trend bias present in log-additive models.

Keywords: Nonlinear dynamic models; X-11; seasonal time series; seasonal adjustment.

*To appear in the *Journal of Forecasting*.

†Formerly at the Institute of Statistics and Operations Research, Victoria University of Wellington, New Zealand.

1 Introduction

The X-11 seasonal adjustment method developed by the US Census Bureau (see Shiskin et al. (1967)) is the standard method used by most of the official statistical agencies in the world, either directly or as an integral part of its successors X-11-ARIMA (Dagum (1980)) and X-12-ARIMA (Findley et al. (1988)). X-11 is based on a simple nonparametric multiplicative model many of whose details can only be inferred implicitly from the sequence of procedural operations that it comprises. This makes formal comparison with other models difficult and, as a consequence, many attempts have been made to replace X-11 by explicit model-based seasonal adjustment methods. See for example Durbin and Murphy (1975), Akaike (1980), Akaike and Ishuguro (1980), Hillmer and Tiao (1982), Gersch and Kitagawa (1983), Maravall and Pierce (1987), Harvey (1989), Cleveland et al. (1990), Ord et al. (1997) among many others. However, despite its age and perceived shortcomings, X-11 remains the dominant seasonal adjustment procedure used in practice. This paper introduces alternative parametric multiplicative seasonal time series models that are directly based on X-11, but whose dynamic properties are explicitly formulated.

Before proceeding to describe the model we first briefly review the key features of the X-11 seasonal adjustment procedure proposed in Shisken et al. (1967). This procedure is based on the simple descriptive multiplicative model

$$O(t) = C(t)S(t)I(t)$$

where t indexes discrete time in months, the $O(t)$ denote the original observations and $C(t)$, $S(t)$, $I(t)$ denote the trend cycle, seasonal and irregular components respectively. The trend cycle $C(t)$ is assumed to be smooth over periods of the order of a year whereas the monthly seasonal factors $S(t)$ are assumed to be smooth from year to year and such that they sum to 12 approximately over any consecutive 12 month period. The irregular component $I(t)$ describes non-systematic error and is assumed to have mean unity. On the basis of their extensive practical experience, Shisken et al (1967) commented that "the seasonal, trend-cycle, trading-day, and irregular components are related in a multiplicative fashion for most national economic time series". This observation remains true today partly as an empirical fact and partly because much of the world of economics and finance is driven by multiplicative relationships such as compound growth rates for example.

It is instructive to consider an equivalent reformulation of the traditional X-11 model which expresses the monthly observations $Y(t)$ as

$$Y(t) = T(t)(1 + S(t))(1 + I(t)) \tag{1}$$

where $T(t)$ denotes the trend, the seasonal factor $S(t)$ now satisfies

$$\sum_{j=0}^{11} S(t-j) \approx 0 \tag{2}$$

and the irregular $I(t)$ satisfies

$$E\{I(t)\} = 0. \tag{3}$$

Thus, in terms of interpretation, $S(t)$ now represents the seasonal effect for month t as a proportion of the long-term trend $T(t)$ whereas $I(t)$ represents the non-systematic error as a proportion of the systematic component $T(t)(1+S(t))$. Furthermore $Y(t)$ can be written as

$$Y(t) = T(t) + T(t)S(t) + T(t)(1+S(t))I(t) \tag{4}$$

which is an additive decomposition of $Y(t)$ into trend, evolving seasonal and heteroskedastic error components. It can be argued (see Thomson and Ozaki (1992)) that the X-11 procedure expresses a multiplicative model as an additive model and fits accordingly. A primary advantage of this approach is that estimates of the trend $T(t)$ are properly centred so that they run through the middle of the data. The fact that this is so follows from (2) and (3) so that $T(t)(1+S(t))I(t)$ will have mean zero and $T(t)S(t)$ will approximately sum to zero over any twelve month period provided that the trend is sufficiently smooth. This ensures that the process of seasonal adjustment is essentially one of redistribution of seasonal variation so that, on a moving annual basis, observed totals are neither created or destroyed.

By contrast, a standard way of handling multiplicative seasonal time series is to apply an additive trend-seasonal decomposition procedure to the logarithms of the observations. However the use of the logarithmic transformation introduces a trend bias in the original scale of the observations which is primarily due to the inequality between the geometric and arithmetic means. (See Young (1968), and Thomson and Ozaki (1992) for further details.) Consider, for example, a model of the form

$$Y(t) = e^{T(t)+S(t)+I(t)}$$

where $S(t)$ and $I(t)$ satisfy (2) and (3) respectively. Then the multiplicative components $\exp S(t)$, $\exp I(t)$ will not in general correspond to $1+S(t)$ and $1+I(t)$ of the X-11 model unless $S(t)$ and $I(t)$ are small. In particular

$$\sum_{j=0}^{11} e^{S(t-j)} \geq 12 \qquad\qquad E\{e^{I(t)}\} \geq 1$$

and it is these inequalities that lie at the heart of the bias problem. In such cases bias correction procedures can be devised and this is the approach taken by Young (1968) and Thomson and Ozaki (1992).

There is no bias problem if (1) is directly fitted to the data. This is done non-parametrically by X-11 and closely related procedures. Among those in the latter category, the model proposed by Durbin and Murphy (1975) is of particular interest. They directly addressed the multiplicative structure of trend and seasonal components by modelling

$$Y(t) = T(t) + S_1(t) + T(t)S_2(t) + I(t)$$

where $S_1(t)$, $S_2(t)$ satisfy (2) and $T(t)$, $I(t)$ satisfy much the same conditions as before. This hybrid model is purely additive if $S_2(t) = 0$ and purely multiplicative if $S_1(t) = 0$. However, even in the latter case, the model is not the same as the X-11 model (4) whose irregular has a particular heteroskedastic form.

There appear to be few published procedures that fit parametric dynamic versions of (1). A recent paper by Ord et al. (1997) does consider (1) as a special case of a class of dynamic non-linear models. However this particular class of models has only one source of randomness unlike the models considered here. In summary, the primary objective of the current paper is to provide a flexible framework for fitting parametric nonlinear dynamic models for the X-11 model (1).

2 A dynamic model for X-11

We take the X-11 model (1) as our starting point and assume that $Y(t)$ is always positive. Appropriate dynamic state-space models for $T(t)$ and $S(t)$ will be defined and then combined into one overall non-linear state-space model for $Y(t)$. Our approach will be to retain as much linear structure as possible to enhance the simplicity and stability of our recursive algorithms which are given in Section 3.

Given the multiplicative nature of the X-11 model, an appropriate dynamic model for the trend $T(t)$ is

$$\Delta^p \log T(t) = \eta(t) \qquad (t = 0, \pm 1, \ldots) \tag{5}$$

where p is a positive integer, $\eta(t)$ is Gaussian white noise and Δ is the backwards difference operator satisfying $\Delta X(t) = X(t) - X(t-1)$. This model, with $\log T(t)$ replaced by $T(t)$, has been commonly used to model trends (see Akaike (1980), Gersch and Kitagawa (1983), Harvey (1989) among many others). Typically $p = 1$ or 2 and the variance of $\eta(t)$ is small so that realisations of $T(t)$ are smooth by comparison to the original observations. Since $\log T(t) - \log T(t-1)$ represents the compound growth rate of the trend over the interval $(t-1, t]$, the continuous time version of (5) for $p = 1$ is extensively used in the finance literature. The model (5) and its variants are readily put in the linear state-space form

$$\begin{aligned} \log T(t) &= \mathbf{c}_T^T \mathbf{m}(t) \\ \mathbf{m}(t) &= \mathbf{A}_T \mathbf{m}(t-1) + \mathbf{B}_T \mathbf{u}(t) \end{aligned} \tag{6}$$

where $\mathbf{m}(t)$, $\mathbf{c}_T$ are p dimensional vectors, $\mathbf{A}_T$, $\mathbf{B}_T$ are matrices with dimensions $p \times p$, $p \times q$ respectively, and $\mathbf{u}(t)$ is a q dimensional Gaussian white noise process. When $p = 2$ for example,

$$\mathbf{m}_T = \begin{bmatrix} \log T(t) \\ \log T(t-1) \end{bmatrix}, \qquad \mathbf{A}_T = \begin{bmatrix} 2 & -1 \\ 1 & 0 \end{bmatrix}, \qquad \mathbf{B}_T = \begin{bmatrix} 1 \\ 0 \end{bmatrix},$$

$\mathbf{c}_T = \mathbf{B}_T$ and $\mathbf{u}(t) = \eta(t)$. Note that (5) and (6) model the trend $T(t)$ indirectly through the logarithms of $T(t)$ rather than through $T(t)$ itself.

