BUREAU OF THE CENSUS
STATISTICAL RESEARCH DIVISION
Statistical Research Report Series
No. RR96/01

Design of Moving-Average Trend Filters using Fidelity, Smoothness and Minimum Revisions Criteria

Alistair Gray and Peter Thomson

Institute of Statistics and Operations Research
Victoria University of Wellington
PO Box 600, Wellington, New Zealand

Report Issued: 2/7/97

Abstract

Many seasonal adjustment procedures decompose time series into trend, seasonal, irregular and other components using simple non-seasonal finite moving-average trend filters. This report considers the design of such filters, both in the body and at the ends of series, based on specified criteria and simple dynamic models operating locally within the span of the filter.

In the body of the series a flexible family of finite moving-average trend filters is developed from specified smoothness and fidelity criteria. These filters are based on local dynamic models and generalise the standard Macaulay and Henderson filters used in practice. The properties of these central filters are determined and evaluated both in theory and in practice.

At the ends of the series the central moving-average trend filter used in the body needs to be extended to handle missing observations. A family of end filters is constructed using a minimum revisions criterion and based on the 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 generalise those developed by Musgrave for the central X-11 Henderson filters and include the BLUP end filters as a special case.

The properties of these end filters are determined both in theory and practice. In particular, they are compared to the Musgrave end filters used by X-11 and to the case where the central filter is evaluated with unknown observations predicted by global ARIMA models. The latter parallels the forecast extension method used in X-11-ARIMA.

Keywords: Moving-average filters; local trend estimation; dynamic models; fidelity; smoothness; minimum revisions; best linear unbiased prediction; best linear biased prediction; X-11; seasonal time series; seasonal adjustment.

Preface

The origins of this research project began in the early 1990's having been stimulated to a large extent by our contribution to the discussion of Cleveland et al. (1990) (see Gray and Thomson (1990)). Following some prior development work, the project benefited in 1993 from two research grants. An application to the New Zealand Foundation for Research Science and Technology made in 1992 was successful and the second author was awarded a Research Fellowship at the US Bureau of the Census which he held for six months during 1993 and 1994. Both forms of research support were important, but the period spent at the US Bureau of the Census with their Time Series Group proved particularly invaluable. Our ongoing interactions with that group and with staff within Statistics New Zealand continue to provide us with an important source of research support and guidance for which we are most grateful.

To a large extent the title reflects the emphasis of our initial ideas. We had thought that once we developed better central filters, end filter design would be straightforward using best linear unbiased prediction and the resulting BLUP end filters would be an improvement on X-11 end filters. To our surprise X-11 end filters often had smaller revisions than the BLUP end filters. This lead us to study Musgrave's work more closely (see Musgrave (1964) and Doherty (1991)) and develop BLIP end filters based on best linear biased prediction where the bias is time invariant. The investigation of BLIP end filters took more time than we had anticipated and was a contributing factor to the delay in the publication of this report.

The work reported here constitutes a beginning and much remains to be done including more comprehensive case studies and further analytical development. An ultimate objective is the construction of a robust semi-parametric seasonal decomposition procedure which might better handle the more volatile time series experienced by countries such as New Zealand. This is unlikely to be straightforward. Indeed, our preliminary experience suggests that embedding these filters within a seasonal adjustment framework such as X-11 will be a challenging exercise if significant gains are to be realised.

This research project was partially supported by the New Zealand Foundation for Research Science and Technology (Project Number VIC-93-36-039). The second author also 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 very helpful advice and encouragement he has provided throughout the project. The data was kindly provided by Statistics New Zealand and the US Bureau of the Census.

Contents

List of Figures

1 Introduction

This report is concerned with the design of local trend estimation filters for non-seasonal time series using finite moving-averages. A particular objective is to try and improve current methods of trend estimation within seasonal adjustment procedures where non-seasonal trend filters play a key role, especially at the ends of series. For example, the seasonal adjustment procedure X-11 (see Shiskin et al. (1967)) uses trend filters originally developed by Henderson (1916) whereas the seasonal-trend decomposition procedure STL uses loess filters (see Cleveland et al. (1990)). However local trend estimation is also important in its own right and is routinely used, for example, in the analysis of financial and other time series.

The design and use of finite moving-averages for local trend estimation has a long history, particularly within the actuarial literature where it is referred to as graduation. The seminal work of Henderson (1916) and Macaulay (1931) is an example of the latter. Useful background material on this topic from a time series perspective is also given in Kendall (1973) and the references contained therein. Much of this development has been posited on simple structural models of the form

$$\phi(y_t) = T_t + S_t + \epsilon_t \tag{1}$$

where y_t denotes the observed time series, $\phi(x)$ denotes some appropriate transformation, T_t denotes the trend, S_t the seasonal and ϵ_t the irregular or noise. Typically the unobserved trend and seasonal components are assumed to evolve slowly with more abrupt changes, including calendar effects, handled via simple adjustments. Recently, after somewhat of a lull, more attention has been focused on both structural models of the form (1) and the need for improved finite moving-average trend filters.

Structural models have been exploited in a number of global parametric seasonal adjustment procedures. These include BAYSEA (Akaike (1980)) and other related model-based methods (see Gersch and Kitagawa (1983), Harvey (1989) and Schlicht (1981) for example). Bell and Hillmer (1984) also provide a useful review and discussion of unobserved component seasonal ARIMA models. An explicit and important feature of much of this work is the use of criteria that optimally weight specific measures of fidelity and trend smoothness. The latter is in the spirit of Henderson (1924) and Whittaker (1923).

Impetus to the design of finite moving-average filters has recently come from work in scatterplot smoothing and local regression models. In particular the work of Cleveland (1979) has led to the development of lowess and loess which, in the time series context, constitute finite moving-average trend filters. The seasonal decomposition procedures SABL (Cleveland et al. (1978)) and more recently STL, are built around the non-seasonal lowess and loess filters respectively. See also the discussion following Cleveland et al. (1990) and, in particular, Gray and Thomson (1990). Other work on finite moving-average filter design has been motivated by the need to control the revision of trend estimates at the ends of series as new data comes to hand (see Dagum and Laniel (1987), Dagum (1996), Doherty

(1991), Kenny and Durbin (1982), Laniel (1986) and Wallis (1983) for example). The work of Cholette (1981) and Leser (1963) is also of interest. They build finite moving-average trend filters for seasonal time series using a criterion based approach that involves explicit measures of both fidelity and smoothness.

In practice the use of finite moving-average trend filters in seasonal adjustment and trend estimation remains the dominant technology. This emphasis needs some justification and explanation. Firstly, the use of moving data windows forms the basis for simple non-parametric models that apply locally and whose assumptions can, in many cases, be justified by simple graphical analysis. As a consequence, they have served official statisticians and many others well over a long period of time. Secondly, they directly control and limit the revisions to historical trend values as additional data is obtained. This is an important and over-riding requirement for most official statisticians. Finally, moving local non-parametric models have the potential to capture both evolutionary change and non-evolutionary structural change in time series in a more direct and transparent way than global parametric models. However the two technologies should not be seen as competing. The non-parametric local procedures are frequently used as exploratory tools prior to the fitting of a more sophisticated parametric global model.

In the case of economic and official time series it is important to identify trends and their turning points accurately and so filter design at the ends of the series is particularly important. The construction of end filters from the ones that apply in the body and the principles by which they should be designed have been discussed by a number of authors including Cleveland (1979), Doherty (1991), Kenny and Durbin (1982), Geweke (1978), Greville (1979), Kendall (1973), Lane (1972), Laniel (1986), Musgrave (1964), Pierce (1980) and Wallis (1983). In the case of X-11 the basis of the Henderson filters that apply in the body of the series is relatively well understood, but the derivation of the filters that apply at the ends of the series is less clear and has been the subject of some debate in the literature (see Kenny and Durbin (1982) and Doherty (1991) for example). Because of known defects such as failure to pass linear trends at the ends of series, the X-11 end filters are often avoided in practice by using the original series augmented by predictions, and then filtering using the central filters. In the absence of this device, the X-11 end filters continue to be used in X-11 despite their apparent shortcomings.

The principle of using prediction at the ends of series seems a key one which goes back to DeForest (1877). The review article by Cleveland (1983) provides a useful discussion on this and other issues in seasonal adjustment. See also Geweke (1978), Greville (1979) and Pierce (1980). The X-11-ARIMA method, for example, uses relatively sophisticated ARIMA models to provide predictions (see Dagum (1980)). This sophistication is in contrast to the relatively unsophisticated local trend smoothing that goes on in the body of the series. Moreover, by virtue of the order of differencing used, many of the ARIMA based models effectively assume only evolving local linearity for the trend whereas the Henderson filters, for example, are based on local quadratic or cubic models. In the body of the series this mismatch does not seem to have been a major cause for concern. However, at the ends, the use of higher order local polynomial trends appears to be one of the primary sources of difficulty in terms of controlling revisions. These difficulties are typically alleviated in practice by the use of prediction augmented methods such as X-11-ARIMA.

This report attempts to combine some of the virtues of the global parametric models within the framework of a finite moving-average trend estimation procedure. In particular a consistent and flexible family of filters is developed which are derived from local dynamic models that employ lower order polynomial trends and have stochastic error structures that enhance short-term prediction at the ends of series. A major thrust of the paper is the design of variants of standard moving-average trend filters that are derived from specified fidelity and smoothness criteria.

2 Trend filter design

We follow the conventional paradigm and consider a moving window of $n = 2r + 1$ consecutive observations within which an estimate of the trend is to be calculated for the central time point. This central reference point is chosen presumably because of symmetry and because that is typically where the trend can be determined with greatest precision. At the ends of the series the window will, by necessity, be truncated due to missing observations. We first describe the form of the local dynamic model adopted and then consider filter design both in the body and at the ends.

2.1 Local dynamic model

Within the finite window, model the observations as

$$y_t = g_t + \epsilon_t \tag{2}$$

where ϵ_t is white noise with variance σ^2 and the trend g_t is given by

$$g_t = \sum_{j=0}^{p} \beta_j t^j + \xi_t. \tag{3}$$

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.) Note that the conventional model is recovered by eliminating the random walk component ξ_t which is achieved by setting its innovation variance to zero.

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

$$\Delta g_t = \beta + \eta_t \tag{4}$$

where β and g_0 are constants, and ϵ_t, η_t are mutually uncorrelated white noise processes with different variances. In terms of (3) the β_j (j= 0, 1) are given by g_0, β respectively and $\xi_t = \sum_{j=1}^{t} \eta_j$. The local constant case $p = 0$ is the same as (4), but with β constrained to be zero. For the local quadratic case $p = 2$ a suitable dynamic model for g_t is

$$\Delta g_t = h_t + \nu_t, \qquad \Delta h_t = \beta + \eta_t \tag{5}$$

where β, g_0 and h_0 are constants, and ϵ_t, ν_t, η_t are mutually uncorrelated white noise processes all with different variances. Again this can be represented in the form (3) with the β_j involving g_0, h_0, β and $\xi_t = \sum_{j=1}^{t} \nu_j + \sum_{j=1}^{t} \sum_{k=1}^{j} \eta_k$. In this case ξ_t comprises two components, one a simple random walk and the other a doubly integrated random walk.

Clearly many other dynamic specifications for g_t and hence ξ_t are possible. For example, the constant β in both (4) and (5) could be allowed to depend on t and evolve as a simple random walk without drift. However this leads to an additional random walk component being incorporated in ξ_t and with a further level of integration. Given the typically modest number of observations in the window and the need to determine the variances of the various random walk components in ξ_t, it is important to keep the models as parsimonious as possible. As a consequence we have focused our attention on models for ξ_t which involve one or at most two variance parameters. Furthermore, we have chosen to keep the level of integration of these components as low as possible. This is in keeping with the empirical evidence for ARIMA structural models where the variances of the higher-order random walk innovations are typically very small by comparison to the simple random walk term. Thus, for example, setting the variance of η_t equal to zero in (5) yields a simpler version of this model which may prove to be more useful in practice.

Since the global parametric ARIMA structural models can be viewed as generalisations of exponential smoothing models (see Harvey (1989)), the models proposed here can be regarded as variants of exponential smoothing models tailored to a finite window. Note that simple truncated exponential smoothing models were considered for g_t in Kenny and Durbin (1982) and met with some success, especially for current trend estimates.

2.2 In the body

Given a finite window of width $n = 2r + 1$ points centred at time point t in the body of the time series, we now choose to estimate g_t by the finite moving-average

$$\hat{g}_t = \sum_{s=-r}^{r} w_s y_{t+s} \tag{6}$$

where the w_s are constrained by the requirement that

$$E\{\hat{g}_t - g_t\} = 0$$

and y_t follows the local dynamic model given by (2) and (3). Thus $\hat{g}_t$ is an unconditionally unbiased predictor of g_t. Note that this condition is equivalent to the requirement that the w_s satisfy

$$\sum_{s=-r}^{r} w_s = 1, \qquad \sum_{s=-r}^{r} s^j w_s = 0 \qquad\qquad (0 < j \leq p) \tag{7}$$

so that the moving average filter passes polynomials of degree p.

To assist in the choice of filter weights w_j we consider the criteria

$$F = E\{(\hat{g}_t - g_t)^2\}, \qquad S = E\{(\Delta^{p+1}\hat{g}_t)^2\} \tag{8}$$

where F measures the *fidelity* of $\hat{g}_t$ as an estimator of g_t and S measures its *smoothness*. Here the expectation operator is with respect to the particular local dynamic model adopted.

The smaller F is the better $\hat{g}_t$ is as an estimator of g_t whereas the smaller S is the closer the $\Delta^{p+1}\hat{g}_t$ are to zero and the closer $\hat{g}_t$ is to a smooth polynomial of degree p in t. In particular, the measure of smoothness S has been explicitly tailored to the degree of curvature (the degree of the underlying deterministic polynomial local trend) present in the data. Note that F is the familiar mean squared error criterion whereas S, appropriately normalised, is referred to as the R^2_{p+1} criterion in the actuarial graduation literature. See London (1985) for example and also Ramsay (1991) who considered minimising S in the presence of correlated y_t.

In the spirit of Henderson (1924) and Whittaker (1923) we further define the compromise criterion

$$Q = \theta F + (1 - \theta)S \tag{9}$$

where $0 \leq \theta \leq 1$ and θ is some user specified value. Note that Q includes both F and S as special cases. We shall adopt the principle that the smaller these three criteria are the better and determine the filter weights that optimise Q.

While minimising F is clearly a reasonable criterion to adopt, it is not so clear that this is appropriate for S. It could be argued that the aim should be to simply control the level of S rather than minimise it. However this approach again leads to Q as will be shown in Theorem 3.

Each local dynamic model describes data with its own intrinsic fidelity and smoothness. Thus the measures defined in (8) are model specific and would not normally be used to compare across different, possibly competing, models. Alternative measures that can be used to compare across models and which will be used later are now defined. First note that the identity filter with coefficients $w_0 = 1$ and $w_s = 0$ $(s \neq 0)$ satisfies (7) whatever the choice of p in the local dynamic model. Thus, for any given window, the fidelity of the data y_t to the trend g_t is given by

$$F_0 = E\{(y_t - g_t)^2\} = \sigma^2$$

and the smoothness of y_t as an estimate of the trend is given by

$$\begin{aligned} S_0 = E\{(\Delta^{p+1} y_t)^2\} &= E\{(\Delta^{p+1}\xi_t)^2\} + E\{(\Delta^{p+1}\epsilon_t)^2\} \\ &= E\{(\Delta^{p+1}\xi_t)^2\} + {}^{2p+2}C_{p+1}\sigma^2. \end{aligned}$$

The normalised quantities

$$F^* = \frac{F}{F_0}, \qquad S^* = \frac{S}{S_0} \tag{10}$$

measure the gains in fidelity and smoothness of $\hat{g}_t$ relative to the fidelity and smoothness of the data and can be used to compare across models. If $\xi_t = 0$ then F^* and S^* are identical to the R_0^2 and R_{p+1}^2 criteria respectively that are used in the actuarial graduation literature. Note also that replacing F by F^* and S by S^* introduces no essential change in Q apart from a simple one-to-one transformation of θ.

A primary advantage of defined measures of fidelity and smoothness is that we can now begin to quantify and classify the effects of competing moving-average trend filters and to consider trade-offs between fidelity and smoothness. In particular we can determine the values of w_j that satisfy (7) and minimise either F, S or Q. Before establishing these general results we first consider some special cases.

Consider minimising F in the case where g_t is the conventional model given by (3) with ξ_t equal to zero. From the Gauss-Markov theorem, minimising F subject to (7) is equivalent to estimating g_t by

$$\hat{g}_t = \sum_{j=0}^{p} \hat{\beta}_j t^j$$

where the estimates $\hat{\beta}_j$ are obtained from fitting the local polynomial model

$$y_{t+s} = \sum_{j=0}^{p} \beta_j (t+s)^j + \epsilon_{t+s} \qquad (-r \leq s \leq r)$$

by ordinary least squares. The latter procedure was initially advocated by Macaulay (1931) and forms the basis of many trend estimation filters. (See Kendall (1973) and Cleveland (1979) for example.) Thus minimising F with ξ_t equal to zero yields the Macaulay filters many of which are tabulated in Kendall (1973). The case where ξ_t is not equal to zero leads to local polynomial fitting using generalised least squares. An example of the latter is loess (see Cleveland et al. (1990)) which uses the tricube function for least squares weights. *In general we shall refer to trend filters obtained by minimising F as Macaulay filters irrespective of the choice of local dynamic model.*

Again assume that the model is given by (3) with ξ_t equal to zero and consider the case $p = 2$. Then minimising S gives the central Henderson filters tabulated, for example, in Dagum (1980). Our proof that this is so follows the treatment given in Kenny and Durbin (1982) with some key differences, the most important of which is that the w_s are not assumed to be symmetric à priori. Given these more general assumptions (7) and S become

$$\sum_{s=-r}^{r} w_s = 1, \qquad \sum_{s=-r}^{r} s w_s = 0, \qquad \sum_{s=-r}^{r} s^2 w_s = 0 \tag{11}$$

and

$$S = \sigma^2 \sum_{s=-r}^{r+3} (\Delta^3 w_s)^2 \tag{12}$$

respectively. Here σ^2 is the variance of ϵ_t and the w_s satisfy the boundary conditions

$$w_s = 0 \qquad (s = \pm(r+1), \pm(r+2), \pm(r+3)). \tag{13}$$

The criterion S can now be minimised with respect to the w_s $(-r \le s \le r)$ using Lagrange multipliers to account for the constraints given by (11). This yields

$$\Delta^6 w_{s+3} = a_1 + a_2 s + a_3 s^2 \qquad\qquad (-r \le s \le r) \tag{14}$$

where the a_j denote the Lagrange multipliers and these equations must now be solved subject to the boundary conditions (13). As in Kenny and Durbin (1982) the solution to (14) is a polynomial in s of order 8 with roots $\pm(r+1)$, $\pm(r+2)$, $\pm(r+3)$ and so

$$w_s = ((r+1)^2 - s^2)(r+2)^2 - s^2)(r+3)^2 - s^2)(b_1 + b_2 s + b_3 s^2) \tag{15}$$

with the b_j determined from (11). The w_s obtained are the central Henderson filters used in X-11.

Note that the Henderson filters derived above follow directly from a simple measure of smoothness and the assumption that the local model is quadratic (rather than cubic). No symmetry has been assumed, this property being a consequence of the optimisation process. The fact that the filter also passes a cubic is to be regarded as serendipitous and arising from the fact that this optimal central filter is symmetric. This latter property will not hold at the ends of the series where the filters are asymmetric. *In general we refer to trend filters obtained by minimising S as Henderson filters irrespective of the choice of local dynamic model.*

We now determine the filter weights w_s that optimise F, S and Q in the case where ξ_t is not necessarily zero. Clearly it is sufficient to consider the compromise criterion Q since F and S can be recovered from Q by setting $\theta = 0$ or $\theta = 1$ respectively.

Further notation is needed to establish and present the results. For the local dynamic model specified by (2) and (3) define covariance matrices $\mathbf{\Omega}$ and $\mathbf{\Gamma}$ with typical elements

$$\mathbf{\Omega}_{jk} = \mathrm{cov}(\xi_{t+j} - \xi_t, \xi_{t+k} - \xi_t), \qquad \mathbf{\Gamma}_{jk} = \mathrm{cov}(\Delta^{p+1}\xi_{t+j}, \Delta^{p+1}\xi_{t+k}) \tag{16}$$

where $-r \leq j, k \leq r$ and t indexes the central time point of the window concerned. Furthermore, define the $n \times (p+1)$ dimensional matrix $\mathbf{C}$ and the $p+1$ dimensional vector $\mathbf{c}$ by

$$\mathbf{C} = \begin{pmatrix} 1 & -r & \ldots & (-r)^p \\ 1 & -r+1 & \ldots & (-r+1)^p \\ . & . & \ldots & . \\ . & . & \ldots & . \\ 1 & r-1 & \ldots & (r-1)^p \\ 1 & r & \ldots & r^p \end{pmatrix}, \qquad \mathbf{c} = \begin{pmatrix} 1 \\ 0 \\ . \\ . \\ 0 \\ 0 \end{pmatrix} \tag{17}$$

and let $\sigma^2 \mathbf{B}_k$ denote the covariance matrix of a sequence of n observations from the stationary moving average process $\Delta^k \epsilon_t$. Observe that $\mathbf{B}_k$ does not involve σ^2 the variance of ϵ_t. Finally, let $\mathbf{I}$ denote the n-dimensional identity matrix. We now have the following result.

Theorem 1 *Let y_t follow the local dynamic model specified by (2) and (3). Then the values of w_s that minimise Q subject to (7) are given by $\mathbf{w} = (w_{-r}, \ldots, w_r)^T$ where*

$$\mathbf{w} = \mathbf{E}_\theta^{-1}\mathbf{C}(\mathbf{C}^T\mathbf{E}_\theta^{-1}\mathbf{C})^{-1}\mathbf{c}$$

and

$$\mathbf{E}_\theta = \theta(\sigma^2\mathbf{I} + \mathbf{\Omega}) + (1-\theta)(\sigma^2\mathbf{B}_{p+1} + \mathbf{\Gamma}).$$

Proof

First note that (7) is equivalent to the requirement that $\mathbf{w}$ satisfy

$$\mathbf{C}^T\mathbf{w} = \mathbf{c}. \tag{18}$$

Given this condition,

$$\sum_{s=-r}^{r} w_s y_{t+s} - g_t = \sum_{s=-r}^{r} w_s(\xi_{t+s} - \xi_t) + \sum_{s=-r}^{r} w_s \epsilon_{t+s}$$

and

$$\sum_{s=-r}^{r} w_s \Delta^{p+1} y_{t+s} = \sum_{s=-r}^{r} w_s \Delta^{p+1}\xi_{t+s} + \sum_{s=-r}^{r} w_s \Delta^{p+1}\epsilon_{t+s}.$$

These yield the quadratic forms

$$F = \mathbf{w}^T(\sigma^2\mathbf{I} + \mathbf{\Omega})\mathbf{w}, \qquad S = \mathbf{w}^T(\sigma^2\mathbf{B}_{p+1} + \mathbf{\Gamma})\mathbf{w} \tag{19}$$

and

$$Q = \mathbf{w}^T \mathbf{E}_\theta \mathbf{w}. \tag{20}$$

We now seek to minimise Q subject to (18).

Using Lagrange multipliers and differentiating with respect to the w_s yields

$$\mathbf{w} = \mathbf{E}_\theta^{-1} \mathbf{C} \mu$$

where μ is the vector of Lagrange multipliers. The constraint (18) yields

$$\mu = (\mathbf{C}^T \mathbf{E}_\theta^{-1} \mathbf{C})^{-1} \mathbf{c}$$

and the result follows. □

Given a local dynamic model, it is of interest to consider the nature of the variation of the optimal F and S with θ. The following result shows, in particular, that the optimal measures of fidelity F and smoothness S decrease and increase respectively as θ increases. Thus the filter with the best smoothness measure S has the worst fidelity measure F and vice-versa.

Theorem 2 *Let y_t follow the local dynamic model specified by (2) and (3). If $\mathbf{w}$ is given by Theorem 1 and F, S given by (8) then*

$$\theta \frac{dF}{d\theta} + (1-\theta)\frac{dS}{d\theta} = 0, \qquad \frac{dF}{d\theta} \leq 0, \qquad \frac{dS}{d\theta} \geq 0$$

where $0 \leq \theta \leq 1$. In particular F and S have zero gradients at $\theta = 1$ and $\theta = 0$ respectively.

Proof

Let

$$\mathbf{A} = \sigma^2 \mathbf{I} + \mathbf{\Omega}, \qquad \mathbf{B} = \sigma^2 \mathbf{B}_{p+1} + \mathbf{\Gamma}$$

so that

$$F = \mathbf{w}^T \mathbf{A} \mathbf{w}, \qquad S = \mathbf{w}^T \mathbf{B} \mathbf{w}$$

and $\mathbf{w}$ is given by Theorem 1. Differentiating F and S with respect to θ we obtain

$$\frac{dF}{d\theta} = -2\mathbf{u}^T(\mathbf{I} - \mathbf{P})\mathbf{K}_\theta^T \mathbf{A} \mathbf{K}_\theta \mathbf{v}, \qquad \frac{dS}{d\theta} = -2\mathbf{u}^T(\mathbf{I} - \mathbf{P})\mathbf{K}_\theta^T \mathbf{B} \mathbf{K}_\theta \mathbf{v}$$

where

$$\begin{aligned} \mathbf{u} &= \mathbf{K}_\theta^T (\mathbf{A} - \mathbf{B}) \mathbf{K}_\theta \mathbf{v} \\ \mathbf{v} &= \mathbf{K}_\theta^T \mathbf{C} (\mathbf{C}^T \mathbf{E}_\theta^{-1} \mathbf{C})^{-1} \mathbf{c} \\ \mathbf{P} &= \mathbf{K}_\theta^T \mathbf{C} (\mathbf{C}^T \mathbf{E}_\theta^{-1} \mathbf{C})^{-1} \mathbf{C}^T \mathbf{K}_\theta \end{aligned}$$

and $\mathbf{E}_\theta^{-1} = \mathbf{K}_\theta \mathbf{K}_\theta^T$. Thus

$$\theta \frac{dF}{d\theta} + (1-\theta)\frac{dS}{d\theta} = 0 \tag{21}$$

and

$$\frac{dS}{d\theta} - \frac{dF}{d\theta} = 2\mathbf{u}^T(\mathbf{I} - \mathbf{P})\mathbf{u} \geq 0 \tag{22}$$

since $\mathbf{P}$ is a projection matrix. From (21) we observe that F has gradient zero at $\theta = 1$ and S has gradient zero at $\theta = 0$. Moreover, from (21) and (22), we obtain

$$0 \leq \theta(\frac{dS}{d\theta} - \frac{dF}{d\theta}) = \frac{dS}{d\theta} \qquad (0 < \theta \leq 1)$$

and

$$0 \geq (1-\theta)(\frac{dF}{d\theta} - \frac{dS}{d\theta}) = \frac{dF}{d\theta} \qquad (0 \leq \theta < 1)$$

as required. □

As mentioned before, if the aim is to control or impose smoothness on the trend estimate $\hat{g}_t$, then an alternative strategy might be to minimise F subject to (7) and a given level of smoothness $S = s$. This is equivalent to minimising Q as we shall see in Theorem 3.

First, however, note that minimising F subject to (7) alone determines the best linear unbiased predictor (BLUP) of g_t from the observations y_t in the window. Thus, minimising F subject to (7) and $S = s$ gives a constrained BLUP for g_t. This estimator will have the desired smoothness property, but at the cost of a higher value of F than before. Now let s_0 denote the smoothness of the optimum smoothness filter and s_1 denote the smoothness of the optimum fidelity filter so that

$$s_0 = \mathbf{w}_0^T(\sigma^2\mathbf{B}_{p+1} + \mathbf{\Gamma})\mathbf{w}_0, \qquad s_1 = \mathbf{w}_1^T(\sigma^2\mathbf{B}_{p+1} + \mathbf{\Gamma})\mathbf{w}_1$$

where $\mathbf{w}_0$ and $\mathbf{w}_1$ are the values of $\mathbf{w}$ given by Theorem 1 for $\theta = 0$ and $\theta = 1$ respectively. If a trend estimator $\hat{g}_t$ is sought with better smoothness properties then the user will need to constrain the smoothness S to lie in the interval $[s_0, s_1]$.

Theorem 3 *Let y_t follow the local dynamic model specified by (2) and (3). Then minimising F subject to both (7) and $S = s$ in $[s_0, s_1]$ is equivalent to minimising Q subject to (7) alone where θ is a solution of*

$$\mathbf{c}^T(\mathbf{C}^T\mathbf{E}_\theta^{-1}\mathbf{C})^{-1}\mathbf{C}^T\mathbf{E}_\theta^{-1}(\sigma^2\mathbf{B}_{p+1} + \mathbf{\Gamma})\mathbf{E}_\theta^{-1}\mathbf{C}(\mathbf{C}^T\mathbf{E}_\theta^{-1}\mathbf{C})^{-1}\mathbf{c} = s,$$

$\mathbf{E}_\theta$ is given by Theorem 1 and θ lies in $[0, 1]$. The minimisations are with respect to the filter weights w_s.

Proof

We seek to minimise

$$\tilde{\mathbf{Q}} = F + \phi(S - s) - 2\mu^T(\mathbf{C}^T\mathbf{w} - \mathbf{c})$$

where the scalar ϕ and the vector μ are Lagrange multipliers. Differentiating and solving the resulting equations yields

$$\mathbf{w} = \tilde{\mathbf{E}}_\phi^{-1}\mathbf{C}(\mathbf{C}^T\tilde{\mathbf{E}}_\phi^{-1}\mathbf{C})^{-1}\mathbf{c}$$

where ϕ is a solution of

$$\mathbf{c}^T(\mathbf{C}^T\tilde{\mathbf{E}}_\phi^{-1}\mathbf{C})^{-1}\mathbf{C}^T\tilde{\mathbf{E}}_\phi^{-1}(\sigma^2\mathbf{B}_{p+1}+\mathbf{\Gamma})\tilde{\mathbf{E}}_\phi^{-1}\mathbf{C}(\mathbf{C}^T\tilde{\mathbf{E}}_\phi^{-1}\mathbf{C})^{-1}\mathbf{c}=s \tag{23}$$

and

$$\tilde{\mathbf{E}}_\phi=(\sigma^2\mathbf{I}+\mathbf{\Omega})+\phi(\sigma^2\mathbf{B}_{p+1}+\mathbf{\Gamma}).$$

In almost exactly the same way as was done in Theorem 2, it can be shown that the left hand side of (23) is a non-increasing function of ϕ whose limit, as $\phi\to\infty$, is s_0. Moreover $\phi=0$ yields $s=s_1$. Since $s_0\le s\le s_1$ we conclude that a solution of (23) can be found for $\phi\ge 0$. The result of Theorem 1 is now retrieved by setting $\theta=1/(1+\phi)$. □

In fact, in much the same way as for Theorem 3, it can be shown that minimising smoothness subject to (7) and a given level of fidelity $F=f$ is also equivalent to minimising Q. However this seems a less relevant criterion to adopt in practice.

