A Non-Gaussian Airline Model for Seasonal Adjustment

John A.D. Aston, US Census Bureau
Siem Jan Koopman, Free University Amsterdam
Contact-John Aston, SRD, US Census Bureau, Washington, DC, USA (jaston@niss.org)

Key Words: Airline Model, ARIMA, Non-Gaussian Models, Seasonal Adjustment, Importance Sampling.

Abstract:

The Airline model, introduced by Box and Jenkins in their seminal book *Time Series Analysis: Forecasting and Control*, is routinely used to model economic time series. This model is parameterized by two factors, and gaussianity is usually assumed for the underlying noise component. Here, this model is generalised to include a non-Gaussian component to model outliers in the data. The model is examined using a state-space modelling approach, and importance sampling (see Durbin and Koopman). It utilises the decomposition method for ARIMA models developed by Hillmer and Tiao. This is necessary in order to preserve the airline structure whilst allowing a flexibility to include non-Gaussian noise terms for different components in the model. Different forms for the generalisation of the noise term are investigated. The models are interrogated through the use of a real series, the US Automobile Retail Series. The new models allow outliers to be accounted for, whilst keeping the underlying structures that are currently used to aid reporting of economic data.

1. Introduction

The aim of seasonal adjustment is straightforward: remove seasonal variations from time series observations. For many economic time series the seasonal cycle is clearly visible in a time series graph but it can not be observed. An important task of the U.S. Bureau of the Census is to extract the seasonal component in economic time series and to remove it from the series. Years of experience from experts and users have culminated in the current Census X-12-Arima procedure which is described and documented in Findley, Monsell, Bell, Otto, and Chen (1998) and Ladiray and Quenneville (2001). The Census X-12 program has become the standard seasonal adjustment method for many statistical agencies worldwide.

A simple approach to estimating seasonal effects that are additive is to consider seasonal means by computing the sample mean of observations associated with a particular season (e.g. calender months or quarters). With regression and other statistical models, seasonal effects can be estimated by including so-called seasonal dummy variables which in effect allow different means for different seasons. Alternatively, seasonal effects can be removed directly by considering the sum of s observations of the last year where s is the number of seasons in a year. When focus is on the growth rate of seasonally adjusted time series, this approach amounts to taking annual differences of the time series. For both approaches seasonal effects are determined by weighting the appropriate observations using zeros and ones as unnormalised weights. As a result the observation weights are not discounted when they lie further away. In the dummy case the weighting patterns are too wide for practical purposes while in the seasonal-sum case they are too short. More advanced methods of seasonal adjustment aim at discounting weighting patterns which are not too wide and not too short. Weighting patterns for seasonally adjusted data can be determined explicitly by choosing a set of moving average filters or implicitly by estimating a seasonal time series model. One of the most commonly used seasonal models is the airline model introduced by Box and Jenkins (1976) who used it to study a time series of monthly number of US airline passengers. The airline model is given by

$$(1-B)(1-B^s)y_t = (1-\theta B)(1-\Theta B^s)\xi_t \qquad (1)$$

where $\xi_t \sim \mathcal{N}(0,\sigma^2), t=1,\ldots,n$ with observation y_t, backshift operator B so that $By_t = y_{t-1}$ and $B^s y_t = y_{t-s}$, seasonal length s ($s=4$ for quarterly data and $s=12$ for monthly data) and white noise disturbance ξ_t. The model requires non-seasonal, $\Delta(B) = 1-B$, and seasonal, $\Delta_s(B) = 1-B^s$, differencing and is based on a moving average polynomial of degree $s+1$. The dynamic characteristics of the model depend on two parameters Θ and θ, which essentially describe the seasonal and non-seasonal structure of the data, respectively, although not completely independently of one another. The airline model falls within the class of seasonal autoregressive integrated moving average (ARIMA) models.