There are a number of dynamic models for $S(t)$ in the literature that satisfy the constraint (2). We consider a selection of these with some variations. A commonly used model which builds directly on (2) is

$$\sum_{j=0}^{11} S(t-j) = \xi(t) \qquad (t = 0, \pm 1, \ldots) \tag{7}$$

where $\xi(t)$ is a zero mean stationary process. The case where $\xi(t)$ is white noise is often used (see Box et al. (1978), Gersch and Kitagawa (1983), Harvey (1989) for example). When $\xi(t)$ is an AR(1) process with parameter $\rho > 0$ then (7) is equivalent to the seasonal model used by the seasonal adjustment procedure BAYSEA (Akaike (1980), Akaikie and Ishuguro (1980)) and also den Butter and Mourik (1990). In this case the convex combination

$$(1-\rho)\sum_{j=0}^{11} S(t-j) + \rho(S(t) - S(t-12)) = \xi(t) - \rho\xi(t-1)$$

will be white noise. Thus the choice of ρ strikes a balance between the requirement (2) and the requirement that the annual moving total $\sum_{j=0}^{11} S(t-j)$ is smooth (i.e. that $\Delta \sum_{j=0}^{11} S(t-j) = S(t) - S(t-12) \approx 0$). In den Butter and Mourik (1990) ρ is fixed at $\rho = 11/12$ when

$$S(t) = S(t-12) - \frac{1}{12}\sum_{j=1}^{12} S(t-j) + w(t) \tag{8}$$

and $w(t)$ is white noise. We shall refer to (7) with white noise $\xi(t)$ as the *BHT seasonal model*, since it would appear that Box et al. (1978) were the first to introduce it to the literature, and we shall refer to (7) with $\xi(t)$ an AR(1) as the *BAYSEA seasonal model* for similar reasons. Note that the BAYSEA seasonal model includes the BHT model as a special case.

Defining L as the lag operator satisfying $LX(t) = X(t-1)$, we observe that

$$\sum_{j=0}^{11} S(t-j) = \sum_{j=0}^{11} L^j S(t) = \prod_{j=1}^{6} \phi_j(L) S(t)$$

where

$$\phi_j(L) = \begin{cases} 1 - 2\cos\lambda_j L + L^2 & (j = 1, \ldots, 5) \\ 1 + L & (j = 6) \end{cases} \tag{9}$$

and $\lambda_j = 2\pi j/12$. Thus $\sum_{j=0}^{11} L^j$ can be regarded as the sequential application or cascade of the individual operators $\phi_j(L)$, each of which corresponds to the 6 different harmonics λ_j of the fundamental frequency $\lambda_1 = 2\pi/12$. This decomposition and the previous development suggest building alternative, more flexible, models for $S(t)$ based on individual models for the harmonics of the form

$$\phi_j(L) S_j(t) = \xi_j(t) \qquad (t = 0, \pm 1, \ldots) \tag{10}$$

where the $\xi_j(t)$ are mutually independent Gaussian white noise processes. These harmonic components operate in parallel and are combined additively to yield an overall seasonal pattern

$$S(t) = \sum_{j=1}^{6} S_j(t). \tag{11}$$

We shall refer to (11) as the *parallel seasonal model*. Note that there is a close relationship between the parallel seasonal model and similar models proposed by Hannan (1968), Harvey (1989) and Ng and Young (1990). There the components $S_j(t)$ satisfy (10), but with $\xi_j(t)$ typically following a particular MA(1) process. The links between these various models is the subject of ongoing research (see Dongfeng et al (1997)).

All the seasonal models described above can all be put in the linear state-space form

$$\begin{aligned} S(t) &= \mathbf{c}_S^T \mathbf{s}(t) \\ \mathbf{s}(t) &= \mathbf{A}_S \mathbf{s}(t-1) + \mathbf{B}_S \mathbf{v}(t) \end{aligned} \tag{12}$$

where $\mathbf{s}(t)$, $\mathbf{c}_S$ are p dimensional vectors, $\mathbf{A}_S$, $\mathbf{B}_S$ are matrices with dimensions $p \times p$, $p \times q$ respectively, $\mathbf{v}(t)$ is a p dimensional Gaussian white noise process, and p is the dimension of the system. For the BAYSEA model $p = 12$,

$$\mathbf{s}(t) = \begin{bmatrix} S(t) \\ S(t-1) \\ S(t-2) \\ \vdots \\ S(t-11) \end{bmatrix}, \qquad \mathbf{A}_S = \begin{bmatrix} \rho-1 & \rho-1 & \dots & \rho-1 & \rho \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix}, \qquad \mathbf{B}_S = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

$\mathbf{c}_S = \mathbf{B}_S$ and $\mathbf{u}(t) = \xi(t)$. For the parallel model the jth harmonic seasonal component ($j < 6$) follows the state-space model (12) with

$$\mathbf{s}_j(t) = \begin{bmatrix} S_j(t) \\ S_j(t-1) \end{bmatrix}, \qquad \mathbf{A}_{S_j} = \begin{bmatrix} 2\cos\lambda_j & -1 \\ 1 & 0 \end{bmatrix}, \qquad \mathbf{B}_{S_j} = \begin{bmatrix} 1 \\ 0 \end{bmatrix},$$

$\mathbf{c}_{S_j} = \mathbf{B}_{S_j}$ and $\mathbf{u}_j(t) = \xi_j(t)$. Note that $S_6(t)$ is already in state-space form with $\mathbf{s}_6(t) = S_6(t)$, $\mathbf{A}_{S_6} = -1$, $\mathbf{c}_{S_6} = \mathbf{B}_{S_6} = 1$ and $\mathbf{u}_6(t) = \xi_6(t)$. These component models can now be stacked in the obvious way to yield a state transition equation of the form given in (12) and an observation equation equivalent to (11).

Finally, we set

$$1 + I(t) = e^{\epsilon(t) - \sigma^2/2} \tag{13}$$

where $\epsilon(t)$ is Gaussian white noise with variance σ^2. Thus $1 + I(t)$ is proportional to a lognormal random variable with the constant of proportionality chosen to ensure that (3) holds. Moreover, in terms of the state-space model, $1 + I(t)$ plays the role of a multiplicative measurement error.

The overall model can now be written in state-space form with the state transition equations following the linear form

$$\mathbf{x}(t) = \mathbf{A}\mathbf{x}(t-1) + \mathbf{B}\mathbf{n}(t). \tag{14}$$

In the case of the BHT and BAYSEA seasonal models

$$\mathbf{x}(t) = \begin{bmatrix} \mathbf{m}(t) \\ \mathbf{s}(t) \end{bmatrix}, \qquad \mathbf{A} = \begin{bmatrix} \mathbf{A}_T & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_S \end{bmatrix}, \qquad \mathbf{B} = \begin{bmatrix} \mathbf{B}_T & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_S \end{bmatrix},$$

and $\mathbf{n}(t) = (\eta(t), \xi(t))^T$. Here the quantities are as defined following (5) and (7). For the parallel seasonal model

$$\mathbf{x}(t) = \begin{bmatrix} \mathbf{m}(t) \\ \mathbf{s}_1(t) \\ \vdots \\ \mathbf{s}_6(t) \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} \mathbf{A}_T & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{S_1} & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{A}_{S_6} \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} \mathbf{B}_T & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_{S_1} & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{B}_{S_6} \end{bmatrix},$$

and $\mathbf{n}(t) = (\eta(t), \xi_1(t), \dots, \xi_6(t))^T$. Here the quantities are as defined following (10). From (1) the observation equation is now given directly by

$$Y(t) = T(t)(1 + S(t))e^{\epsilon(t) - \sigma^2/2} \tag{15}$$

or, indirectly, by

$$\log Y(t) = M(t) + \log(1 + S(t)) + \epsilon(t) \tag{16}$$

where $M(t) = \log T(t) - \sigma^2/2$ satisfies (5) and (6) with $\log T(t)$ replaced by $M(t)$. Because of its predominantly additive structure, we choose to use the indirect observation equation (16) for model fitting and for extracting the underlying trend $T(t)$, seasonal $S(t)$ and seasonally adjusted series $T(t)(1 + I(t))$ from the data.

A key aspect of the model formulation adopted is that the nonlinearity resides wholly in the observation equation (16). Moreover, apart from a noise bias correction which is likely to be small in practice, it is the use of $\log(1 + S(t))$ and a linear model for $S(t)$ that is the primary difference between the model and the familiar log-additive model. As is demonstrated in Section 4, it is this difference that will typically lie at the root of eliminating trend bias.