In practice ξ_t will comprise one or other or possibly the sum of the random walk components $\xi_t^{(1)}$ and $\xi_t^{(2)}$ where

$$\xi_t^{(1)}=\sum_{j=1}^{t}\nu_j, \qquad\qquad \xi_t^{(2)}=\sum_{j=1}^{t}\sum_{k=1}^{j}\eta_k \tag{24}$$

and ν_t, η_t are mutually uncorrelated white noise processes with $E(\nu_t^2)=\sigma_\nu^2$, $E(\eta_t^2)=\sigma_\eta^2$. If ξ_t is a linear combination of random walk components such as the above whose levels of integration do not exceed $p+1$, then $\mathbf{\Gamma}$ will be a linear combination of covariance matrices $\mathbf{B}_k$ where k is a nonnegative integer such that $k\le p$.

The matrix $\mathbf{\Omega}$ can be constructed directly from the covariance matrices of the random walk components that make up ξ_t. However some simplification is possible as we now show for the two cases of primary interest.

In the case of model (4) or (5) with σ_η^2 set to zero, ξ_t is just $\xi_t^{(1)}$. Then we have

$$\xi_{t+s}^{(1)}-\xi_t^{(1)}=\begin{cases}\sum_{j=1}^{|s|}\nu_{t+j} & (s>0)\\ 0 & (s=0)\\ -\sum_{j=1}^{|s|}\nu_{t+1-j} & (s<0)\end{cases}$$

and $\mathbf{\Omega}=\mathbf{\Omega}^{(1)}$ has typical element

$$\mathbf{\Omega}_{jk}^{(1)}=\sigma_\nu^2\min(|j|,|k|) \qquad (-r\le j,k\le -1, 1\le j,k\le r) \tag{25}$$

and zero otherwise.

Now suppose that $\xi_t=\xi_t^{(2)}$ as would be the case for model (5) with σ_ν^2 set to zero and consider the quadratic form

$$\mathbf{w}^T\mathbf{\Omega}\mathbf{w}=E\{(\sum_{-r}^{r}w_s(\xi_{t+s}^{(2)}-\xi_t^{(2)}))^2\}.$$

For $s > 0$

$$\begin{aligned}
\xi_{t+s}^{(2)} - \xi_t^{(2)} &= \sum_{j=1}^{s} \sum_{k=1}^{t+j} \eta_k \\
&= s \sum_{k=1}^{t} \eta_k + s\eta_{t+1} + \sum_{j=1}^{s-1} \sum_{k=1}^{j} \eta_{t+1+k} \qquad (26)
\end{aligned}$$

and for $s < 0$

$$\begin{aligned}
\xi_{t+s}^{(2)} - \xi_t^{(2)} &= -\sum_{j=0}^{|s|-1} \sum_{k=1}^{t-j} \eta_k \\
&= s \sum_{k=1}^{t} \eta_k + \sum_{j=1}^{|s|-1} \sum_{k=1}^{j} \eta_{t+1-k} \qquad (27)
\end{aligned}$$

where the last term on the right hand side of (26) and (27) is interpreted as zero when $|s| = 1$. The first term on the right hand side of (26) and (27) is eliminated by the filter since $\sum_{s=-r}^{r} s w_s = 0$ and the remaining terms of (26) and (27) are uncorrelated. Since the weights w_s satisfy (18) we obtain

$$\begin{aligned}
\mathbf{w}^T \boldsymbol{\Omega} \mathbf{w} &= \sigma_\eta^2 (\sum_{s=1}^{r} s w_s)^2 + \mathbf{w}^T \boldsymbol{\Psi} \mathbf{w} \\
&= \frac{\sigma_\eta^2}{2} ((\sum_{s=1}^{r} s w_s)^2 + (\sum_{s=1}^{r} s w_{-s})^2) + \mathbf{w}^T \boldsymbol{\Psi} \mathbf{w}
\end{aligned}$$

where $\boldsymbol{\Psi}$ has typical element

$$\boldsymbol{\Psi}_{jk} = \sigma_\eta^2 \sum_{s=1}^{\min(|j|,|k|)} (|j| - s)(|k| - s) \qquad (-r \le j, k \le -1, 1 \le j, k \le r) \qquad (28)$$

and zero otherwise. Thus, when $\xi_t = \xi_t^{(2)}$, the matrix $\boldsymbol{\Omega}$ is given by $\boldsymbol{\Omega}^{(2)}$ which has typical element

$$\boldsymbol{\Omega}_{jk}^{(2)} = \tfrac{1}{2} \sigma_\eta^2 |j||k| + \boldsymbol{\Psi}_{jk} \qquad (-r \le j, k \le -1, 1 \le j, k \le r) \qquad (29)$$

and zero otherwise.

Note, in particular, that $\boldsymbol{\Omega}^{(1)}$ and $\boldsymbol{\Omega}^{(2)}$ do not depend on t, the absolute value of time indexing the origin of the window. This natural and important invariance property holds for general $\boldsymbol{\Omega}$ and is a consequence of (7) and the assumption that the levels of integration of the random walk components that make up ξ_t do not exceed $p + 1$.

So far no requirement has been made that the filters we consider have symmetric weights. The following result shows that the filter with weights given by Theorem 1 is symmetric.

Theorem 4 *The values of w_s given by Theorem 1 correspond to a symmetric filter of length $n = 2r + 1$.*

Proof

Let $\mathbf{u} = (w_{-1}, \ldots, w_{-r})^T$, $\mathbf{v} = (w_1, \ldots, w_r)^T$ and $\mathbf{d}_j = (1^j, 2^j, \ldots, r^j)$ for j an integer such that $0 \leq j \leq p$.

Consider first the fidelity measure F and the particular case where $\xi_t = \xi_t^{(2)}$. Let the matrix $\mathbf{\Omega}_+^{(2)}$ have typical element $\mathbf{\Omega}_{jk}^{(2)}$ $(1 \leq j, k \leq r)$. Then the constraints (18) become

$$\begin{aligned} w_0 + \mathbf{d}_0^T\mathbf{u} + \mathbf{d}_0^T\mathbf{v} &= 1 \\ \mathbf{d}_j^T\mathbf{u} &= \mathbf{d}_j^T\mathbf{v} \qquad (0 < j \leq p) \end{aligned}$$

and F can be written as

$$\sigma^2(w_0^2 + \mathbf{u}^T\mathbf{u} + \mathbf{v}^T\mathbf{v}) + \sigma_\eta^2(\tfrac{1}{2}(\mathbf{d}_1^T\mathbf{u})^2 + \tfrac{1}{2}(\mathbf{d}_1^T\mathbf{v})^2 + \mathbf{u}^T\mathbf{\Omega}_+^{(2)}\mathbf{u} + \mathbf{v}^T\mathbf{\Omega}_+^{(2)}\mathbf{v}).$$

Note that both the constraints and F are symmetric in the vectors $\mathbf{u}$ and $\mathbf{v}$. A similar argument holds for more general forms of ξ_t involving sums of integrated random walks of any finite order.

Now consider the smoothness measure S. Note that S is of the form

$$S = \mathbf{w}^T\mathbf{A}\mathbf{w}$$

where $\mathbf{A}$ is a Toeplitz matrix with spectral density $f(\omega)$ and typical element

$$\mathbf{A}_{jk} = \int_{-\pi}^{\pi} e^{i(j-k)\omega} f(\omega) d\omega \qquad (-r \leq j, k \leq r).$$

Thus

$$S = \int_{-\pi}^{\pi} |w_0 + \sum_{j=1}^{r} u_j e^{-ij\lambda} + \sum_{j=1}^{r} v_j e^{ij\lambda}|^2 f(\omega) d\omega.$$

is also symmetric in the vectors $\mathbf{u}$ and $\mathbf{v}$. Since $Q = \theta F + (1-\theta)S$ inherits the same symmetry property we conclude that the value of $\mathbf{w}$ minimising Q satisfies $\mathbf{u} = \mathbf{v}$ and corresponds to a symmetric filter. $\square$

The properties of these filters are considered and discussed in Section 3.1.

2.3 At the ends

Consider now the ends of the series and, in particular, the most recent time points. Here the need for high quality estimates of the trend and hence the seasonal is of dominating and over-riding importance.

Various strategies have been adopted at the ends of series. For example Macaulay advocates using a best-fitting polynomial determined for the last complete window to provide the required trend values at the ends of the series. This approach is also adopted by Kendall (1973) and by Cleveland (1979) for the loess and lowess smoothers. Thus the moving window paradigm used in the body of the series is supplanted by a static window paradigm at the ends. This inconsistency can lead to smoothness problems at the ends of the series (see Gray and Thomson (1990)). The alleviation of these difficulties no doubt lies behind Macaulay's suggestion (Macaulay (1931), Chapter VIII) that the fitted polynomial for the last complete window should share a common slope with the most recent trend estimates. However this suggestion is not adopted by either Kendall (1973) or Cleveland (1979) for their end filters.

An alternative strategy is to maintain the same moving window paradigm at the ends of the series as was adopted in the body. This consistency comes at a price since the windows adopted at the ends now include 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). As noted in Section 1, 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

$$E\{(\hat{g}_t - \tilde{g}_t)^2\} = E\{(\sum_{s=-r}^{r} w_s y_{t+s} - \tilde{g}_t)^2\} \tag{30}$$

where $\tilde{g}_t$ is a predictor of $\hat{g}_t$ based on past data and the w_s are the weights of the central filter that applies once all the data is to hand. In general, given a history of observations $y_1, \ldots, y_T$, it is evident that (30) is minimised when

$$\tilde{g}_t = \sum_{s=-r}^{r} w_s \hat{y}_{t+s} \tag{31}$$

and $\hat{y}_{t+s} = E(y_{t+s}|y_1, \ldots, y_T)$ 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 (31) 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 (31) shows that in general $\hat{y}_{t+s}$, and hence $\tilde{g}_t$, need not necessarily be linear in the observations.

The expectation operator applying in (33) is with respect to the global or local model adopted for the time series concerned. In this regard note that a local model is implicitly used by the X-11-ARIMA procedure (Dagum (1980)) to determine trends from forecast augmented data where the forecasts are based on a global parametric model. Although this mismatch does not seem to matter in practice, in principle one would like to be consistent and work within the same model framework. This is our objective here.

In this paper we adopt the minimum revisions strategy based on the moving window paradigm and the local dynamic trend model of Section 2.1. At the ends of the series we choose to predict $\hat{g}_t = \sum_{s=-r}^{r} w_s y_{t+s}$ by the linear predictor

$$\tilde{g}_t = \sum_{s=-r}^{q} u_s y_{t+s} \tag{32}$$

which minimises

$$R_q = E\{(\hat{g}_t - \tilde{g}_t)^2\} \tag{33}$$

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)).

2.3.1 Unbiased predictors

If $\tilde{g}_t$ is to be an unbiased predictor of $\sum_{s=-r}^{r} w_s y_{t+s}$ then the u_s must satisfy

$$\sum_{s=-r}^{q} u_s = 1, \qquad \sum_{s=-r}^{q} s^j u_s = 0 \qquad\qquad (0 < j \leq p). \tag{34}$$

Thus the asymmetric moving average filter implied by (32) passes polynomials of degree p. Moreover R_q is now given by

$$R_q = \mathbf{v}^T \mathbf{E}_1 \mathbf{v} \tag{35}$$

where $\mathbf{v}$ has typical element

$$v_s = \begin{cases} w_s - u_s & (-r \leq s \leq q) \\ w_s & (q < s \leq r) \end{cases} \tag{36}$$

and the central filter weights w_s that apply in the body of the series satisfy (7).

For each q, appropriate values of the u_s can now be determined by minimising R_q subject to (34). As we show below, this results in an end filter that satisfies a particular form of (31) involving optimal prediction.

First consider predicting

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

from $y_{t-r}, \dots y_{t+q}$ 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.

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

$$\mathbf{I} = [\mathbf{L_1}, \mathbf{L_2}] \tag{37}$$

where $\mathbf{I}$ denotes the n-dimensional identity matrix. We now establish the following result.

Theorem 5 *Let y_t follow the local dynamic model specified by (2) and (3). Given $\delta = (\delta_{-r}, \dots, \delta_r)^T$ and observations $y_{t-r}, \dots, y_{t+q}$ for $0 \le q < r$, the BLUP of $\sum_{s=-r}^{r} \delta_s y_{t+s}$ is $\sum_{s=-r}^{q} u_s y_{t+s}$ where $\mathbf{u} = (u_{-r}, \dots, u_q)^T$ 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 $\mathbf{E}_\theta$ defined in Theorem 1 evaluated at $\theta = 1$ and $\mathbf{C}$ is given by (17).

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 $q < s \le r$ and y_{t+s} otherwise.

Proof

For $\hat{Y} = \sum_{s=-r}^{q} u_s y_{t+s}$ to be an unbiased predictor of $Y = \sum_{s=-r}^{r} \delta_s y_{t+s}$ the u_s must satisfy

$$\sum_{-r}^{q} s^j u_s = \sum_{-r}^{r} s^j \delta_s \qquad (0 \le j \le p). \tag{38}$$

Given (38)

$$Y - \hat{Y} = \sum_{s=-r}^{r} v_s(\xi_{t+s} - \xi_t + \epsilon_{t+s})$$

where

$$v_s = \begin{cases} \delta_s - u_s & (-r \le s \le q) \\ \delta_s & (q < s \le r) \end{cases} \tag{39}$$

and

$$\sum_{-r}^{r} s^j v_s = 0 \qquad\qquad (0 \le j \le p). \tag{40}$$

In terms of the v_s these conditions can be written as

$$\mathbf{L}_2^T \mathbf{v} = \mathbf{L}_2^T \delta, \qquad\qquad \mathbf{C}^T \mathbf{v} = \mathbf{0} \tag{41}$$

where $\mathbf{v} = (v_{-r}, \ldots, v_r)^T$ and $\mathbf{C}$ is given by (17). We now need to minimise

$$E\{(Y - \hat{Y})^2\} = \mathbf{v}^T(\sigma^2 \mathbf{I} + \mathbf{\Omega})\mathbf{v} = \mathbf{v}^T \mathbf{E}_1 \mathbf{v} \tag{42}$$

subject to (41).

Using Lagrange multipliers, optimising with respect to the v_s, and incorporating the constraints (41) yields the equations

$$\begin{aligned} \mathbf{v} &= \mathbf{E}_1^{-1}\mathbf{C}\mu + \mathbf{E}_1^{-1}\mathbf{L}_2\nu \\ \mathbf{L}_2^T\mathbf{E}_1^{-1}\mathbf{C}\mu + \mathbf{L}_2^T\mathbf{E}_1^{-1}\mathbf{L}_2\nu &= \mathbf{L}_2^T\delta \\ \mathbf{C}^T\mathbf{E}_1^{-1}\mathbf{C}\mu + \mathbf{C}^T\mathbf{E}_1^{-1}\mathbf{L}_2\nu &= \mathbf{0} \end{aligned}$$

where μ, ν are the vectors of Lagrange multipliers. These together with (39) yield

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

as stated.

Now the BLUP of y_{t+s} is obtained by setting $\delta = \mathbf{0}$ with the exception of the s-th element which is set to unity. Then the BLUP of y_{t+s} is given by

$$\hat{y}_{t+s} = \begin{cases} y_{t+s} & (-r \le s \le q) \\ \mathbf{h}_s^T\mathbf{y} & (q < s \le r) \end{cases}$$

where $\mathbf{h}_s$ is the s-th column of $-\mathbf{L}_1^T\mathbf{G}\mathbf{L}_2(\mathbf{L}_2^T\mathbf{G}\mathbf{L}_2)^{-1}\mathbf{L}_2^T$ and $\mathbf{y} = (y_{t-r}, \ldots, y_{t+q})^T$. Thus, for arbitrary choice of δ,

$$\sum_{s=-r}^{q} u_s y_{t+s} = \delta^T(\mathbf{I} - \mathbf{G}\mathbf{L}_2(\mathbf{L}_2^T\mathbf{G}\mathbf{L}_2)^{-1}\mathbf{L}_2^T)^T\mathbf{L}_1\mathbf{y} = \sum_{s=-r}^{r} \delta_s \hat{y}_{t+s}$$

as required. $\square$

Since the v_s in Theorem 5 satisfy (40), the same comments that were made concerning the form of $\mathbf{\Omega}$ following Theorem 3 apply here also. In particular, $\mathbf{\Omega}$ and hence (42) do not depend on t, the absolute value of time indexing the origin of the window. If ξ_t is $\xi_t^{(1)}$ or $\xi_t^{(2)}$ then $\mathbf{\Omega}$ can be replaced by $\mathbf{\Omega}^{(1)}$ or $\mathbf{\Omega}^{(2)}$ respectively.

Replacing the arbitrary δ_s by the weights w_s corresponding to some central filter chosen for the body of the series yields the following result.

Corollary 6 *Let y_t follow the local dynamic model specified by (2), (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 $\tilde{g}_t = \sum_{s=-r}^{q} u_s y_{t+s}$ be an unbiased predictor of $\sum_{s=-r}^{r} w_s y_{t+s}$ with the u_s satisfying (34). Then, for $0 \leq q < r$, the values of u_s that minimise R_q subject to (34) are given by Theorem 5 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.

Proof

Since $\delta = \mathbf{w}$ it follows from (34) that the u_s satisfy (38) and $R_q = E\{(Y - \hat{Y})^2\}$. The result follows. □

Because of their dependence on BLUP predictors we shall henceforth refer to the end filters specified by Corollary 6 as *BLUP end filters*. Their properties are investigated in Section 3.2.

Theorem 5 and Corollary 6 establish a direct relationship between optimal linear prediction and minimum revisions. However the nature of these results suggests that there may be a special link between these two criteria and optimal fidelity where $\theta = 1$. The following result makes this connection more precise.

Theorem 7 *Let y_t follow the local dynamic model specified by (2), (3) and let* $\mathbf{w}$ *specify the vector of weights for the optimum fidelity central filter given by Theorem 1 with $\theta = 1$. Then, for $0 \leq q < r$ and $\tilde{g}_t$ given by (32), minimising the revisions R_q subject to (34) is equivalent to optimising the fidelity*

$$F = E\{(\tilde{g}_t - g_t)^2\} = E\{(\sum_{s=-r}^{q} u_s y_{t+s} - g_t)^2\}$$

subject to (34) where both optimisations are with respect to the end filter weights u_s. In this case the BLUP of $\sum_{s=-r}^{r} w_s y_{t+s}$ equals the BLUP of g_t.

Proof

Note that

$$F = \mathbf{v}^T \mathbf{E}_1 \mathbf{v}$$

and

$$R_q = (\mathbf{w} - \mathbf{v})^T \mathbf{E}_1 (\mathbf{w} - \mathbf{v}) = \mathbf{w}^T \mathbf{E}_1 \mathbf{w} + \mathbf{v}^T \mathbf{E}_1 \mathbf{v} - 2\mathbf{v}^T \mathbf{E}_1 \mathbf{w}$$

where $\mathbf{E}_1$ is given by Theorem 1 and $\mathbf{v}$ has typical element

$$v_s = \begin{cases} u_s & (-r \leq s \leq q) \\ 0 & (q < s \leq r) \end{cases}.$$

Moreover $\tilde{g}_t$ is an unbiased predictor of g_t since the u_s satisfy (34) or, equivalently, $\mathbf{v}$ satisfies (18).

Let δ denote the vector with typical element δ_s, $-r \leq s \leq r$, where $\delta_0 = 1$ and $\delta_s = 0$, $s \neq 0$. Then, from Theorem 1,

$$\begin{aligned}\mathbf{v}^T\mathbf{E}_1\mathbf{w} &= \mathbf{v}^T\mathbf{C}(\mathbf{C}^T\mathbf{E}_1^{-1}\mathbf{C})^{-1}\mathbf{c} \\ &= \delta^T\mathbf{C}(\mathbf{C}^T\mathbf{E}_1^{-1}\mathbf{C})^{-1}\mathbf{c} = \delta^T\mathbf{E}_1\mathbf{w}\end{aligned}$$

since $\mathbf{v}$ and δ satisfy (18). Similarly $\mathbf{w}^T\mathbf{E}_1\mathbf{w} = \delta^T\mathbf{E}_1\mathbf{w}$. Furthermore $\mathbf{E}_1 = \sigma^2\mathbf{I} + \boldsymbol{\Omega}$ and $\boldsymbol{\Omega}_{s0} = \boldsymbol{\Omega}_{0s} = 0$ for $-r \leq s \leq r$ so that $\delta^T\mathbf{E}_1\mathbf{w} = \sigma^2 w_0$. The result now follows since

$$R_q = F - \sigma^2 w_0$$

and $\sigma^2 w_0$ does not depend on the u_s. □

Before concluding this section we briefly comment on an alternative approach to the above that derives from the development given in Section 2.1. Consider directly estimating the trend g_t at the ends of the series by the end filter (32). Now, maintaining the same balance of smoothness and fidelity that was used in the body of the series, consider determining the values of u_s that minimise Q given by (8) and (9) subject to the requirement that $\tilde{g}_t$ be an unconditionally unbiased predictor of g_t or, equivalently, (34).

Superficially this seems a reasonable strategy. However a number of caveats apply. First, it is not at all clear that one should maintain the same balance of smoothness and fidelity at the ends as in the body. End filters basically deal with a transition problem that ultimately goes away as time points are subsumed into the body of the series. A criterion, such as the minimum revisions criterion, that minimises the total cost of this transition has much to commend it. Secondly, the values of the fidelity and smoothness measures F and S are time invariant in the body, but dependent on time at the ends. Thus controlling the balance of fidelity and smoothness in the body indirectly controls the absolute levels of F or S (see Theorem 3 for example). This is no longer true at the ends where the end filters are time dependent. Furthermore, the smoothness measure S is defined in terms of backwards differences. Thus the variation of S with time at the ends may well result in end filters with significant phase shifts. Despite these reservations, the requisite end filters were derived and these are presented without proof in Appendix A.1.

2.3.2 Biased predictors

Again consider a window of $n = 2r + 1$ observations $y_{t-r}, \ldots, y_{t+r}$ centred at time t of which the $r - q$ observations $y_{t+q+1}, \ldots, y_{t+r}$ are missing and $0 \leq q < r$. Assume, as before, that $\tilde{g}_t$ is a linear predictor of the form (32), 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. Unlike the end filters based on BLUP predictors, the end filters we construct will no longer be independent of these parameters. The requirement that the bias be time invariant does, however, lead to relatively straightforward procedures for estimating the parametric quantities involved. By virtue of the local dynamic models adopted, the end filters derived generalise and extend the current X-11 end filters which were developed by Musgrave (1964) and placed in a prediction context by Doherty (1991).

Following the development in Section 2.3.1 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 mean squared error criterion $E\{(Y-\hat{Y})^2\}$ is given by

$$E\{(Y-\hat{Y})^2\} = (\sum_{s=-r}^{r} v_s E y_{t+s})^2 + \mathrm{Var}\{\sum_{s=-r}^{r} v_s y_{t+s}\} \tag{43}$$

where $\mathbf{v}$ is given by (39) with typical element $v_s = \delta_s - u_s$ when $-r \le s \le q$ and δ_s otherwise. The bias term in (43) can be written as

$$\sum_{s=-r}^{r} v_s E y_{t+s} = \sum_{s=-r}^{r} v_s \sum_{j=0}^{p} \beta_j (t+s)^j = \sum_{k=0}^{p} (\sum_{j=0}^{p-k} \beta_{j+k} \, {}^{j+k}C_k \sum_{s=-r}^{r} v_s s^j) t^k \tag{44}$$

and this will be invariant to the location of the window's time origin t if $p=0$ or if

$$\sum_{s=-r}^{r} s^j v_s = 0 \qquad\qquad (0 \le j < p). \tag{45}$$

when $p > 0$. Note that (45) is equivalent to

$$\sum_{s=-r}^{q} s^j u_s = \sum_{s=-r}^{r} s^j \delta_s \qquad\qquad (0 \le j < p). \tag{46}$$

If $p > 0$ and the u_s satisfy (46) then

$$\mathrm{Var}\{\sum_{s=-r}^{r} v_s y_{t+s}\} = E\{(\sum_{s=-r}^{r} v_s(\xi_{t+s} - \xi_t + \epsilon_{t+s}))^2\}$$

and (43) becomes

$$E\{(Y-\hat{Y})^2\} = \beta_p^2 (\sum_{s=-r}^{r} s^p v_s)^2 + \mathbf{v}^T \mathbf{E}_1 \mathbf{v} \tag{47}$$

with $\mathbf{v}$ given by (39) and $\mathbf{E}_1 = \sigma^2 \mathbf{I} + \mathbf{\Omega}$ given by Theorem 5. As before, it is desirable for the mean squared error $E\{(Y-\hat{Y})^2\}$ as well as the bias to be time invariant. However, as indicated in the discussions following Theorems 3 and 5, this will only be the case if (40) holds and the levels of integration of the random walk components that make up ξ_t do not exceed $p+1$. Here the v_s satisfy (45) rather than (40). Thus, to ensure that the mean squared error for the biased predictors is time invariant, we need to impose the stronger condition that 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 so (47) continues to hold for $p=0$ with $\mathbf{E}_1 = \sigma^2\mathbf{I}$.

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. For such a model we define $\hat{Y}$ to be the best linear time invariant predictor (BLIP) of Y if the u_s are chosen to satisfy (46) and the mean squared error criterion $E\{(Y-\hat{Y})^2\}$ is a minimum. Thus 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 8

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 5. The $n \times p$ matrix $\mathbf{C}_{p-1}$ and the n 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 (17) 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.

Proof

The proof of this result proceeds in much the same way as Theorem 5. Consider first the case $p>0$. Using Lagrange multipliers we optimise (47) subject to (46) and obtain

$$\begin{aligned} \mathbf{v} &= \tilde{\mathbf{E}}_1^{-1}\mathbf{C}_{p-1}\mu + \tilde{\mathbf{E}}_1^{-1}\mathbf{L}_2\nu \\ \mathbf{L}_2^T\tilde{\mathbf{E}}_1^{-1}\mathbf{C}_{p-1}\mu + \mathbf{L}_2^T\tilde{\mathbf{E}}_1^{-1}\mathbf{L}_2\nu &= \mathbf{L}_2^T\delta \\ \mathbf{C}_{p-1}^T\tilde{\mathbf{E}}_1^{-1}\mathbf{C}_{p-1}\mu + \mathbf{C}_{p-1}^T\tilde{\mathbf{E}}_1^{-1}\mathbf{L}_2\nu &= \mathbf{0} \end{aligned}$$

where μ, ν are the vectors of Lagrange multipliers. Solving these equations together with (39) yields the stated result. The case $p = 0$ follows similarly.

Setting $\delta = \mathbf{0}$ with the exception of the s-th element which is set to unity, we see that the BLIP of y_{t+s} is

$$\hat{y}_{t+s} = \begin{cases} y_{t+s} & (-r \leq s \leq q) \\ \tilde{\mathbf{h}}_s^T \mathbf{y} & (q < s \leq r) \end{cases}$$

where $\tilde{\mathbf{h}}_s$ is the s-th column of $-\mathbf{L}_1^T \tilde{\mathbf{G}} \mathbf{L}_2 (\mathbf{L}_2^T \tilde{\mathbf{G}} \mathbf{L}_2)^{-1} \mathbf{L}_2^T$ and $\mathbf{y} = (y_{t-r}, \ldots, y_{t+q})^T$. Thus, for arbitrary choice of δ,

$$\sum_{s=-r}^{q} u_s y_{t+s} = \delta^T (\mathbf{I} - \tilde{\mathbf{G}} \mathbf{L}_2 (\mathbf{L}_2^T \tilde{\mathbf{G}} \mathbf{L}_2)^{-1} \mathbf{L}_2^T)^T \mathbf{L}_1 \mathbf{y} = \sum_{s=-r}^{r} \delta_s \hat{y}_{t+s}$$

as required. □

Note that the BLIP predictor given by Theorem 8 has the form of a shrinkage estimator since it is exactly the same as the BLUP predictor given by Theorem 5, 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 (33) 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 consider predicting $\sum_{s=-r}^{r} w_s y_{t+s}$ by a linear predictor $\tilde{g}_t$ of the form (32) subject to the requirement that the bias of $\tilde{g}_t$ is time invariant. The latter quantity is given by (44) with $w_s = \delta_s$ $(-r \leq s \leq r)$. Thus, from (46) and (47), $\tilde{g}_t$ is a linear time invariant predictor of $\sum_{s=-r}^{r} w_s y_{t+s}$ if

$$\sum_{s=-r}^{q} u_s = 1, \qquad \sum_{s=-r}^{q} s^j u_s = 0 \qquad (0 < j < p) \tag{48}$$

when $p > 0$ and, for $p \geq 0$,

$$R_q = \beta_p^2 (\sum_{s=-r}^{q} s^p v_s)^2 + \mathbf{v}^T \mathbf{E}_1 \mathbf{v} \tag{49}$$

with

$$\beta_p^2 (\sum_{s=-r}^{q} s^p v_s)^2 = \begin{cases} \beta_p^2 (\sum_{s=-r}^{q} s^p u_s)^2 & (p > 0) \\ \beta_0^2 (1 - \sum_{s=-r}^{q} u_s)^2 & (p = 0) \end{cases}.$$

Here $\mathbf{v}$ is given by (36) and R_q is also time invariant. This leads to the following result.

Corollary 9 *Let y_t follow the restricted local dynamic model specified in Theorem 8 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 a linear predictor of $\sum_{s=-r}^{r} w_s y_{t+s}$ with time invariant bias so that the u_s satisfy (48) when $p > 0$. Then, for $0 \leq q < r$, the values of u_s that minimise R_q subject to these conditions are given by Theorem 8 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 $q < s \leq r$ and y_{t+s} otherwise.

Proof

Since the u_s and w_s satisfy (46) when $p > 0$, the result is a direct application of Theorem 8. □

Because of their dependence on BLIP predictors we shall henceforth refer to the end filters specified by Corollary 9 as *BLIP end filters.* Their properties are investigated in Section 3.2.

