ON A FAMILY OF MOVING-AVERAGE TREND FILTERS FOR THE ENDS OF SERIES

Alistair Gray and Peter Thomson
Institute of Statistics and Operations Research
Dr Peter Thomson, Victoria University of Wellington, PO Box 600, Wellington, New Zealand

Keywords: Local Dynamic Model; Revisions; Linear Unbiased Prediction; Linear Biased Prediction.

1 Introduction

Many seasonal adjustment procedures decompose time series into trend, seasonal, irregular and other components using non-seasonal moving-average trend filters. This paper is concerned with the extension of the central moving-average trend filter used in the body of the series to the ends where there are missing observations.

For any given central moving-average trend filter, a family of end filters is constructed using a minimum revisions criterion and a local dynamic model operating within the span of the central filter. These end filters are equivalent to evaluating the central filter with unknown observations replaced by constrained optimal linear predictors. Two prediction methods are considered. Best Linear Unbiased Prediction (BLUP) and Best Linear Biased Prediction, where the bias is time invariant (BLIP). The BLIP end filters are shown to be a generalisation of those developed by Musgrave (1964) for the central X-11 Henderson filters and include the BLUP end filters as a special case.

The theoretical properties of BLUP's and BLIP's are examined. In particular, it is established that the BLIP end filters generally have smaller mean squared revisions than the BLUP end filters. However, unlike the BLUP filters, the BLIP filters are no longer independent of the parameters in the local dynamic model and so, in practice, it is possible that a mis-specification of these parameters will lead to BLIP end filters with greater mean squared revisions than BLUP end filters. The effects of such mis-specification are discussed. Comparisons are also made between these end filters and the Musgrave end filters used by X-11, and the end filters obtained when the central filter is evaluated with unknown observations predicted by global ARIMA models. The latter parallels the ARIMA forecast extension method used in X-11-ARIMA. Finally these filters are evaluated on some New Zealand time series.

2 Local dynamic model

The conventional paradigm for trend filter design is to consider a moving window of $n = 2r + 1$ observations within which an estimate of the trend is to be calculated for the central time point.

Within the finite window we choose to model the observations as

$$y_t = g_t + \epsilon_t \tag{1}$$

where the trend g_t is given by

$$g_t = \sum_{k=0}^{p} \beta_j t^j + \xi_t. \tag{2}$$

The zero mean stochastic process ξ_t is assumed to be correlated, but uncorrelated with ϵ_t, and ξ_t, σ^2 are assumed to be not both zero. In particular we consider the situation where the β_j and σ^2 are parameters local to the window, but p, n and the model for ξ_t/σ involve global parameters which are constant across windows. Thus, although the parameters involved with the mean and variance of y_t vary across windows, the autocorrelation structure of y_t will be a function of time invariant parameters in addition to time itself.

Loosely speaking, the finite polynomial is intended to capture deterministic low order polynomial trend whereas ξ_t is intended to capture smooth deviations from the polynomial trend. Note that it is the incorporation of ξ_t which distinguishes this local model from the standard situation where it is zero. Among the anticipated benefits of including ξ_t are lower values of p and improved performance at the ends of series.

Because the window is not likely to be large the model will need to involve as few parameters as possible on the one hand, while allowing for a sufficiently flexible family of forms for g_t on the other. *With these points in mind we choose to model ξ_t as a (possibly integrated) random walk with initial value zero.* In particular, if Δ denotes the backwards difference operator satisfying $\Delta X_t = X_t - X_{t-1}$, we have in mind the situation where $\Delta^{p+1} g_t = \Delta^{p+1} \xi_t$ is a zero mean stationary process. In keeping with this rationale, we shall always assume that the levels of integration of the random walk components that make up ξ_t do not exceed $p + 1$.

This seems an appropriate and parsimonious model which should account for smooth deviations from the deterministic polynomial trend component. It also provides a dynamic trend model for g_t which is essentially of the same form as that used in the ARIMA structural models that have been successfully applied to economic and official data. (See Bell (1993) and Kenny and Durbin (1982) for example.)

In the local linear case $p = 1$ a simple dynamic model for g_t is given by

$$y_t = g_t + \epsilon_t = \beta_0 + \beta_1 t + \xi_t + \epsilon_t \qquad (3)$$

where β_0, β_1 are constants and ξ_t is a simple random walk satisfying

$$\xi_t = \xi_{t-1} + \eta_t$$

with $\xi_0 = 0$. Moreover ϵ_t and η_t are mutually uncorrelated white noise processes with variances σ^2 and $\sigma_\eta^2 = \lambda\sigma^2$ respectively.

In the local quadratic case $p = 2$ a simple dynamic model for g_t is given by

$$y_t = g_t + \epsilon_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \xi_t + \epsilon_t \qquad (4)$$

where β_0, β_1, β_2 are constants and ξ_t is a simple random walk satisfying the same conditions as in the local linear model.

Preliminary analysis indicates that these models have properties that can be regarded as representative of other more general models of the type discussed above.

3 Trend filter design at the ends

Now consider a finite window of width $n = 2r + 1$ points centred at time point t and within which the observations follow the local dynamic model given in Section 2. We consider the case where the trend g_t is to be estimated by a given central moving-average trend filter

$$\hat{g}_t = \sum_{s=-r}^{r} w_s y_{t+s}. \qquad (5)$$

In keeping with standard practice, we assume that the w_s are constrained by the requirement that

$$E\{\hat{g}_t - g_t\} = 0 \qquad (6)$$

so that $\hat{g}_t$ is an unbiased estimator of g_t. Note that this condition is equivalent to the requirement that the w_s satisfy

$$\sum_{s=-r}^{r} w_s = 1 \qquad \sum_{s=-r}^{r} s^j w_s = 0 \qquad (7)$$

where $0 < j \leq p$ so that the moving average filter passes polynomials of degree p.

At the ends of series the central moving average filter (5) will involve future unknown observations. How these missing observations should be treated is open to question.

A common and natural approach involves forecasting the missing values, either implicitly or explicitly, and then applying the desired central filter. The forecasting methods used range from simple extrapolation to model based methods, some based on the local trend model adopted, others based on global models for the series as a whole. The latter include the fitting of ARIMA models to produce forecasts (see Dagum (1980) in particular). The principle of using prediction at the ends of series seems a key one which goes back to DeForest (1877). See also the discussion in Cleveland (1983), Greville (1979) and Wallis (1983).