Finally we examine the nature of the dynamic evolution of the model defined by (14) and (15). It is sufficient to consider the simple trend model (5) with $p = 1$, $E(\eta(t)^2) = \sigma_\eta^2$ and the BHT seasonal model (7) with $E(\xi(t)^2) = \sigma_\xi^2$. Then

$$\begin{aligned} Y(t) &= T(t-1)e^{\sigma_\eta^2/2}(1 - \sum_{j=1}^{11} S(t-j))(1 + z_1(t)) + T(t-1)e^{\sigma_\eta^2/2} z_2(t) \\ &= E\{T(t)(1 + S(t)) | \mathbf{x}(t-1)\}(1 + z_1(t)) + E\{T(t) | \mathbf{x}(t-1)\} z_2(t) \end{aligned}$$

where

$$z_1(t) = (1 + I(t))e^{\eta(t) - \sigma_\eta^2/2} - 1, \qquad z_2(t) = (1 + I(t))e^{\eta(t) - \sigma_\eta^2/2}\xi(t)$$

are mutually uncorrelated white noise processes with zero means and variances

$$E(z_1(t)^2) = e^{\sigma^2 + \sigma_\eta^2} - 1, \qquad E(z_2(t)^2) = \sigma_\xi^2 e^{\sigma^2 + \sigma_\eta^2}.$$

Similar results hold for the other trend and seasonal models. Thus, given state information at time $t - 1$, the general model specified by (14) and (15) has an evolution which allows for heteroskedastic errors of the form indicated in (4). However these come not only from the (multiplicative) observation error $I(t)$, but also from the propagation errors $\eta(t)$ and $\xi(t)$. Note that the last error term is proportional to the level of the trend only and is not present in the case of a deterministic seasonal pattern.

3 Computational procedures

From the previous section, the non-linear dynamic state-space model for X-11 is now given by

$$\begin{aligned} \mathbf{x}(t) &= \mathbf{A}\mathbf{x}(t-1) + \mathbf{B}\mathbf{n}(t) \\ \log Y(t) &= h(\mathbf{x}(t)) + \epsilon(t) \end{aligned} \tag{17}$$

where $\mathbf{x}(t)$, $\mathbf{A}$, $\mathbf{B}$, $\mathbf{n}(t)$ are as given following (14) with $\mathbf{m}(t) = (M(t), M(t-1), \ldots)^T$, and $\epsilon(t)$ is as given following (13). Moreover

$$h(\mathbf{x}(t)) = \mathbf{a}^T\mathbf{x}(t) + \log(1 + \mathbf{b}^T\mathbf{x}(t)) \tag{18}$$

where $\mathbf{a}$ and $\mathbf{b}$ are vectors with the same dimension as $\mathbf{x}(t)$. In the case of the parallel seasonal model

$$\mathbf{a} = \begin{bmatrix} \mathbf{c}_T \\ \mathbf{0} \end{bmatrix}, \qquad \mathbf{b} = \sum_{j=1}^{6} \mathbf{b}_j$$

where $\mathbf{c}_T$ is given by (6), the vectors $\mathbf{b}_j$ have the same dimension as $\mathbf{b}$ and

$$\mathbf{b}_1 = \begin{bmatrix} \mathbf{0} \\ \mathbf{c}_{S_1} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix}, \qquad \mathbf{b}_2 = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{c}_{S_2} \\ \vdots \\ \mathbf{0} \end{bmatrix}, \qquad \ldots, \qquad \mathbf{b}_6 = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \mathbf{c}_{S_6} \end{bmatrix}$$

with the $\mathbf{c}_{S_j}$ given following (12). For the BHT and BAYSEA seasonal models $\mathbf{a}$ is as given for the parallel model, but $\mathbf{b} = (\mathbf{0}^T, \mathbf{c}_S^T)^T$ with $\mathbf{c}_S$ given by (12).

The non-linearity of $h(\mathbf{x}(t))$ in the observation equation means that the observations will in general be non-Gaussian and so conventional Kalman filtering is inappropriate. One could proceed with the model directly (see Jazwinski (1970)) and determine the various conditional distributions involved through a mix of simulation (see Kitagawa (1987), Pole and West (1990)) and exact results (see Harvey and Shephard (1993) for example). A detailed discussion of techniques for nonlinear dynamic model fitting and prediction is given in Ozaki et al (1997). However, for reasons of simplicity and to minimise computational cost, we shall adopt a simpler approximate strategy using a variant of the extended Kalman filter.

We now consider the recursive model given by (17) with $h(\mathbf{x}(t))$ replaced by the affine linear approximation

$$\tilde{h}(\mathbf{x}(t)) = \alpha(\hat{\mathbf{x}}_{t|t-1}) + \beta(\hat{\mathbf{x}}_{t|t-1})^T\mathbf{x}(t) \tag{19}$$

where

$$\hat{\mathbf{x}}_{s|t} = \tilde{E}(\mathbf{x}(s) | \log Y(u), 1 \le u \le t) \qquad (s, t > 0)$$

with $\hat{\mathbf{x}}_{s|0}$ defined to be $E(\mathbf{x}(s))$ for all $s > 0$ and $\tilde{E}$ denotes expectation with respect to the approximate model. Here $\alpha(\hat{\mathbf{x}}_{t|t-1})$, $\beta(\hat{\mathbf{x}}_{t|t-1})$ are suitably chosen functions of $\hat{\mathbf{x}}_{t|t-1}$

with $\alpha(\hat{\mathbf{x}}_{t|t-1})$ a scalar and $\beta(\hat{\mathbf{x}}_{t|t-1})$ a vector with the same dimension as $\mathbf{x}(t)$. Note that, in general, $\hat{\mathbf{x}}_{t|t-1}$ is not linear with respect to the $\log Y(s)$ $(s < t)$ and is not the optimal one-step ahead predictor of $\mathbf{x}(t)$ for the original model (17) although it is for the approximating model. However one might expect that such quantities calculated for both models will be close if $\tilde{h}(\mathbf{x}(t))$ is a good approximation of $h(\mathbf{x}(t))$ at each time point t.

The primary virtue of the approximating model is that it is conditionally Gaussian for which we can construct an exact version of the Kalman filter, the fixed-interval smoother and the likelihood of the observations based on the one-step ahead prediction errors. A full discussion of conditionally Gaussian state space models is given in Lipster and Shiryayev (1978). In the case where the data follow the approximating model (17) with $h(\mathbf{x}(t))$ replaced by (19), but $\mathbf{n}(t)$ and $\epsilon(t)$ are non-Gaussian with finite mean square, then the $\hat{\mathbf{x}}_{s|t}$ are no longer the best predictors of $\mathbf{x}(s)$. However they are the best linear unbiased predictors of $\mathbf{x}(s)$.

From Chapter 13 of Lipster and Shiryayev (1978) the filtering formulae for the approximate X-11 state-space model are now given by

$$\begin{aligned}
\hat{\mathbf{x}}_{t|t} &= \hat{\mathbf{x}}_{t|t-1} + \mathbf{P}_t\beta(\hat{\mathbf{x}}_{t|t-1})\nu(t)/\sigma^2_{t|t-1} \\
\hat{\mathbf{x}}_{t|t-1} &= \mathbf{A}\hat{\mathbf{x}}_{t-1|t-1} \\
\nu(t) &= \log Y(t) - \alpha(\hat{\mathbf{x}}_{t|t-1}) - \beta(\hat{\mathbf{x}}_{t|t-1})^T\hat{\mathbf{x}}_{t|t-1} \\
\mathbf{P}_t &= \mathbf{A}\mathbf{V}_{t-1}\mathbf{A}^T + \mathbf{B}\mathbf{\Gamma}\mathbf{B}^T \\
\mathbf{V}_t &= \mathbf{P}_t - \mathbf{P}_t\beta(\hat{\mathbf{x}}_{t|t-1})\beta(\hat{\mathbf{x}}_{t|t-1})^T\mathbf{P}_t/\sigma^2_{t|t-1} \\
\sigma^2_{t|t-1} &= \beta(\hat{\mathbf{x}}_{t|t-1})^T\mathbf{P}_t\beta(\hat{\mathbf{x}}_{t|t-1}) + \sigma^2
\end{aligned} \tag{20}$$

where $\sigma^2_{t|t-1}$ is the variance of the one-step ahead prediction error $\nu(t)$ and the covariance matrix of $\mathbf{n}(t)$ is given by $\mathbf{\Gamma}$. The mean squared error matrices $\mathbf{V}_t$, $\mathbf{P}_t$ are given by