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 6, the BLIP end filters specified by Corollary 9 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 10 and also Sections 3.2.2 and 4.1 where the properties of these end filters and the estimation of β_p^2/σ^2 are discussed.

Note that Corollary 9 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 9 provides a generalisation and extension of the current X-11 end filters to any of the local dynamic models specified in Section 2.1 provided the levels of integration of the random walk components that make up ξ_t do not exceed p.

The following result considers an alternative form of the BLIP end filters given by Corollary 9 which explicitly builds on the corresponding BLUP end filters given by Corollary 6. In addition the result provides a means of exploring the effects of mis-specification of β_p^2/σ^2.

Theorem 10

Let y_t follow the restricted local dynamic model specified in Theorem 8 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 $0 \leq q < r$, let $\tilde{g}_t(\phi) = \sum_{s=-r}^{q} u_s(\phi) y_{t+s}$ 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 5, and the $p+1$ dimensional vector $\mathbf{d}$ is zero save for the last element which is unity. Then

(a) $\tilde{g}_t(\phi)$ 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 (48) 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 6 and Corollary 9 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 $R_q(\phi)$.

Proof

First observe from Theorem 5 that $\tilde{g}_t(\phi)$ is the BLUP of $\sum_{s=-r}^{r}(w_s + \phi\gamma_s)y_{t+s}$ and $\mathbf{C}^T\gamma = \mathbf{d}$ so that

$$\sum_{s=-r}^{q} s^p\gamma_s = 1, \qquad \sum_{s=-r}^{q} s^j\gamma_s = 0 \qquad (0 \leq j < p).$$

Thus, from (38),

$$\sum_{s=-r}^{q} s^j u_s(\phi) = \sum_{s=-r}^{r} s^j(w_s + \phi\gamma_s) = \begin{cases} 1 & (j = 0) \\ 0 & (0 < j < p) \\ \phi & (j = p) \end{cases}$$

when $p > 0$ and $\sum_{s=-r}^{q} u_s(\phi) = 1 + \phi$ when $p = 0$. Moreover

$$E\{\sum_{s=-r}^{r} w_s y_{t+s} - \tilde{g}_t(\phi)\} = -\phi\sum_{s=-r}^{r}\gamma_s E(y_{t+s}) = -\phi\beta_p$$

since $\mathbf{C}^T\gamma = \mathbf{d}$. This establishes (a).

Now, from (49) and (36),

$$R_q(\phi) = \beta_p^2\phi^2 + \mathbf{v}^T\mathbf{E}_1\mathbf{v}$$

where

$$\mathbf{v} = \mathbf{w} - \mathbf{L}_1\mathbf{u}(\phi).$$

Noting that $\mathbf{L}_2^T(\mathbf{I}-\mathbf{GH})=\mathbf{0}$ we can write $\mathbf{v}$ as

$$\begin{aligned}\mathbf{v} &= \mathbf{w}-(\mathbf{I}-\mathbf{GH})(\mathbf{w}+\phi\gamma) \\ &= \mathbf{GHw}-\phi(\mathbf{I}-\mathbf{GH})\gamma \qquad (50)\end{aligned}$$

and $R_q(\phi)$ becomes

$$\begin{aligned}R_q(\phi) &= \mathbf{w}^T\mathbf{HGE}_1\mathbf{GHw}-2\phi\gamma^T(\mathbf{I}-\mathbf{HG})\mathbf{E}_1\mathbf{GHw} \qquad (51)\\ &\quad +\phi^2(\beta_p^2+\gamma^T(\mathbf{I}-\mathbf{HG})\mathbf{E}_1(\mathbf{I}-\mathbf{GH})\gamma^T). \qquad (52)\end{aligned}$$

Since $\mathbf{GE}_1\mathbf{G}=\mathbf{G}$, $\mathbf{HGH}=\mathbf{H}$ and $\gamma^T\mathbf{E}_1\mathbf{G}=\mathbf{0}$, the above now reduces to the expression for $R_q(\phi)$ given by (b).

From the form of $\mathbf{u}(\phi)$ it follows directly that $\mathbf{u}(0)$ is the optimal end filter given by Corollary 6. All that remains to prove is that $\mathbf{u}(\phi_0)$ is the optimal end filter given by Corollary 9. Consider again minimising (49) with respect to the u_s where now $\mathbf{v}$ is defined by (36) and the u_s satisfy (48) if $p>0$. This is equivalent to optimising

$$\tilde{R}_q=\beta_p^2\phi^2+\mathbf{v}^T\mathbf{E}_1\mathbf{v}-2\mu^T(\mathbf{C}^T\mathbf{v}+\phi\mathbf{d})-2\nu^T\mathbf{L}_2^T(\mathbf{v}-\mathbf{w})$$

with respect to $\mathbf{v}$, ϕ and the Lagrange multipliers μ, ν. Here ϕ is a function of $\mathbf{v}$ defined by

$$\phi=-\sum_{s=-r}^{r}s^p v_s=\begin{cases}\sum_{s=-r}^{q}s^p u_s & (p>0)\\ \sum_{s=-r}^{q}u_s-1 & (p=0)\end{cases}. \qquad (53)$$

where the latter equality follows from (7) and (36). Optimising $\tilde{R}_q$ first with respect to $\mathbf{v}$, μ and ν we obtain the equations

$$\begin{aligned}\mathbf{v} &= \mathbf{E}_1^{-1}\mathbf{C}\mu+\mathbf{E}_1^{-1}\mathbf{L}_2\nu\\ \mathbf{L}_2^T\mathbf{E}_1^{-1}\mathbf{C}\mu+\mathbf{L}_2^T\mathbf{E}_1^{-1}\mathbf{L}_2\nu &= \mathbf{L}_2^T\mathbf{w}\\ \mathbf{C}^T\mathbf{E}_1^{-1}\mathbf{C}\mu+\mathbf{C}^T\mathbf{E}_1^{-1}\mathbf{L}_2\nu &= -\phi\mathbf{d}\end{aligned}$$

and these, together with (36), yield $\mathbf{u}=\mathbf{u}(\phi)$. Substituting this solution back into $\tilde{R}_q$ gives $R_q(\phi)$ which must now be optimised with respect to ϕ. Since $R_q(\phi)$ is a quadratic in ϕ and takes its minimum value when $\phi=\phi_0$ result (c) follows. □

An immediate consequence of Theorem 10 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

$$R_q(\phi_0)=\mathbf{w}^T\mathbf{Hw}-\frac{(\gamma^T\mathbf{Hw})^2}{\beta_p^2+\mathbf{d}^T(\mathbf{C}^T\mathbf{E}_1^{-1}\mathbf{C})^{-1}\mathbf{d}+\gamma^T\mathbf{H}\gamma}=R_q(0)-|\phi_0)||\gamma^T\mathbf{Hw}|$$

so that $R_q(\phi_0)$ is a monotonically increasing function of β_p^2/σ^2 and a linearly decreasing function in $|\phi_0|$. In particular

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

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 8, the BLIP end filters become BLUP end filters for the local dynamic model with order $p-1$, but the same stochastic structure.

From the representation (50) and the properties of G and H, observe in passing that

$$\text{cov}\{\sum_{s=-r}^{r} w_s y_{t+s} - \tilde{g}_t(0), \sum_{s=-r}^{r} w_s y_{t+s} - \tilde{g}_t(\phi_0)\} = \mathbf{w}^T\mathbf{H}\mathbf{w} + \phi_0\gamma^T\mathbf{H}\mathbf{w} = R_q(\phi_0)$$

so that the normalised quantity

$$\frac{R_q(\phi_0)}{R_q(0)} = 1 - |\phi_0|\frac{|\gamma^T\mathbf{H}\mathbf{w}|}{\mathbf{w}^T\mathbf{H}\mathbf{w}}$$

represents the regression coefficient of the BLIP revisions $\sum_{s=-r}^{r} w_s y_{t+s} - \tilde{g}_t(\phi_0)$ on the BLUP revisions $\sum_{s=-r}^{r} w_s y_{t+s} - \tilde{g}_t(0)$. Moreover, as noted in the proof to Theorem 10, $\tilde{g}_t(\phi)$ is the BLUP of $\sum_{s=-r}^{r} w_s y_{t+s} + \phi\sum_{s=-r}^{r}\gamma_s y_{t+s}$. Here $\sum_{s=-r}^{r}\gamma_s y_{t+s}$ is the best linear unbiased estimator (BLUE) of β_p given $y_{t-r}, \ldots, y_{t+r}$.

Both the BLIP and BLUP end filters are dependent on the global parameters specified by p, n and the model for ξ_t/σ. 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 10 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, so that $\hat{\phi}_0$ lies between 0 and $-\gamma^T\mathbf{H}\mathbf{w}/(\mathbf{d}^T(\mathbf{C}^T\mathbf{E}_1^{-1}\mathbf{C})^{-1}\mathbf{d} + \gamma^T\mathbf{H}\gamma)$. 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

$$\frac{\beta_p^2}{\sigma^2} < 2\frac{\hat{\beta}_p^2}{\hat{\sigma}^2} + (\mathbf{d}^T(\mathbf{C}^T\mathbf{E}_1^{-1}\mathbf{C})^{-1}\mathbf{d} + \gamma^T\mathbf{H}\gamma)/\sigma^2. \tag{54}$$

In particular, if

$$\frac{\beta_p^2}{\sigma^2} < (\mathbf{d}^T(\mathbf{C}^T\mathbf{E}_1^{-1}\mathbf{C})^{-1}\mathbf{d} + \gamma^T\mathbf{H}\gamma)/\sigma^2,$$

then any value of $\hat{\beta}_p^2/\hat{\sigma}^2$ will yield BLIP end filters which have better mean squared revisions than the corresponding BLUP end filters.

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, over-estimate 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

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

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

Further discussion on the properties of these end filters is given in Section 3.2.

3 Properties of the filters

In this section we focus attention on two particular local dynamic models likely to be used in practice. Preliminary analysis indicates that these models have properties that can be regarded as representative of other more general models of the type discussed in Section 2.1. The models considered for the window are the *local linear model* (p =1) given by

$$y_t = g_t + \epsilon_t = \beta_0 + \beta_1 t + \xi_t + \epsilon_t \tag{55}$$

and the *local quadratic model* (p =2) given by

$$y_t = g_t + \epsilon_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \xi_t + \epsilon_t. \tag{56}$$

In both cases ξ_t is a simple random walk satisfying

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

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

Unless otherwise stated, we also restrict attention to the case where the window is of length 13 so that $r = 6$.

3.1 In the body

The impulse response functions of the filters given by Theorem 1 for the local linear model (55) and the local quadratic model (56) are plotted in Figure 1 for selected values of θ and λ. Note that the 13 point X-11 Henderson filter corresponds to the case where $\theta = 0$, $\lambda = 0$, $p = 2$ and is shown in the top right hand plot of Figure 1.

As λ varies the Henderson filters that optimise smoothness S ($\theta = 0$) differ only slightly whereas the Macaulay filters that optimise fidelity F ($\theta = 1$) show much greater variation. In particular the Macaulay filters become more adaptive as λ increases with the weights at central lags being progressively increased at the expense of those for extreme lags. The mixture of smoothness and fidelity corresponding to $\theta = 0.5$ yields a compromise between the Henderson and Macaulay filters as expected.

The optimal impulse response function given by Theorem 1 converges to a limit $\bar{\mathbf{w}}$ as λ increases where $\bar{\mathbf{w}}$ is given by Theorem A2 in Appendix A.2. This limit corresponds to the case where $\sigma^2 = 0$ or, in practice, where σ_η^2 is very much larger than σ^2. In such cases the model has no noise component so that y_t is just the trend g_t. The limiting impulse response functions are plotted in Figure 2.

Thus, as λ increases and for given θ, the impulse response functions of the filters given by Theorem 1 smoothly vary from a mixture of Henderson and Macaulay filters through to the impulse response functions of the filters specified by Theorem A2.

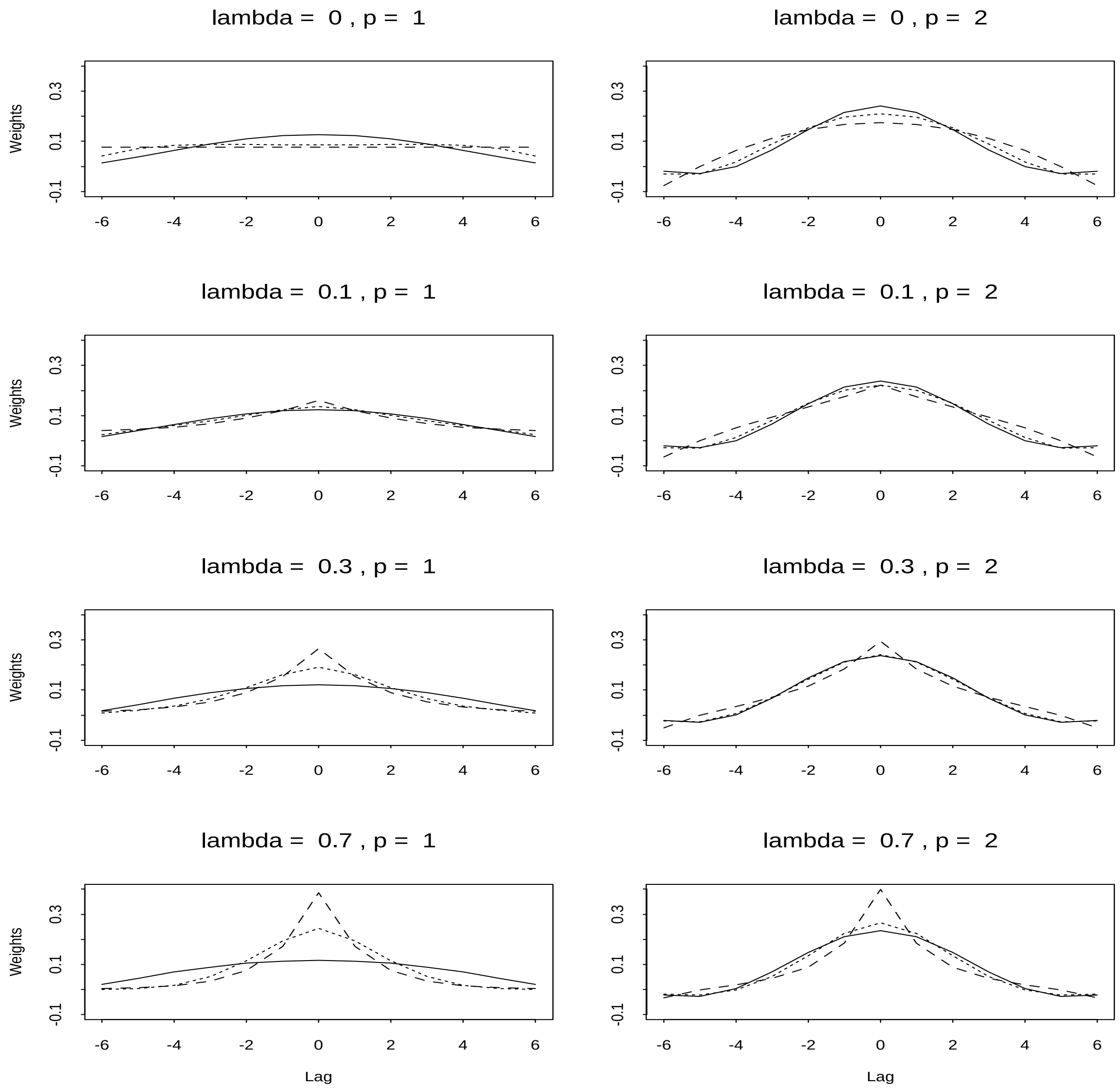


Figure 1: Plots of the impulse response functions of the 13 point filters given by Theorem 1 for the specified local linear and quadratic models and selected values of θ and λ. The solid lines correspond to $\theta = 0$ (smoothness criterion only), the dashed lines to $\theta = 1$ (fidelity criterion only) and the dotted lines to $\theta = 0.5$ which gives a comprise between smoothness and fidelity. The 13 point X-11 Henderson filter corresponds to the case where $\theta = 0$, $\lambda = 0$, $p = 2$ and is shown in the top right hand plot.

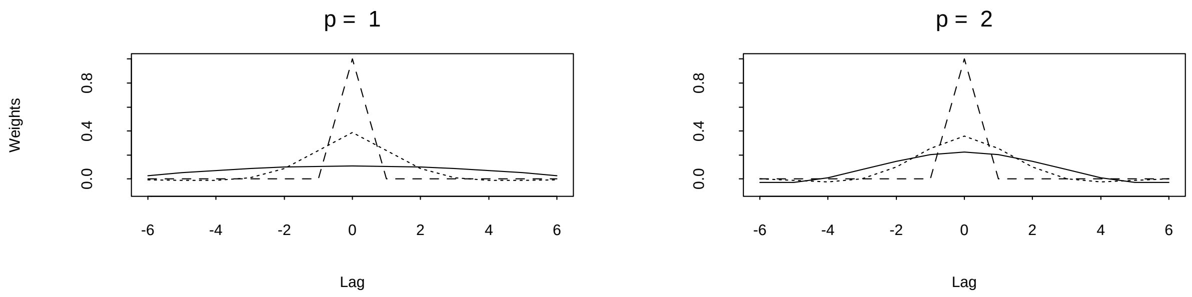


Figure 2: Plots of the limiting ($\lambda = \infty$) impulse response functions of the 13 point filters given by Theorem 1 for the specified local linear and quadratic models and selected values of θ. The solid lines correspond to $\theta = 0$ (smoothness criterion only), the dashed lines to $\theta = 1$ (fidelity criterion only) and the dotted lines to $\theta = 0.5$ which gives a comprise between smoothness and fidelity.

Figure 3 shows the gain functions of the filters given by Theorem 1 for the local linear model (55) and the local quadratic model (56) for selected values of θ and λ. The gain function of the 13 point X-11 Henderson is shown in the top right hand plot of Figure 3 and corresponds to the case where $\theta = 0$, $\lambda = 0$ and $p = 2$.

The broader pass band in the case of $p = 2$ follows from the fact that any linear filter constrained to pass a polynomial of degree p has a gain function whose first p derivatives are zero at the origin. (See Brillinger (1965).) This leads to greater restrictions on the curvature of the gain function at the origin as p increases.

Note that the Henderson filters that optimise smoothness S ($\theta = 0$) show little leakage over the frequency band [0.1,0.5] when $p = 1$ and [0.2,0.5] when $p = 2$. By comparison the Macaulay filters that optimise fidelity F ($\theta = 1$) have significant side lobes in these frequency bands. Clearly optimal fidelity comes at a price. Again the mixture corresponding to $\theta = 0.5$ has the expected compromise effect.

The phase functions of the filters were also examined for a variety of values of λ. It would appear that the Henderson and mixture filters in general together with the Macaulay filters for $\lambda = 0$ are not positive definite. However, for $\lambda > 0$, the Macaulay filters are frequently positive definite and this improvement in phase characteristics carries through in part to the mixture filters.

For the purposes of graphical display we now consider alternative measures of smoothness and fidelity based on F^* and S^* given by (8), (19) and (10). Recall that F^* and S^* measure the gains in fidelity and smoothness of the trend estimate $\hat{g}_t$ relative to the fidelity and smoothness of the data. Thus, given a vector of central filter weights $\mathbf{w}$ and a local dynamic

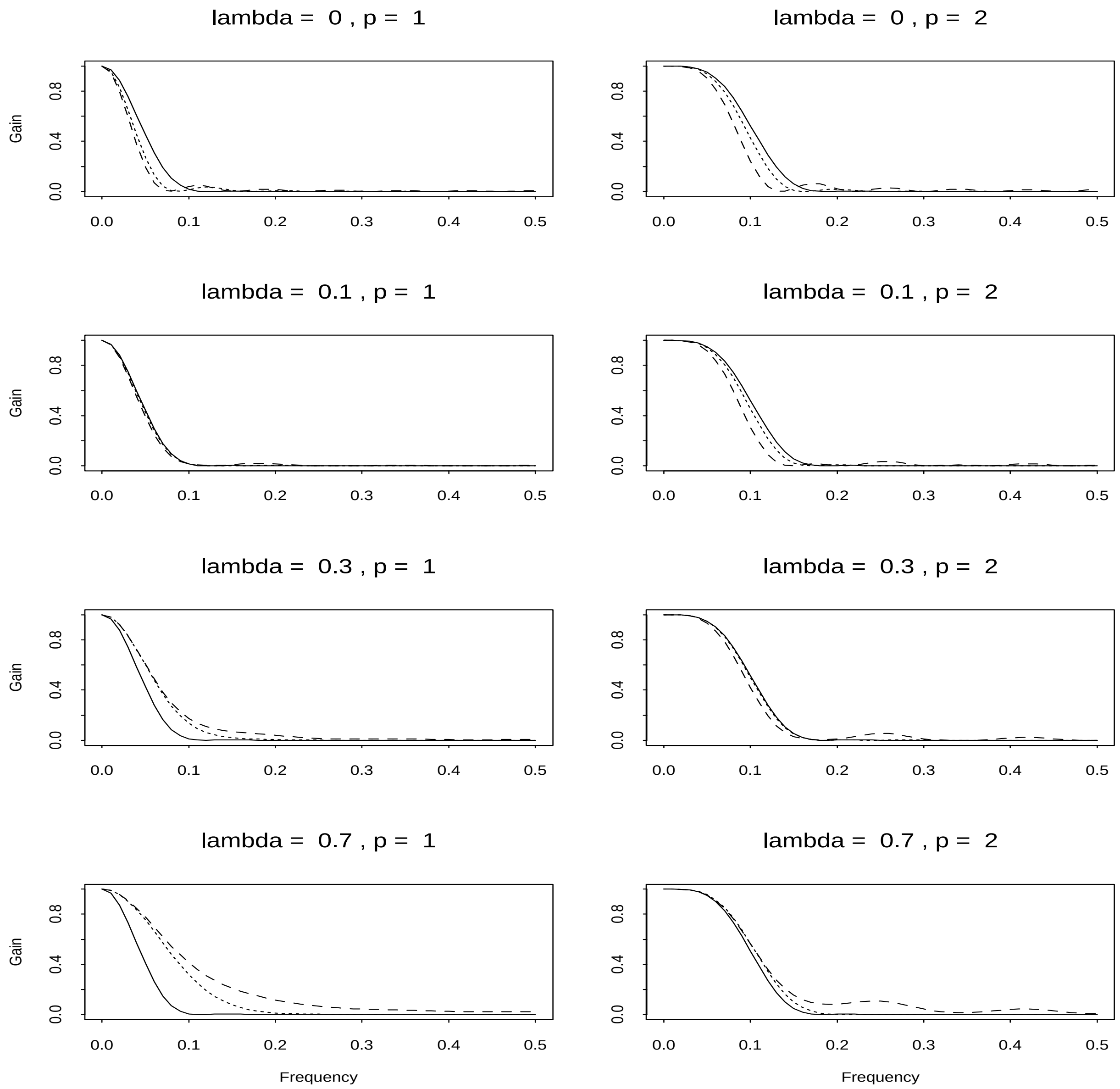


Figure 3: Plots of the gain functions of the 13 point filters given by Theorem 1 for the specified local linear and quadratic models and selected values of θ and λ. The solid lines correspond to $\theta = 0$ (smoothness criterion only), the dashed lines to $\theta = 1$ (fidelity criterion only) and the dotted lines to $\theta = 0.5$ which gives a comprise between smoothness and fidelity. The gain function of the 13 point X-11 Henderson filter is shown in the top right hand plot of Figure 3 and corresponds to the case where $\theta = 0$, $\lambda = 0$ and $p = 2$.

model, F^* and S^* are given by the quadratic forms

$$F^* = \mathbf{w}^T(\sigma^2\mathbf{I} + \mathbf{\Omega})\mathbf{w}/\sigma^2 \qquad S^* = \mathbf{w}^T(\sigma^2\mathbf{B}_{p+1} + \mathbf{\Gamma})\mathbf{w}/S_0 \tag{57}$$

where $S_0 = E\{(\Delta^{p+1}y_t)^2\}$. Since the local models considered in this section have fixed window length $n = 13$ and are specified by p and λ, we denote the values of F^* and S^* evaluated at the filter specified by Theorem 1 as $F^*(\theta|p,\lambda)$ and $S^*(\theta|p,\lambda)$ respectively where the notation reflects the fact that θ is user specified.

To compare gains in fidelity and smoothness across different values of θ and different models indexed by λ, we consider the efficiency measures

$$F_{eff}(\theta|p,\lambda) = \frac{F^*(1|p,0)}{F^*(\theta|p,\lambda)}, \qquad S_{eff}(\theta|p,\lambda) = \frac{S^*(0|p,0)}{S^*(\theta|p,\lambda)}. \tag{58}$$

Here F_{eff} measures the efficiency of the trend estimate $\hat{g}_t$ in terms of its fidelity gains relative to those of the Macaulay filter with $\theta = 1$, $\lambda = 0$. Similarly S_{eff} measures the efficiency of $\hat{g}_t$ in terms of its smoothness gains relative to those of the Henderson filter with $\theta = 0$, $\lambda = 0$. The reference Macaulay ($\theta = 1, \lambda = 0$) and Henderson filters ($\theta = 0, \lambda = 0$) are used as convenient, but somewhat arbitrary, benchmarks.

In practice we will typically have only one model available. Thus we also consider the following relative efficiency measures which compare gains in fidelity and smoothness across different values of θ for the same model. These measures are defined as

$$F_{rel}(\theta|p,\lambda) = \frac{F^*(1|p,\lambda)}{F^*(\theta|p,\lambda)}, \qquad S_{rel}(\theta|p,\lambda) = \frac{S^*(0|p,\lambda)}{S^*(\theta|p,\lambda)}. \tag{59}$$

In this case F_{rel} measures the efficiency of the trend estimate $\hat{g}_t$ in terms of its fidelity gains relative to those of the optimal fidelity or Macaulay filter for the model indexed by λ. Similarly S_{rel} measures the efficiency of $\hat{g}_t$ in terms of its smoothness gains relative to the optimal smoothness or Henderson filter for the same model.

Perspective plots of F_{eff} and S_{eff} are given in Figure 4. For given θ greater than 0.1 ($p = 1$) or 0.3 ($p = 2$) smoothness efficiency improves as λ increases from 0 to around 0.1 ($p = 1$) or 0.2 ($p = 2$), and then deteriorates thereafter as λ increases. Thus a modest amount of positive correlation in y_t allows for greater gains in the smoothness of the trend estimate. However fidelity efficiency appears to uniformly deteriorate as λ increases.

Perspective plots of F_{rel} and S_{rel} are given in Figure 5. For both $p = 1$ and $p = 2$ relative fidelity efficiency is less variable than relative smoothness efficiency. In particular, relative fidelity efficiency exceeds 90%, approximately, for $\theta > 0.3$. In general relative smoothness efficiency decreases more rapidly than relative fidelity efficiency as θ increases with the effect more marked for $p = 1$ than $p = 2$. For $\theta < 0.3$ relative smoothness efficiency exceeds 80% when λ is near 0.1 ($p = 1$) and for all values of λ considered ($p = 2$). A more detailed picture of the trade-off between relative fidelity efficiency and relative smoothness efficiency for given θ and selected values of λ is illustrated in Figure 6.

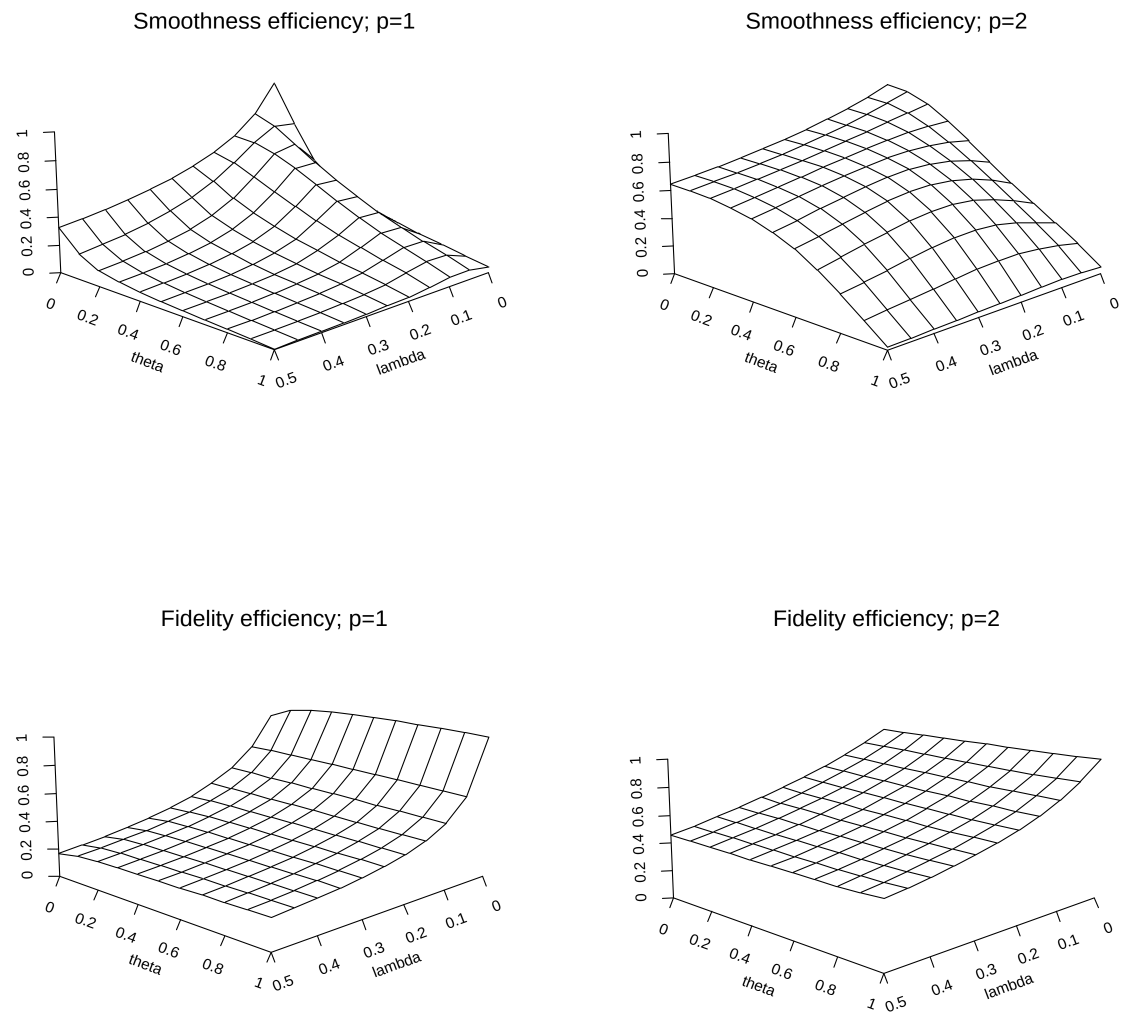


Figure 4: Perspective plots of the fidelity efficiency F_{eff} and the smoothness efficiency S_{eff} for the 13 point filters given by Theorem 1 and the specified local linear and quadratic models. Here F_{eff} measures the efficiency of the trend estimate $\hat{g}_t$ in terms of its fidelity gains relative to those of the Macaulay filter with $\theta = 1$, $\lambda = 0$. Similarly S_{eff} measures the efficiency of $\hat{g}_t$ in terms of its smoothness gains relative to those of the Henderson filter with $\theta = 0$, $\lambda = 0$. .