Yet another way to handle the missing values in the window is to employ additional criteria specific to the ends of the series. An important requirement, especially among official statisticians, is to keep seasonal adjustment revisions and therefore trend revisions to a minimum as more data comes to hand. Thus, at the ends of series, a natural criterion to consider is

$$R_q = E\{(\sum_{s=-r}^{r} w_s y_{t+s} - \tilde{g}_t)^2\} \qquad (8)$$

where $\tilde{g}_t$ is a predictor of $\hat{g}_t = \sum_{s=-r}^{r} w_s y_{t+s}$ based on past data. In general, given a history of observations $y_1, \ldots, y_T$, it is evident that R_q is minimised when

$$\tilde{g}_t = \sum_{s=-r}^{r} w_s \hat{y}_{t+s} \qquad (9)$$

where $\hat{y}_{t+s} = E(y_{t+s}|y_1, \ldots, y_T)$ denotes the best predictor of y_{t+s} in the usual mean squared error sense. Thus there is a close relationship between the minimum revisions approach and that of forecasting the missing values in the window.

The minimum revisions strategy appears to have been originally proposed by Musgrave (1964) for the case where $\tilde{g}_t$ is restricted to be linear in the observations within the window. (See the discussion in Doherty (1991).) This approach has also been adopted by Lane (1972), Laniel (1986) and will also be adopted here. Geweke (1978) and Pierce (1980) established the result (9) for the case where $\tilde{g}_t$ is linear in past values of the time series (not just those within the window) and where the time series follows an appropriate global model. However the argument leading to (9) shows that in general $\hat{y}_{t+s}$, and hence $\tilde{g}_t$, need not necessarily be linear in the observations.

In this paper we adopt the minimum revisions strategy based on the moving window paradigm and the local dynamic trend model of Section 2. At the ends of the series we choose to predict $\hat{g}_t = \sum_{s=-r}^{r} w_s y_{t+s}$ by a linear predictor of the form

$$\tilde{g}_t = \sum_{s=-r}^{q} u_s y_{t+s} \qquad (10)$$

where $q = T - t$ with $0 \leq q < r$, T denotes the time point of the last observation and the u_s are dependent on q. We shall consider two cases. The first imposes the condition that $\tilde{g}_t$ be an unbiased predictor of $\sum_{s=-r}^{r} w_s y_{t+s}$. The second weakens this requirement by considering biased predictors such as those developed by Musgrave (1964) for X-11 (see in particular Doherty (1991)).

3.1 Unbiased predictors

If $\tilde{g}_t$ is to be an unbiased predictor of $\sum_{s=-r}^{r} w_s y_{t+s}$ given by (5) and (7), then the u_s must satisfy

$$\sum_{s=-r}^{q} u_s = 1 \qquad \sum_{s=-r}^{q} s^j u_s = 0 \qquad (11)$$

where $0 < j \leq p$. Thus the asymmetric moving average filter implied by (10) passes polynomials of degree p. Moreover R_q is now given by

$$R_q = \mathbf{v}^T \mathbf{E}_1 \mathbf{v} \qquad (12)$$

where $\mathbf{v}$ has typical element

$$v_s = \begin{cases} w_s - u_s & (-r \leq s \leq q) \\ w_s & (q < s \leq r) \end{cases} \qquad (13)$$

and

$$\mathbf{E}_1 = \sigma^2 \mathbf{I} + \Omega. \qquad (14)$$

Here $\mathbf{I}$ denotes the identity matrix and the covariance matrix Ω has typical element

$$\Omega_{jk} = \text{cov}(\xi_{t+j} - \xi_t, \xi_{t+k} - \xi_t) \qquad (15)$$

for $-r \leq j, k \leq r$. Note, in particular, that Ω does not depend on t, the absolute value of time indexing the origin of the window. This natural and important invariance property is a consequence of (7), (11) and the assumption that the levels of integration of the random walk components that make up ξ_t do not exceed $p + 1$.

For each q appropriate values of the u_s can now be determined by minimising R_q subject to (11). As we show below, this results in an end filter that satisfies a particular form of (9) involving optimal prediction.

First consider predicting

$$Y = \sum_{s=-r}^{r} \delta_s y_{t+s}$$

from $y_{t-r}, \ldots y_{t+q}$ by a linear predictor of the form

$$\hat{Y} = \sum_{s=-r}^{q} u_s y_{t+s}$$

where $q < r$ and the δ_s are arbitrary known values. Then, in terms of the local dynamic model that applies in the window, $\hat{Y}$ is the best linear unbiased predictor (BLUP) of Y if the u_s are chosen so that $E(Y - \hat{Y}) = 0$ (unbiased prediction error) and the mean squared error criterion $E\{(Y - \hat{Y})^2\}$ is a minimum. Note that the condition $E(Y - \hat{Y}) = 0$ is equivalent to the requirement that

$$\sum_{-r}^{r} s^j v_s = 0 \qquad (0 \leq j \leq p) \qquad (16)$$

where

$$v_s = \begin{cases} \delta_s - u_s & (-r \leq s \leq q) \\ \delta_s & (q < s \leq r). \end{cases} \qquad (17)$$

Additional notation is needed to present the results. For the model specified by (1) and (2) define the $n \times (p+1)$ dimensional matrix $\mathbf{C}$ and the $p+1$ dimensional vector $\mathbf{c}$ by

$$\mathbf{C} = \begin{pmatrix} 1 & -r & \ldots & (-r)^p \\ 1 & -r+1 & \ldots & (-r+1)^p \\ \cdot & \cdot & \ldots & \cdot \\ \cdot & \cdot & \ldots & \cdot \\ 1 & r-1 & \ldots & (r-1)^p \\ 1 & r & \ldots & r^p \end{pmatrix} \qquad (18)$$

and $\mathbf{c} = (1, 0, \ldots, 0)^T$. Furthermore, define the $n \times (q + r + 1)$ matrix $\mathbf{L_1}$ and the $n \times (r - q)$ matrix $\mathbf{L_2}$ by the relation

$$\mathbf{I} = [\mathbf{L_1}, \mathbf{L_2}]. \qquad (19)$$

We can now state the following result.