$$\begin{aligned}
\mathbf{V}_t &= \tilde{E}\{(\mathbf{x}(t) - \hat{\mathbf{x}}_{t|t})(\mathbf{x}(t) - \hat{\mathbf{x}}_{t|t})^T | \log Y(u), 1 \le u \le t\} \\
\mathbf{P}_t &= \tilde{E}\{(\mathbf{x}(t) - \hat{\mathbf{x}}_{t|t-1})(\mathbf{x}(t) - \hat{\mathbf{x}}_{t|t-1})^T | \log Y(u), 1 \le u \le t-1\}
\end{aligned}$$

and the recursions (20) are initialised in the same way as the conventional Kalman filter (see Harvey (1989) for example). Given a finite sequence of observations $\log Y(1), \ldots, \log Y(T)$, the smoothed estimates $\hat{\mathbf{x}}_{t|T}$ for the approximating model follow the recursions

$$\begin{aligned}
\hat{\mathbf{x}}_{t|T} &= \hat{\mathbf{x}}_{t|t} + \mathbf{J}_t(\hat{\mathbf{x}}_{t+1|T} - \hat{\mathbf{x}}_{t+1|t}) \\
\mathbf{P}_{t|T} &= \mathbf{V}_t + \mathbf{J}_t((\mathbf{P}_{t+1|T} - \mathbf{P}_{t+1})\mathbf{J}_t^T \\
\mathbf{J}_t &= \mathbf{V}_t\mathbf{A}^T\mathbf{P}^{-1}_{t+1}
\end{aligned} \tag{21}$$

where

$$\mathbf{P}_{t|T} = \tilde{E}\{(\mathbf{x}(t) - \hat{\mathbf{x}}_{t|T})(\mathbf{x}(t) - \hat{\mathbf{x}}_{t|T})^T | \log Y(u), 1 \le u \le T\}$$

and $\mathbf{P}_{T|T} = \mathbf{V}_T$. These are the classical smoothing formulae for the fixed interval smoother. Other smoothers are possible with enhanced computational properties (see for example Koopman (1993) and Young et al. (1998)) and these could also be used here in place of the classical smoother.

The approximating model is fitted to the data by first using the filter (20), appropriately initialised, to determine the innovations $\nu(t)$ and the innovation variances $\sigma^2_{t|t-1}$. The likelihood L is then formed by calculating

$$-2\log L = T\log 2\pi + \sum_{t=1}^{T}(\log\sigma^2_{t|t-1} + \nu(t)^2/\sigma^2_{t|t-1}).$$

Although a variety of ways are available to initialise the filter and calculate the $\nu(t)$, $\sigma^2_{t|t-1}$ (see Harvey and Shephard (1993) for example), we have chosen in the examples that follow to treat the initial values as fixed parameters. If the approximating model is close to the true model, then the estimates should be Gaussian with the usual asymptotic properties. Based on this likelihood, the AIC criterion

$$AIC = -2\log L + 2d \tag{22}$$

can be used to select between competing models and assess goodness of fit. Here d denotes the number of parameters in the model.

It is clear that $\alpha(\mathbf{x}(t|t-1))$ and $\beta(\mathbf{x}(t|t-1))$ must be selected so that $\tilde{h}(\mathbf{x}(t))$ is close to $h(\mathbf{x}(t))$ and the approximating model is close to the original model (17). The conventional *extended Kalman filter* is obtained by expanding $h(\mathbf{x}(t))$ in a first-order Taylor series expansion about $\mathbf{x}_{t|t-1}$ which results in values of $\alpha(\hat{\mathbf{x}}_{t|t-1})$ and $\beta(\hat{\mathbf{x}}_{t|t-1})$ given by

$$\begin{aligned}\hat{\alpha}_0(\hat{\mathbf{x}}_{t|t-1}) &= \log(1+\mathbf{b}^T\hat{\mathbf{x}}_{t|t-1}) - \mathbf{b}^T\hat{\mathbf{x}}_{t|t-1}/(1+\mathbf{b}^T\hat{\mathbf{x}}_{t|t-1})\\ \hat{\beta}_0(\hat{\mathbf{x}}_{t|t-1}) &= \mathbf{a}+\mathbf{b}/(1+\mathbf{b}^T\hat{\mathbf{x}}_{t|t-1})\end{aligned} \tag{23}$$

with $\tilde{h}(\mathbf{x}(t))$ given by (19). Unfortunately this form of $\tilde{h}(\mathbf{x}(t))$ always overestimates $h(\mathbf{x}(t))$ due to the concavity of $h(\mathbf{x})$.

To improve on the extended Kalman filter approximation we note that $h(\mathbf{x}(t))$ comprises two terms, one linear in $\mathbf{x}(t)$ and the other a function of the linear form $\mathbf{b}^T\mathbf{x}(t)$. This leads us to restrict attention to approximations $\tilde{h}(\mathbf{x}(t))$ where $\beta(\hat{\mathbf{x}}_{t|t-1}) = \mathbf{a} + \gamma(\hat{\mathbf{x}}_{t|t-1})\mathbf{b}$ and $\gamma(\hat{\mathbf{x}}_{t|t-1})$ is some appropriately chosen scalar function of $\hat{\mathbf{x}}_{t|t-1}$. Setting

$$g(\mathbf{x}(t)) = \log(1+\mathbf{b}^T\mathbf{x}(t))$$

we now consider minimising

$$\tilde{E}_{t-1}\{(h(\mathbf{x}(t)) - \tilde{h}(\mathbf{x}(t)))^2\} = \tilde{E}_{t-1}\{(g(\mathbf{x}(t)) - \alpha(\hat{\mathbf{x}}_{t|t-1}) - \gamma(\hat{\mathbf{x}}_{t|t-1})\mathbf{b}^T\mathbf{x}(t))^2\} \tag{24}$$

where $\tilde{E}_{t-1}$ denotes expectation with respect to the approximating model and conditional on the data up to and including time $t-1$. However the minimising values of $\alpha(\hat{\mathbf{x}}_{t|t-1})$ and $\gamma(\hat{\mathbf{x}}_{t|t-1})$ given by

$$\begin{aligned}\alpha_1(\hat{\mathbf{x}}_{t|t-1}) &= \tilde{E}_{t-1}\{g(\mathbf{x}(t))\} - \gamma_0(\hat{\mathbf{x}}_{t|t-1})\mathbf{b}^T\hat{\mathbf{x}}_{t|t-1}\\ \gamma_1(\hat{\mathbf{x}}_{t|t-1}) &= \tilde{E}_{t-1}\{\mathbf{b}^T(\mathbf{x}(t)-\hat{\mathbf{x}}_{t|t-1})g(\mathbf{x}(t))\}/\mathbf{b}^T\mathbf{P}_t\mathbf{b}\end{aligned} \tag{25}$$

are of limited practical use since they incorporate expectations of terms involving the non-linear function $g(\mathbf{x}(t))$. To make these formulae operational we now approximate $g(\mathbf{x}(t))$ by a Taylor series expansion about $\hat{\mathbf{x}}_{t|t-1}$ and obtain

$$\begin{aligned} \hat{\alpha}_1(\hat{\mathbf{x}}_{t|t-1}) &= \hat{\alpha}_0(\hat{\mathbf{x}}_{t|t-1}) - \tfrac{1}{2}\mathbf{b}^T\mathbf{P}_t\mathbf{b}^T/(1+\mathbf{b}^T\hat{\mathbf{x}}_{t|t-1})^2 \\ \hat{\gamma}_1(\hat{\mathbf{x}}_{t|t-1}) &= 1/(1+\mathbf{b}^T\hat{\mathbf{x}}_{t|t-1}) \\ \hat{\beta}_1(\hat{\mathbf{x}}_{t|t-1}) &= \hat{\beta}_0(\hat{\mathbf{x}}_{t|t-1}) \end{aligned} \tag{26}$$

where we have retained terms to second order. These formulae are readily computed and differ from those of the extended Kalman filter only by a bias correction term.