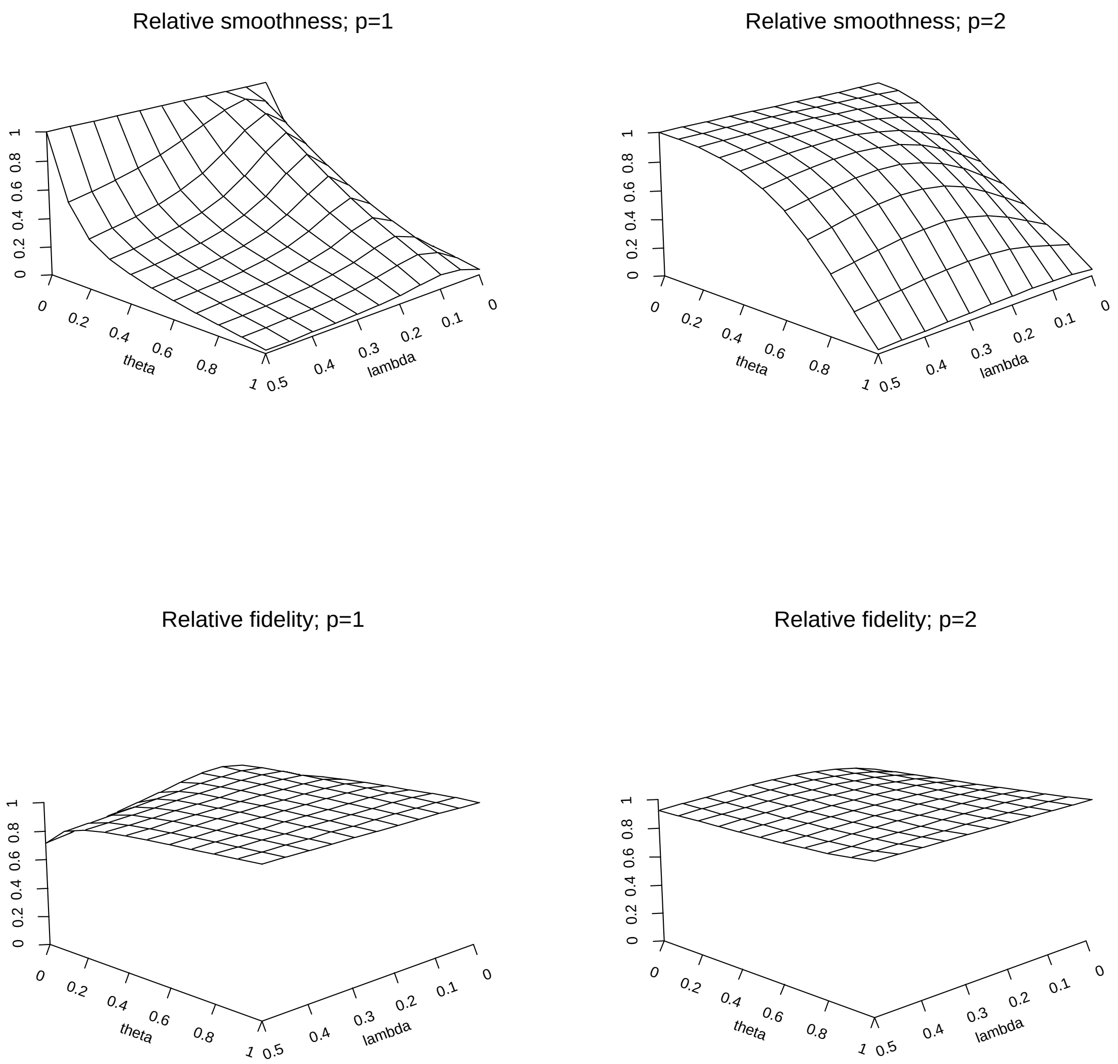


Figure 5: Perspective plots of the fidelity relative efficiency F_{rel} and the smoothness relative efficiency S_{rel} for the 13 point filters given by Theorem 1 and the specified local linear and quadratic models. Here F_{rel} measures the efficiency of the trend estimate $\hat{g}_t$ in terms of its fidelity gains relative to those of the optimal fidelity or Macaulay filter for the model indexed by λ. Similarly S_{rel} measures the efficiency of $\hat{g}_t$ in terms of its smoothness gains relative to the optimal smoothness or Henderson filter for the same model.

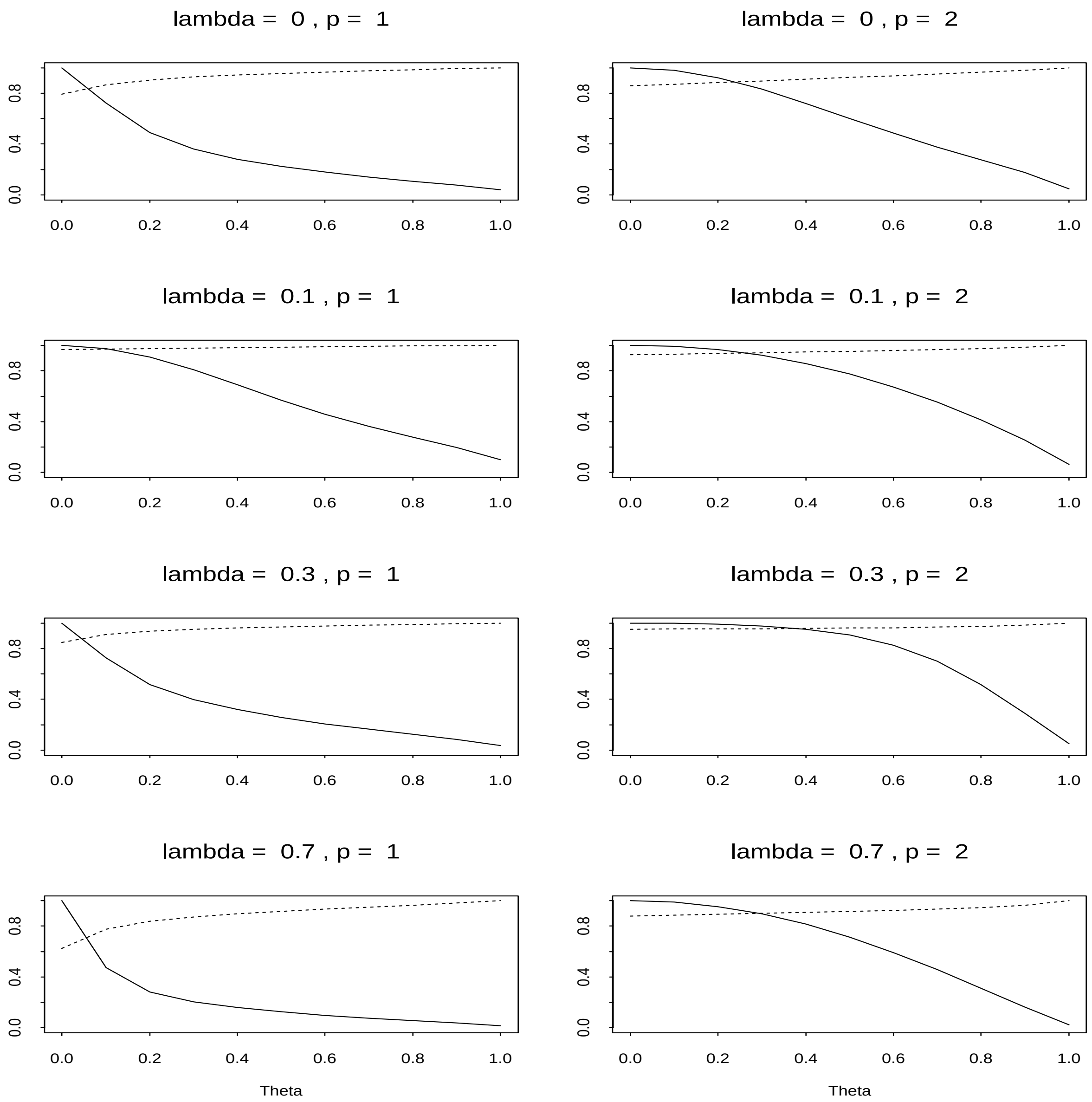


Figure 6: Plots of the fidelity relative efficiency F_{rel} and the smoothness relative efficiency S_{rel} of the 13 point filters given by Theorem 1 for the specified local linear and quadratic models, and selected values of θ and λ. Here F_{rel} (dashed lines) measures the efficiency of the trend estimate $\hat{g}_t$ in terms of its fidelity gains relative to those of the optimal fidelity or Macaulay filter for the model indexed by λ. Similarly S_{rel} (solid lines) measures the efficiency of $\hat{g}_t$ in terms of its smoothness gains relative to the optimal smoothness or Henderson filter for the same model. The plots illustrate the trade-off between relative fidelity efficiency and relative smoothness efficiency for given θ.

Finally, we consider the performance of the X-11 Henderson filter ($\lambda = 0$, $p = 2$) in terms of both fidelity and smoothness. The fidelity F^* and smoothness S^* of the X-11 Henderson filter are given by (57) for each of the local models $p = 1$ and $p = 2$ considered and with $\mathbf{w}$ containing the vector of X-11 Henderson filter weights. Plots of these measures relative to the optimal $S^*(0|p,\lambda)$ and $F^*(1|p,\lambda)$ are given in Figure 7. Note that $S^*(0|p,\lambda)$ gives the smoothness of the optimal Henderson filter for the particular local dynamic model concerned and $F^*(1|p,\lambda)$ gives the fidelity of the optimal Macaulay filter for the same model.

In terms of smoothness, it is clear in the case $p = 2$ that there is little to differentiate the X-11 Henderson ($\lambda = 0$) from its generalisation based on the local dynamic model where $\lambda > 0$. However, in the case $p = 1$, the smoothness of the X-11 Henderson filter is substantially less than that of the Henderson filter tailored for the local linear model concerned. This is not entirely unexpected since the conventional Henderson filter is designed to accommodate quadratic rather than linear trends and so its trends exhibit greater variation in curvature.

In terms of fidelity, the X-11 Henderson filter performs poorly when λ is less than 0.1 and $p = 1$, but otherwise its relative fidelity exceeds 80% for λ in excess of around 0.2 ($p = 1$) and for all λ considered ($p = 2$).

3.2 At the ends

This section considers the properties of the end filters specified by Corollary 6 and Corollary 9 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. As stated previously, 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.

In parallel with Section 2.3, the following subsections address the two cases where the end filters are based on unbiased and biased predictors respectively.

3.2.1 Unbiased predictors

Here we consider the properties of the BLUP end filters based on Corollary 6 and specified central filter weights w_s.

Examples of BLUP end filters based on the central filters given by Theorem 1 and the local dynamic models (55) and (56) are given in Figures 8 and 9 for the cases $q = 3$, $q = 0$ and selected values of λ. In much the same way as the central filters that they are based on, the BLUP end filters for the central Henderson filters that optimise smoothness ($\theta = 0$) differ only slightly whereas those for the central Macaulay filters that optimise fidelity ($\theta = 1$) show greater variability especially for the case $q = 3$. In the latter case note that, as λ increases, progressively more weight is placed on the observation at the origin of the

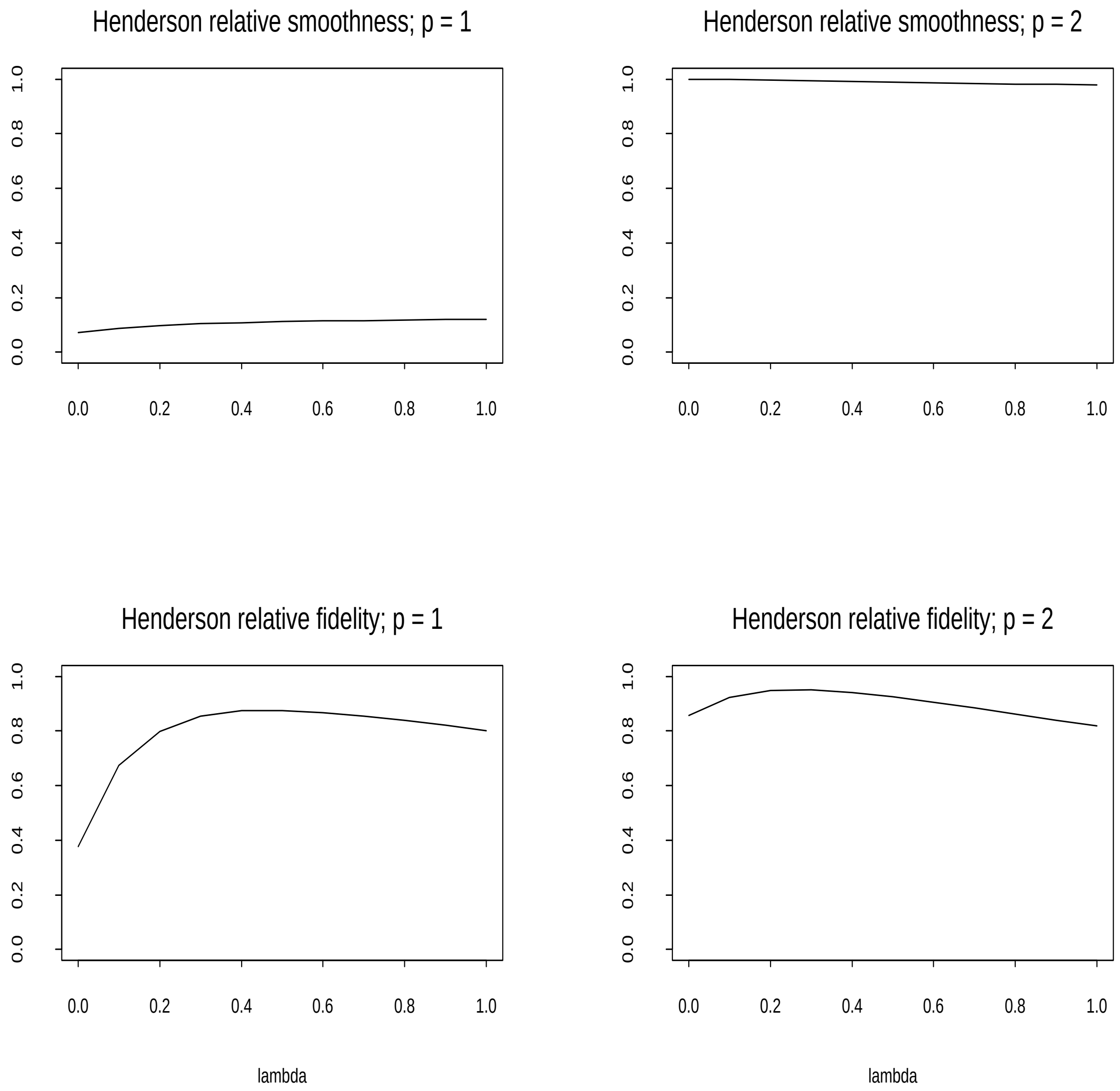


Figure 7: Plots of the fidelity and smoothness of the X-11 Henderson 13 point filter relative to the optimal fidelity and smoothness filters respectively for the specified local dynamic models considered. The latter are indexed by p and λ. Note that the smoothness of the X-11 Henderson filter is close to 100% in the case of $p = 2$ as expected, but is far from being smooth by comparison to the optimal smoothness filter tailored for the local linear model $p = 1$. For λ in excess of around 0.2 ($p = 1$) and for all λ ($p = 2$), the relative fidelity of the X-11 Henderson filter exceeds 80%.

window so that the end filters become more adaptive.

For any given central filter with weights given by $\mathbf{w}$, the BLUP end filters converge as λ increases to a limit specified by Corollary 6 with $\mathbf{G}$ replaced by $\bar{\mathbf{G}}$. Here

$$\lim_{\lambda\to\infty} \lambda\mathbf{G} = \bar{\mathbf{G}}$$

is given by (71) of Appendix A.2 where the proof of this result is also given. The limiting impulse response functions of the BLUP end filters for $q = 3$ and $q = 0$ are plotted in Figure 10. These are based on the limiting central filters of Theorem 1 given by Theorem A2. Note that, as in the case of the limiting central filters, the pure fidelity case ($\theta = 1$) yields a limiting end filter with all its weight concentrated at the origin of the window so that $\tilde{g}_t = y_t$.

Now consider the mean squared revisions criterion R_q given by (35) evaluated for the BLUP end filters based on specified central filters.

Figure 11 shows R_q/σ^2 for the BLUP end filters based on the optimum central filters given by Theorem 1 and the local dynamic models (55) and (56). For comparative purposes and because of its central role within X-11, Figure 11 also includes a plot of R_q/σ^2 for the BLUP end filters based on the central X-11 Henderson filter. 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 typically negligible by comparison.

Consider the plots of R_q/σ^2 for the BLUP end filters based on the optimum central filters of Theorem 1. Note that individual plots are based on different central filters which are shown in Figure 1. For local dynamic models with λ less than 0.1 ($p = 1$) and 0.3 ($p = 2$), the BLUP end filters based on central Macaulay filters with optimal fidelity ($\theta = 1$) show higher mean squared revisions than those based on central Henderson filters with optimal smoothness ($\theta = 0$). Given the comments made in Section 3.1 concerning the properties of the central filters on which these end filters are based, it would appear that the use of central Henderson filters is favoured in this case.

For values of $\lambda \geq 0.1$ ($p = 1$) and $\lambda \geq 0.3$ ($p = 2$) the situation is reversed with the BLUP end filters based on the central Henderson filters ($\theta = 0$) showing higher (in some cases significantly higher) mean squared revisions than those based on the central Macaulay filters ($\theta = 1$). Thus, if y_t exhibits significant local positive correlation (λ sufficiently large), it would appear that there is a need for more adaptive end filters whose central filters have good fidelity properties if revisions are to be minimised.

As a general observation, it would appear from Figure 11 that the mean squared revisions R_q/σ^2 increase as λ increases for the BLUP end filters based on central Henderson filters ($\theta = 0$), and decrease as λ increases for the BLUP end filters based on central Macaulay filters ($\theta = 1$). Moreover, although smoothness plays the influential role in the body (see Figures 5 and 6), it would appear that it is fidelity that plays the dominant role in minimising expected revisions at the ends.

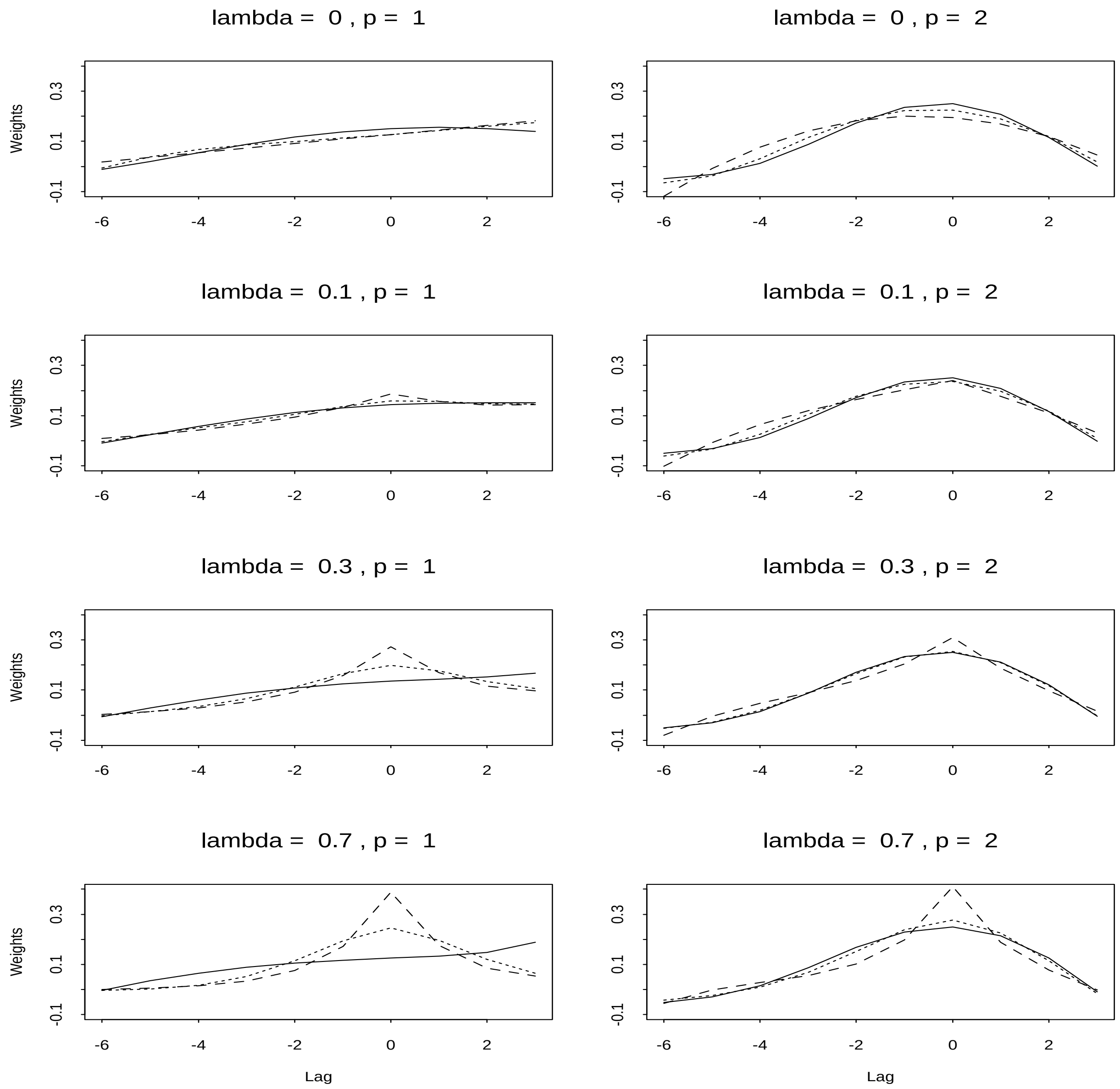


Figure 8: Plots of the impulse response functions of the BLUP end filters for $q = 3$ based on the 13 point central filters given by Theorem 1. These end filters are evaluated for selected values of θ and the local linear and quadratic models specified by p and λ. The solid lines correspond to $\theta = 0$ (smoothness criterion only), the dashed lines to $\theta = 1$ (fidelity criterion only) and the dotted lines to $\theta = 0.5$ which gives a compromise between smoothness and fidelity.

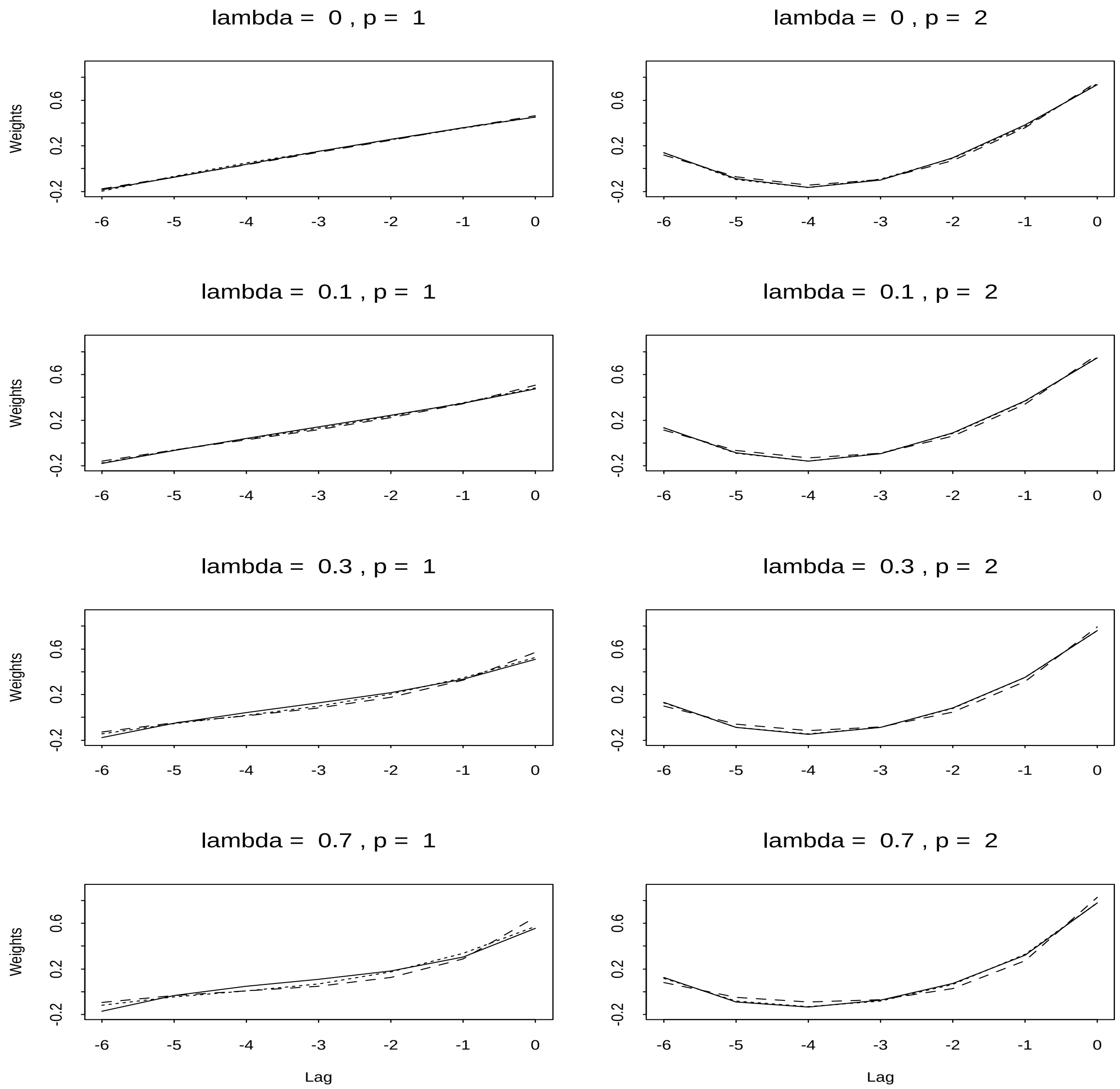


Figure 9: Plots of the impulse response functions of the BLUP end filters for $q = 0$ based on the 13 point central filters given by Theorem 1. These end filters are evaluated for selected values of θ and the local linear and quadratic models specified by p and λ. The solid lines correspond to $\theta = 0$ (smoothness criterion only), the dashed lines to $\theta = 1$ (fidelity criterion only) and the dotted lines to $\theta = 0.5$ which gives a compromise between smoothness and fidelity.

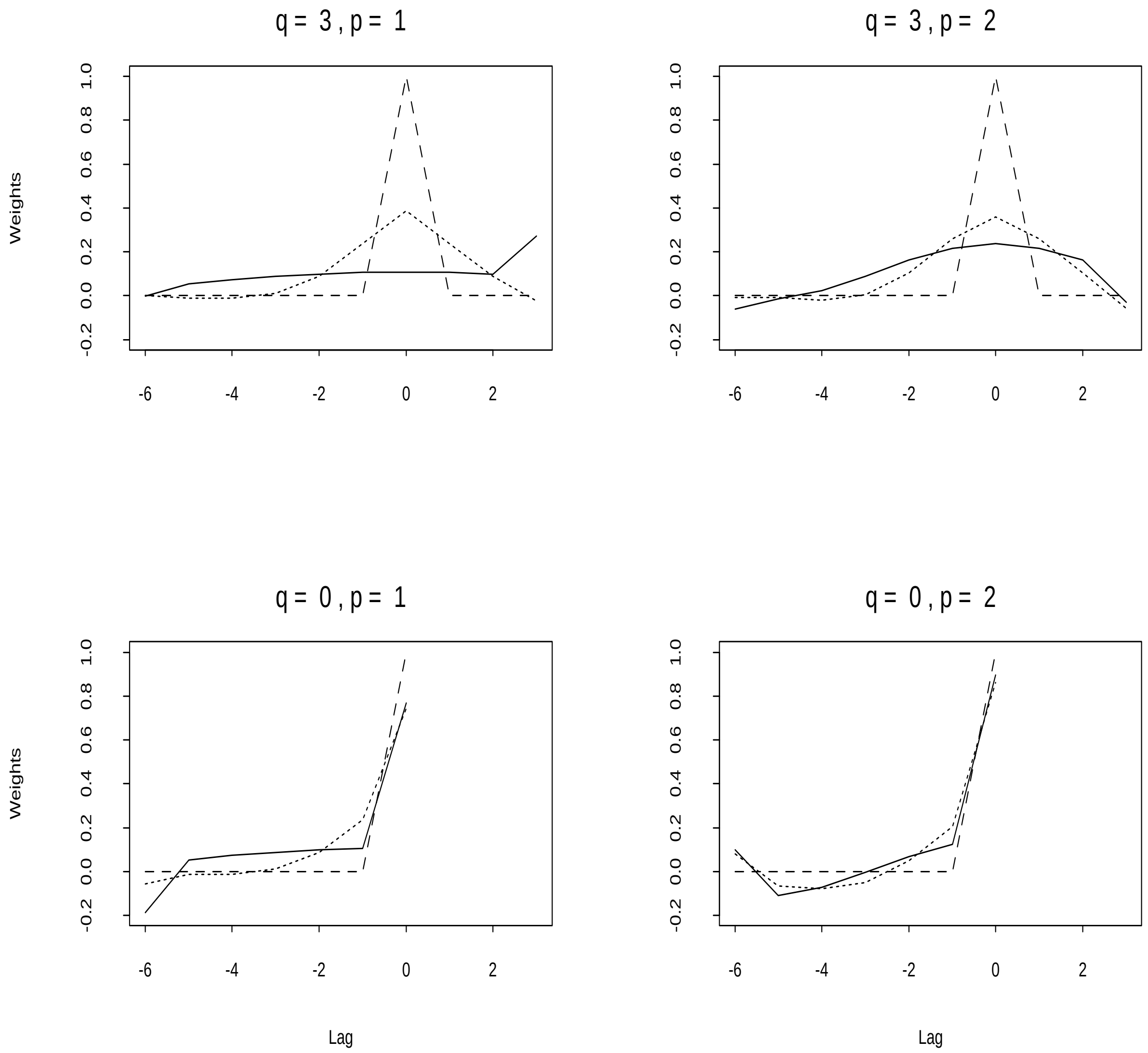


Figure 10: Plots of the limiting ($\lambda = \infty$) impulse response functions for the end filters given by Corollary 6 based on the limiting central filters of Theorem 1. Here the cases $q = 0$ and $q = 3$ are given for selected values of θ and the local linear and quadratic models specified by p. The solid lines correspond to $\theta = 0$ (smoothness criterion only), the dashed lines to $\theta = 1$ (fidelity criterion only) and the dotted lines to $\theta = 0.5$ which gives a compromise between smoothness and fidelity.