Theorem 1 *Let y_t follow the local dynamic model specified by (1) and (2). Given $\delta = (\delta_{-r}, \ldots, \delta_r)^T$ and observations $y_{t-r}, \ldots, y_{t+q}$ for $0 \leq q < r$, the BLUP of $\sum_{s=-r}^{r} \delta_s y_{t+s}$ is $\sum_{s=-r}^{q} u_s y_{t+s}$ where $\mathbf{u} = (u_{-r}, \ldots, u_q)^T$ is given by*

$$\mathbf{u} = \mathbf{L}_1^T(\mathbf{I} - \mathbf{G}\mathbf{L}_2(\mathbf{L}_2^T\mathbf{G}\mathbf{L}_2)^{-1}\mathbf{L}_2^T)\delta.$$

Here

$$\mathbf{G} = \mathbf{E}_1^{-1} - \mathbf{E}_1^{-1}\mathbf{C}(\mathbf{C}^T\mathbf{E}_1^{-1}\mathbf{C})^{-1}\mathbf{C}^T\mathbf{E}_1^{-1},$$

$\mathbf{E}_1$ *is given by (14) and* $\mathbf{C}$ *is given by (18).*

In particular the BLUP of $\sum_{s=-r}^{r}\delta_s y_{t+s}$ *is given by* $\sum_{s=-r}^{r}\delta_s\hat{y}_{t+s}$ *where* $\hat{y}_{t+s}$ *is the BLUP of* y_{t+s} *for* $q < s \leq r$ *and* y_{t+s} *otherwise.*

Replacing the arbitrary δ_s by the weights w_s of the central moving-average trend filter yields the following result.

Corollary 2 *Let* y_t *follow the local dynamic model specified by (1), (2) and let* $\mathbf{w}$ *denote the vector of weights* w_s *for the central filter used in the body of the series with the* w_s *satisfying (7). Furthermore let* $\tilde{g}_t = \sum_{s=-r}^{q} u_s y_{t+s}$ *be an unbiased predictor of* $\sum_{s=-r}^{r} w_s y_{t+s}$ *with the* u_s *satisfying (11). Then, for* $0 \leq q < r$*, the values of* u_s *that minimise* R_q *subject to (11) are given by Theorem 1 with* $\delta = \mathbf{w}$ *and*

$$\sum_{s=-r}^{q} u_s y_{t+s} = \sum_{s=-r}^{r} w_s \hat{y}_{t+s}.$$

Here $\hat{y}_{t+s}$ *is the BLUP of* y_{t+s} *for* $q < s \leq r$ *and* y_{t+s} *otherwise.*

The properties of these end filters are investigated in Section 4.

3.2 Biased predictors

Again consider the situation where $\tilde{g}_t$ is a linear predictor of the form (10), but now no longer required to be unbiased. Instead we require the bias to be time invariant in the sense that it does not depend on the absolute time t indexing the origin of the window, whatever the parameters of the local dynamic model adopted.

Following the development in Section 3.1 we first consider predicting $Y = \sum_{s=-r}^{r}\delta_s y_{t+s}$ using a linear predictor of the form $\hat{Y} = \sum_{s=-r}^{q} u_s y_{t+s}$ where $q < r$ is given and the δ_s are arbitrary known values. In general, the bias term $E(Y - \hat{Y})$ will be invariant to the location of the window's time origin t if $p = 0$ or, if $p > 0$, when

$$\sum_{s=-r}^{r} s^j v_s = 0 \qquad (0 \leq j < p) \tag{20}$$

where the v_s are given by (17). In the latter case

$$E\{(Y - \hat{Y})^2\} = \beta_p^2(\sum_{s=-r}^{r} s^p v_s)^2 + \mathbf{v}^T\mathbf{E}_1\mathbf{v} \tag{21}$$

with $\mathbf{E}_1$ given by (14). It is desirable for the mean squared error $E\{(Y - \hat{Y})^2\}$ as well as the bias to be time invariant. However, because the v_s now satisfy (20) rather than (16), we need to impose stronger conditions on ξ_t.

These observations lead us to consider a *restricted local dynamic model* for the window centred at t where the levels of integration of the random walk components that make up ξ_t do not exceed p. In the case where $p = 0$ this necessarily leads to the requirement that $\xi_t = 0$ and $\mathbf{E}_1 = \sigma^2\mathbf{I}$. Given a restricted local dynamic model we define $\hat{Y}$ to be the best linear time invariant predictor (BLIP) of Y if the u_s are chosen to satisfy (20) and the mean squared error criterion $E\{(Y - \hat{Y})^2\}$ is a minimum. Here the expected prediction error $E(Y - \hat{Y})$ and the mean squared error $E\{(Y - \hat{Y})^2\}$ do not depend on t whatever the parameters of the local dynamic model concerned.

Theorem 3 *Let* y_t *follow the restricted local dynamic model specified above and let* $Y = \sum_{s=-r}^{r}\delta_s y_{t+s}$ *where* $\delta = (\delta_{-r}, \ldots, \delta_r)^T$ *is known. Given observations* $y_{t-r}, \ldots, y_{t+q}$ *and* q *satisfying* $0 \leq q < r$*, the BLIP of* Y *is* $\hat{Y} = \sum_{s=-r}^{q} u_s y_{t+s}$ *where* $\mathbf{u} = (u_{-r}, \ldots, u_q)^T$ *is given by*

$$\mathbf{u} = \mathbf{L}_1^T(\mathbf{I} - \tilde{\mathbf{G}}\mathbf{L}_2(\mathbf{L}_2^T\tilde{\mathbf{G}}\mathbf{L}_2)^{-1}\mathbf{L}_2^T)\delta$$

with

$$\tilde{\mathbf{G}} = \tilde{\mathbf{E}}_1^{-1} - \tilde{\mathbf{E}}_1^{-1}\mathbf{C}_{p-1}(\mathbf{C}_{p-1}^T\tilde{\mathbf{E}}_1^{-1}\mathbf{C}_{p-1})^{-1}\mathbf{C}_{p-1}^T\tilde{\mathbf{E}}_1^{-1}.$$

Here

$$\tilde{\mathbf{E}}_1 = \begin{cases} \mathbf{E}_1 + \beta_p^2\mathbf{c}_p\mathbf{c}_p^T & (p > 0) \\ \sigma^2\mathbf{I} + \beta_0^2\mathbf{c}_0\mathbf{c}_0^T & (p = 0) \end{cases}.$$

and $\mathbf{E}_1$ *is as given by Theorem 1. The* $n \times p$ *matrix* $\mathbf{C}_{p-1}$ *and the* p *dimensional vector* $\mathbf{c}_p$ *are defined implicitly by the partitioned matrix*

$$\mathbf{C} = [\mathbf{C}_{p-1}, \mathbf{c}_p]$$

where $\mathbf{C}$ *is given by (18) and* $\mathbf{C}_{p-1}$ *is null when* $p = 0$.