Observing that $h(\mathbf{x}(t)) = 0$ when $\mathbf{x}(t) = 0$, an alternative strategy is to minimise (24) subject to the constraint $\alpha(\hat{\mathbf{x}}_{t|t-1}) = 0$. This is equivalent to approximating $h(\mathbf{x}(t))$ by the simple linear form

$$\tilde{h}(\mathbf{x}(t)) = \beta(\hat{\mathbf{x}}_{t|t-1})^T\mathbf{x}(t)$$

where $\beta(\hat{\mathbf{x}}_{t|t-1}) = \mathbf{a} + \gamma(\hat{\mathbf{x}}_{t|t-1})\mathbf{b}$ as before. In this case the minimising value for $\gamma(\hat{\mathbf{x}}_{t|t-1})$ is given by

$$\gamma_2(\hat{\mathbf{x}}_{t|t-1}) = \theta(\hat{\mathbf{x}}_{t|t-1})\frac{\tilde{E}_{t-1}\{g(\mathbf{x}(t)\}}{\mathbf{b}^T\hat{\mathbf{x}}_{t|t-1}} + (1-\theta(\hat{\mathbf{x}}_{t|t-1}))\gamma_1(\hat{\mathbf{x}}_{t|t-1}) \tag{27}$$

where

$$\theta(\hat{\mathbf{x}}_{t|t-1}) = \frac{(\mathbf{b}^T\hat{\mathbf{x}}_{t|t-1})^2}{\mathbf{b}^T\mathbf{P}_t\mathbf{b} + (\mathbf{b}^T\hat{\mathbf{x}}_{t|t-1})^2} = (1 + \frac{\mathbf{b}^T\mathbf{P}_t\mathbf{b}}{(\mathbf{b}^T\hat{\mathbf{x}}_{t|t-1})^2})^{-1}$$

and $\mathbf{b}^T\mathbf{P}_t\mathbf{b}/(\mathbf{b}^T\hat{\mathbf{x}}_{t|t-1})^2$ is the square of the coefficient of variation of $\mathbf{b}^T\mathbf{x}(t)$ given data up to and including time $t-1$. Since $0 \leq \theta(\hat{\mathbf{x}}_{t|t-1}) \leq 1$, $\gamma_2(\hat{\mathbf{x}}_{t|t-1})$ is a weighted average of $\gamma_1(\hat{\mathbf{x}}_{t|t-1})$ and

$$\gamma_3((\hat{\mathbf{x}}_{t|t-1}) = \frac{\tilde{E}_{t-1}\{g(\mathbf{x}(t)\}}{\mathbf{b}^T\hat{\mathbf{x}}_{t|t-1}}$$

which is a simple ratio of the average value of $g(\mathbf{x}(t))$ to the average value of $\mathbf{x}(t)$. To make these formulae operational, we expand $g(\mathbf{x}(t))$ in a Taylor series to second order as before to obtain

$$\hat{\gamma}_2(\hat{\mathbf{x}}_{t|t-1}) = \theta(\hat{\mathbf{x}}_{t|t-1})\hat{\gamma}_3(\hat{\mathbf{x}}_{t|t-1}) + (1-\theta(\hat{\mathbf{x}}_{t|t-1}))\hat{\gamma}_1(\hat{\mathbf{x}}_{t|t-1}) \tag{28}$$

and

$$\hat{\gamma}_3(\hat{\mathbf{x}}_{t|t-1}) = \frac{\log(1+\mathbf{b}^T\hat{\mathbf{x}}_{t|t-1}) - \frac{1}{2}\mathbf{b}^T\mathbf{P}_t\mathbf{b}^T/(1+\mathbf{b}^T\hat{\mathbf{x}}_{t|t-1})^2}{\mathbf{b}^T\hat{\mathbf{x}}_{t|t-1}} \tag{29}$$

where $\hat{\gamma}_3(\hat{\mathbf{x}}_{t|t-1})$ is defined to be zero when $\hat{\mathbf{x}}_{t|t-1} = 0$. The corresponding values of $\alpha(\hat{\mathbf{x}}_{t|t-1})$ and $\beta(\hat{\mathbf{x}}_{t|t-1})$ in $\tilde{h}(\mathbf{x}(t))$ are now given by

$$\begin{aligned} \hat{\alpha}_j(\hat{\mathbf{x}}_{t|t-1}) &= 0 \\ \hat{\beta}_j(\hat{\mathbf{x}}_{t|t-1}) &= \mathbf{a} + \hat{\gamma}_j(\hat{\mathbf{x}}_{t|t-1})\mathbf{b} \end{aligned} \qquad (j = 2, 3). \tag{30}$$

The various alternative approximations $\tilde{h}(\mathbf{x}(t))$ for $h(\mathbf{x}(t))$ are assessed in the next section.

In the above it has been tacitly assumed that both $1+\mathbf{b}^T\mathbf{x}(t)$ and $1+\mathbf{b}^T\hat{\mathbf{x}}_{t|t-1}$ are strictly positive so that their logarithms are well defined and $Y(t)$ remains positive. This will almost always be the case in practice. However, in cases where $T(t)(1+S(t)$ is close to zero it may be necessary to base the approximating model on $g(\mathbf{x}) = \log(1+\delta(\mathbf{b}^T\mathbf{x}(t)))$ where $\delta(x)$ is a suitably smooth function of x with $\delta(x) > -1$ for all x and $\delta(x) = x$ for $x > -1+\delta$, some small $\delta > 0$. This revised approximating model will yield filtering and smoothing formulae which are essentially the same as the above, but with a few minor adjustments.

Finally all that remains is to form the trend and seasonally adjusted series for the original observations $Y(1), \ldots, Y(T)$. The estimated trend is given by

$$\begin{aligned}\tilde{E}_T\{T(t)\} &= e^{\sigma^2/2}\tilde{E}_T\{e^{M(t)}\} \\ &= e^{\mathbf{a}^T\hat{\mathbf{x}}_{t|T}+\mathbf{a}^T\mathbf{P}_{t|T}\mathbf{a}/2+\sigma^2/2} \\ &= e^{\tilde{E}_T\{\log T(t)\}+\mathbf{a}^T\mathbf{P}_{t|T}\mathbf{a}/2} \qquad (31)\end{aligned}$$

where we have taken advantage of the conditional Gaussianity of the approximating model. Similarly the seasonally adjusted data is now given in terms of the approximating model by

$$\begin{aligned}\tilde{E}_T\{Y(t)/(1+S(t))\} &= Y(t)\tilde{E}_T\{e^{-\alpha(\hat{\mathbf{x}}_{t|t-1})-(\beta(\hat{\mathbf{x}}_{t|t-1})-\mathbf{a})^T\mathbf{x}(t)}\} \\ &= \frac{Y(t)}{1+\tilde{E}_T\{S(t)\}}e^{(\beta(\hat{\mathbf{x}}_{t|t-1})-\mathbf{a})^T\mathbf{P}_{t|T}(\beta(\hat{\mathbf{x}}_{t|t-1})-\mathbf{a})} \qquad (32)\end{aligned}$$

where we have again taken advantage of conditional Gaussianity to determine $\tilde{E}_T\{1/(1+S(t))\}$ and $\tilde{E}_T\{S(t)\}$. Thus the division by the seasonal factor $1+\tilde{E}_T\{S(t)\}$ again forms a direct part of the seasonal adjustment as it does for the X-11 procedures.

In the following section we apply these procedures to time series data. The various approximations are evaluated and the results compared with standard log-additive model-based procedures. Ideally we would like the estimates of the components and their standard errors obtained from the approximating model to be essentially the same as those obtained by fitting the original model. This is difficult to achieve in practice and we shall instead use measures such as AIC to determine quality of fit between competing models.

4 Application

We apply the procedures described in the previous sections to the monthly time series of short-term visitor arrivals to New Zealand over the 12 year period January 1980 to December 1991 inclusive. This series has been chosen because of its marked seasonal character, its amenability to global multiplicative parametric modelling and because it is part of a longer data set used in Ozaki and Thomson (1992) where further analyses were undertaken. It is plotted in Figure 1. The results of applying the nonlinear dynamic X-11 model to the visitor arrivals data are compared to X-11 itself and also to log-additive model-based methods. Using X-11 as our benchmark, we are concerned with determining any trend and seasonal biases, and with overall quality of fit.

Following an exploratory analysis, the nonlinear dynamic X-11 model was fitted to the data using the three seasonal models given in Section 2, the approximations $\tilde{h}(\mathbf{x}(t))$ described in Section 3 and a dynamic local linear model for the logarithms of the trend. The latter was chosen to be the random walk model

$$\Delta \log T(t) = \mu + \eta(t) \tag{33}$$

where $\eta(t)$ is Gaussian white noise and μ is a fixed drift parameter. This model for the $\log T(t)$, termed the "local linear trend" model by Harvey (1989), has been widely used in practice and is a member of the class of generalised random walk models (see Young et al. (1998)). Its state space representation is given by (6) with

$$\mathbf{m}_T = \begin{bmatrix} \log T(t) \\ \mu(t) \end{bmatrix}, \qquad \mathbf{A}_T = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \qquad \mathbf{B}_T = \begin{bmatrix} 1 \\ 0 \end{bmatrix},$$

$\mu(0) = \mu$ and $\mathbf{u}(t) = \eta(t)$. We denote the various approximations to $h(\mathbf{x}(t))$ by

$$\tilde{h}_j(\mathbf{x}(t)) = \hat{\alpha}_j(\hat{\mathbf{x}}_{t|t-1}) + \hat{\beta}_j(\hat{\mathbf{x}}_{t|t-1})^T \mathbf{x}(t) \qquad (j = 0, 1, 2, 3) \tag{34}$$

where $\tilde{h}_0(\mathbf{x}(t))$ denotes the extended Kalman filter given by (23), $\tilde{h}_1(\mathbf{x}(t))$ denotes the bias-corrected extended Kalman filter given by (26), and $\tilde{h}_2(\mathbf{x}(t))$, $\tilde{h}_3(\mathbf{x}(t))$ denote the simple linear and simple ratio approximations respectively given by (30).