For any seasonal adjustment procedure, a difficult problem to handle in practice is the treatment of outliers in a time series. The identification of outliers for a given time series and a given model specification leads to two particular problems. The first is that an outlier can only be identified with respect to a specific model. An observation may be an outlier for one model but not for another model while both models can be nested or even fall within the same class. The second problem is related to the sample choice. An observation common to two different samples may be identified as an outlier in one sample while it is not an outlier for another sample. When an outlier is identified in a time series, it is not guaranteed that the observation will again be identified as an outlier when the time series is extended with more recent observations even if the same model and the same outlier detection method are used. Such outlier detection practices lead to seasonal

---

JA is a National Institute of Statistical Sciences Postdoctoral Fellow at US Census Bureau. Some of this work was carried out when SJK was an American Statistical Association visiting fellow at the US Census Bureau. This paper reports the results of research and analysis undertaken at the U.S. Census Bureau. It has undergone a Census Bureau review more limited in scope than that given to official Census Bureau publications. This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress.

adjustment procedures for which new outliers may be identified while previously identified outliers may later be regarded as regular observations. Therefore revisions of seasonal adjustments and sliding-spans statistics are highly sensitive to outliers; (see Findley, Monsell, Bell, Otto, and Chen (1998)). Since outliers and ways in which they are treated in seasonal adjustment procedures are highly influential with respect to the resulting seasonally adjusted series and its properties, official statisticians regard these as important problems that deserve attention.

Here, outliers in time series undergoing seasonal adjustment are considered utilising heavy tailed distributions to handle outliers within seasonal adjustment procedures. In this approach the detection and the seasonal adjustment itself are not considered as separate parts of the procedure but they are treated simultaneously. Whilst the airline model will be explored explicitly, this approach can be used for a wide variety of model classes in a similar way.

Earlier contributions in the treatment of outliers in the context of model-based seasonal adjustment are given by Chang, Tiao, and Chen (1988) who consider the estimation of general ARIMA models in the presence of outliers and by Hillmer, Bell, and Tiao (1983) who consider these methods in the context of seasonal adjustment. A different approximation technique is developed by Durbin and Cordero (1993) who present a treatment for outliers in the context of general state space models. The estimation techniques that we use in this paper rely on state space and importance sampling methods which are explored in Durbin and Koopman (2001). These methods lead to exact maximum likelihood estimates subject to Monte Carlo error.

2. The canonical airline model in state space form

The airline model is an example of the wider ARIMA class of models. Box and Jenkins (1976) discusses the airline model as a special case of seasonal ARIMA models. These are models which can be written as stationary ARMA models when appropriate seasonal and non-seasonal differences are taken. When the Box and Jenkins methodology of identifying ARIMA models is taken, many seasonal time series in economics and business can be effectively described by an airline model. Further the airline model is parsimonious and is relatively easy to estimate.

Model-based seasonal adjustment relies on the principle that the airline model can be decomposed into different unobserved components. The seasonal component is removed from the data to obtain seasonally adjusted data. The seasonal component is not observed and hence has to be inferred from the data itself through the use of a model.

2.1 Canonical decompositions

The aim is to decompose time series y_t into components for seasonal S_t, trend T_t and irregular I_t under the assumption that $y_t = S_t + T_t + I_t$ is modelled by the airline model. Typical models for the components are

$$\begin{aligned} \Delta(B)^2 T_t &= \theta_T(B)\omega_t \\ (1+B+\ldots+B^{s-1})S_t &= \theta_S(B)\eta_t \\ I_t &= \varepsilon_t \end{aligned} \tag{2}$$

and ω_t, η_t and ε_t are assumed to be white noise with variances σ^2_ω, σ^2_η and σ^2_ε respectively. The moving average polynomials $\theta_S(B)$ and $\theta_T(B)$ refer to the seasonal and trend components respectively. However, unless other assumptions are made the decompositions will not be unique. It is often advantageous to have known noise properties for the different components and as such one property that is commonly proposed is to completely remove white noise from the seasonal component. In the case of separate trend and irregular components, all the white noise from the trend component is removed as well. This results in a unique decomposition if the decomposition is possible (the possibility exists that there may not be enough white noise in the system to do this). The noise processes for each of the components are all assumed to be independent and as such this is one reason for the restriction on the parameter space that can be represented in the canonical form.