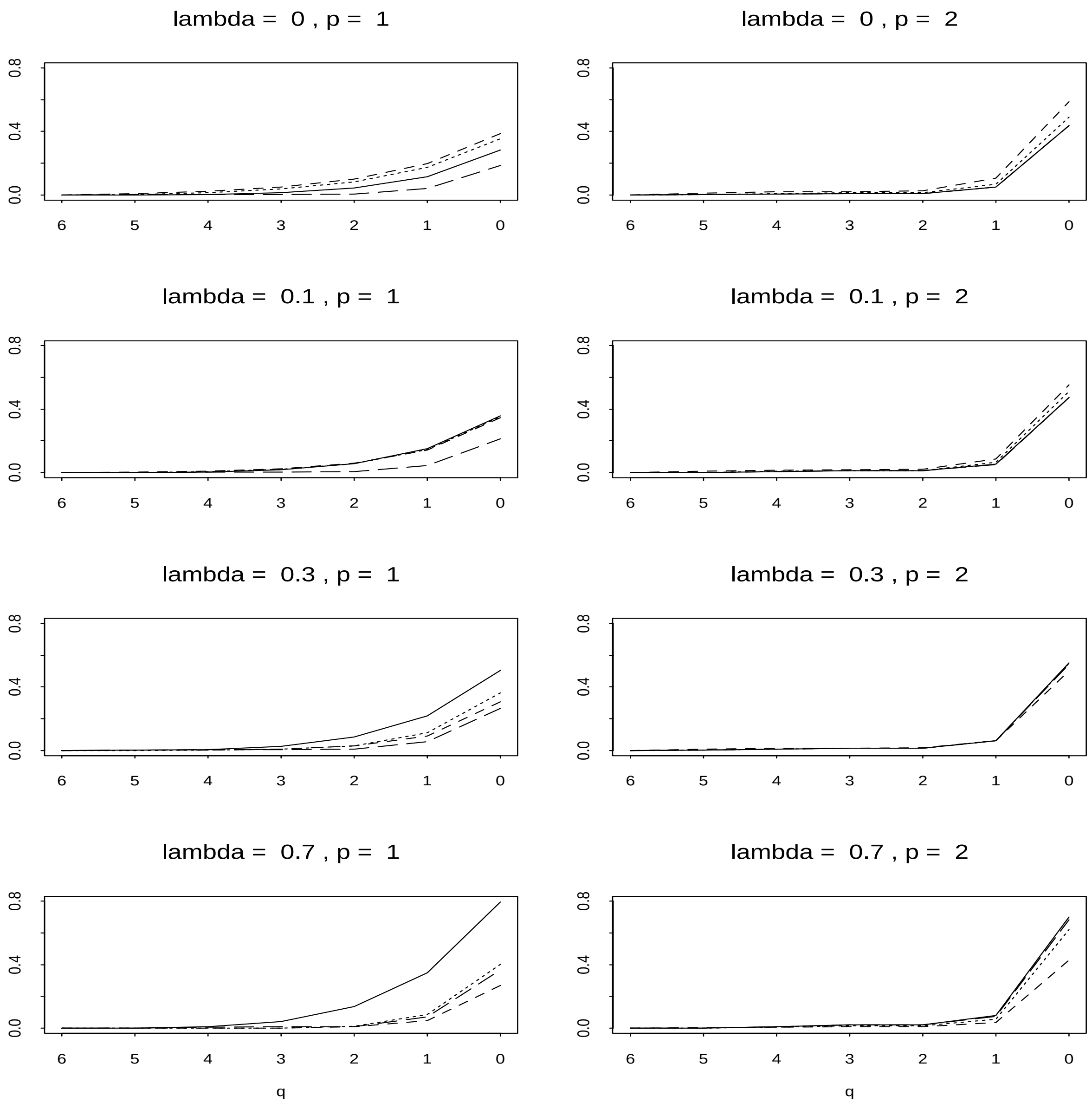


Figure 11: Plots of the mean squared revisions criterion R_q/σ^2 for the BLUP end filters based on the optimum central filters given by Theorem 1 and selected values of θ and λ. Here the local linear and quadratic models are specified by λ and p, the solid lines correspond to the central filter with $\theta = 0$ (smoothness criterion only), the short dashed lines to the central filter with $\theta = 1$ (fidelity criterion only) and the dotted lines to the central filter with $\theta = 0.5$ which gives a compromise between smoothness and fidelity. The long dashed lines correspond to R_q/σ^2 for the BLUP end filters based on the central X-11 Henderson filter. This has been included for comparative purposes.

A common feature of all the plots is the smaller values of R_q/σ^2 for the local quadratic model ($p = 2$) when $q > 0$ by comparison to the local linear model ($p = 1$). This is a direct consequence of the central filters involved. To see this note that

$$R_q = E\{(\sum_{s=-r}^{r} w_s(y_{t+s} - \hat{y}_{t+s}))^2\}$$

where the w_s are the central filter weights and $\hat{y}_{t+s}$ is the BLUP of y_{t+s} based on observations $y_{t-r}, \ldots, y_{t+q}$. From Figure 1 note that the w_s for $p = 1$ are positive whereas those for $p = 2$ are typically negative for $|s| \geq 5$. In addition, the w_s for $p = 2$ are typically small for $|s| = 4$ and are identically zero in this case for the X-11 Henderson filter. Since the forecast errors $y_{t+s} - \hat{y}_{t+s}$ will typically be positively correlated, there is more likely to be a cancellation of errors when $p = 2$ and q is 1 or 2 than for the corresponding case when $p = 1$. This difference decreases as λ increases due to the increasing adaptivity of the $p = 1$ filters which place progressively more weight on the observations at the origin of the window. Thus, although the $p = 1$ local model may well provide better predictions $\hat{y}_{t+s}$, the mean squared revisions will typically be better for $p = 2$ and $q > 0$ due to the shape of the central filters involved.

It is clear that the choice of the central filter has a direct bearing on the size of the expected revisions at the ends of the series. In particular, the choice of a central filter with an appropriate balance of smoothness and fidelity may well result in lower mean squared revisions. However care must be exercised since better mean squared revisions obtained by using high values of θ may well lead to central filters with inferior smoothness properties (see Figure 6 for example). A balance must be struck between the desired smoothness and fidelity properties in the body of the series and the desire for minimum revisions at the ends. For example, in practice it may well prove that a local linear model with large λ will provide a comparable explanation of the data by comparison to a local quadratic model with small λ. In this case the plots of R_q/σ^2 for $p = 1$, $\lambda = 0.7$, $\theta = 0.5$ and $p = 2$, $\lambda = 0$, $\theta = 0$ suggest that the local linear model is to be preferred since it has better mean squared revisions.

Now consider the plots of R_q/σ^2 for the BLUP end filters based on the same central X-11 Henderson filter. Examination of these particular plots in Figure 11 reveals that mean squared revisions increase as p and λ increase. This is what one would expect given standard results from prediction theory with the scale of the increases moderated by the common central filter adopted.

For $p = 2$ and all values of λ it would appear that the BLUP end filters based on the central X-11 Henderson filter give much the same mean squared revisions as the BLUP end filters based on the central Henderson filters with optimal smoothness ($\theta = 0$). This is as expected since the central filters concerned are identical when $\lambda = 0$ and differ only slightly when $\lambda > 0$ (see Figure 1 for example). When $p = 1$ and $\lambda > 0.4$ the BLUP end filters based on the central filters given by Theorem 1 turn out to have better mean squared revisions than those based on the central X-11 Henderson filter. However, for $p = 1$ and

$\lambda < 0.4$ the reverse is true and the BLUP end filters based on the central X-11 Henderson filter perform significantly better. Note, however, that this improvement in performance at the ends comes at the expense of decreased performance in the body of the series as Figure 7 attests. Once again an appropriate balance must be struck, in this case between the desired mean squared revisions properties of the BLUP end filters and the fidelity and smoothness properties of the central filter on which they are based.

3.2.2 Biased predictors

Here we consider the properties of the BLIP end filters based on Corollary 9 and specified central filter weights w_s. To evaluate the BLIP end filters we need to know the parameters required to specify the BLUP end filters together with the value of β_p^2/σ^2. Since the latter is typically unknown, in practice the BLIP end filters are evaluated with β_p^2/σ^2 replaced by some estimated or target value $\hat{\beta}_p^2/\hat{\sigma}^2$.

Three end filters are considered: $\hat{\beta}_p^2/\hat{\sigma}^2 = \infty$ (this yields the BLUP end filters), $\hat{\beta}_p^2/\hat{\sigma}^2 = 4/(\pi(3.5)^2)$ (this value gives the X-11 end filter in the case where $p = 1$, $\lambda = 0$ and the central filter concerned is the X-11 13 point Henderson filter), and $\hat{\beta}_p^2/\hat{\sigma}^2 = 0$. When $\beta_p^2 = 0$ the latter choice for $\hat{\beta}_p^2/\hat{\sigma}^2$ corresponds to the BLUP end filter for the reduced model of order $p - 1$. As noted in the discussion following Theorem 10, this particular end filter yields the best mean squared revisions possible in the sense that 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 (54) 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 $R_q(\hat{\phi}_0)$.

Examples of the three BLIP end filters based on the central filters given by Theorem 1 and the local dynamic models (55) and (56) are given in Figures 12–15 for the cases $q = 3$, $q = 0$ and selected values of θ and λ. In terms of the general shape of the impulse response functions, it is clear that θ and λ are the dominant parameters when $q = 3$, but $\hat{\beta}_p^2/\hat{\sigma}^2$ is the dominant parameter when $q = 0$ and prediction within the window is most important. In general it would appear that $\hat{\beta}_p^2/\hat{\sigma}^2$ has greatest effect when λ is small. Note also that the X-11 based value of $\hat{\beta}_p^2/\hat{\sigma}^2 = 4/(\pi(3.5)^2)$ would seem to be too close to $\hat{\beta}_p^2/\hat{\sigma}^2 = \infty$ to make any real difference. In terms of variation with θ and λ for a given value of $\hat{\beta}_p^2/\hat{\sigma}^2$, similar comments apply here as were made for the BLUP end filters given in Figures 8 and 9.

In much the same way as before, the limiting impulse response functions when $\lambda \to \infty$ can be determined. Details are given in Theorem A4 of Appendix A.2. A key feature of these end filters is that they do not depend on the choice of $\hat{\beta}_p^2/\hat{\sigma}^2$. The limiting BLIP end filters are given in Figure 16 for selected values of θ, q and the local linear and quadratic models specified by p. Although the central filters on which they are based could have been held fixed, the limiting 13 point central filters given by Theorem 1 have been used. Again, the pure fidelity case ($\theta = 1$) yields a limiting end filter with all its weight concentrated at the

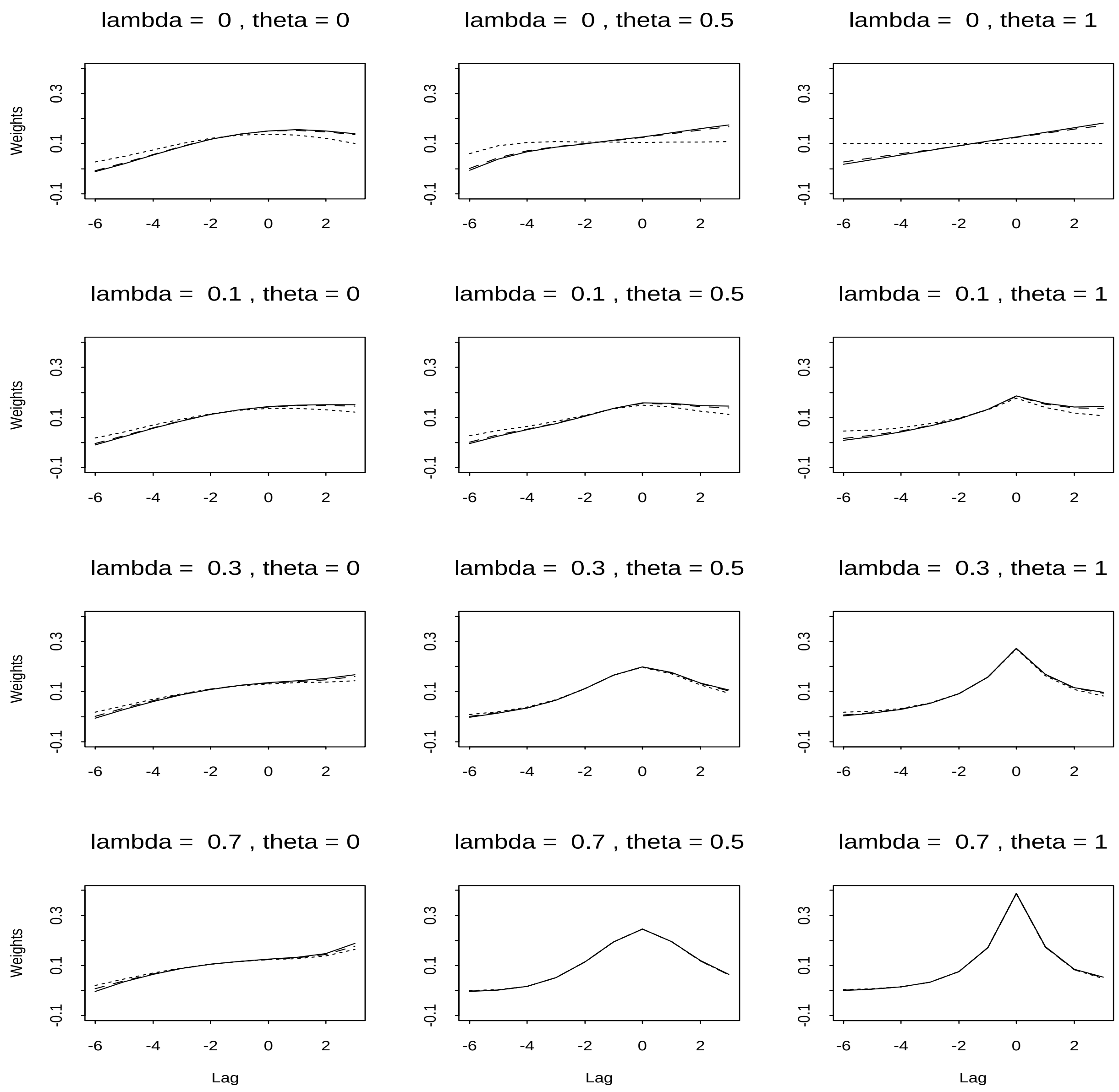


Figure 12: Plots of the impulse response functions of the BLIP end filters for $q = 3$ based on the local linear model ($p = 1$) and the 13 point central filters given by Theorem 1. These end filters are evaluated for selected values of θ and λ. The solid lines correspond to $\hat{\beta}_p^2/\hat{\sigma}^2 = \infty$ (this yields the BLUP end filters), the dashed lines to $\hat{\beta}_p^2/\hat{\sigma}^2 = 4/(\pi(3.5)^2)$ (this value gives the X-11 end filter in the case where $p = 1$, $\lambda = 0$ and the central filter concerned is the X-11 13 point Henderson filter), and the dotted lines to $\hat{\beta}_p^2/\hat{\sigma}^2 = 0$. When $\beta_p^2 = 0$ the latter choice for $\hat{\beta}_p^2/\hat{\sigma}^2$ corresponds to the BLUP end filter for the reduced model of order $p - 1$ and represents the best mean squared revisions possible.

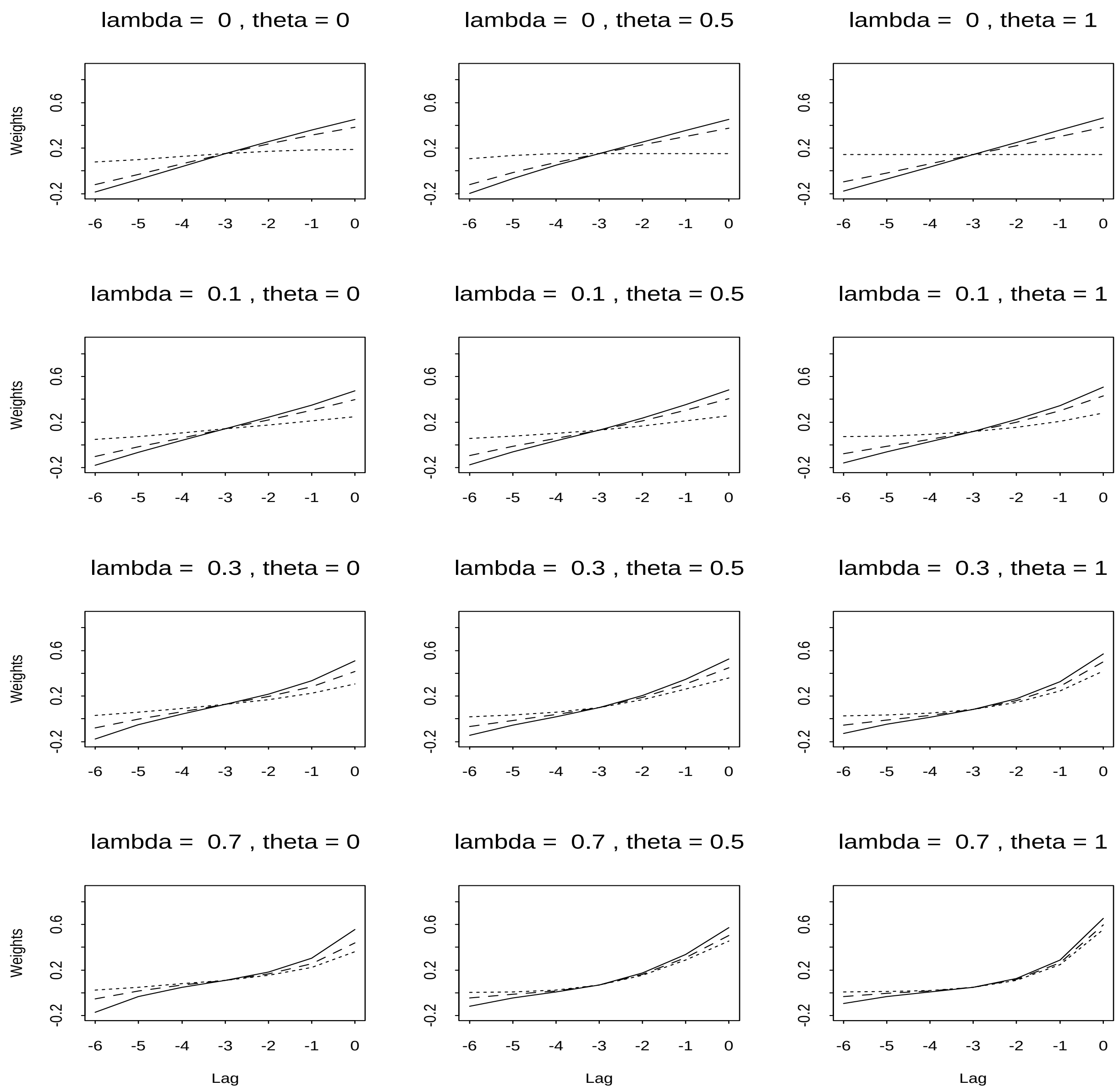


Figure 13: Plots of the impulse response functions of the BLIP end filters for $q = 0$ based on the local linear model ($p = 1$) and the 13 point central filters given by Theorem 1. These end filters are evaluated for selected values of θ and λ. The solid lines correspond to $\hat{\beta}_p^2/\hat{\sigma}^2 = \infty$ (this yields the BLUP end filters), the dashed lines to $\hat{\beta}_p^2/\hat{\sigma}^2 = 4/(\pi(3.5)^2)$ (this value gives the X-11 end filter in the case where $p = 1$, $\lambda = 0$ and the central filter concerned is the X-11 13 point Henderson filter), and the dotted lines to $\hat{\beta}_p^2/\hat{\sigma}^2 = 0$. When $\beta_p^2 = 0$ the latter choice for $\hat{\beta}_p^2/\hat{\sigma}^2$ corresponds to the BLUP end filter for the reduced model of order $p-1$ and represents the best mean squared revisions possible.

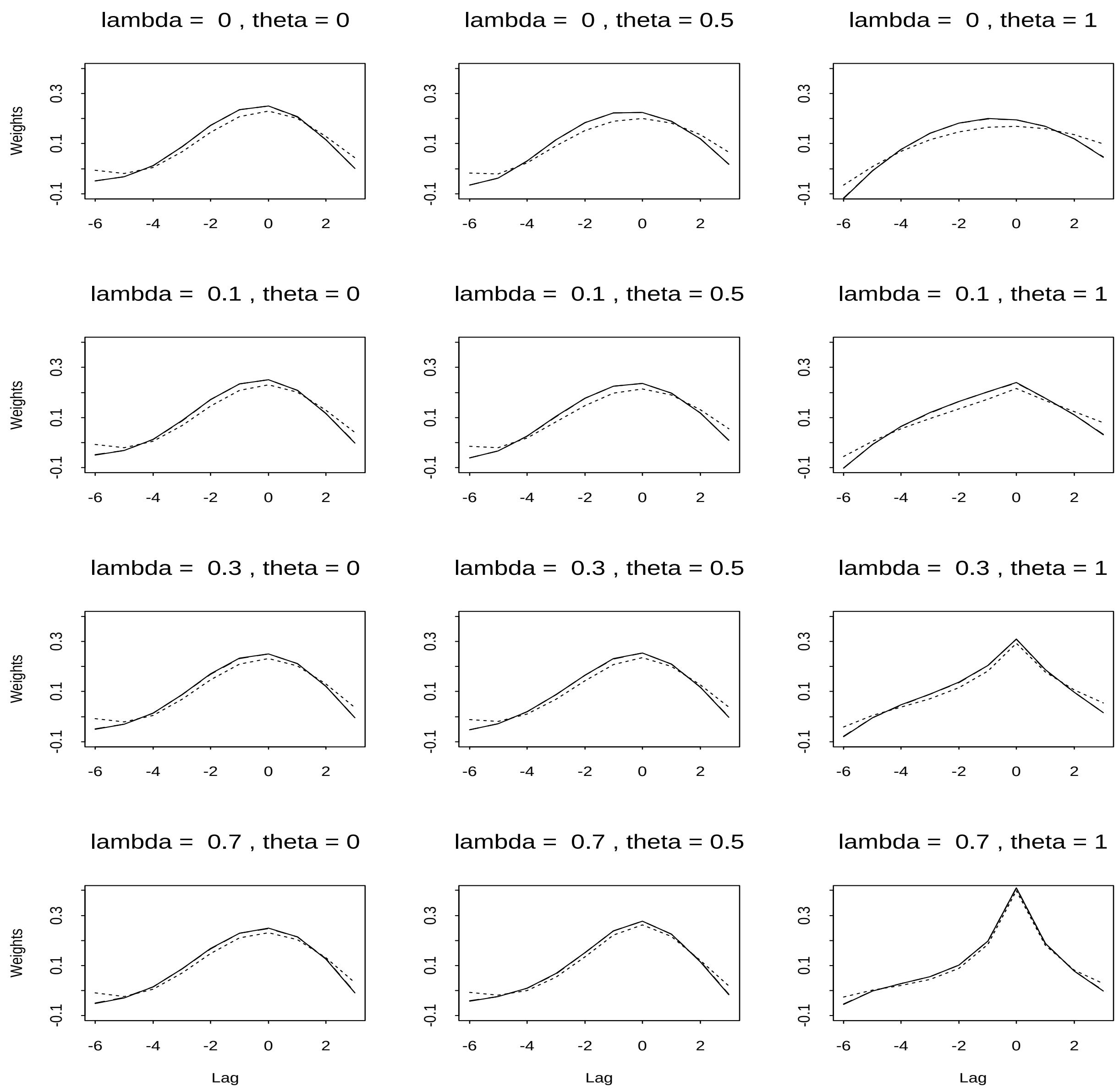


Figure 14: Plots of the impulse response functions of the BLIP end filters for $q = 3$ based on the local quadratic model ($p = 2$) and the 13 point central filters given by Theorem 1. These end filters are evaluated for selected values of θ and λ. The solid lines correspond to $\hat{\beta}_p^2/\hat{\sigma}^2 = \infty$ (this yields the BLUP end filters), the dashed lines to $\hat{\beta}_p^2/\hat{\sigma}^2 = 4/(\pi(3.5)^2)$ (this value gives the X-11 end filter in the case where $p = 1$, $\lambda = 0$ and the central filter concerned is the X-11 13 point Henderson filter), and the dotted lines to $\hat{\beta}_p^2/\hat{\sigma}^2 = 0$. When $\beta_p^2 = 0$ the latter choice for $\hat{\beta}_p^2/\hat{\sigma}^2$ corresponds to the BLUP end filter for the reduced model of order $p - 1$ and represents the best mean squared revisions possible.

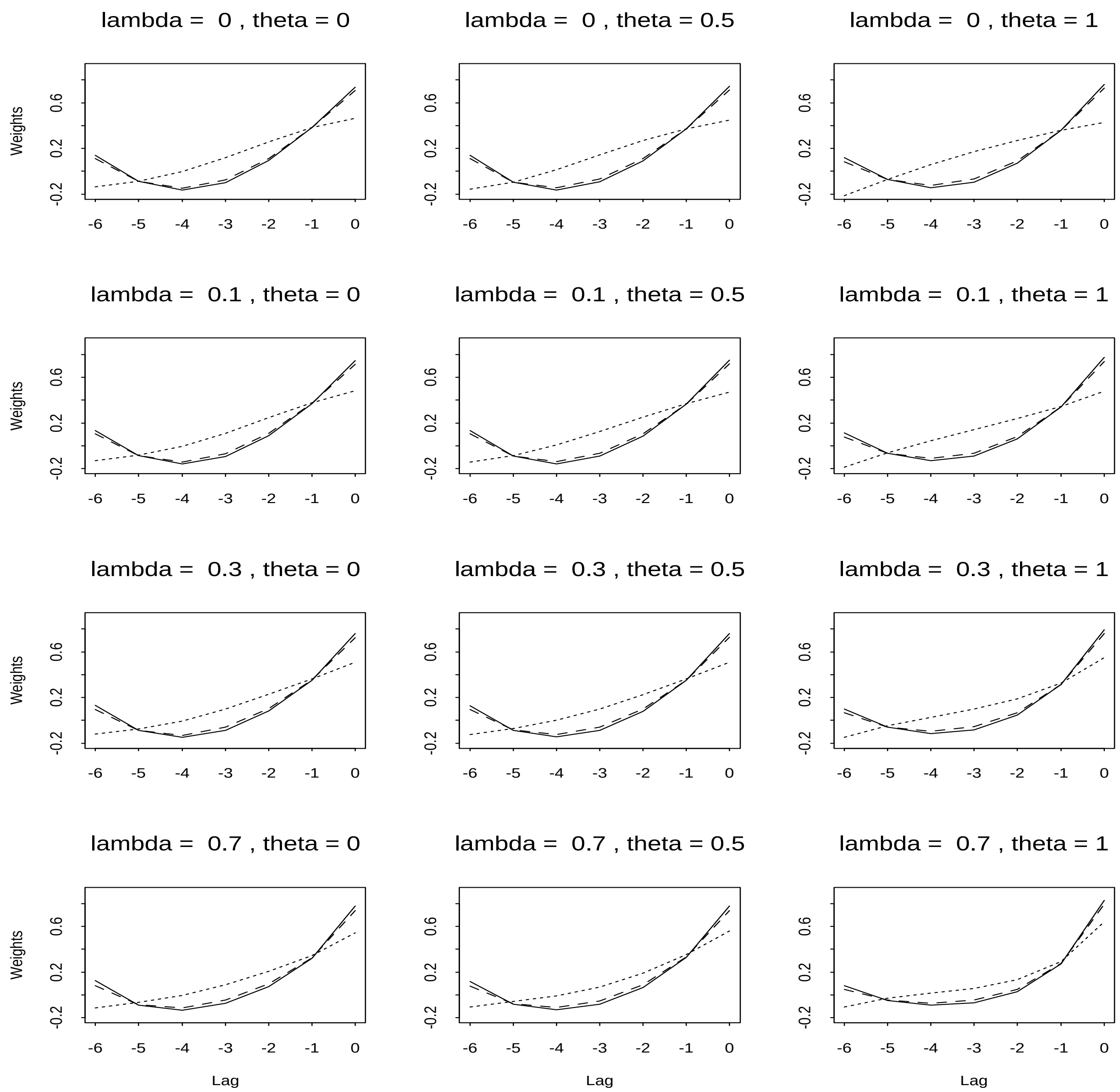


Figure 15: Plots of the impulse response functions of the BLIP end filters for $q = 0$ based on the local quadratic model ($p = 2$) and the 13 point central filters given by Theorem 1. These end filters are evaluated for selected values of θ and λ. The solid lines correspond to $\hat{\beta}_p^2/\hat{\sigma}^2 = \infty$ (this yields the BLUP end filters), the dashed lines to $\hat{\beta}_p^2/\hat{\sigma}^2 = 4/(\pi(3.5)^2)$ (this value gives the X-11 end filter in the case where $p = 1$, $\lambda = 0$ and the central filter concerned is the X-11 13 point Henderson filter), and the dotted lines to $\hat{\beta}_p^2/\hat{\sigma}^2 = 0$. When $\beta_p^2 = 0$ the latter choice for $\hat{\beta}_p^2/\hat{\sigma}^2$ corresponds to the BLUP end filter for the reduced model of order $p - 1$ and represents the best mean squared revisions possible.

origin of the window so that $\tilde{g}_t = y_t$.