No major difficulties were experienced with fitting the models and the optimisation procedures were all stable. However in the case of the BAYSEA seasonal models, estimates of the AR parameter ρ converged to unity, the boundary of the stationarity region. The choice of a seasonal component with a root on the unit circle leads to ambiguities in the overall seasonal-trend decomposition with the seasonal taking up variation which should belong to the trend. This was interesting as in an earlier analysis of the same series from April 1956 to January 1991 (see Ozaki and Thomson (1992)) a value of $\rho = 0.658$ was obtained. It is likely that the difference is due to the limited evolution of the seasonal pattern over the 12 year period considered here. Following den Butter and Mourik (1990), who noted similar difficulties with the unconstrained estimation of ρ, we fixed $\rho = 11/12$ for all the BAYSEA seasonal models so that $S(t)$ followed (8).

The AIC values (22) that result when fitting the various models are reported in Table 1. For each approximation method, the BAYSEA model gives the best AIC value and, of the approximations, there is a preference for $\tilde{h}_2(\mathbf{x}(t))$ or $\tilde{h}_3(\mathbf{x}(t))$ in terms of AIC value.

Seasonal	Approximation				Log-
Model	$\tilde{h}_0(\mathbf{x}(t))$	$\tilde{h}_1(\mathbf{x}(t))$	$\tilde{h}_2(\mathbf{x}(t))$	$\tilde{h}_3(\mathbf{x}(t))$	additive
BHT	-643.4	-643.4	-644.2	-643.8	-643.8
BAYSEA	-661.0	-661.0	-663.0	-662.7	-662.5
Parallel	-657.2	-657.3	-658.5	-658.5	-657.0

Table 1: The AIC values that result when the nonlinear dynamic X-11 model is fitted to the monthly short-term visitor arrivals to New Zealand. The models considered fit the local linear trend model (33) with each combination of the three seasonal models given in Section 2, and each approximation $\tilde{h}_j(\mathbf{x}(t))$ given by (34).

The differences between approximations, although small, are consistent. In general the choice of approximation will be largely dependent on the nature of the data. If strong seasonality is present, such as the case here, then $\tilde{h}_2(\mathbf{x}(t))$ or $\tilde{h}_3(\mathbf{x}(t))$ will be preferable. However for moderate or weak seasonality then any of the approximations could be used.

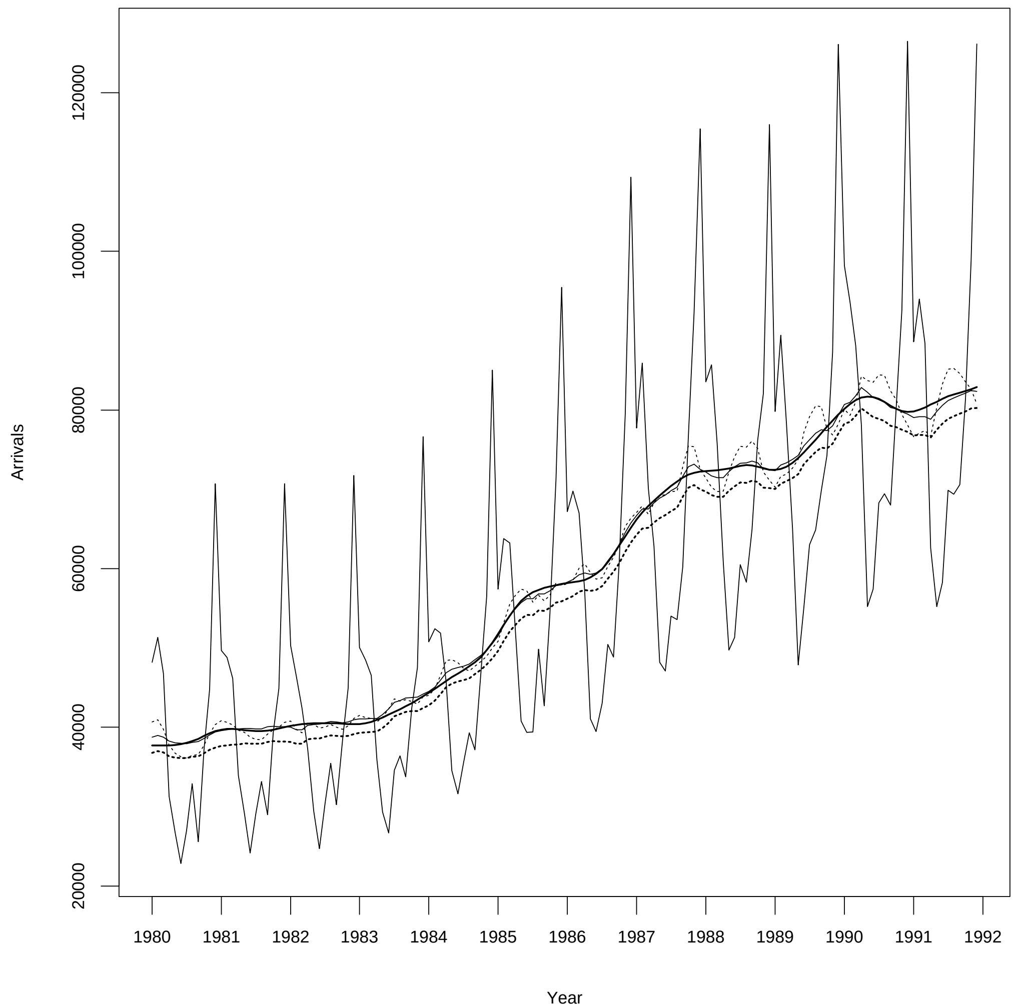

Figure 1: Short-term visitor arrivals to New Zealand by month with four trends superimposed: the trends for the nonlinear dynamic X-11 model with BAYSEA (thin solid line) and BHT (thin dotted line) seasonal models and approximation (34) with $j = 2$, the X-11 trend (thick solid line), and the trend for the log-additive model with BAYSEA seasonal (thick dotted line).

The estimated trends and seasonal components from the various models were compared to the trend and seasonal components obtained from X-11. To aid comparison, the latter were calculated using the standard multiplicative version of X-11 without ARIMA forecast extension and without trading day adjustment. When plotted as time series, all estimated trend and seasonal components were generally similar to those of X-11 with the exception of the BHT trend which was much more curvilinear. An illustration is given in Figure 1