In particular the BLIP of $\sum_{s=-r}^{r}\delta_s y_{t+s}$ *is given by* $\sum_{s=-r}^{r}\delta_s\hat{y}_{t+s}$ *where* $\hat{y}_{t+s}$ *is the BLIP of* y_{t+s} *for* $q < s \leq r$ *and* y_{t+s} *otherwise.*

Note that the BLIP predictor given by Theorem 3 has the form of a shrinkage estimator since it is exactly the same as the BLUP predictor given by Theorem 1, but with $\mathbf{E}_1$ replaced by $\mathbf{E}_1 + \beta_p^2\mathbf{c}_p\mathbf{c}_p^T$ and $\mathbf{C}$ replaced by $\mathbf{C}_{p-1}$. Indeed, when $\beta_p = 0$, the BLIP predictor becomes the BLUP predictor for the reduced model where g_t is replaced by

$$g_t = \begin{cases} \sum_{j=0}^{p-1}\beta_j t^j + \xi_t & (p > 0) \\ 0 & (p = 0) \end{cases},$$

but ξ_t and ϵ_t remain the same.

We now return to the minimisation of the revisions criterion R_q given by (8) where the w_s are the known central filter weights that apply in the body of the series and the w_s satisfy (7). Given a restricted local dynamic model, we choose to predict $\sum_{s=-r}^{r} w_s y_{t+s}$ by a linear predictor $\tilde{y}_t$ of the form (10) subject to the requirement that the bias of $\tilde{y}_t$ is time invariant. The latter is true if $p = 0$ or, when $p > 0$, if

$$\sum_{s=-r}^{q} u_s = 1 \qquad \sum_{s=-r}^{q} s^j u_s = 0 \qquad (22)$$

for $0 < j < p$. Replacing the arbitrary δ_s in Theorem 3 by the weights w_s of the central moving-average trend filter yields the following result.

Corollary 4 *Let y_t follow the restricted local dynamic model specified in Theorem 3 and let $\mathbf{w}$ denote the vector of weights w_s for the central filter used in the body of the series with the w_s satisfying (7). Furthermore, let $\tilde{y}_t = \sum_{s=-r}^{q} u_s y_{t+s}$ be a linear predictor of $\sum_{s=-r}^{r} w_s y_{t+s}$ with time invariant bias so that the u_s satisfy (22). Then, for $0 \le q < r$, the values of u_s that minimise R_q subject to (22) are given by Theorem 3 with $\delta = \mathbf{w}$ and*

$$\sum_{s=-r}^{q} u_s y_{t+s} = \sum_{s=-r}^{r} w_s \hat{y}_{t+s}.$$

Here $\hat{y}_{t+s}$ is the BLIP of y_{t+s} for $q < s \le r$ and y_{t+s} otherwise.

Now $\mathbf{E}_1/\sigma^2$ does not depend on σ^2 and $\tilde{\mathbf{E}}_1$ need only be known up to a constant of proportionality. Thus, unlike the BLUP end filters specified by Corollary 2, the BLIP end filters specified by Corollary 4 can only be made operational when β_p^2/σ^2 is known. Since this will rarely, if ever, be the case, estimates of β_p^2/σ^2 of one form or another need to be determined from the data. Such estimates will, necessarily, differ from their true values and it is therefore important to determine the effects of mis-specification of β_p^2/σ^2. This issue is addressed below in Theorem 5 and also Section 4 where the properties of these end filters and the estimation of β_p^2/σ^2 are discussed.

Note that Corollary 4 yields the X-11 end filters derived by Musgrave (1964) and Doherty (1991) when $\xi_t = 0$, $p = 1$, $\beta_p^2/\sigma^2 = 4/(\pi(3.5)^2)$ and $\mathbf{w}$ contains the X-11 Henderson central filter weights. Thus Corollary 4 provides a generalisation and extension of the current X-11 end filters.

The following result considers an alternative form of the BLIP end filters given by Corollary 4 which explicitly builds on the corresponding BLUP end filters given by Corollary 2. In addition the result provides a means of exploring the effects of mis-specification of β_p^2/σ^2.

Theorem 5 *Let y_t follow the restricted local dynamic model specified in Theorem 3 and let $\mathbf{w}$ denote the vector of weights w_s for the central filter used in the body of the series with the w_s satisfying (7). Given observations $y_{t-r}, \ldots, y_{t+q}$ and q satisfying $0 \le q < r$, let $\tilde{y}_t = \sum_{s=-r}^{q} u_s(\phi) y_{t+s}$ be a linear predictor of $\sum_{s=-r}^{r} w_s y_{t+s}$ where ϕ is an arbitrary scalar, $\mathbf{u}(\phi) = (u_{-r}(\phi), \ldots, u_q(\phi))^T$ is given by*

$$\mathbf{u}(\phi) = \mathbf{L}_1^T(\mathbf{I} - \mathbf{G}\mathbf{L}_2(\mathbf{L}_2^T\mathbf{G}\mathbf{L}_2)^{-1}\mathbf{L}_2^T)(\mathbf{w} + \phi\gamma)$$

and

$$\gamma = \mathbf{E}_1^{-1}\mathbf{C}(\mathbf{C}^T\mathbf{E}_1^{-1}\mathbf{C})^{-1}\mathbf{d}.$$

Here $\mathbf{L}_1$, $\mathbf{L}_2$, $\mathbf{G}$, $\mathbf{E}_1$ and $\mathbf{C}$ are as given in Theorem 1, and the $p+1$ dimensional vector $\mathbf{d}$ is zero save for the last element which is unity. Then

(a) $\tilde{y}_t(\phi)$ is a linear time invariant predictor of $\sum_{s=-r}^{r} w_s y_{t+s}$ with squared bias $\beta_p^2\phi^2$ and the $u_s(\phi)$ satisfy (22) when $p > 0$;