The decomposition was originally derived in work by Burman (1980) and Hillmer and Tiao (1982) who demonstrated how to calculate the separate components and the noise ratios for each of the new resulting components. Essentially this is based on a partial fraction decomposition and minimization,

$$\begin{aligned} \frac{\theta(B)\theta(B^{-1})}{\Delta(B)\Delta_s(B)\Delta(B^{-1})\Delta_s(B^{-1})}\sigma^2 &= \frac{\theta_S(B)\theta_S(B^{-1})}{U(B)U(B^{-1})}\sigma^2_\omega \\ &\quad + \frac{\theta_T(B)\theta_T(B^{-1})}{\Delta(B)^2\Delta(B^{-1})^2}\sigma^2_\eta + \sigma^2_\varepsilon \end{aligned} \tag{3}$$

where $\theta(B)$, σ^2 refer to the Airline model as defined in (1), that is considered for time series y_t. $U(B)$ is the seasonal sum operator, namely $U(B) = 1+B+\ldots B^{s-1}$. The partial fraction decomposition can be found by coefficient matching or any other suitable method, yielding a set of partial fractions and a remainder term. It was found that for the airline model, it was very efficient to find the coefficients using a matrix system of equations. The denominators are easy to construct for the airline model as the multiplied out polynomials are easy to compute ($x^{-2} - 4x^{-1} + 6 - 4x + x^2$ and $x^{-(s-1)} + 2x^{-(s-2)} + \ldots + s + \ldots + 2x^{s-2} + x^{s-1}$) and thus the matrix is easy to set up:

$$\left[\begin{array}{cccccccc} 1 & -4 & 6 & -4 & 1 & 0 & \ldots & 0 \\ 0 & 1 & -4 & 6 & -4 & 1 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & \ldots & 0 & 1 & -4 & 6 & -4 & 1 \\ \hline 1 & \ldots & s & \ldots & 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & s & \ldots & 1 & 0 & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & \ldots & 0 & 1 & \ldots & s & \ldots & 1 \end{array}\right]. \tag{4}$$

This matrix describes how to convert each of the two components, seasonal (top block) and non-seasonal (lower block), to the MA component of the original airline model. Thus the inverse of this matrix applied to the squared (where here squared is multiplying the forward and backward MAs) MA coefficients of the airline model then results in the squared MA coefficients for the individual components. The squaring of the original airline model MA parameters can easily be accomplished using the fast fourier transform or other polynomial multiplier.

In the following we adopt the frequency domain approach of signal extraction, see Bell (1984). The partial fraction components contain white noise and as such are not the canonical form that we require. To remove white noise, it must be possible for the components to have a minimum value, across all frequencies, of zero, *i.e.* the pseudo-spectrum must have a minimum value of zero. This is because white noise has a flat spectrum, and as such the addition or subtraction of a constant to the spectrum can be seen as changing the white noise component. Hence if

$$\gamma_S = min_w \frac{\theta_S(e^{iw})\theta_S(e^{-iw})}{U(e^{iw})U(e^{-iw})}\sigma^2_\omega \quad (5)$$

$$\gamma_T = min_w \frac{\theta_T(e^{iw})\theta_T(e^{-iw})}{\Delta(e^{iw})^2\Delta(e^{-iw})^2}\sigma^2_\eta, \quad (6)$$

then by subtracting the minimum from each of the components and adding $\gamma_S+\gamma_T$ to σ^2_ε, the canonical decomposition is formed. However, when $\gamma_S + \gamma_T + \sigma^2_\varepsilon < 0$, the decomposition is said to be inadmissible and a model decomposition with the desired properties does not exist. Hillmer and Tiao (1982) showed that the decomposition is always admissible for the airline model if $\Theta > 0$. Assuming the canonical form is admissible, then the minimums are subtracted from the original decompositions and the new decompositions are formed. These are still in the form of squared components so these are then solved to find the underlying MA models for each of the components. Several methods have been proposed to do this, including simple root solving (the roots come in inverse pairs due to the nature of the polynomial) or a method developed by Maravall and Mathis (1994) which allows more numerical stability in the root finding for large polynomials.