We now consider the performance of the BLIP end filters as measured by the mean squared revisions $R_q(\hat{\phi}_0)$. This is a function of both β_p^2/σ^2 and $\hat{\beta}_p^2/\hat{\sigma}^2$ as well as other parameters. 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 ($p = 1$) and

$$\frac{I}{C} = \frac{E|\Delta\epsilon_t|}{E|\Delta g_t|} \tag{60}$$

which can be thought of as the population parameter that $\bar{I}/\bar{C}$ is estimating. Under Gaussian assumptions

$$\frac{I}{C} = \frac{2}{\sqrt{\pi\tilde{\beta}_1^2}}(\Phi(\frac{|\tilde{\beta}_1|}{\sqrt{\lambda}}) - \Phi(-\frac{|\tilde{\beta}_1|}{\sqrt{\lambda}}) + \sqrt{\frac{2\lambda}{\pi\tilde{\beta}_1^2}}\exp(-\frac{\tilde{\beta}_1^2}{2\lambda}))^{-1} \tag{61}$$

where $\tilde{\beta}_1 = \beta_1/\sigma$, $\Phi(x)$ is the standard normal cumulative distribution function and the result follows from Theorem A5 given in the Appendix. A plot of I/C is given in Figure 17 as a function of $|\tilde{\beta}_1|$ for selected values of λ. The horizontal lines correspond to the boundaries $I/C = 1$ and $I/C = 3.5$. From the graph and (61) it can be seen that the sets of feasible values for $|\tilde{\beta}_1|$ lie in $[0, 2/\sqrt{\pi}]$.

It would appear (see Musgrave (1964) and Doherty (1991)) that the guidelines given by X-11's $\bar{I}/\bar{C}$ ratio were based on the local linear model with $\lambda = 0$. To extend this measure to higher order models consider

$$\frac{I_p}{C_p} = \frac{E|\Delta^p\epsilon_t|}{E|\Delta^p g_t|}$$

which is identical to I/C when $p = 1$. It can be thought of as a noise to signal ratio for the differenced series $\Delta^p y_t$. For the local quadratic model ($p = 2$) and under Gaussian assumptions

$$\frac{I_2}{C_2} = \sqrt{\frac{3}{\pi\tilde{\beta}_2^2}}(\Phi(|\tilde{\beta}_2|\sqrt{\frac{2}{\lambda}}) - \Phi(-|\tilde{\beta}_2|\sqrt{\frac{2}{\lambda}}) + \sqrt{\frac{\lambda}{\pi\tilde{\beta}_2^2}}\exp(-\frac{\tilde{\beta}_2^2}{\lambda}))^{-1} \tag{62}$$

where $\tilde{\beta}_2 = \beta_2/\sigma$. This result also follows from Theorem A5 in the Appendix. A plot of I_2/C_2 is given in Figure 17 as a function of $|\tilde{\beta}_2|$ and for selected values of λ. If the same I/C limits are retained (a possibly dubious presumption) it would seem that the interval $[0, 2/\sqrt{\pi}]$ again provides a feasible set of values for $|\tilde{\beta}_2|$.

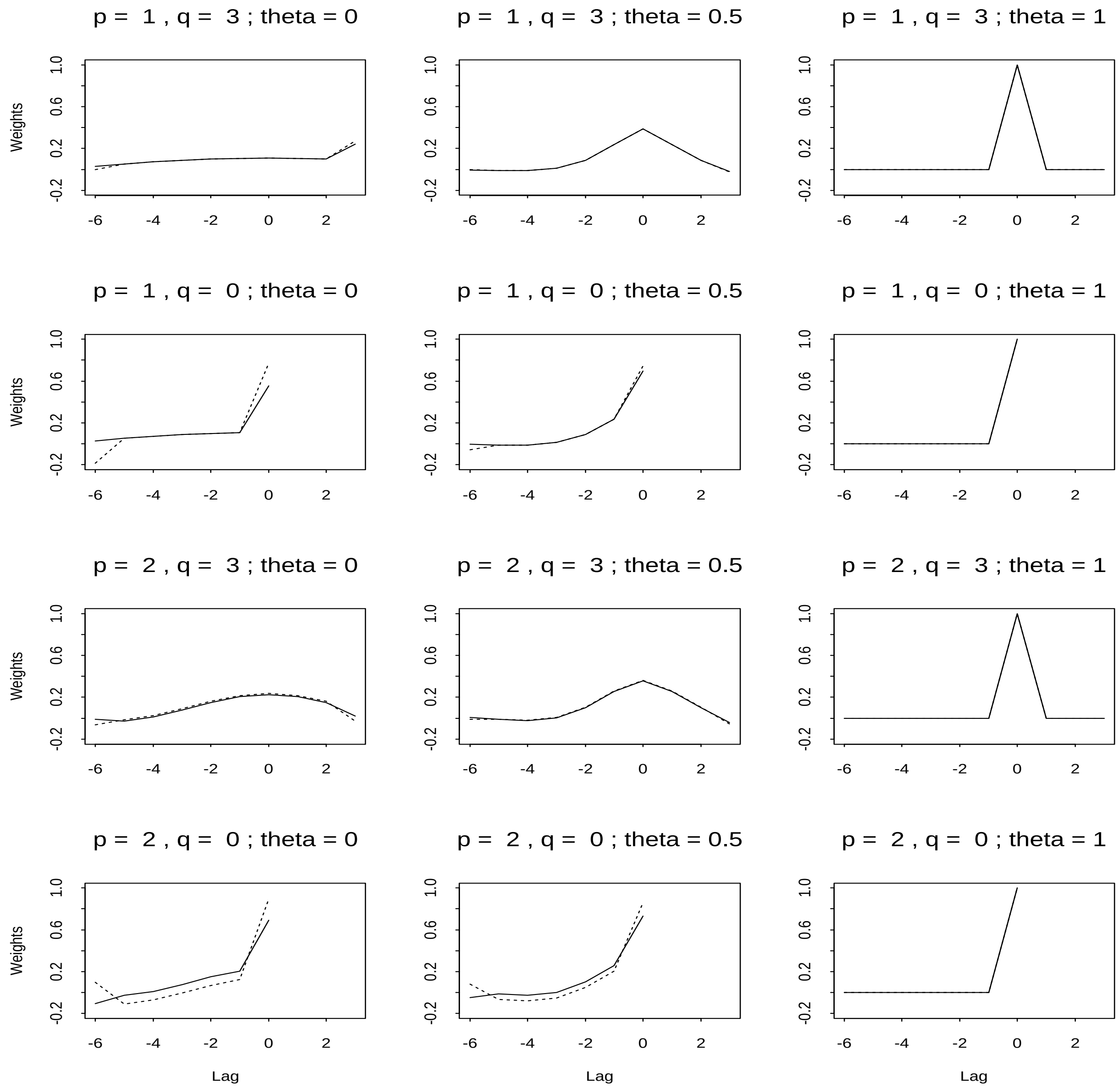
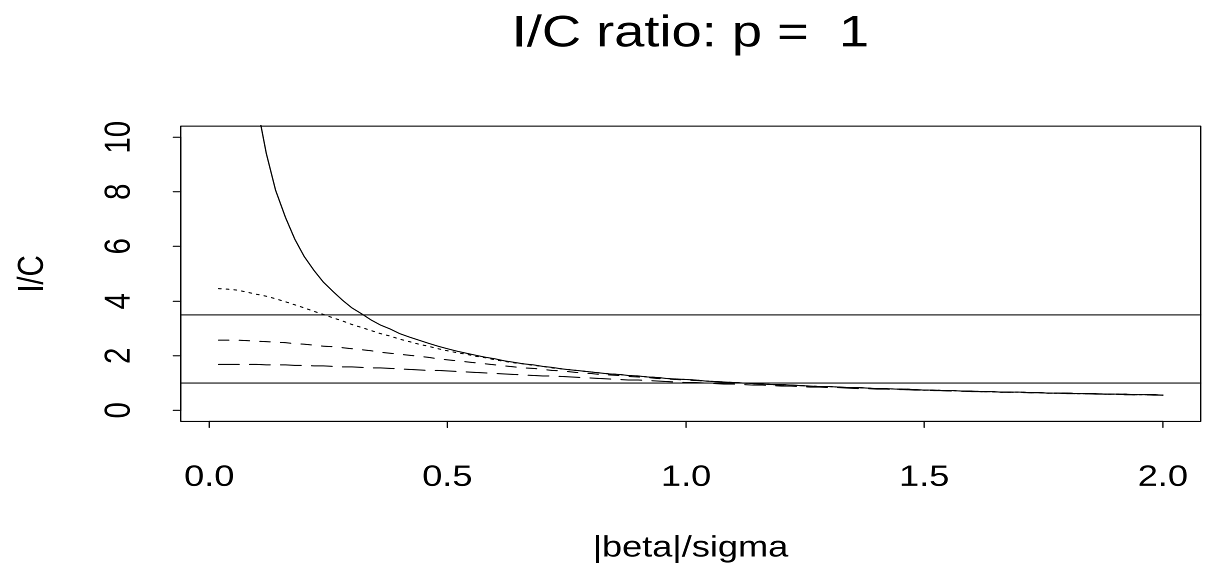


Figure 16: Plots of the limiting ($\lambda = \infty$) impulse response functions of the BLIP end filters for selected values of p, θ, q and the limiting 13 point central filters given by Theorem 1. The solid lines correspond to the limiting BLIP end filters and, for comparison, the dotted lines to the limiting BLUP end filters.



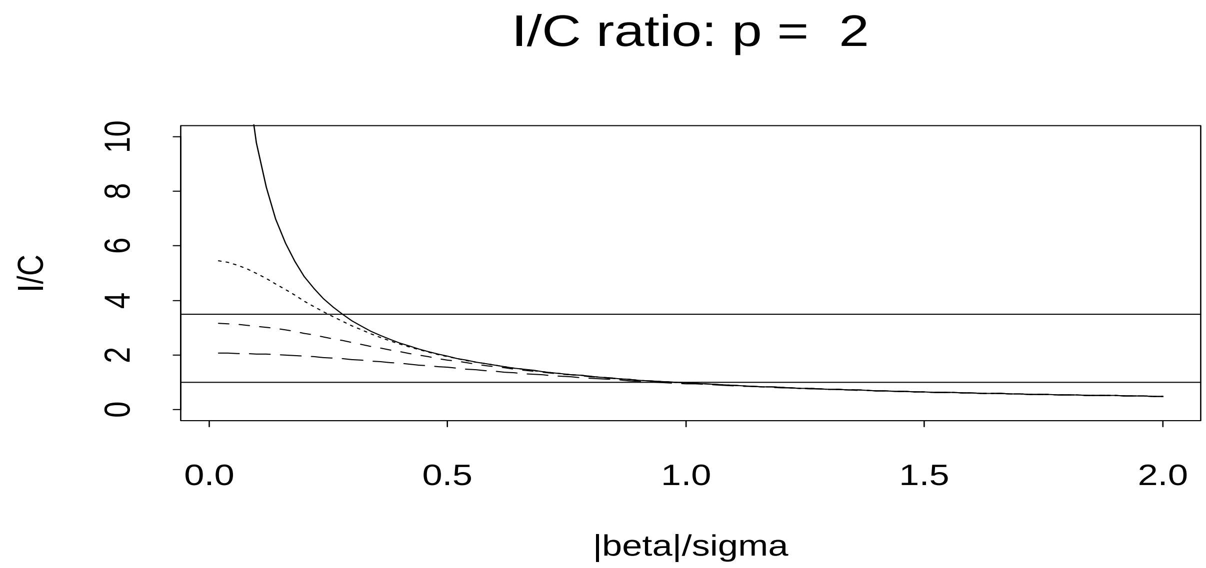


Figure 17: Plots of the I/C ratio against $|\beta_1|/\sigma$ and the I_2/C_2 ratio against $|\beta_2|/\sigma$ for the local linear ($p = 1$) and quadratic ($p = 2$) models respectively and selected values of λ. The solid line corresponds to $\lambda = 0$, the dotted line to $\lambda = 0.1$, the short dashed line to $\lambda = 0.3$ and the long dashed line to $\lambda = 0.7$ The horizontal lines correspond to the boundaries $I/C = 1$ and $I/C = 3.5$.

On the basis of the above analysis 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 10, $R_q(\phi)$ 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 as a function of q in Figures 18–19 for each of the three BLIP end filters concerned. These are based on the central filters given by Theorem 1, the local dynamic models (55) and (56), and selected values of θ and λ. In each case the mean squared revisions are plotted for $\beta_p^2/\sigma^2 = 0$ (a lower limit representing the best obtainable result for that particular end filter) and $\beta_p^2/\sigma^2 = 4/\pi$ (a nominal upper limit based on X-11's I/C ratio). For each model specified by p and λ the least possible revisions occur when $\hat{\beta}_p^2/\hat{\sigma}^2 = \beta_p^2/\sigma^2 = 0$. Note also that the mean squared revisions for the BLIP end filters with $\hat{\beta}_p^2/\hat{\sigma}^2 = \infty$ do not depend on the choice of β_p^2/σ^2 since they are BLUP end filters which are unbiased.

The same general comments made before concerning Figure 11 and the BLUP end filters apply here also. Bearing in mind that individual plots are based on different central filters (see Figure 1), mean squared revisions generally increase as λ increases for end filters based on central Henderson filters ($\theta = 0$), and decrease as λ increases for end filters based on central Macaulay filters ($\theta = 1$). For low values of λ, end filters based on central Macaulay filters with optimal fidelity ($\theta = 1$) show higher mean squared revisions than those based on central Henderson filters with optimal smoothness ($\theta = 0$). Again, it would appear that whereas smoothness plays the influential role in the body of the series, it is fidelity that plays the dominant role in minimising expected revisions at the ends.

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 typically negligible by comparison. Although not entirely 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 filters 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 $\beta_p^2/\sigma^2 = \hat{\beta}_p^2/\hat{\sigma}^2 = 0$. From Figures 18–19 these bounds show that there are gains to be had using BLIP end filters in preference to BLUP end filters. When λ is small (say $\lambda \leq 0.1$) these gains are likely to be greatest for BLIP end filters based on central Macaulay filters ($\theta = 1$). For larger values of λ (say $\lambda > 0.1$) the reverse is true with the greatest gains occurring for BLIP end filters based on central Henderson filters ($\theta = 0$).

If $\hat{\beta}_p^2/\hat{\sigma}^2$ has been selected inappropriately then the mean squared revisions $R_q(\hat{\phi}_0)/\sigma^2$ can become very large indeed. This effect is best seen in the plots of $R_q(\hat{\phi}_0)/\sigma^2$ evaluated at the upper limit $\beta_p^2/\sigma^2 = 4/\pi$. Thus a good prior knowledge of β_p^2/σ^2 is necessary if significant gains are to be made over the BLUP end filters.

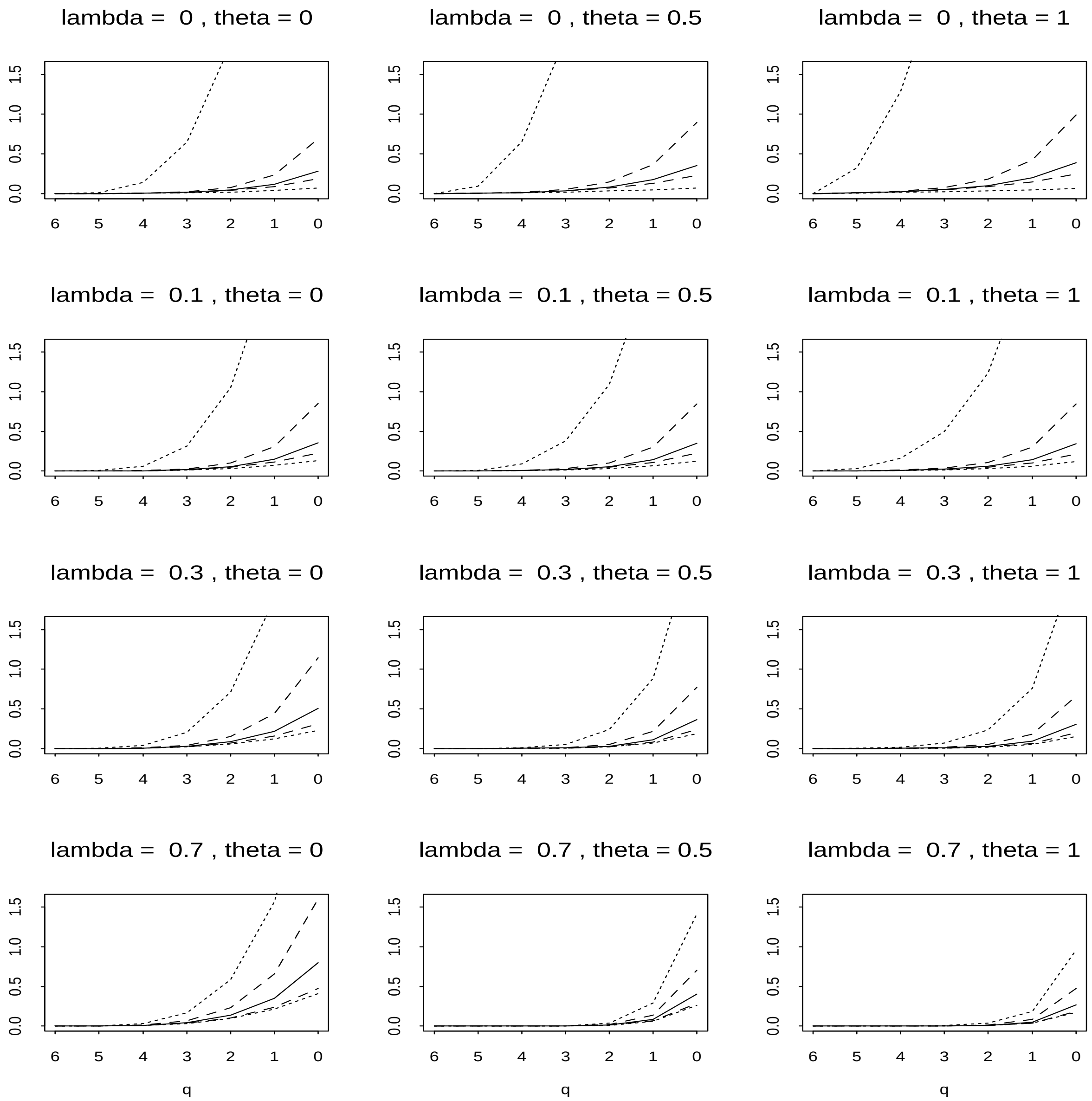


Figure 18: Plots of the mean squared revisions criterion $R_q(\hat{\phi}_0)$ for the BLIP end filters based on the local linear model ($p = 1$), the optimum central filters given by Theorem 1 and selected values of θ and λ. Here the solid line corresponds to $\hat{\beta}_p^2/\hat{\sigma}^2 = \infty$ (this yields the BLUP end filters), the dashed lines to $\hat{\beta}_p^2/\hat{\sigma}^2 = 4/(\pi(3.5)^2)$ (this value gives the X-11 end filter in the case where $p = 1$, $\lambda = 0$ and the central filter concerned is the X-11 13 point Henderson filter), and the dotted lines to $\hat{\beta}_p^2/\hat{\sigma}^2 = 0$ (this value gives the BLUP end filter for the reduced model of order $p - 1$). In each case the mean squared revisions are plotted for the $\beta_p^2/\sigma^2 = 0$ (lower limit) and $\beta_p^2/\sigma^2 = 4/\pi$ (upper limit). Since it is unbiased, the mean squared revisions for the BLUP end filter do not depend on the choice of β_p^2/σ^2.

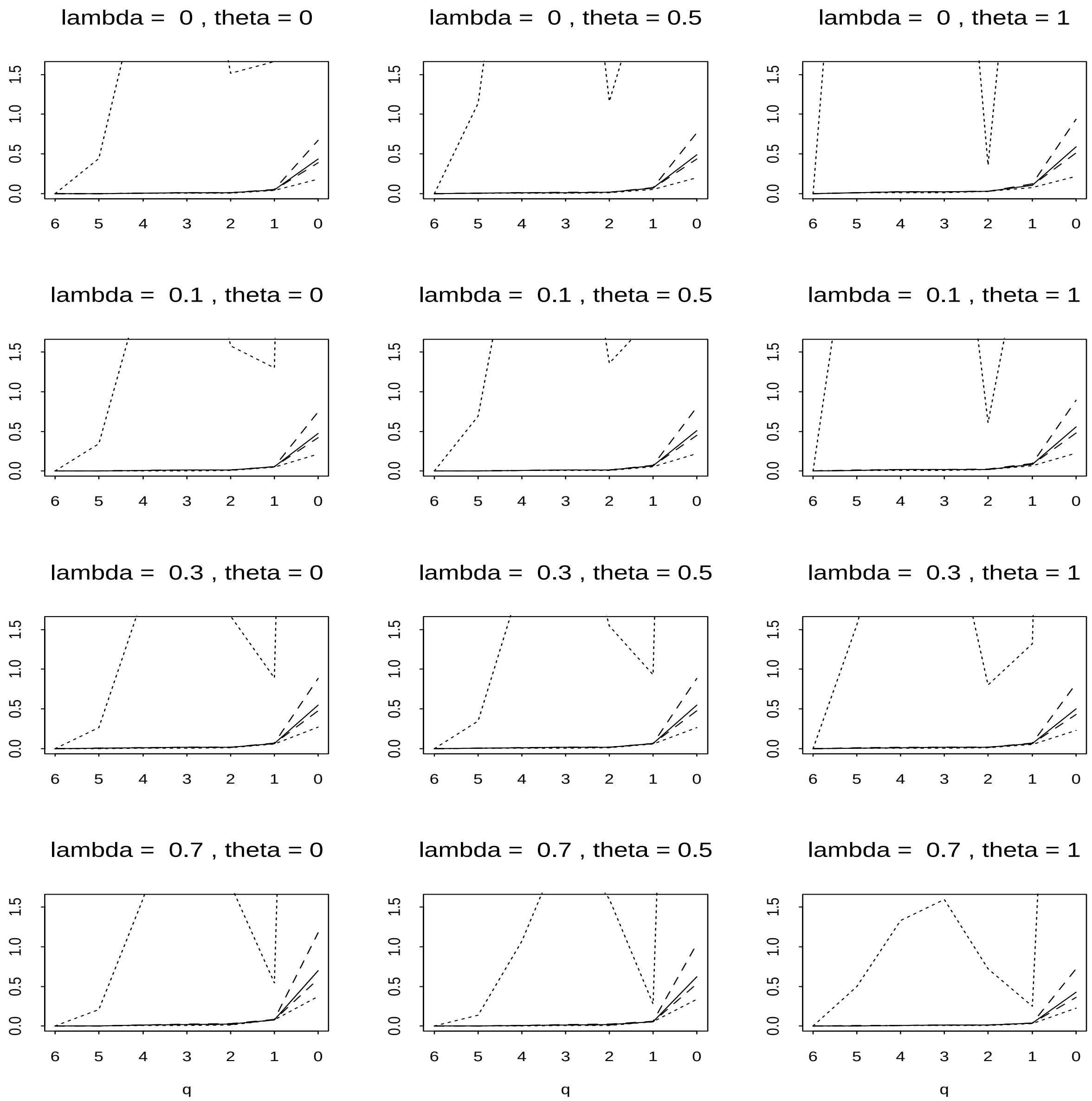


Figure 19: Plots of the mean squared revisions criterion $R_q(\hat{\phi}_0)$ for the BLIP end filters based on the local quadratic model ($p = 2$), the optimum central filters given by Theorem 1 and selected values of θ and λ. Here the solid line corresponds to $\hat{\beta}_p^2/\hat{\sigma}^2 = \infty$ (this yields the BLUP end filters), the dashed lines to $\hat{\beta}_p^2/\hat{\sigma}^2 = 4/(\pi(3.5)^2)$ (this value gives the X-11 end filter in the case where $p = 1$, $\lambda = 0$ and the central filter concerned is the X-11 13 point Henderson filter), and the dotted lines to $\hat{\beta}_p^2/\hat{\sigma}^2 = 0$ (this value gives the BLUP end filter for the reduced model of order $p - 1$). In each case the mean squared revisions are plotted for the $\beta_p^2/\sigma^2 = 0$ (lower limit) and $\beta_p^2/\sigma^2 = 4/\pi$ (upper limit). Since it is unbiased, the mean squared revisions for the BLUP end filter do not depend on the choice of β_p^2/σ^2.

For the remainder of this section we restrict attention to the important case where the central filter used in the body of the series is the central X-11 Henderson filter. Plots of the impulse response functions of the three BLIP end filters based on the central X-11 Henderson filter are given in Figures 20–21 for $q = 3$, $q = 0$ and the local dynamic models (55) and (56). As before, it would appear that the difference between the BLIP end filters based on the X-11 value $\hat{\beta}_p^2/\hat{\sigma}^2 = 4/(\pi(3.5)^2)$ and the BLUP end filters is modest, especially when $p = 1$. In general, the BLIP end filters show greater variation in the important case where $q = 0$ and also when λ is small.

The mean squared revisions criterion $R_q(\hat{\phi}_0)/\sigma^2$ for the various BLIP end filters based on the central X-11 Henderson filter is plotted in Figure 22 as a function of q for a selection of models. Recall that when $p = 1$, $\lambda = 0$ and $\hat{\beta}_p^2/\hat{\sigma}^2 = 4/(\pi(3.5)^2)$, the BLIP end filters given by Corollary 9 are just the X-11 end filters derived by Musgrave (1964) and Doherty (1991).

Similar comments apply as before. If $\hat{\beta}_p^2/\hat{\sigma}^2$ has been selected appropriately then $R_q(\hat{\phi}_0)/\sigma^2$ 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 indicate 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.

Note that the mean squared revisions generally increase as λ and p increase. This is of marginal utility in practice, since the local dynamic 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 20 it would appear that, in such cases, the BLIP end filter for the quadratic model may have greater capacity to achieve lower revisions.

Consider now the BLIP end filters based on the X-11 value $\hat{\beta}_p^2/\hat{\sigma}^2 = 4/(\pi(3.5)^2)$. For the local linear model, these end filters appear to offer only marginal gains over the BLUP end filters when β_p^2/σ^2 satisfies (54). In the local quadratic case, they are 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 $\beta_p^2/\sigma^2 = 4/\pi$ for these particular end filters lead to end filters with unacceptably high revisions. This may explain, in part, the reason why forecast extension using ARIMA models 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 (54).

It is of interest to consider the ranges of values of $(|\hat{\beta}_p|/\hat{\sigma}, |\beta_p|/\sigma)$ that satisfy (54) and to ascertain the general shape of $R_q(\hat{\phi}_0)/\sigma^2$ as a function of $(|\hat{\beta}_p|/\hat{\sigma}, |\beta_p|/\sigma)$ or, equivalently,

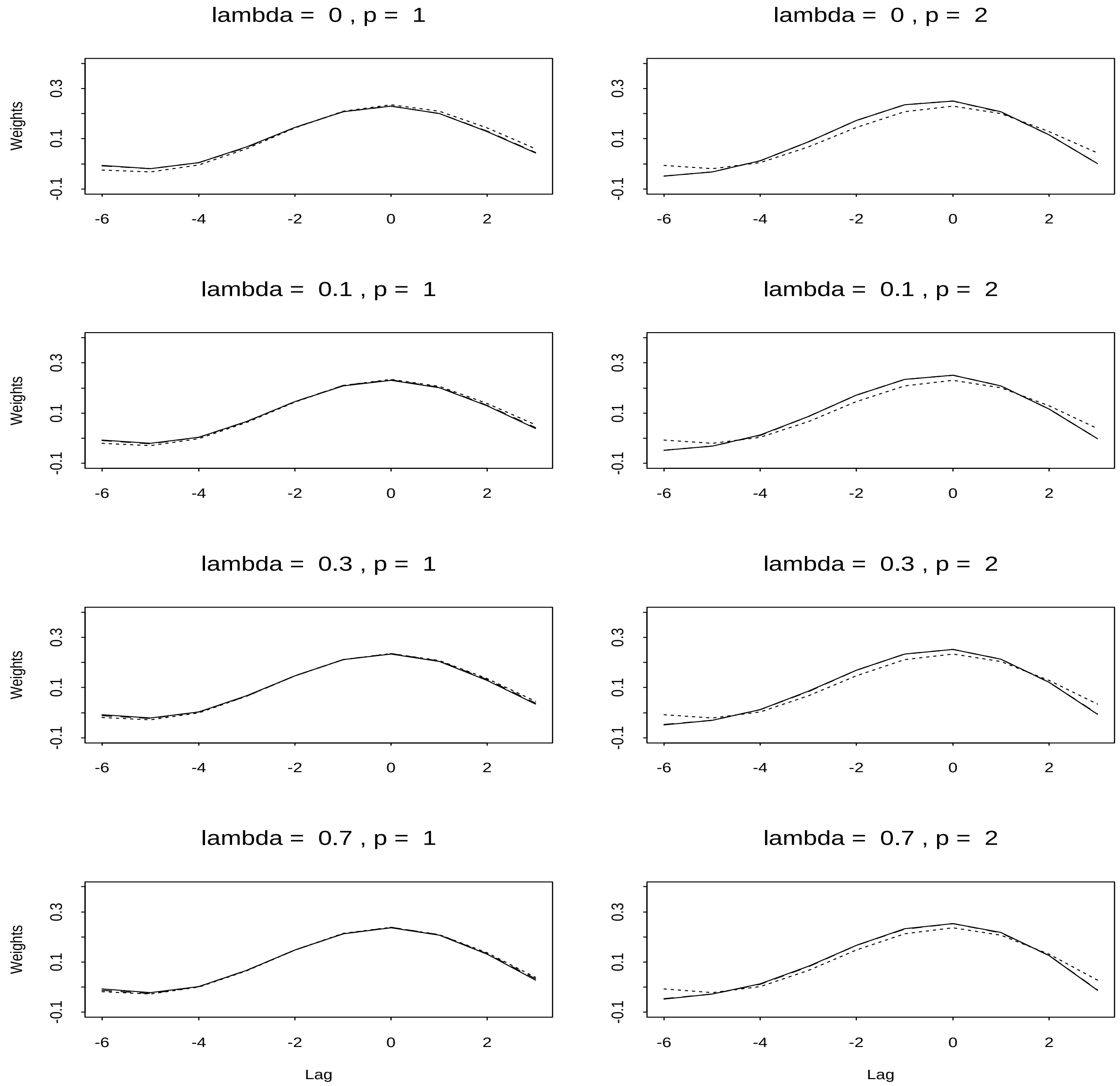


Figure 20: Plots of the impulse response functions of the BLIP end filters for $q = 3$ based on the central X-11 Henderson filter and the local linear and quadratic models specified by λ and p. The solid lines correspond to $\hat{\beta}_p^2/\hat{\sigma}^2 = \infty$ (this yields the BLUP end filters), the dashed lines to $\hat{\beta}_p^2/\hat{\sigma}^2 = 4/(\pi(3.5)^2)$ (this value gives the X-11 end filter in the case where $p = 1$, $\lambda = 0$), and the dotted lines to $\hat{\beta}_p^2/\hat{\sigma}^2 = 0$ (this value gives the BLUP end filter for the reduced model of order $p - 1$).