(b) the revisions criterion R_q is given in this case by

$$R_q(\phi) = \mathbf{w}^T\mathbf{H}\mathbf{w} + 2\phi\gamma^T\mathbf{H}\mathbf{w} + \phi^2(\beta_p^2 + \mathbf{d}^T(\mathbf{C}^T\mathbf{E}_1^{-1}\mathbf{C})^{-1}\mathbf{d} + \gamma^T\mathbf{H}\gamma)$$

where

$$\mathbf{H} = \mathbf{L}_2(\mathbf{L}_2^T\mathbf{G}\mathbf{L}_2)^{-1}\mathbf{L}_2^T;$$

(c) the optimal end filters of Corollary 2 and Corollary 4 are given by $\mathbf{u}(0)$ and $\mathbf{u}(\phi_0)$ respectively where

$$\phi_0 = -\frac{\gamma^T\mathbf{H}\mathbf{w}}{\beta_p^2 + \mathbf{d}^T(\mathbf{C}^T\mathbf{E}_1^{-1}\mathbf{C})^{-1}\mathbf{d} + \gamma^T\mathbf{H}\gamma}$$

minimises $R_q(\phi)$.

An immediate consequence of Theorem 5 is that end filters based on BLIP predictors will generally have smaller mean squared revisions than those based on the corresponding BLUP predictors since $R_q(\phi_0) \le R_q(0)$. Moreover

$$\lim_{\beta_p^2/\sigma^2 \to \infty} R_q(\phi_0) = R_q(0), \qquad \lim_{\beta_p^2/\sigma^2 \to \infty} \phi_0 = 0$$

so that BLIP end filters converge to their corresponding BLUP end filters as β_p^2/σ^2 increases. The best possible mean squared revisions are achieved when $\beta_p^2/\sigma^2 = 0$.

Then $R_q(\phi_0)$ is least and, as noted following Theorem 3, the BLIP end filters become BLUP end filters for the local dynamic model with order $p-1$, but the same stochastic structure.

Both the BLIP and BLUP end filters are dependent on the global parameters specified by p, n and the model for ξ_t/σ_t. Although these global parameters are sufficient to determine the BLUP end filters, the BLIP end filters further require knowledge of β_p^2/σ^2. The latter is a function of local parameters whose values will not normally be known in practice and which will need to be estimated from the data. In this case it is possible that a mis-specified value of β_p^2/σ^2 could result in a BLIP end filter which has greater mean squared revisions than its corresponding BLUP end filter.

Now $\tilde{g}_t(\phi_0)$ depends only on β_p^2/σ^2 through ϕ_0 which is a one-to-one function of β_p^2/σ^2. Thus Theorem 5 enables us to consider the effects of mis-specification of β_p^2/σ^2. Let $\hat{\phi}_0$ denote ϕ_0 evaluated at $\hat{\beta}_p^2/\hat{\sigma}^2$, some estimated or target value of β_p^2/σ^2. Then, since $R_q(\phi)$ is a quadratic in ϕ, it is evident that $R_q(\hat{\phi}_0) \leq R_q(0)$ if and only if $\hat{\phi}_0$ lies between 0 and $2\phi_0$. The BLIP end filters will therefore have better mean squared revisions than their corresponding BLUP end filters when

$$\frac{\beta_p^2}{\sigma^2} < 2\frac{\hat{\beta}_p^2}{\hat{\sigma}^2} + (\mathbf{d}^T(\mathbf{C}^T\mathbf{E}_1^{-1}\mathbf{C})^{-1}\mathbf{d} + \gamma^T\mathbf{H}\gamma)/\sigma^2. \tag{23}$$

In particular, since $R_q(\phi_0+\phi) = R_q(\phi_0-\phi)$, it is sufficient to select $\hat{\beta}_p^2/\hat{\sigma}^2$ so that $\hat{\phi}_0$ is between 0 and ϕ_0 or, equivalently, $\hat{\beta}_p^2/\hat{\sigma}^2 \geq \beta_p^2/\sigma^2$. In practice this choice should lead to values for $\hat{\beta}_p^2/\hat{\sigma}^2$ that, if anything, overestimate β_p^2/σ^2 thus controlling the mis-specification error by shrinking the BLIP end filter towards its BLUP counterpart. Note that if $\hat{\beta}_p^2/\hat{\sigma}^2 \geq \beta_p^2/\sigma^2$ then

$$\tilde{g}_t(\hat{\phi}_0) = (1-\frac{\hat{\phi}_0}{\phi_0})\tilde{g}_t(0)+\frac{\hat{\phi}_0}{\phi_0}\tilde{g}_t(\phi_0) \qquad (0 \leq \frac{\hat{\phi}_0}{\phi_0} \leq 1)$$

so that $\tilde{g}_t(\hat{\phi}_0)$ is a convex combination of the two optimal end filters.

The properties of these end filters are investigated in Section 4.

4 Properties of the Filters

This section considers the properties of the end filters specified by Corollary 2 and Corollary 4 which are designed to minimise the expected mean squared revisions between the output of these filters and that of the central filters on which they are based. These end filters deal with a transition problem that ultimately goes away as the current time points are subsumed into the body of the series. The minimum revisions criterion therefore provides a measure of the total cost of this transition.

We restrict attention to the important case where the window is of length 13 with $r = 6$ and the central filter used in the body of the series is the central X-11 Henderson filter. Furthermore, we focus on two particular local dynamic models likely to be used in practice. These are the *local linear model* ($p = 1$) given by (3) and the *local quadratic model* ($p = 2$) given by (4).

Three end filters are considered: the BLUP end filter (or BLIP end filter with $\hat{\beta}_p^2/\hat{\sigma}^2 = \infty$), the BLIP end filter with $\hat{\beta}_p^2/\hat{\sigma}^2 = 4/(\pi(3.5)^2)$ (this gives the X-11 end filter in the case where $p = 1$ and $\lambda = 0$), and the BLIP end filter with $\hat{\beta}_p^2/\hat{\sigma}^2 = 0$. As noted in the discussion following Theorem 5, the smallest value of $R_q(\phi)$ is achieved when $\phi = \phi_0$ and $\beta_p^2/\sigma^2 = \hat{\beta}_p^2/\hat{\sigma}^2 = 0$. Also, if $\hat{\beta}_p^2/\hat{\sigma}^2$ is chosen so that (23) is satisfied or $\hat{\beta}_p^2/\hat{\sigma}^2 \geq \beta_p^2/\sigma^2$, then the largest value of $R_q(\hat{\phi}_0)$ is $R_q(0)$ which is the minimum revisions for the BLUP end filter. These values serve as useful bounds for $R_q(\hat{\phi}_0)$.