2.2 State space representation of decomposition model

The ARMA model can be put in state space in different ways; see, for example, Problem 2.6 of Anderson and Moore (1979). Also any ARIMA model can be written in state space form. The general state space model is given by

$$y_t = Z\alpha_t + u_t, \qquad \alpha_{t+1} = T\alpha_t + Rv_t, \qquad t = 1, \ldots, n, \quad (7)$$

where

$$u_t \sim \mathcal{N}(0, \sigma^2 H), \qquad v_t \sim \mathcal{N}(0, \sigma^2 Q), \qquad \alpha_1 \sim \mathcal{N}(a, \sigma^2 P), \quad (8)$$

with $m \times 1$ state vector α_t, modelled as a vector autoregressive process, and observation y_t, depending linearly on α_t. The system matrices Z, T and R are fixed and known but some elements may depend on, for example, the coefficients of the autoregressive and moving average polynomials. The ARMA model, is given in state space form as (7) and (8) with the first element of the state vector α_t equal to y_t (remaining elements are auxiliary variables) and with

$$Z' = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}, T = \begin{bmatrix} \phi_1 & 1 & 0 & \cdots & 0 \\ \phi_2 & 0 & 1 & & 0 \\ \vdots & \vdots & & \ddots & \\ \phi_{m-1} & 0 & 0 & & 1 \\ \phi_m & 0 & 0 & \cdots & 0 \end{bmatrix}, R = \begin{pmatrix} 1 \\ \theta_1 \\ \vdots \\ \theta_{m-2} \\ \theta_{m-1} \end{pmatrix} \quad (9)$$

with $m = \max(p, q+1)$ and further $H = 0$ and $Q = 1$. The mean vector of the initial state vector is $a = 0$ and the initial variance matrix, P, is the solution of

$$(I - T \otimes T)\,\text{vec}(P) = \text{vec}(RR'). \quad (10)$$

The general ARIMA model can also be formulated in state space form in similar ways. State elements associated with $D(B)$, the non-stationary part of the ARIMA model, require special initialisation conditions.

The ARIMA components model can be formulated in state space form by incorporating the individual components (i.e. Trend, Seasonal and Irregular) in one state vector; see discussion in Bell (2003). For example, consider the decomposition model with the ARIMA components S_t and N_t as implied by the canonical decomposition of the airline model. The state space form for this model is a combined version of the individual models for the components.

3. Treatment of outliers for seasonal adjustment

Outliers can be problematic for seasonal adjustment since the estimated seasonal component, whether model-based methods are used or not, can be heavily influenced by the presence of outliers in the time series. A common method for identifying outliers is to include an outlier dummy for a specific observation at time t. The corresponding coefficient is estimated simultaneously with the other parameters of the model. When the dummy coefficient is significant for a certain confidence level, the associating observation is then treated as an outlier. This procedure is repeated for every observation. When the model is cast in state space form, a more efficient method of computing outlier diagnostic tests for all observations can be adopted using the disturbance smoothing algorithm; see Harvey and Koopman (1992). In the case of the ARIMA components model, the outlier dummy variable can be incorporated within the framework of regression ARIMA components models of Bell (2003).

The aforementioned method of detecting outliers based on outlier dummies is not satisfactory since it relies on a variety of subjective decisions of which usually little reference can be made to underlying statistical properties. Also, once an observation is detected to be an outlier and the associated dummy variable is kept in the model, the observation is effectively removed from the time series. It is an all or nothing strategy. A more compelling strategy would be to weight observations in some effective way so that less weight is given to potential outlying observations while more weight is given to other observations. It is therefore preferable to model outliers as a component and to include them as part of the model. This leads to the introduction of a component for outliers that is part of the ARIMA components model. Various distributional assumptions for this component can be considered including heavy tailed distributions such as t and mixture of normals.