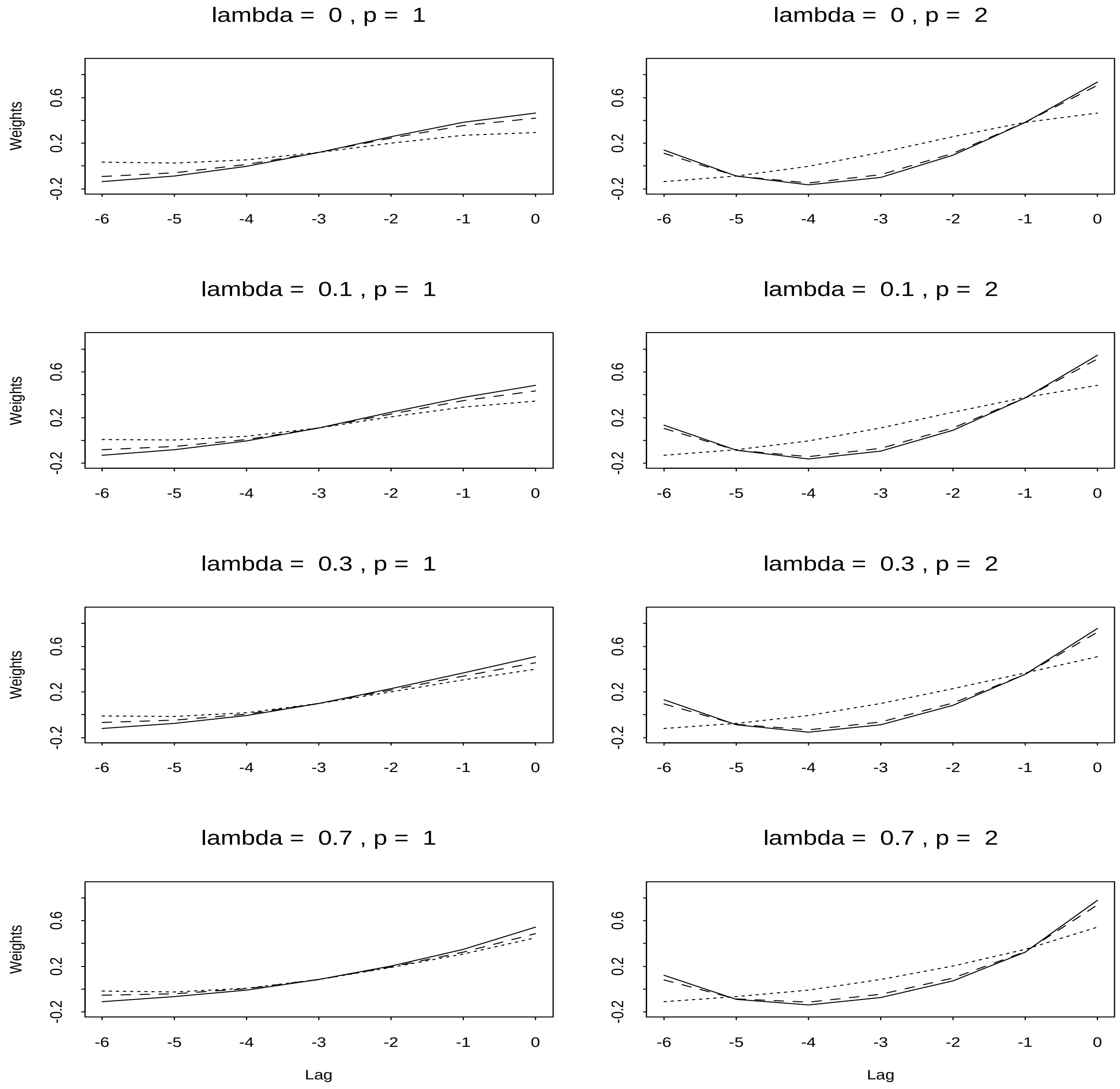


Figure 21: Plots of the impulse response functions of the BLIP end filters for $q = 0$ based on the central X-11 Henderson filter and the local linear and quadratic models specified by λ and p. The solid lines correspond to $\hat{\beta}_p^2/\hat{\sigma}^2 = \infty$ (this yields the BLUP end filters), the dashed lines to $\hat{\beta}_p^2/\hat{\sigma}^2 = 4/(\pi(3.5)^2)$ (this value gives the X-11 end filter in the case where $p = 1$, $\lambda = 0$), and the dotted lines to $\hat{\beta}_p^2/\hat{\sigma}^2 = 0$ (this value gives the BLUP end filter for the reduced model of order $p - 1$).

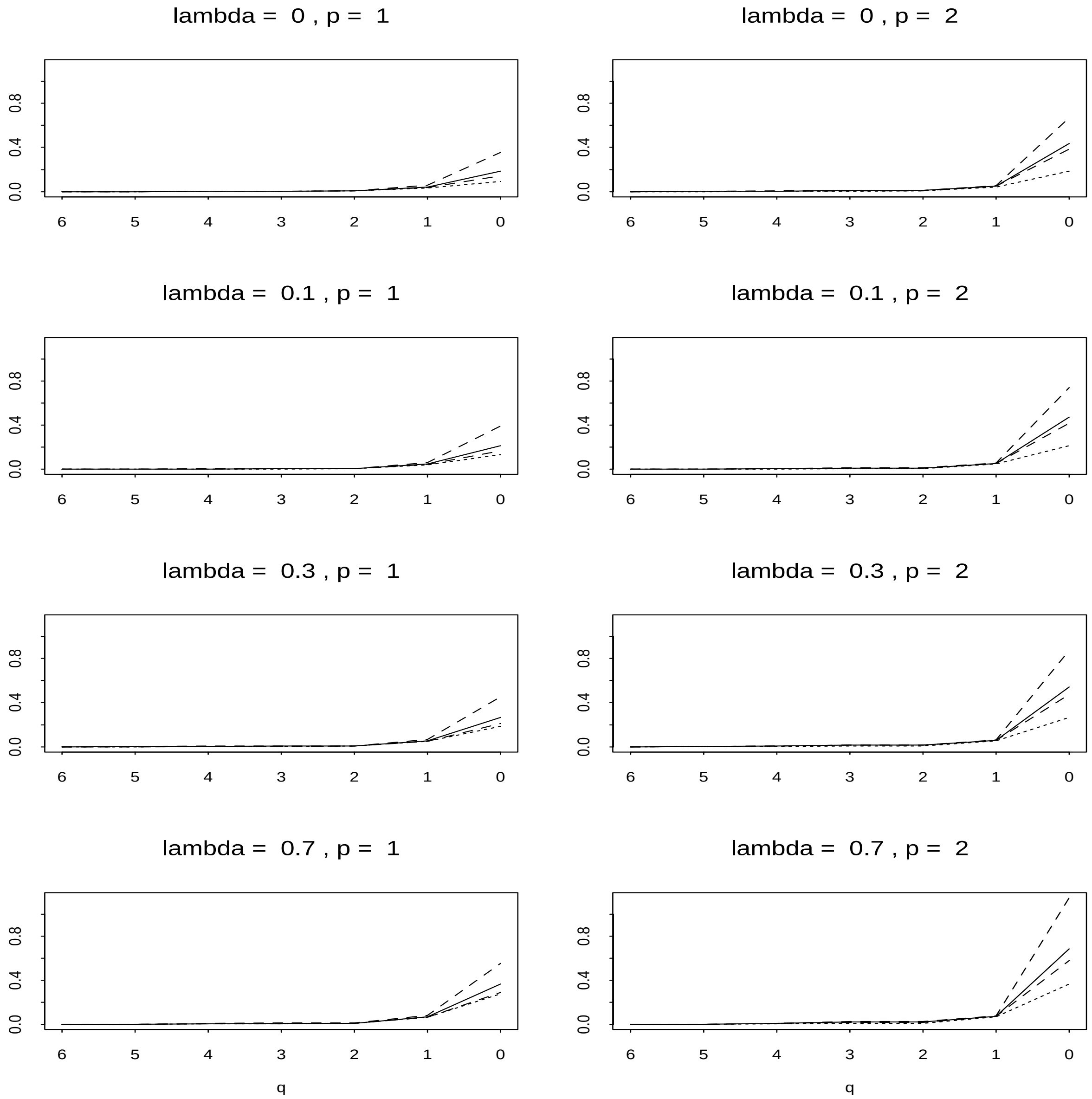


Figure 22: Plots of the mean squared revisions criterion R_q/σ^2 for the BLIP end filters based on the central X-11 Henderson filter and the local linear and quadratic models specified by λ and p. The solid lines correspond to the BLUP end filters or the BLIP end filters with $\hat{\beta}_p^2/\hat{\sigma}^2 = \infty$. The dashed lines correspond to $\hat{\beta}_p^2/\hat{\sigma}^2 = 4/(\pi(3.5)^2)$ (this value gives the X-11 end filters when $p = 1$, $\lambda = 0$) in the two cases where $\beta_p^2/\sigma^2 = 0$ and $\beta_p^2/\sigma^2 = 4/\pi$. The dotted lines correspond to $\hat{\beta}_p^2/\hat{\sigma}^2 = 0$ 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.

$(\hat{\beta}_p^2/\hat{\sigma}^2, \beta_p^2/\sigma^2)$. Plots of the lines

$$\frac{|\beta_p|}{\sigma} = \sqrt{2\frac{\hat{\beta}_p^2}{\hat{\sigma}^2} + (\mathbf{d}^T(\mathbf{C}^T\mathbf{E}_1^{-1}\mathbf{C})^{-1}\mathbf{d} + \gamma^T\mathbf{H}\gamma)/\sigma^2}$$

are given in Figure 23 for $q = 3$, $q = 0$ and the local dynamic models (55) and (56) based on the central X-11 Henderson filter and selected values of λ. From (54) the BLIP end filters have better mean squared revisions when $(|\hat{\beta}_p|/\hat{\sigma}, |\beta_p|/\sigma)$ lies below the lines. For large values of $|\hat{\beta}_p|/\hat{\sigma}$ the BLIP end filters approach the BLUP end filters and the boundary becomes $|\beta_p|/\sigma = \sqrt{2}|\hat{\beta}_p|/\hat{\sigma}$. However, for smaller values of $|\hat{\beta}_p|/\hat{\sigma}$ a somewhat greater range of possibilities is evident, especially for the larger values of λ. There is less latitude when $p = 2$ as might be expected. Given a range of (time-varying) values of $|\beta_p|/\sigma$ one would wish to select $|\hat{\beta}_p|/\hat{\sigma}$ as low as possible in order to maximise the gains in terms of expected revisions.

Now consider Figure 24 which plots $R_q(\hat{\phi}_0)/R_q(0)$, the mean squared revisions of the BLIP end filters normalised by the mean squared revisions of the BLUP end filters, as a function of $(|\hat{\beta}_p|/\hat{\sigma}, |\beta_p|/\sigma)$. Here only $q = 0$ is shown and the BLIP end filters are based on the local dynamic models (55) and (56), the central X-11 Henderson filter and selected values of λ, $|\beta_p|/\sigma$. From these plots and Theorem 10 a three dimensional picture of $R_q(\hat{\phi}_0)$ can be visualised. Looking out from the origin along the line $|\beta_p|/\sigma = |\hat{\beta}_p|/\hat{\sigma}$, $R_q(\hat{\phi}_0)$ has the appearance of a rising valley with a sharply increasing left hand side (representing the unacceptably high mean squared revisions incurred when $(|\hat{\beta}_p|/\hat{\sigma}, |\beta_p|/\sigma)$ does not satisfy (54)) and a right hand side that levels out at the BLUP limit $R_q(0)$. Note that the optimum mean squared revisions $R_q(\phi_0)$ occur when $|\beta_p|/\sigma = |\hat{\beta}_p|/\hat{\sigma}$ and, as noted in the discussion following Theorem 10, $R_q(\phi_0)$ is a monotonically increasing function of $|\beta_p|/\sigma$. Clearly the greatest gains are to be had when β_p^2/σ^2 and hence $\hat{\beta}_p^2/\hat{\sigma}^2$ are small.

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

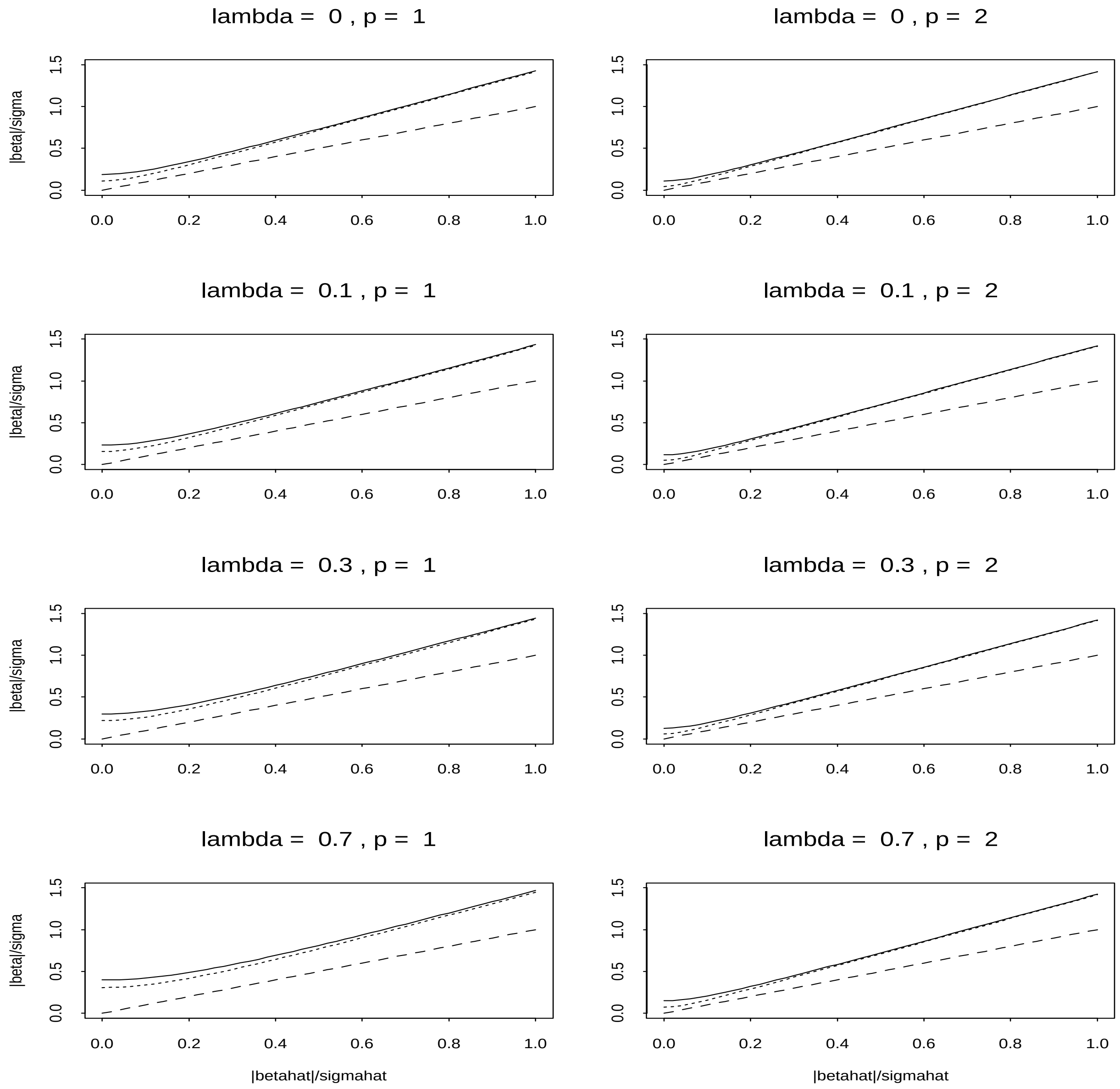


Figure 23: Plots of the lines in the $(|\hat{\beta}_p|/\hat{\sigma}, |\beta_p|/\sigma)$ plane below which BLIP end filters have better mean squared revisions than BLUP end filters. Here the BLIP end filters are based on the central X-11 Henderson filter, and the local linear and quadratic models are specified by p and λ. The solid lines correspond to $q = 0$, the dotted lines to $q = 3$ and the dashed line to the case $|\beta_p|/\sigma = |\hat{\beta}_p|/\hat{\sigma}$ when the optimum revisions are achieved for a given value of $|\beta_p|/\sigma$.

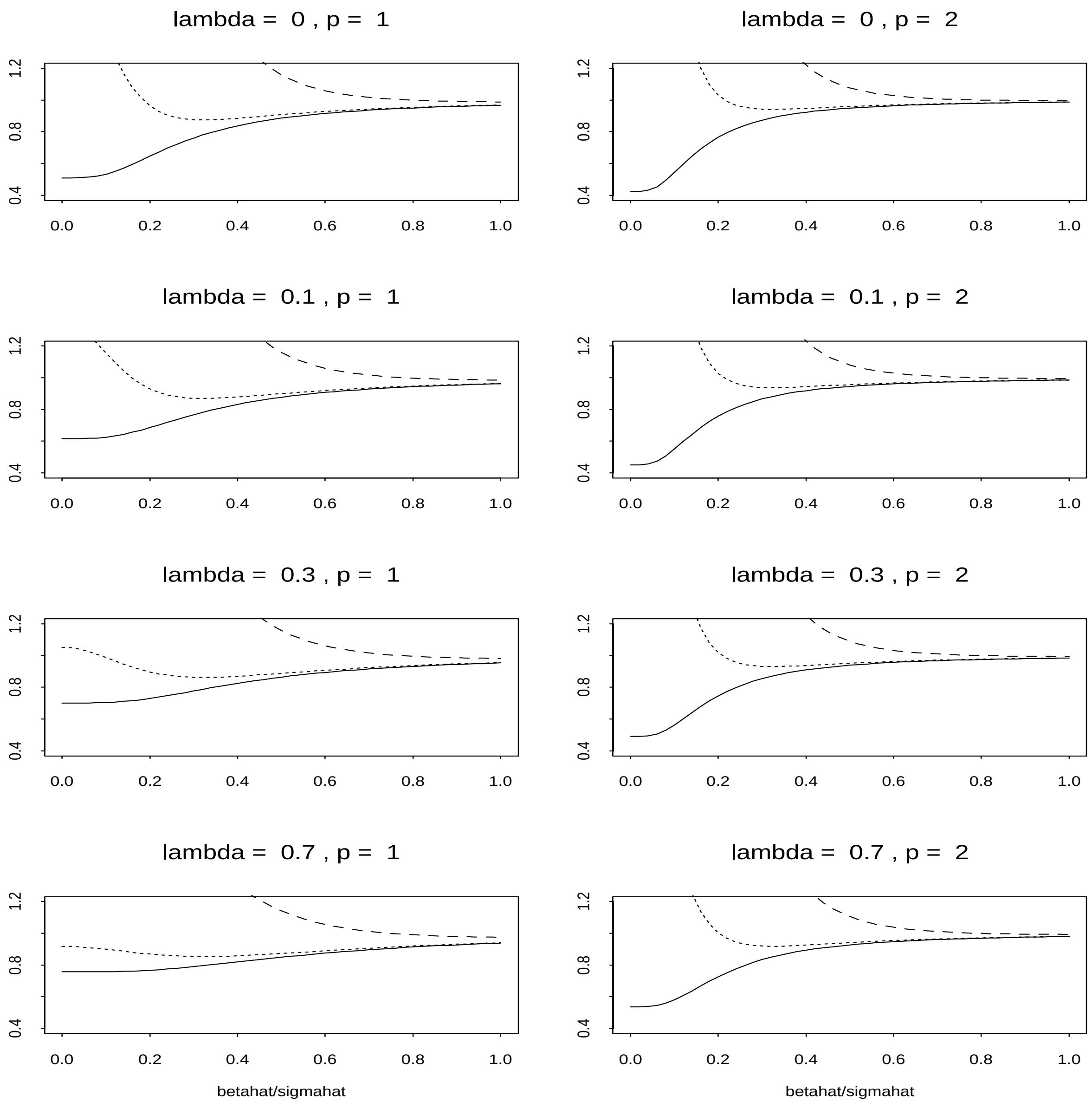


Figure 24: Plots of $R_q(\hat{\phi}_0)/R_q(0)$, the mean squared revisions of the BLIP end filters normalised by the mean squared revisions of the BLUP end filters, as a function of $(|\hat{\beta}_p|/\hat{\sigma}, |\beta_p|/\sigma)$. Here $q = 0$ and the BLIP end filters are based on the local linear $(p = 1)$ and quadratic $(p = 2)$ models, the central X-11 Henderson filter and selected values of λ, $|\beta_p|/\sigma$. The solid lines correspond to $\beta_p^2/\sigma^2 = 0$ (this yields the best mean squared revisions possible for any given value of $\hat{\beta}_p^2/\hat{\sigma}^2$), the dotted lines to $\beta_p^2/\sigma^2 = 4/(\pi(3.5)^2)$ and the dashed lines to $\beta_p^2/\sigma^2 = 4/\pi$. These values of β_p^2/σ^2 are derived from recommendations made by X-11 concerning I/C ratios.

4 Practical Study

We briefly report on the outcomes of a study on selected official time series where the central filters given by Theorem 1 were compared with conventional Macaulay or Henderson filters, and where the end filters given by Corollary 6 and Corollary 9 were compared with X-11 end filters and ARIMA forecast extension.

All the time series we examined were seasonal and some of them had large outliers. Since the trend filters under study are for non-seasonal time series and have not yet been modified to handle outliers, the first task in this analysis was to remove the seasonal component and large outliers from the time series so that we could examine the relative performance of these filters. For simplicity we used the modified seasonally adjusted series from the X-11 seasonal adjustment method as the data for this study. For the New Zealand series we chose a window of the series which we believed would give a good X-11 decomposition. Other than choosing the commonly used limits of 1.8 and 2.8 for treating outliers we used the default options of X-11 to produce the seasonally adjusted data.

The resulting non-seasonal times series presented different types of short and long term trends and their local dynamic models included random walk components with a range of variances. In this section we shall discuss five monthly series representing this range of trend behaviour. These are as follows.

Building Permits New Zealand data on the number of Building Permits issued by Local Authorities for the construction of private houses and flats. A subset of the data has been taken in an attempt to get a consistent series where for example there was no Government intervention in Housing Policy.

Merchandise Exports New Zealand data on the value of Merchandise Exports. A subset of the data has been taken in an attempt to get a consistent series where the economy has adjusted to the impact of deregulation in the mid 1980s.

NZ Furniture Sales New Zealand data on the value of Retail Sales in Furniture Stores.

Permanent Migration New Zealand data on the Net Permanent and Long Term Migration.

US Furniture Sales United States of America data on the value of Retail Sales in Furniture Stores in the North West Region.

4.1 Fitting the Local Dynamic Model

To study the filters given by Theorem 1, the window length $n = 2r + 1$, the order of the polynomial p, and a value of λ must be determined from the data. These parameters specify the local dynamic model concerned. Once the local dynamic model has been identified, the user will need to decide on a value for θ, the balance between fidelity and smoothness. Although subjective, the choice of θ will typically be influenced by the overall objectives of the smoothing, as is to some extent, the choice of n.

Typically n and p are identified from a simple graphical analysis of the data although more formal methods could be used. Given n and p, there are a variety of methods for determining λ. These range from trial and error and simple variational arguments based on quantities like X-11's I/C ratio, through to likelihood analysis based on fitting the local dynamic model within non-overlapping windows, etc. Which approach turns out to be most useful is still very much an open question and the subject of further research.

Consider for example the local dynamic model given by (55). For this model the first difference of the data is given by

$$\Delta y_t = \beta_1 + \eta_t + \Delta\epsilon_t, \tag{63}$$

which is locally an MA(1) process with non-zero mean β_1. Here the local parameters β_1 and σ^2 are regarded as constant within any window, but can evolve slowly across windows, and $\lambda = \sigma_\eta^2/\sigma^2$ is a fixed global parameter. If there were no evolution in β_1 and σ^2 across windows then one could fit a global MA(1) model to the data to estimate λ. However, any significant time evolution in β_1 and σ^2 will bias the estimates of λ obtained in this way.

More generally one might write (63) as

$$\Delta y_t = m_t + s_t u_t, \tag{64}$$

where the location and scale parameters $m_t = \beta_1$ and s_t are evolving slowly over time, but u_t is a a global MA(1) process with time invariant parameters and lag one autocorrelation $-1/(\lambda+2)$. This suggests recovering λ by fitting a global MA(1) model to the standardised series $(\Delta y_t - \hat{m}_t)/\hat{s}_t$ where $\hat{m}_t$ and $\hat{s}_t$ are simple non-parametric estimates of m_t and s_t. Typically, we would expect m_t to be highly correlated and the variance of m_t to be much smaller than the variance of Δy_t.

To get an idea of the global behaviour of m_t and s_t we estimated m_t non-parametrically by smoothing Δy_t with a standard low pass filter (a 2×12 moving average) to give $\hat{m}_t$, and then estimated s_t by smoothing $|\Delta y_t - \hat{m}_t|$ with the same filter to obtain $\hat{s}_t$. These quantities and their autocorrelation functions are given in Figures 25–27 for Building Permits, Exports and Permanent Migration respectively. Of course these plots should be interpreted with care due to the filtering that has taken place. However, they do provide a good indication of the potential fit of the local dynamic model in practice. In particular, and as might be expected, m_t is typically smooth with strong positive autocorrelation and the autocorrelation of Δy_t is modified to a greater or lesser degree depending on the relative magnitudes of the variances of m_t and $\Delta y_t - \hat{m}_t$ (i.e. the signal to noise ratio). The Building Permits data (see Figure 25) is a good example of a highly correlated m_t with low signal to noise ratio. Here the global ARIMA model proved hard to fit and had poor fit, whereas the local ARIMA structure is simple and seems appropriate. The Permanent Migration data (Figure 27) again has a highly correlated m_t, but a higher signal to noise ratio and more variable s_t.

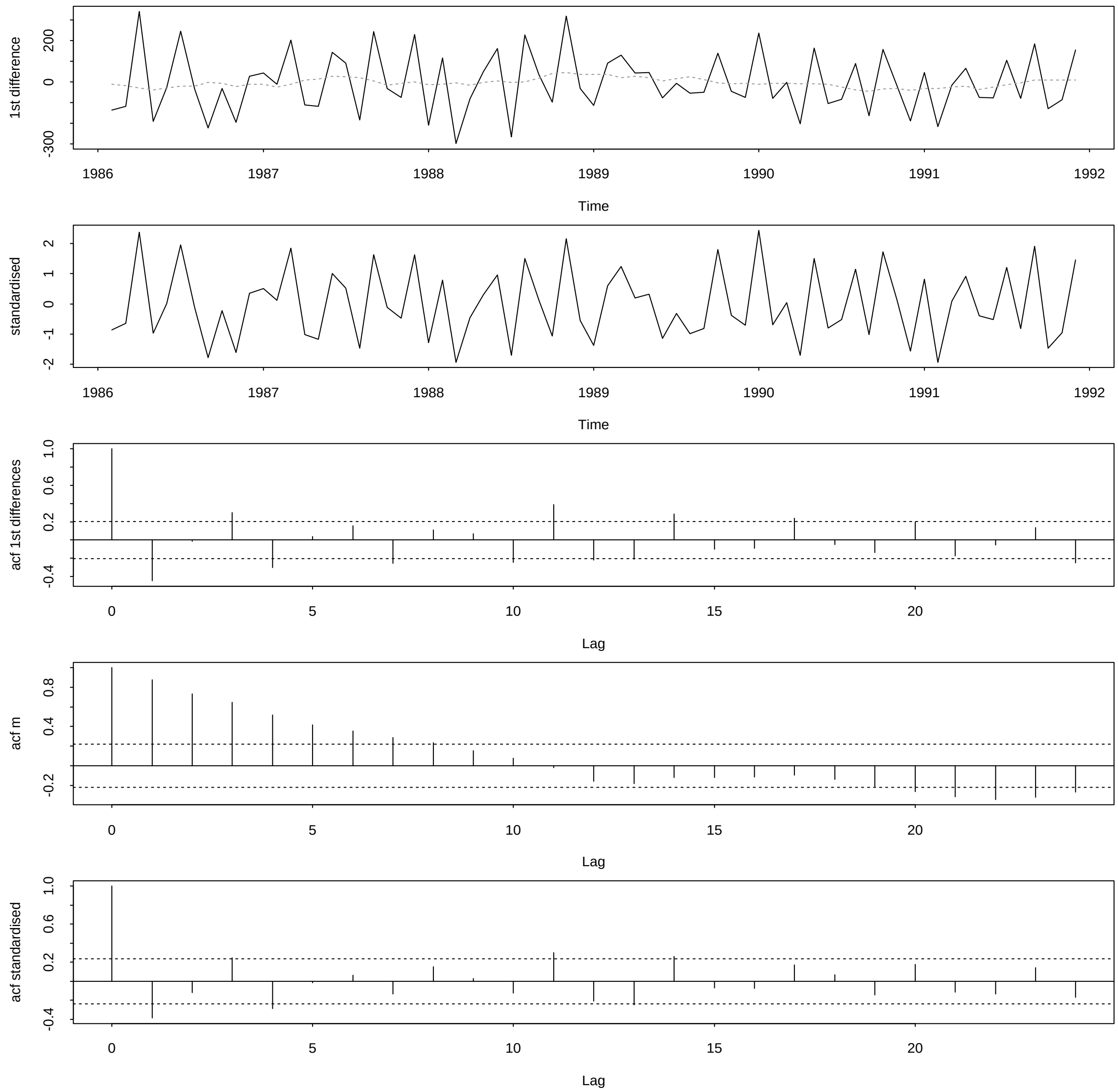


Figure 25: Building Permits data. The top graph shows the location estimate $\hat{m}_t$ superimposed on the the first differences of the data and the second graph shows the standardised series $(\Delta y_t - \hat{m}_t)/\hat{s}_t$. Here $\hat{m}_t$ and $\hat{s}_t$ are obtained by smoothing Δy_t and $|\Delta y_t - \hat{m}_t|$ respectively with a standard low pass filter (a 2×12 moving average). The remaining graphs are the autocorrelation functions of Δy_t, $\hat{m}_t$ and $(\Delta y_t - \hat{m}_t)/\hat{s}_t$ respectively.