To provide a range of possible values of β_p^2/σ^2 we briefly consider the so-called $\bar{I}/\bar{C}$ ratio used by X-11 to specify the length of the trend moving average adopted. Here $\bar{I}$ and $\bar{C}$ are the respective averages of the absolute values of the month to month changes in the (estimated) irregular ϵ_t and trend g_t. Thus the $\bar{I}/\bar{C}$ ratio measures the importance of month to month changes in ϵ_t relative to those in the trend g_t. X-11 recommends that the central 9 point Henderson filter be used when $\bar{I}/\bar{C} < 1$, the central 13 point Henderson trend filter when $1 \leq \bar{I}/\bar{C} < 3.5$, and the central 23 point Henderson trend filter when $\bar{I}/\bar{C} \geq 3.5$. Following Musgrave (1964) and Doherty (1991) we consider the local linear model and

$$\frac{I}{C} = \frac{E|\Delta\epsilon_t|}{E|\Delta g_t|} \tag{24}$$

which can be thought of as the population parameter that $\bar{I}/\bar{C}$ is estimating. This is readily determined under Gaussian assumptions as a function of β_p^2/σ^2 and λ. Simple evaluations show that the values of β_p^2/σ^2 for which $1 \leq I/C < 3.5$ lie in $[0, 4/\pi]$. This is the interval of values for β_p^2/σ^2 that we choose to adopt.

We now examine the mean squared revisions of the given end filters for β_p^2/σ^2 in $[0, 4/\pi]$. Note that it is sufficient to consider the extremes of this interval since, from Theorem 5, $R_q(\phi)$ is linearly increasing in β_p^2 over the interval for given ϕ. The mean squared revisions criterion $R_q(\hat{\phi}_0)/\sigma^2$ is plotted in Figure 1 as a function of q for a selection of models and for the particular end filters considered.

As expected, the mean squared revisions are greatest when q is least with $q = 0$ yielding the greatest revisions followed by $q = 1$. The mean squared revisions for the other values of q are negligible by comparison. Although not apparent on this scale, $R_q(\hat{\phi}_0)/\sigma^2$ is not necessarily a monotonic function of q. For example the local quadratic model yields mean squared revisions when $q = 2$ that are typically smaller than those for $q = 3$. This is a consequence of the shape of the central X-11 Henderson filter adopted.

If $\hat{\beta}_p^2/\hat{\sigma}^2$ has been selected appropriately, $R_q(\hat{\phi}_0)/\sigma^2$ will be bounded above by the mean squared revisions for the BLUP end filter and below by the mean squared revisions for the BLIP end filter with $\beta_p^2/\sigma^2 = \hat{\beta}_p^2/\hat{\sigma}^2 = 0$. These bounds are plotted in Figure 1 and show that there are gains to be had using BLIP end filters, although these are likely to be modest in the case of the local linear model. As evident from the form of $R_q(\hat{\phi}_0)/\sigma^2$ given by Theorem 5, the gains are greatest when β_p^2/σ^2 is least.

Note that the mean squared revisions generally increase as λ and p increase. This is of marginal utility in practice, since the local model chosen is determined from the particular time series concerned. However it is possible that a local linear model with large λ might describe a time series as well as a local quadratic model with small λ. From Figure 1 it would appear that, in such cases, the BLIP end filter for the quadratic model may have greater capacity to achieve lower revisions.

For the local linear model, the X-11 BLIP end filter appears to offer only marginal gains over the BLUP end filters when β_p^2/σ^2 satisfies (23). In the local quadratic case, the X-11 BLIP end filter is clearly too conservative and a lower value of $\hat{\beta}_p^2/\hat{\sigma}^2$ might more profitably be considered. For both models, the upper limit of $\beta_p^2/\sigma^2 = 4/\pi$ leads to X-11 end filters with unacceptably high revisions. This may explain, in part, the reason why the use of end filters obtained when the central filter is evaluated with unknown observations predicted by global ARIMA models (ARIMA forecast extension) has largely superseded the use of the X-11 end filters in practice. The former yield (global) BLUP end filters with properties that one might expect are close to the (local) BLUP end filters considered here. On the other hand, for the local linear model, the $\bar{I}/\bar{C}$ guidelines imply that the X-11 end filters are inflexibly applied whenever β_p^2/σ^2 satisfies $\beta_p^2/\sigma^2 \leq 4/\pi$ rather than (23).

If the BLUP end filters yield results that are comparable to ARIMA forecast extension end filters, then it would appear that judiciously selected BLIP end filters may offer modest performance gains in terms of improved revisions. The price of this improvement is a better understanding of the time varying values of β_p^2/σ^2.

We have carried out some analysis of how the BLIP end filters perform on actual data. Some results are given in Figure 2 for three New Zealand series which have been seasonally adjusted by X-11 and have been modified to remove any large outliers. Only results for the local linear model case are presented.

To determine λ in the local linear model we fitted global ARIMA models to the the modified seasonally adjusted series. Using standard diagnostics we selected ARIMA (0,1,1) models for building permits and permanent migration, and an ARIMA (2,1,0) model for exports. Means were fitted in each case. The local linear model is an ARIMA (0,1,1) model with a mean and λ is a function of the moving-average parameter θ and hence, implicitly, a function of the first order autocorrelation. If the global model is ARIMA (0,1,1) then λ can be estimated from the estimate of θ. If the global model is not ARIMA (0,1,1) then one approach is to match the first order autocorrelation and estimate λ via the implicit functional relationship holding in the ARIMA (0,1,1) case.

To determine $\hat{\beta}_p^2/\hat{\sigma}^2$ for the BLIP end filters we considered two methods. In the first method we estimated I/C, as given in (24), over the window of the central filter and then took the average of these estimates and determined $\hat{\beta}_p^2/\hat{\sigma}^2$ from these. This gave values for building permits, exports, and permanent migration of .04, .01 and .28 respectively. In the second method we searched over a range of values of $\hat{\beta}_p^2/\hat{\sigma}^2$ while keeping λ fixed at the value determined by the global ARIMA model to find the one which gave the minimum revisions. This gave values for building permits, exports, and permanent migration of .01, .01 and .16 respectively. So both methods gave similar values. The BLIP end filter in Figure 2 is determined using the second method.