3.1 Outliers in the Airline Model

Modelling outliers as part of the error term ε_t in the airline model (1) is problematic. For example, instead of a Gaussian disturbance term ε_t a heavy tailed distribution for ε_t may be considered. However, the disturbance term of the airline model is associated with all the dynamics, including the ones associated with the seasonal and irregular components, as opposed to just a part of it; see also the discussion in Box and Tiao (1975). It is therefore not easy to discriminate the irregular error from the seasonal and trend components. However, the additive outliers should be associated with the irregular rather than with all components. Thus, the heavy tailed distribution must apply to the irregular but not to the trend and seasonal components. Here the theory of canonical decomposition can come into play. The decomposition allows the addition of an error component into the irregular term which can follow a heavy tailed distribution without that term being associated with the seasonal or the trend. However, care must be taken when outlining the new model so that the structure of the model still remains of an airline form. This is important so that the underlying economic understandings derived from the model are not lost. The general component models allow for a much wider range of options than the underlying airline model and as such estimation must be constrained so that the airline model structure (or a version as close as possible) is still implicit in the model. A similar problem applies to the trend component if a heavy tailed distribution is to be used for its noise component so that breaks in trends, also known as level shifts, can be modelled.

3.2 Outlier component

When the airline model is represented by the ARIMA components model implied by the canonical decomposition, it includes an irregular component. Although this white noise irregular component is independent of the trend and seasonal components, it depends on the airline coefficients θ, Θ and σ^2 via the canonical construction in the same way as the other two ARIMA components depend on them. This suggests that another white noise component, that is independent of the imposed airline structure, may be identified. In the case where the time series y_t is truly generated by airline model (1) without any additional shocks, the outlier component will be zero since it does not come into play and, when it is considered for estimation, it will not be significant. When the generation of y_t is distorted by additive noise that is not related to the implied trend and seasonal dynamics of the airline model, the outlier component will be non-zero to accommodate potential outliers.

The outlier component O_t can be modelled in various ways. An obvious choice is to have it as a Gaussian white noise term with a relative large variance, that is

$$O_t \sim \mathcal{N}(0, \sigma_O^2), \qquad t = 1, \ldots, n, \tag{11}$$

with $\sigma_O^2 >> \sigma_\varepsilon^2$. The outlier component is independently distributed amongst all other disturbances in the model and for all disturbances at different time points. For noisier time series a heavy tailed distribution can be considered with the t or the mixture of normals distributions as obvious examples. The definition of the outlier component in the case of a t distribution is given by

$$O_t \sim t(\nu, \sigma_O^2), \qquad t = 1, \ldots, n, \tag{12}$$

where $\nu > 2$ is the number of degrees of freedom and σ_O^2 is the variance which is constant for any ν. In the other case of an outlier component modelled by a mixture of normals, we have

$$O_t \sim (1-\psi)\mathcal{N}(0, \sigma_O^2) + \psi\mathcal{N}(0, \sigma_O^2\lambda), \qquad t = 1, \ldots, n, \tag{13}$$

where $0 \leq \psi \leq 1$ determines the intensity of outliers in the series and λ measures the magnitude of the outliers.

The inclusion of an outlier component will add at least one (in the Gaussian case) or more (in the non-Gaussian case) coefficients which brings the total number of four or higher parameters to be estimated. The extra variance parameter that needs to be considered can be dropped if it is possible to fix it as a multiple of the irregular variance of the canonical decomposition. This reduces the parameter space at the cost of less flexibility in modelling the time series. If multiple breaks in the trend also need consideration, it would be possible in turn to add a further one or two parameters to the model to account for this but would again increase the problems of identifying unknown parameters from usually short time series.