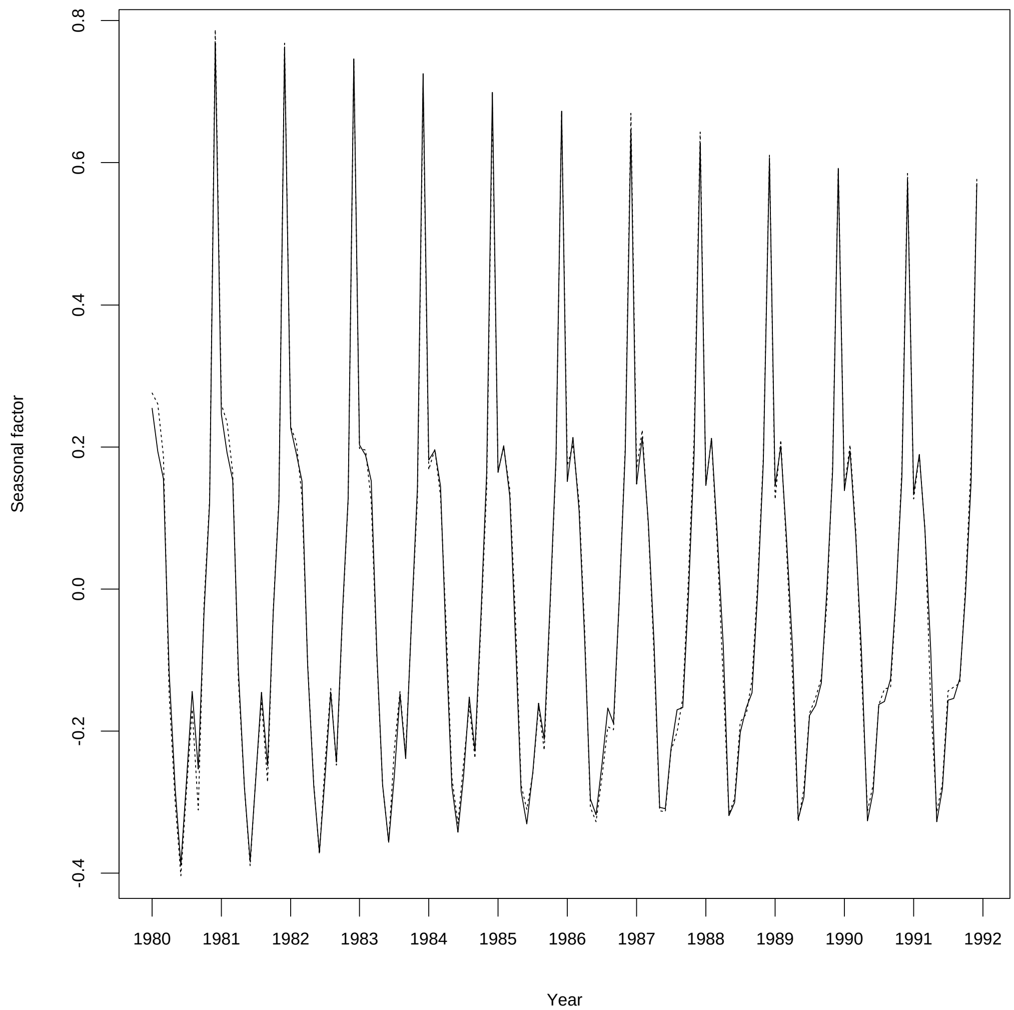

Figure 2: Short-term visitor arrivals to New Zealand by month: the fitted seasonal component for the nonlinear dynamic X-11 model with parallel seasonal (dotted line), and the X-11 seasonal (solid line).

where the trends for the nonlinear dynamic X-11 model with BHT and BAYSEA seasonal models and approximation (34) with $j = 2$ are compared to the X-11 trend. The nonlinear dynamic X-11 model with parallel seasonal yielded a trend which was almost exactly the same as the model with BAYSEA seasonal, although the former slightly more closely followed the X-11 trend at the end of the series.

Figure 2 plots the X-11 seasonal factor as defined in (1) with the fitted seasonal component for the nonlinear dynamic X-11 model with parallel seasonal where the latter is, again, almost exactly the same as the model with BAYSEA seasonal. From Figure 2 it would appear that there is very close agreement between the fitted X-11 and parallel seasonal factors, especially given that the X-11 seasonal pattern is itself not an absolute measure of seasonal variation.

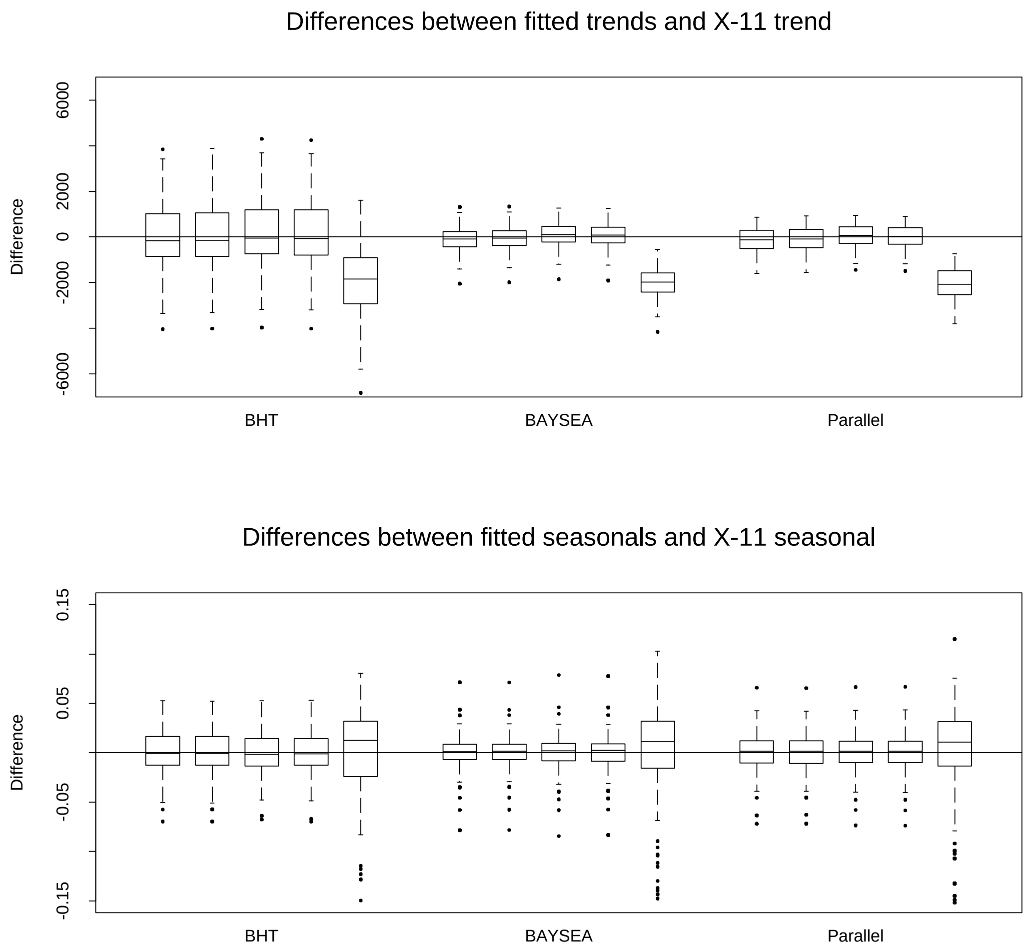

Figure 3: Short-term visitor arrivals to New Zealand by month: box plots of the differences between the various estimated trends and seasonals and the benchmark X-11 trend and seasonal respectively. For each seasonal model the jth box plot from the left ($j < 5$) refers to the approximation $\tilde{h}_{j-1}(\mathbf{x}(t))$ given by (34) and the fifth boxplot from the left within each cluster refers to the log-additive model.

Boxplots of the differences between the various estimated trends and seasonals and the benchmark X-11 trend and seasonal are given in Figure 3. As noted previously, the trend and seasonal differences are least for the nonlinear dynamic X-11 models with BAYSEA and parallel seasonals, where the latter two are difficult to distinguish. Although the difference in seasonal factors appears modest it needs to be born in mind that, as a proportion of the trend, small seasonal differences will translate to more marked differences in the scale of the observations. (2).

Figure 4 gives the boxplots of the moving sums $\sum_{j=0}^{11} S(t-j)$ for all the estimated seasonals. The BHT variation is least, followed by the X-11, BAYSEA and parallel seasonals.

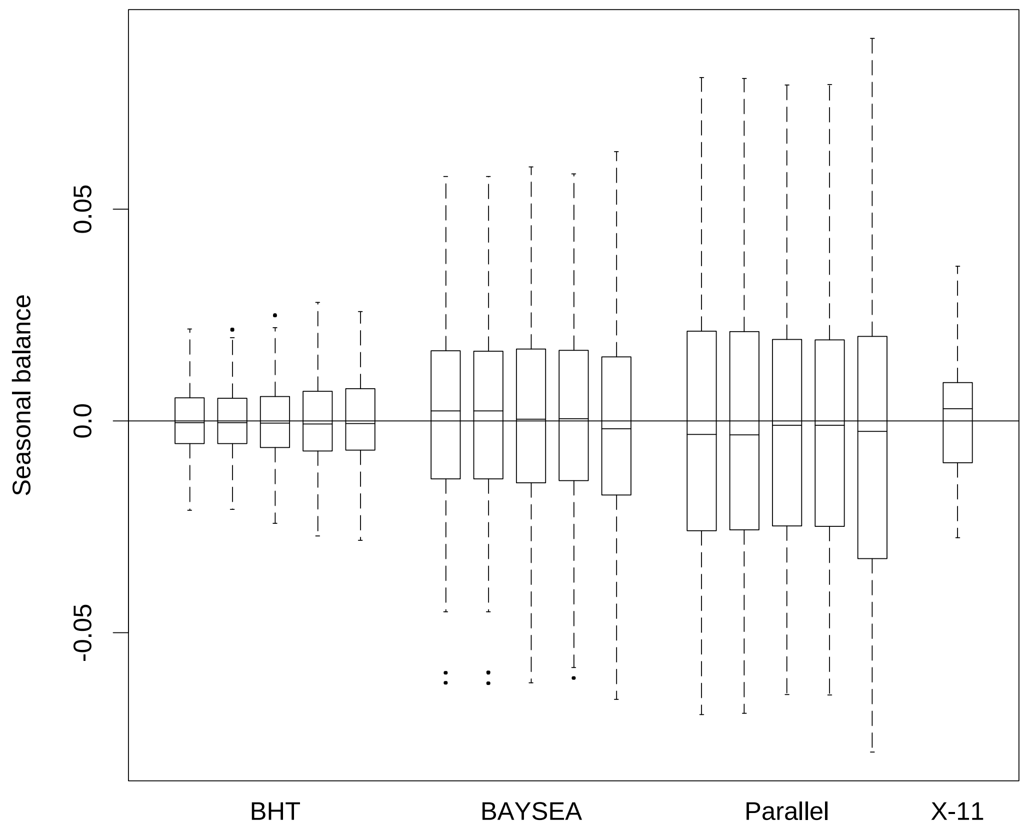

Figure 4: Short-term visitor arrivals to New Zealand by month: box plots of the moving sums $\sum_{j=0}^{11} S(t-j)$ for all the various estimated seasonals. For each seasonal model (with the exception of X-11) the jth box plot from the left ($j < 5$) refers to the approximation $\tilde{h}_{j-1}(\mathbf{x}(t))$ given by (34) and the fifth boxplot from the left within each cluster refers to the log-additive model.