We did the following empirical comparison of the end filters. First, for each local window, we calculated the trend for the central point of that window using the central filter as in (5). Next for each q, for each local window, we calculated the trend using the appropriate end filter. For the BLUP and BLIP cases, this meant predicting the trend as in (10), where the u_s are the weights of the end filter. For the ARIMA forecast extension case this meant predicting the trend as in (9). For now, in (9), $\hat{y}_{t+s}$ is the prediction using the forecast function of the global ARIMA model previously used in determining λ restricted to data up to and including the local window. We did not refit the ARIMA model for each local window as generally both the ARIMA models and their coefficients were fairly stable over time. The difference between these two trend estimates is the revision which occurs as the predicted missing data is replaced with actual data. Finally we made boxplots of the absolute value of the revisions since the median of the absolute value of the revisions provides a robust estimate of the square root of the mean squared revisions, which was the criterion used to evaluate the theoretical performance of the filters.

As expected from Theorem 5, the BLIP end filters where $\hat{\beta}_p^2/\hat{\sigma}^2$ is determined from the data, have smaller revisions than BLUP filters and smaller revisions than the Musgrave end filters used in X-11 (which are BLIP end filters for a local linear model with $\hat{\beta}_p^2/\hat{\sigma}^2$ set to $4/(\pi(3.5)^2)$. For the exports and migration series, the choice of $\hat{\beta}_p^2/\hat{\sigma}^2$ used in the X-11 end filters is inappropriate, and so here the X-11 end filters have larger revisions than the BLUP end filters.

We have also compared the revisions of BLIP end filters and the revisions of the ARIMA forecast extension end filters. In the case of exports and permanent migration our choice of BLIP end filter would seem to perform as well or better than the ARIMA forecast extension end filter. For building permits our choice of BLIP end filter has about 1% higher revisions than the ARIMA forecast extension end filter. Perhaps in practice this can be regarded as the same. In any case, since in this series there are many turning points, this may be a case, as mentioned above in the theoretical discussion, where the BLIP end filter for the quadratic model may have greater capacity to achieve lower revisions, and these revisions may be smaller than the revisions of ARIMA forecast extension end filter.

Acknowledgements

The second author gratefully acknowledges support provided by an ASA/NSF/Census Research Fellowship which he held at the US Bureau of the Census. In particular, both authors wish to record their gratitude to Dr David Findley of that organisation for the helpful advice and encouragement he has provided throughout the project. The data was kindly provided by Statistics New Zealand.

References

Bell, W. (1993). Empirical comparisons of seasonal ARIMA and ARIMA component (structural) time series models. Research Report CENSUS/SRD/RR-93/10, Statistical Research Division, Bureau of the Census, Washington, D.C. 20233-4200.

Cleveland, W. S. (1983). Seasonal and calendar adjustment. In Brillinger, D. and Krishnaiah, P., editors, *Handbook of Statistics* **3**, pages 39–72. North Holland, Amsterdam.

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

DeForest, E. L. (1877). On adjustment formulas. *The Analyst*, 4:79–86, 107–113.

Doherty, M. (1991). Surrogate Henderson filters in X-11. Technical report, NZ Department of Statistics, Wellington, New Zealand.

Geweke, J. (1978). The revision of seasonally adjusted time series. In *Proceedings of the Business and Economic Statistics Section, American Statistical Association*, pages 320–325.

Greville, T. N. E. (1979). Moving-weighted-average smoothing extended to the extremities of the data. Technical Summary Report #2025, Mathematics Research Center, University of Wisconsin, Madison, Wisconsin.

Kenny, P. B. and Durbin, J. (1982). Local trend estimation and seasonal adjustment of economic and social time series. *Journal of the Royal Statistical Society, Series A*, 145:1–41.

Lane, R. O. D. (1972). Minimal revision trend estimates. Technical Report Research Exercise Note 8/72, Central Statistical Office, London.

Laniel, N. (1986). Design criteria for 13 term Henderson end-weights. Technical Report Working Paper TSRA-86-011, Statistics Canada, Ottawa K1A 0T6.

Musgrave, J. C. (1964). A set of end weights to end all end weights. Working paper, Bureau of the Census, US Department of Commerce, Washington, D.C.

Pierce, D. A. (1980). Data revisions with moving average seasonal adjustment procedures. *Journal of Econometrics*, 14:95–114.

Wallis, K. F. (1983). Models for X-11 and X-11-FORECAST procedures for preliminary and revised seasonal adjustments. In Zellner, A., editor, *Applied Time Series Analysis of Economic Data*, pages 3–11, Washington, D. C. U. S. Bureau of the Census.