Apart from adding the outlier component O_t to the decomposition $y_t = S_t + T_t + I_t$, another way of incorporating the outlier component exists. An outlier term attempts to model a discrete function with a continuous approximation. Outliers are points which do not lie within the limited statistically likely space defined by the model. Whilst we are modelling outliers as a continuously distributed function, the known case would just be to exclude an outlying point from the model, a discrete function. One possibility to improve the approximation is to include the underlying irregular term from the airline model along with the discrete outlier function to generate a new inclusive noise term. In effect, the irregular term I_t is then replaced by the outlier component O_t such that

$$y_t = S_t + T_t + O_t, \qquad O_t \sim t(\nu, \sigma_O^2), \qquad t = 1, \ldots, n.$$

where the t-density approximates the density of the sum of the irregular component I_t and the discrete outlier function. This model has the same number of parameters as the earlier specifications, the only difference being whether the two noise components are modelled separately or together. However in both cases the component O_t can be regarded as continuous in the sense that it is defined at every point even in the known case.

Whilst the first formulation of the decomposition model with O_t added and assumed to be an observational error does not require use of the canonical form in its estimation, the second with O_t replacing I_t does require estimation through the canonical form, due to the inability to write this model in terms of the non-canonical airline model.

4. Illustration

A real data set from the US Census Bureau with outliers was examined. All computations have been carried out using the object-oriented matrix programming environment of `Ox`, see Doornik (2001). Extensive use is made of the state space functions in the `SsfPack` library, see Koopman, Shephard, and Doornik (1999).

For Gaussian models, once the model is in state space form and all disturbances are assumed Gaussian, the Kalman filter can be used to construct the exact likelihood function. However, non-gaussian estimation of the model is carried out using the importance sampling and simulation methods described in detail in both Durbin and Koopman (2000) and Durbin and Koopman (2001).

4.1 US Automobile Retail Series

A data set from the US Census Bureau was investigated to evaluate the effectiveness of the procedure. This series was examined and it was determined that there were outliers present using a longer version of the data set. However when a shortened subset of the data was taken, it was found that the outliers, assumed to be in the data, depended on the threshold that was chosen for outlier detection. If a relatively small threshold was used (that is 3.0) then the two of the three outliers determined for the longer series were found, whilst if this number was larger (that is 4.5) then none of the outliers were detected. Typical values of thresholds used in seasonal adjustment tend to be high due to the large number of tests that are being performed. This difference allows illustration of the advantages of using a t-distribution, as a predetermined threshold for the detection of the outliers is not required.

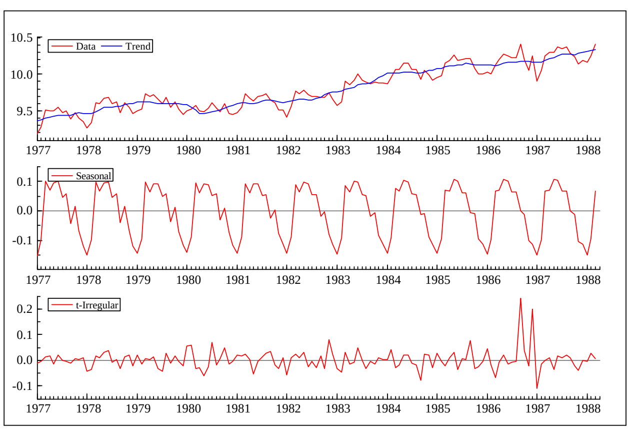

Figure 1: US Retail Sales of Automobiles (Jan 1977-Mar 1988, source: US Census Bureau) and estimated components from a three components model with t-distribution for outlier component: (top) series y_t and trend T_t; (middle) seasonal component S_t; (bottom) outlier component O_t.

The components resulting from fitting the model with the t-distribution to account for outliers can be seen in Figure 1. In the four components model, the outlier component does not only account for the outliers but also accounts for some of the other noise in the system. This is due to the continuous nature of the t-distribution. Whilst the irregular component left in the Gaussian part of the model is still present, it is now different to the component in the original Gaussian only model. However, in the three components model, the irregular component is replaced by the outlier component which does not lead to this problem of the noise being separately modelled. Whilst only the three components model output is shown in Figure 1 the four components model is very similar but has a slight tendency to have problems distinguishing the two different error components.