Of the BAYSEA and parallel seasonal models, the seasonal balance constraint is best observed by the approximations $\tilde{h}_2(\mathbf{x}(t))$ and $\tilde{h}_3(\mathbf{x}(t))$. The X-11 moving seasonal sums are slightly positively biased, but are smaller in magnitude than either the BAYSEA or parallel models.

On balance the nonlinear dynamic X-11 model that best fits the data and most closely agrees with X-11 is the model with BAYSEA seasonal and approximation $\tilde{h}_2(\mathbf{x}(t))$ or $\tilde{h}_3(\mathbf{x}(t))$. However the difference between the BAYSEA and parallel seasonal models is small. Indeed the AIC values for the short-term visitor arrivals series from April 1956 to January 1991 showed a preference for the parallel seasonal model (see Ozaki and Thomson (1992)). Thus, in the case of longer series or where more seasonal evolution is present in the data, the parallel seasonal may well prove preferable.

Finally we compared the non-linear dynamic X-11 model to the log-additive parametric model. For the latter we fitted the local linear trend model (33) to the $\log Y(t)$ and used AIC to select between the BHT, BAYSEA and parallel seasonal models. (The AIC values are given in Table 1.) Once the components were fitted they were then back-transformed using the exponential function into the original scale of the data. The best fitting log-additive model used the BAYSEA seasonal model and its trend is plotted in Figure 1. A marked trend bias is evident. Both trend and seasonal biases from the X-11 benchmark are also clearly evident in Figure 3.

5 Conclusion

A nonlinear dynamic model for multiplicative seasonal time series has been introduced that follows and extends the X-11 paradigm. The model is not limited to the selection of trend and seasonal models considered here and can be further modified to handle outliers and special effects in much the same way as Bruce and Jurke (1992). The results are consistent with X-11 and, in particular, the dynamic nonlinear X-11 model does not suffer from the trend bias present in log-additive models.

Acknowledgements

The authors wish to record their thanks to Statistics New Zealand for providing the data on which the research was based, and to Mr Alistair Gray for helpful comments and assistance with X-11. The work was done under the auspices of a co-operative research agreement between the Institute of Statistical Mathematics and the Institute of Statistics and Operations Research. The first author also acknowledges financial support from the Takahashi Science Foundation. Both authors are grateful to the referees for a number of suggestions that lead to improvements in the paper and, in particular, the comparison with X-11.

References

Akaike, H. (1980) Seasonal adjustment by a Bayesian modelling. *Journal of Time Series Analysis* **1**, 1-13.

Akaike, H. and Ishiguro, M. (1980) BAYSEA, A Bayesian Seasonal Adjustment Program. Computer Science Monographs **13**, Institute of Statistical Mathematics, Tokyo.

Box, G.E.P., Hillmer, S.C. and Tiao, G.C. (1978) Analysis and modelling of seasonal time series. In *Seasonal Analysis of Economic Time Series* (ed. Zellner, A.), 309-334, U.S. Department of Commerce, Bureau of the Census, Washington D.C.

Bruce, A.G. and Jurke, S.R. (1996) Non-Gaussian seasonal adjustment: X-12-ARIMA versus robust structural models. *J. Forecast.* **9**, 305-328.

Cleveland, R.B., Cleveland, W.S., McRae, J.E. and Terpenning, I.J. (1990) STL: A seasonal-trend decomposition procedure based on Loess. *Journal of Official Statistics* **6**, 3-73.

Dagum, E.B. (1980) The X-11-ARIMA seasonal adjustment method. Research paper, Statistics Canada, Ottawa K1A 0T6.

den Butter, F.A.G. and Mourik, T.J. (1990) Seasonal adjustment using structural time series models: an application and a comparison with the Census X-11 method. *Journal of Business and Economic Statistics* **8**, 385-394.

Dongfeng, L., Thomson, P.J. and Haywood, J. (1997) Dynamic seasonal models. *Technical Report* 44, Institute of Statistics and Operations Research, Victoria University of Wellington, New Zealand.

Durbin, J. and Murphy, M.J. (1975) Seasonal adjustment based on a mixed additive-multiplicative model. *Journal of the Royal Statistical Society A* **138**, 385-409.

Findley, D.F., Monsell, B.C., Otto, M., Bell, W.R. and Pugh, M.G. (1988) Toward X-12-ARIMA. Proceedings of the Fourth Annual Research Conference, U.S. Bureau of the Census, 591-622.

Gersch, W. and Kitagawa, G. (1983) The prediction of time series with trends and seasonalities. *Journal of Business and Economic Statistics* **1**, 253-264.

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

Harvey, A.C. and Shephard, N. (1993) Structural time series models. In *Handbook of Statistics* **11** (eds. Maddala, G.S. et al), 261-302, North Holland, Amsterdam.

Hillmer, S.C. and Tiao, G.C. (1982) An ARIMA model based approach to seasonal adjustment. *Journal of the American Statistical Association* **77**, 63-70.

Jazwinski, A.H. (1970) *Stochastic Processes and Filtering Theory.* Academic Press, New York.

Kitagawa, G. (1987) Non-Gaussian state space modeling of nonstationary time series. *Journal of the American Statistical Association* **82**, 1032-1063.

Koopman, S.J. (1993) Disturbance smoother for state-space models. *Biometrika* **80**, 117-126.

Koopman, S.J., Harvey, A.C., Doornik, J.A. and Shepherd, N. (1995) *STAMP 5.0: Structural Time Series Analyser, Modeller and Predicter*, Chapman and Hall, London.

Lipster, R.S. and Shiryayev, A.N. (1978) *Statistics of Random Processes II Applications.* Springer-Verlag, New York.

Maravall, A. and Pierce, D.A. (1987) A prototypical seasonal adjustment model. *Journal of Time Series Analysis* **8**, 177-193.

Ng, C.N. and Young, P.C. (1990) Recursive estimation and forecasting of nonstationary time series. *Journal of Forecasting* **9**, 173-204.

Ozaki, T., Sosa, P.V. and Haggan-Ozaki, V. (1997) Reconstructing nonlinear dynamics from time series with application to epilepsy data analysis. *Technical Report* 1997/02, Manchester Centre for Statistical Science, UMIST, Manchester, UK.

Ord, J.K., Koehler, A.B. and Snyder, R.D. (1997) Estimation and prediction for a class of dynamic nonlinear statistical models. *Journal of the American Statistical Association* **92**, 1621-1629.

Ozaki, T. and Thomson, P.J. (1992) A dynamical systems approach to X-11 type seasonal adjustment. *Research Memorandum* 498, Institute of Statistical Mathematics, Tokyo, Japan.

Pole, A. and West, M. (1990) Efficient Bayesian learning in non-linear dynamic models. *Journal of Forecasting* **9**, 119-136.

Shiskin, J., Young, A.H. and Musgrave, J.C. (1967) The X-11 variant of the Census Method II seasonal adjustment program. *Technical Paper* 15, U.S. Department of Commerce, Bureau of the Census, Washington D.C.

Thomson, P.J. and Ozaki, T. (1992) Transformation and seasonal adjustment. *Research Memorandum* 491, Institute of Statistical Mathematics, Tokyo, Japan.

Young, A.H. (1968) Linear approximations to the Census and BLS seasonal adjustment methods. *Journal of the American Statistical Association* **63**, 445-471.

Young, P.C., Pedregal, D.J. and Tych, W. (1998) Dynamic harmonic regression. To appear in *Journal of Forecasting.*