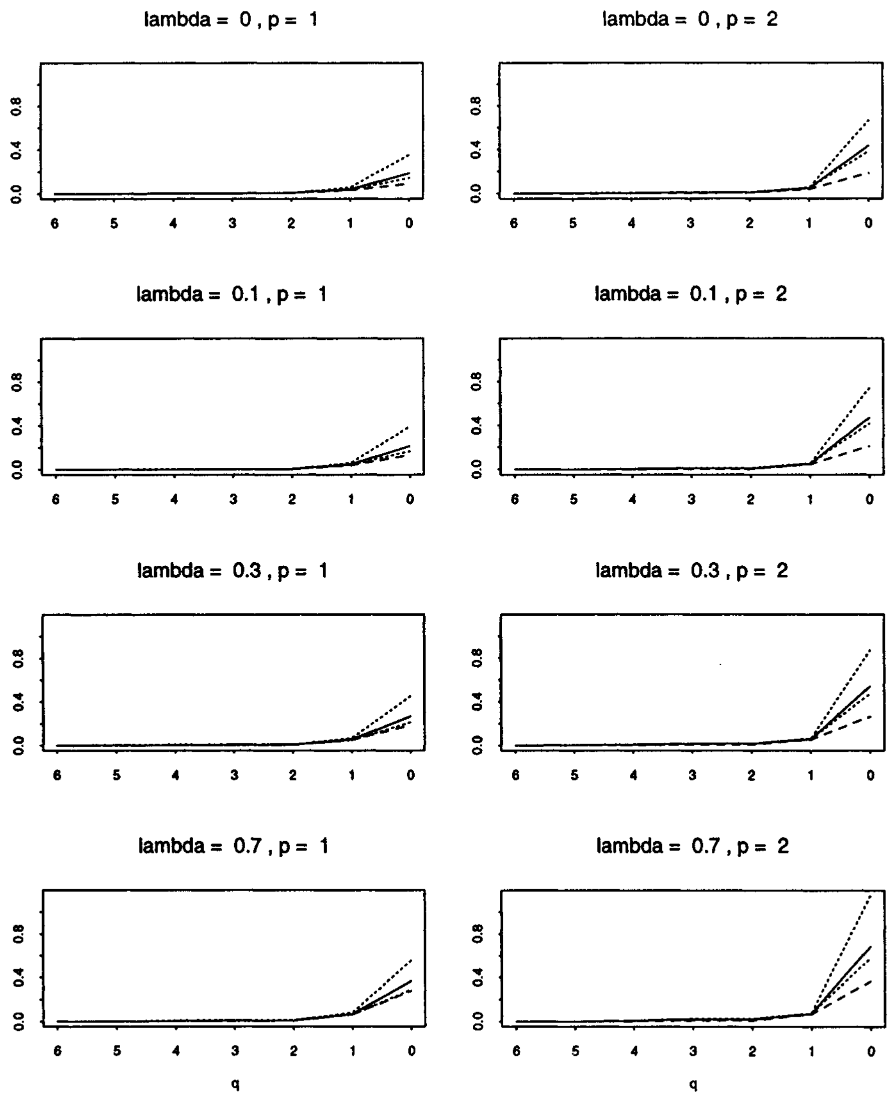

Figure 1: Plots of the mean squared revisions criterion R_q/σ^2 for the BLUP and BLIP end filters based on the central X-11 Henderson filter. Here the local linear and quadratic models are specified by λ and p, and the solid lines correspond to the BLUP end filter. The dashed line corresponds to the BLIP end filter with $\hat{\beta}_p^2/\hat{\sigma}^2 = 0$ in the case where $\beta_p^2/\sigma^2 = 0$. This is the BLUP end filter for the reduced model of order $p - 1$ and represents the best mean squared revisions possible. The dotted lines correspond to the X-11 BLIP end filter ($\hat{\beta}_p^2/\hat{\sigma}^2 = 4/(\pi(3.5)^2)$) in the two cases where $\beta_p^2/\sigma^2 = 0$ and $\beta_p^2/\sigma^2 = 4/\pi$. These limits are derived from recommendations made by X-11 concerning I/C ratios.

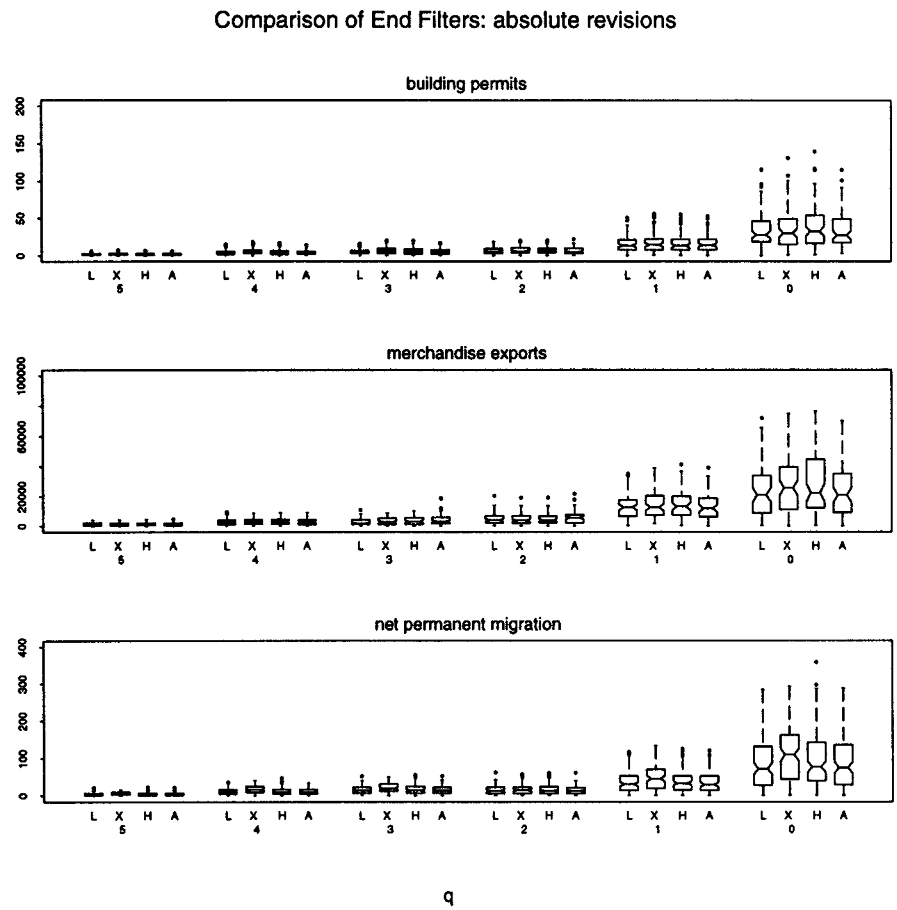

Figure 2: This analysis uses New Zealand data on the number of building permits issued by Local Authorities for the construction of private houses and flats, the value of merchandise exports, and the net number of permanent migrants. The series have been seasonally adjusted by X-11 and been modified to remove any large outliers. Using the central X-11 Henderson filter in the body, comparisons are made between the various end filters and for the various values of q, for the local linear model case. The choice of λ is determined from the global ARIMA model fitted to the modified seasonally adjusted series. The values of λ for building permits, exports, and permanent migration are .6, .22 and 7.5 respectively. The choice of $\hat{\beta}_p^2/\hat{\sigma}^2$ is determined by searching for values which improve the revisions. The values of $\hat{\beta}_p^2/\hat{\sigma}^2$ for building permits, exports, and permanent migration are .01, .01 and .16 respectively. Here L refers to the BLIP end filter based on the local linear model. Likewise, X refers to the standard X-11 end filter, which is a BLIP end filter with $\hat{\beta}_p^2/\hat{\sigma}^2 = 4/(\pi(3.5)^2)$ or .104, H refers to the BLUP end filter based on the local linear model. Finally A refers to the filter obtained by using the central filter with unknown observations predicted by a global ARIMA model.