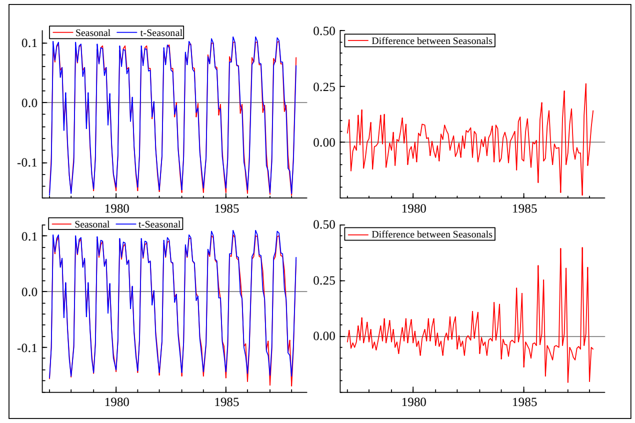

Figure 2: US Retail Sales of Automobiles (Jan 1977-Mar 1988, source: US Census Bureau) and estimated components from a Gaussian model and a three components model with t-distribution for outlier component: (top) seasonal component from Gaussian model with outlier threshold value of 3.0 (seasonal) and seasonal from three components t-model (t-seasonal) and the difference between the two series; (bottom) seasonal component from Gaussian model with outlier threshold value of 5.0 (seasonal) and seasonal from three components t-model (t-seasonal) and the difference between the two series. The differences are normalised to the root mean square of the seasonal signal.

In Figure 2, the seasonal components from the two different models are overlayed with the two different threshold levels on the subsequent graphs. As can be seen, if the threshold is set too high, in this case, the seasonality is markedly different from the t-seasonality as the presence of the outliers is not detected. It is most noticeable towards the end of the series. This can be more easily seen in the difference between the two. It should also be noted that the changes in seasonality are largest (up to 40% of the root mean square of the seasonal signal) near the outliers, they are still present even in months much earlier than the outliers. However, as can also be seen, if the outliers are detected then the seasonal components in the Gaussian and t-case are much more alike. Even subtle changes can have a big effect, especially on yearly changes of seasonally adjusted data. Thus it is likely that the problem of where to set the threshold, so that only outliers are removed and not actual data, can be ignored if a t-distribution is used, where no predetermined threshold is needed.

5. Discussion

This paper discusses the role and treatment of outliers in seasonal adjustment. Although seasonal adjustment can take many forms, here we are concerned with unobserved component ARIMA model based seasonal adjustment where a model is used to seasonally adjust the data, as opposed to a filter based approach such as the X-11 approach. However, the model based ideas can encompass either the structural modelling ideas of STAMP or alternatively the airline and other ARIMA based canonical decomposition models. Traditional treatment of outliers is based on the use of dummy regression variables to account for the outliers in the data evaluated on a model by model basis for each sample of the data. This is an all or nothing approach where the data point under evaluation is either included or excluded from the model based on a subjective threshold and the current model, including any previously identified outliers, under evaluation.

A method of using heavy tailed distributions to account for outliers is investigated. The heavy tailed distributions are applied here to the irregular component in the model decomposition in order to account for additive outliers although this generalises to other outlier types by assigning heavy tailed distributions to other components, for example the trend component to account for level shifts in the data. This method allows for an objective method of determining the outliers through a form of weighting for the data, as opposed to the all or nothing subjective approach. The seasonal adjustment will be less sensitive to the outliers in the series as they will be separately dealt with on a continuum basis as opposed to the their entire inclusion or exclusion. In the automobile series example, it was seen that the inclusion or exclusion of a point outlier can have a marked effect on the seasonal component and thus on the seasonally adjusted series, whilst the seasonal adjustment taking into account the t-distributed outlier component is more stable and closely resembles the series where all outliers are adequately captured.

There are assumptions made in the method and it is necessary to realise the limitations these impose. It is important to remember that any additional components to be estimated in the model make the entire model more difficult to identify and as such accounting for many different types of outliers such as the ones in the seasonal component for example, as well as level shift and additive outliers, may lead to the model becoming unidentifiable for short series. Longer series may also have problems due to the changing nature of the data over the time period, as definitions and other non-statistical effects occur in the data. Thus it is important to determine which components are most effectively modelled, which is a problem for all model based procedures. Also importance sampling, whilst working well for most t-distributed data can break down as the degrees of freedom become too small. In most cases this will not be a problem, and limits can be set during the estimation to ensure the degrees of freedom are only estimated over an estimable range.

The new methods presented here for seasonally adjusting the airline model lead to a more robust adjustment in the presence of outliers whether or not these outliers are well known. Although the focus here was for the airline model, these methods are easily extendable to other model based decompositions and adjustments and will provide a less subjective and more robust framework in which to carry out the seasonal adjustment of many economic time series.

References

Anderson, B. D. O. and J. B. Moore (1979). *Optimal Filtering*. Englewood Cliffs: Prentice-Hall.

Bell, W. R. (1984). Signal extraction for nonstationary time series. *Annals of Statistics 13*, 646–64.

Bell, W. R. (2003). On RegComponent time series models and their applications. working paper, January 2003, Statistics Research Division, U.S. Bureau of the Census.

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

Box, G. E. P. and G. C. Tiao (1975). Intervention analysis with applications to economic and environmental problems. *Journal of American Statistical Association 70*, 70–79.

Burman, J. P. (1980). Seasonal adjustment by signal extraction. *Journal of Royal Statistical Society A 143*, 321–37.

Chang, I., G. C. Tiao, and C. Chen (1988). Estimation of time series parameters in the presence of outliers. *Technometrics 30*, 193–204.

Doornik, J. A. (2001). *Object-Oriented Matrix Programming using Ox 3.0*. London: Timberlake Consultants Press.

Durbin, J. and M. Cordero (1993). Handling structural shifts, outliers and heavy-tailed distributions in state space time series models. working paper, February 1993, Statistics Research Division, U.S. Bureau of the Census.

Durbin, J. and S. J. Koopman (2000). Time series analysis of non-Gaussian observations based on state space models from both classical and Bayesian perspectives (with discussion). *Journal of Royal Statistical Society B 62*, 3–56.

Durbin, J. and S. J. Koopman (2001). *Time Series Analysis by State Space Methods*. Oxford: Oxford University Press.

Findley, D. F., B. C. Monsell, W. R. Bell, M. C. Otto, and B. C. Chen (1998). New capabilities of the X-12-ARIMA seasonal adjustment program (with discussion). *Journal of Business and Economic Statistics 16*, 127–77. http://www.census.gov/ts/papers/jbes98.pdf.

Harvey, A. C. and S. J. Koopman (1992). Diagnostic checking of unobserved components time series models. *Journal of Business and Economic Statistics 10*, 377–89.

Hillmer, S. C., W. R. Bell, and G. C. Tiao (1983). Modeling considerations in the seasonal adjustment of economic time series. In A. Zellner (Ed.), *Applied Time Series Analysis of Economic Data*, pp. 74–100. Washington, D. C: U. S. Bureau of the Census.

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

Koopman, S. J., N. Shephard, and J. A. Doornik (1999). Statistical algorithms for models in state space form using SsfPack 2.2. *Econometrics Journal 2*, 113–66. http://www.ssfpack.com/.

Ladiray, D. and B. Quenneville (2001). *Seasonal Adjustment with the X-11 Method (Lecture Notes in Statistics)*. New York: Springer-Verlag.

Maravall, A. and A. Mathis (1994). Encompassing univariate models in multivariate time series: a case study. *Journal of Econometrics 61*, 197–233.