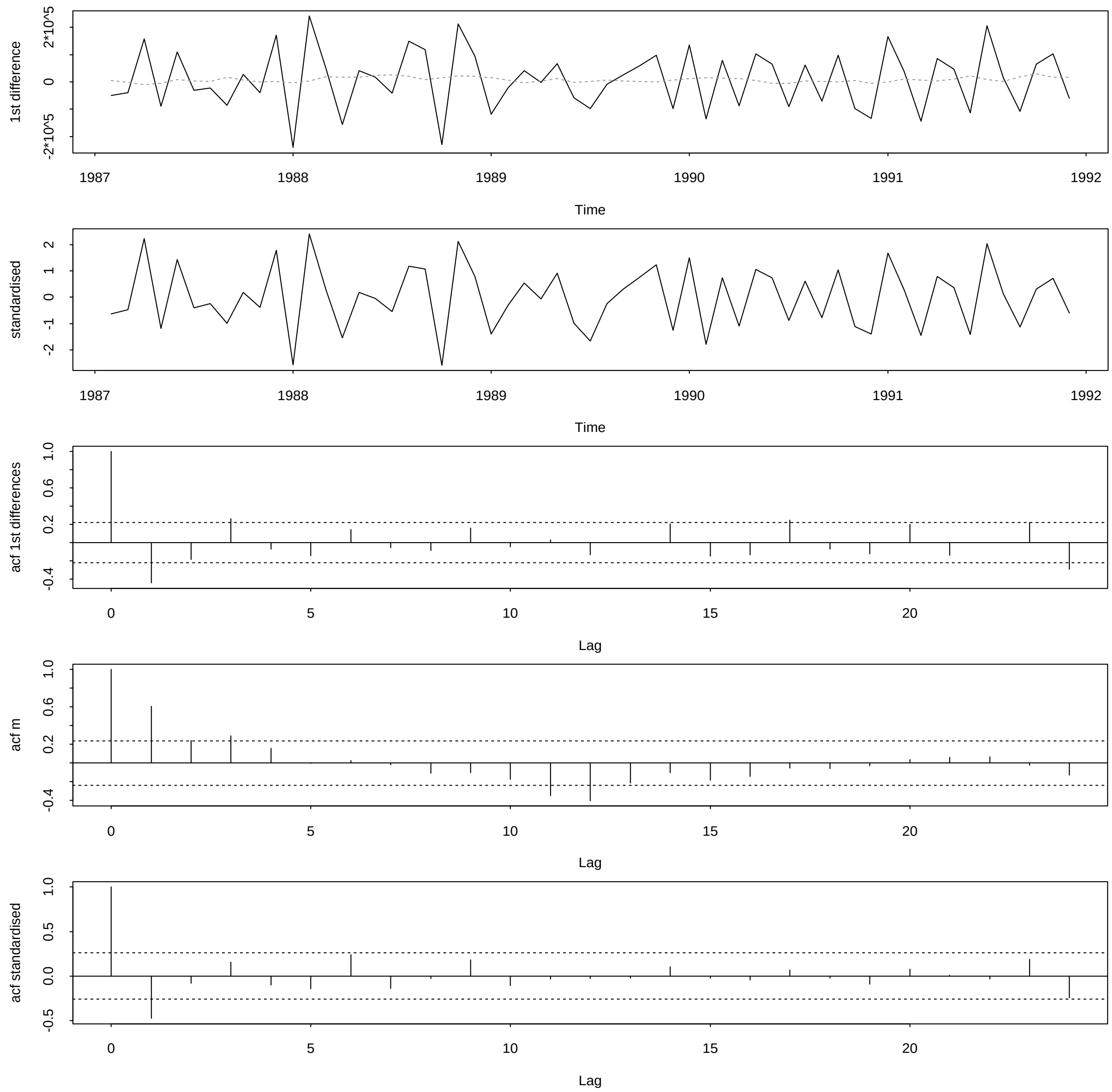


Figure 26: Merchandise Exports data. The top graph shows the location estimate $\hat{m}_t$ superimposed on the the first differences of the data and the second graph shows the standardised series $(\Delta y_t - \hat{m}_t)/\hat{s}_t$. Here $\hat{m}_t$ and $\hat{s}_t$ are obtained by smoothing Δy_t and $|\Delta y_t - \hat{m}_t|$ respectively with a standard low pass filter (a 2×12 moving average). The remaining graphs are the autocorrelation functions of Δy_t, $\hat{m}_t$ and $(\Delta y_t - \hat{m}_t)/\hat{s}_t$ respectively.

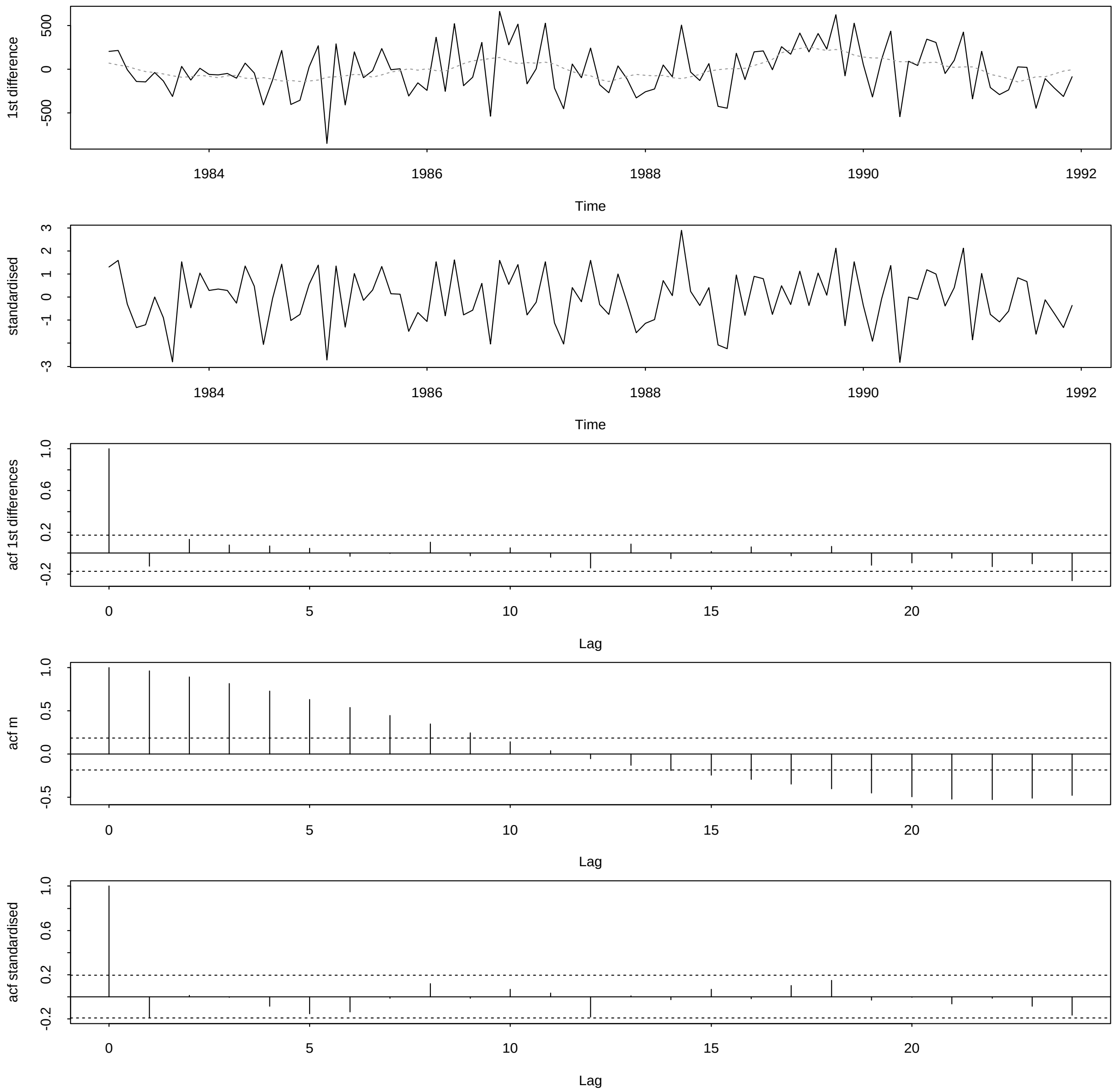


Figure 27: Permanent Migration data. The top graph shows the location estimate $\hat{m}_t$ superimposed on the the first differences of the data and the second graph shows the standardised series $(\Delta y_t - \hat{m}_t)/\hat{s}_t$. Here $\hat{m}_t$ and $\hat{s}_t$ are obtained by smoothing Δy_t and $|\Delta y_t - \hat{m}_t|$ respectively with a standard low pass filter (a 2×12 moving average). The remaining graphs are the autocorrelation functions of Δy_t, $\hat{m}_t$ and $(\Delta y_t - \hat{m}_t)/\hat{s}_t$ respectively.

However, the fitting of a global ARIMA model to data standardised in this way needs to be handled with care. Rather than follow this strategy, we adopted a simpler and more direct method of estimating λ that took advantage of the unbiased BLUP predictors.

Consider the family of BLUP end filters based on Corollary 6 for a given central filter and given values of n and p. Note that since these end filters are unbiased they depend only on q and λ. Applying the end filters to the data yields the revisions

$$r_t(q,\lambda) = \sum_{-r}^{r} w_s y_{t+s} - \sum_{-r}^{q} u_s(\lambda) y_{t+s}$$

where the w_s are the given central filter weights and the $u_s(\lambda)$ are the BLUP end filter weights given by Corollary 6. An appropriate cost function such as

$$P = \sum_{q=0}^{r-1} \alpha_q r_t^2(q,\lambda) \tag{65}$$

can now be constructed with the given positive weights α_q reflecting the relative costs of the respective revisions. Finally λ is determined by minimising P with respect to λ.

For our study $n = 13$, $p = 1$ and the central filter chosen was the central X-11 Henderson filter. Moreover we chose $\alpha_0 = 1$ and $\alpha_q = 0$ for $q \neq 0$ yielding a cost function which focused on the case of worst revisions. However, other central filters and other values for the α_q could have been chosen and would lead to different results. Again, more research is needed, but it is believed that the results presented here are indicative of what might be achieved more generally. With these caveats, this procedure was applied to Building Permits, Exports and Permanent Migration, yielding values for λ of 0.046, 0.52, and 3.8 respectively.

Finally we consider the estimation of the ratio β_p/σ which involves the local parameters β_p and σ. Again, local smoothing should provide reasonable measures of β_p, σ and, in turn, the ratio β_p/σ. Consider the local linear model (55) and the approach described earlier with regard to determining λ from the standardised data $(\Delta y_t - \hat{m}_t)/\hat{s}_t$. In this case $\hat{m}_t$ should give a reasonable estimate of β_1 and $\hat{s}_t$ should estimate

$$E|\Delta y_t - \beta_1| = \sigma\sqrt{\frac{2}{\pi}(\lambda+2)} \tag{66}$$

under (local) Gaussian assumptions. Thus an appropriately scaled form of $\hat{m}_t/\hat{s}_t$ should provide a reasonable measure of β_1/σ and its evolution.

Examples of estimated $|\beta_p|/\sigma$ ratios for the local linear model given by (55) and obtained in this way are plotted in Figures 28–30 for Building Permits, Exports and Permanent Migration respectively. It is clear that these ratios undergo considerable evolution relative to their means with significant variation associated with the inflection points of the original series (see, for example, Permanent Migration in Figure 30). Clearly the greater one's knowledge of this local variation, the greater the gains using BLIP end filters, especially at turning points.

Although local estimates of $|\beta_p|/\sigma$ should be explored in order to minimise revisions from BLIP end filters, this option was left for further study. Instead we adopted the more conservative strategy of estimating a common value for $|\beta_p|/\sigma$ for each of the series considered. A natural first choice would be the mean (or median) of the estimated ratios plotted in Figures 28–30. However, consideration of the plots in Figures 23 and 24 for given values of λ suggests that this will not, in general, be an optimal choice because of the complex and non-symmetric nature of the mean squared revisions as a function of $|\hat{\beta}_p|/\hat{\sigma}$ and $|\beta_p|/\sigma$. Given a range of values for $|\beta_p|/\sigma$ (such as those indicated in Figures 28–30), any increase in $|\hat{\beta}_p|/\hat{\sigma}$ will involve a trade off between the reduction in the (high) revisions experienced by large values of $|\beta_p|/\sigma$ with the increase in (low) revisions experienced by the low values of $|\beta_p|/\sigma$. Thus an overall choice for $|\hat{\beta}_p|/\hat{\sigma}$ requires a complex balancing of marginal gains and losses. This clearly depends on the value of λ as evidenced, for example, by the nature of the mean square revisions function for $p = 1$, $\lambda = 0$ and $p = 1$, $\lambda = 0.7$ given in Figure 24.

To determine a global value for $|\hat{\beta}_p|/\hat{\sigma}$ from the data we adopted a similar approach to that used for determining λ, but now focused on the BLIP end filters given by Corollary 9. Here we determined the revisions

$$\tilde{r}_t(q, |\hat{\beta}_p|/\hat{\sigma}) = \sum_{-r}^{r} w_s y_{t+s} - \sum_{-r}^{q} u_s(|\hat{\beta}_p|/\hat{\sigma}) y_{t+s}$$

where the w_s are the known central filter weights, the $u_s(|\hat{\beta}_p|/\hat{\sigma})$ are the BLIP end filter weights given by Corollary 9 and n, p are given. The value of λ is set at its previously estimated value. Now the cost function $\tilde{P}$ given by (65) with r_t replaced by $\tilde{r}_t$ can be evaluated and minimised with respect to $|\hat{\beta}_p|/\hat{\sigma}$.

As before we chose $\alpha_0 = 1$, $\alpha_q = 0$ for $q \neq 0$ and set $n = 13$, $p = 1$ with the central filter chosen to be the central X-11 Henderson filter. Minimising $\tilde{P}$ in this way for Building Permits, Exports and Permanent Migration, yielded global values for $|\hat{\beta}_p|/\hat{\sigma}$ of 0.22, 0.3, and 0.32 respectively. Note that in the case of Exports we chose between two competing local minima on the basis of Figure 29; in the other two cases the minima were clearly delineated. As was noted before in relation to the estimation of λ, different central filters and different values of α_q could have been chosen and would give different results. The determination of λ, $|\hat{\beta}_p|/\hat{\sigma}$ globally and $|\hat{\beta}_p|/\hat{\sigma}$ locally remain topics for further research.

4.2 In the body

Here we discuss the behaviour of the central trend filters given by Theorem 1 on the data sets described in the introduction to Section 4. The effects on the trend estimates for these series of varying λ for given choice of θ are displayed in Figure 31–35 for the local

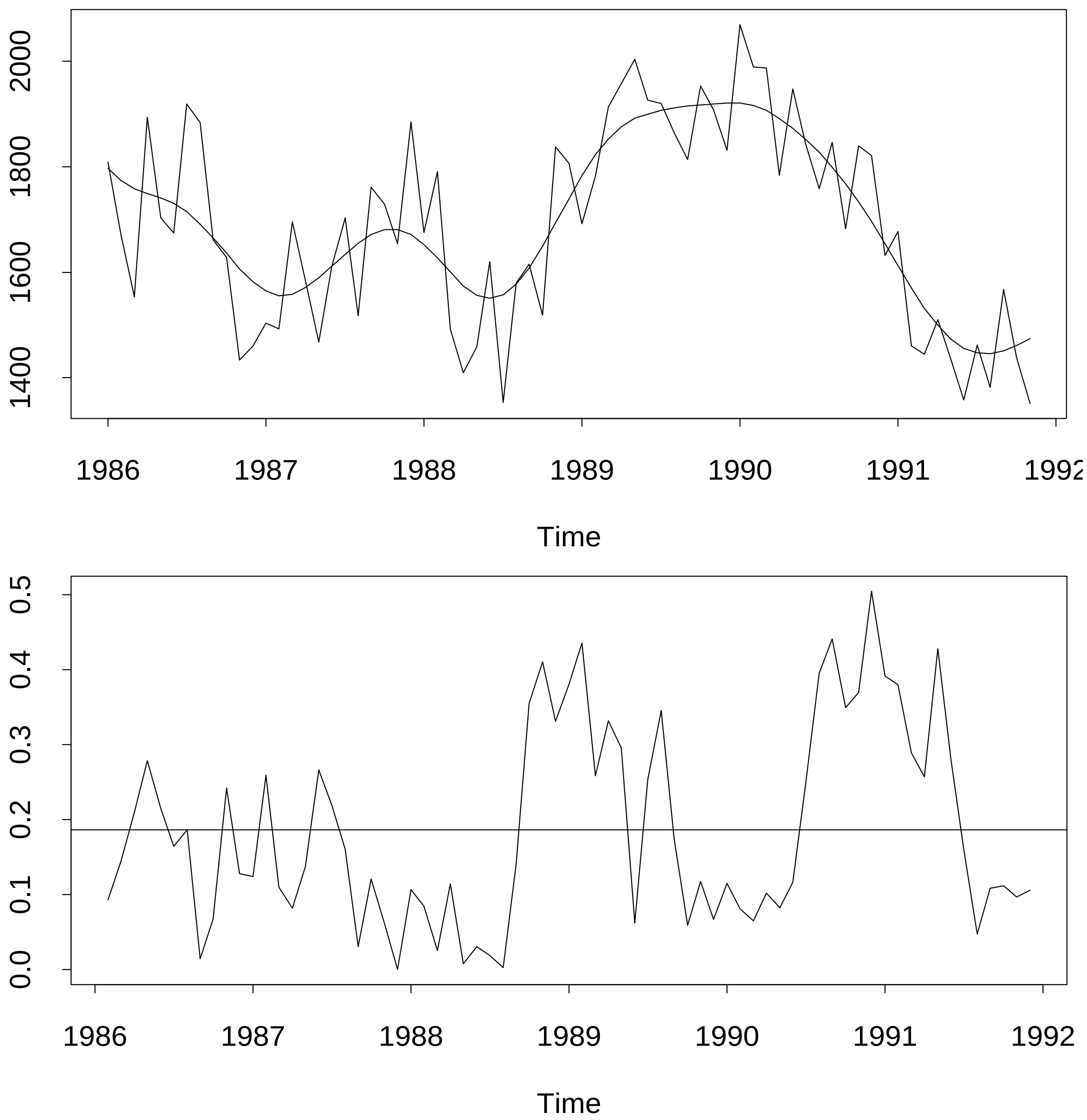


Figure 28: The top graph is Building Permits data and the smooth solid line is the trend from the 13 point filter given by Theorem 1 for the local linear model with $\lambda = .046$ and $\theta = 0$. The bottom graph is the $|\hat{\beta}_p|/\hat{\sigma}$ ratio where β_p is estimated by $\hat{m}_t$, a 2×12 moving average of Δy_t, and σ is estimated by a scaled 2×12 moving average of the absolute residuals $|\Delta y_t - \hat{m}_t|$ where the scale factor is $1/\sqrt{\frac{2}{\pi}(\lambda + 2)}$ (see (66)). The solid line is the mean of the $|\hat{\beta}_p|/\hat{\sigma}$ ratios (0.19) which is well under the default X-11 value of 0.31. These ratios undergo considerable evolution relative to their mean with significant variation associated with the inflection points of the original series.

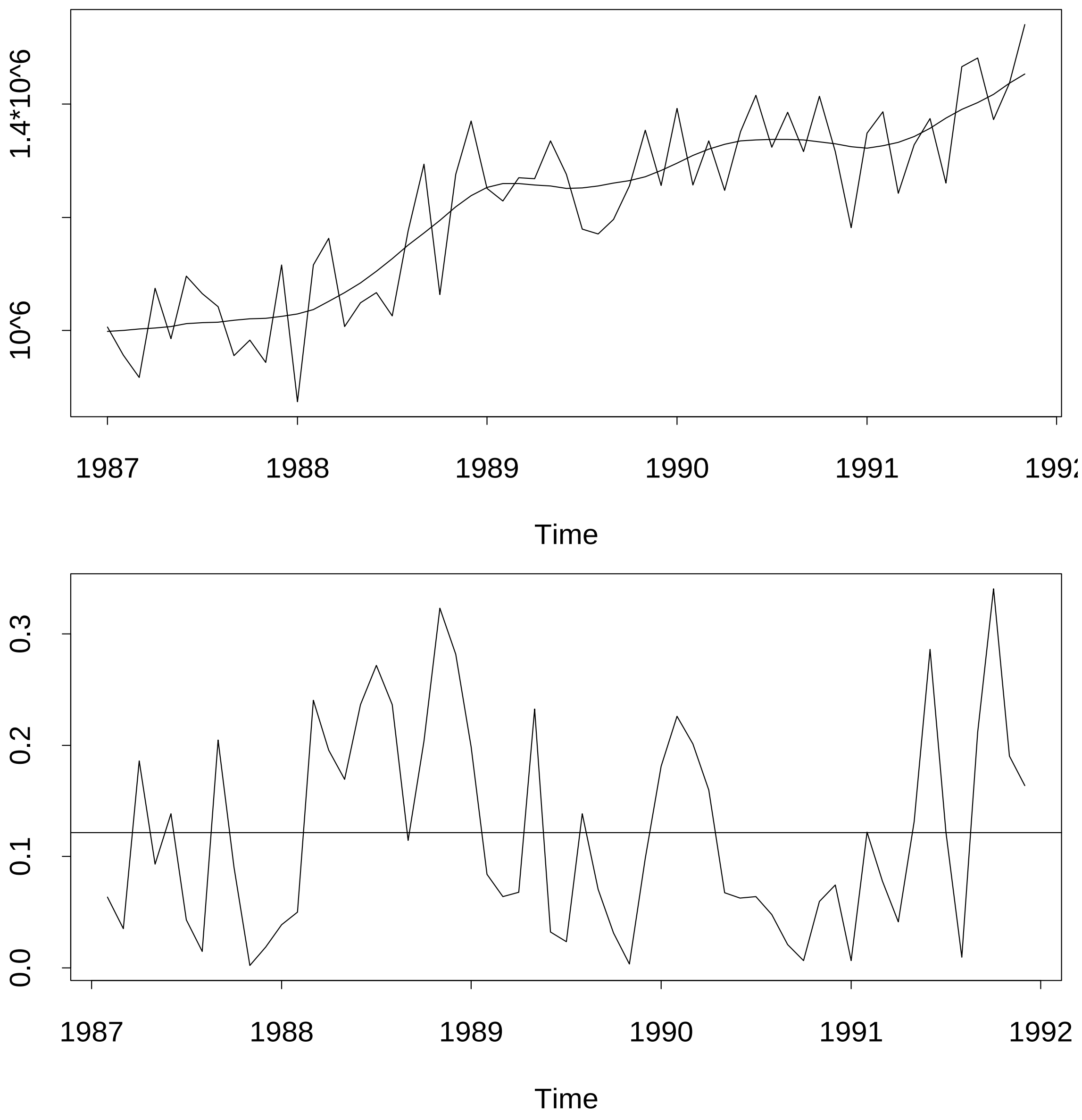


Figure 29: The top graph is Merchandise Exports data and the smooth solid line is the trend from the 13 point filter given by Theorem 1 for the local linear model with $\lambda = .52$ and $\theta = 0$. The bottom graph is the $|\hat{\beta}_p|/\hat{\sigma}$ ratio where β_p is estimated by $\hat{m}_t$, a 2×12 moving average of Δy_t, and σ is estimated by a scaled 2×12 moving average of the absolute residuals $|\Delta y_t - \hat{m}_t|$ where the scale factor is $1/\sqrt{\frac{2}{\pi}(\lambda + 2)}$ (see (66)). The solid line is the mean of the $|\hat{\beta}_p|/\hat{\sigma}$ ratios (0.12) which is well under the default X-11 value of 0.31. These ratios undergo considerable evolution relative to their mean with significant variation associated with the inflection points of the original series.

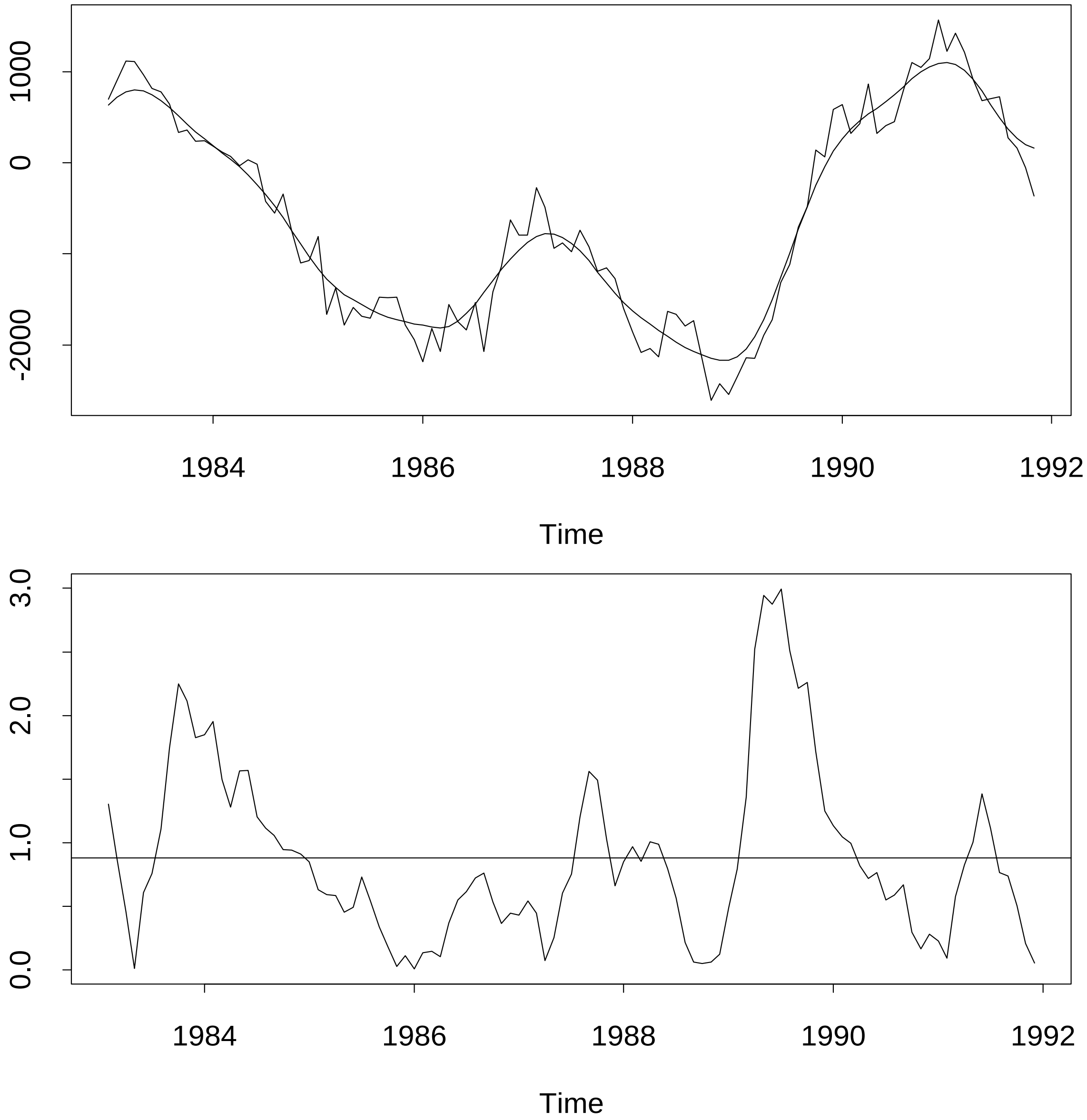


Figure 30: The top graph is Permanent Migration data and the smooth solid line is the trend from the 13 point filter given by Theorem 1 for the local linear model with $\lambda = 3.8$ and $\theta = 0$. The bottom graph is the $|\hat{\beta}_p|/\hat{\sigma}$ ratio where β_p is estimated by $\hat{m}_t$, a 2×12 moving average of Δy_t, and σ is estimated by a scaled 2×12 moving average of the absolute residuals $|\Delta y_t - \hat{m}_t|$ where the scale factor is $1/\sqrt{\frac{2}{\pi}(\lambda + 2)}$ (see (66)). The solid line is the mean of the $|\hat{\beta}_p|/\hat{\sigma}$ ratios (0.88) which is well under the default X-11 value of 0.31. These ratios undergo considerable evolution relative to their mean with significant variation associated with the inflection points of the original series.

linear model given by (55) and for window length $n = 13$. Two choices of θ are displayed; $\theta = 0$ which maximises smoothness and $\theta = .5$ which maximises a compromise between smoothness and fidelity.

The results obtained are as expected from the discussion given in Section 3. An example is given in Figure 31. Here, as λ varies, the Henderson filters which optimise smoothness S ($\theta = 0$) differ only slightly. However, the filters based on a mixture of smoothness and fidelity become more adaptive as λ increases with this effect becoming more marked the closer θ is to 1.

In particular note that the effect of varying λ is greatest at turning points. In Figures 31–35 the central filters based on a mixture of fidelity and smoothness ($\theta = 0.5$) and large λ were more faithful to the data than those based on $\theta = 0.5$ and small λ. Indeed, in Figure 35, the central filter based on the local linear model ($p = 1$) with large λ and $\theta = 0.5$ approaches the central X-11 Henderson filter which is based on a local quadratic model ($p = 2$) with $\theta = 0$ and $\lambda = 0$. This feature is also evident in Figures 31–34. In this sense, the local dynamic model is able to explore trend variation in the window that lies somewhere between the pure linear and pure quadratic case.

4.3 At the ends

In this subsection we consider the comparative performance of the BLIP, BLUP, X-11 and ARIMA forecast extension filters on the data sets under study. Here, as before, the window length is $n = 13$ and only the local liner model ($p = 1$) is considered with λ and a global value for $|\hat{\beta}_p|/\hat{\sigma}$ determined by the methods described in Section 4.1. As mentioned in that section, a more appropriate procedure would have been to determine local values for $|\hat{\beta}_p|/\hat{\sigma}$ and use these in the BLIP end filters. However this was not done and remains a research topic for further study. It might be expected that the use of local values for $|\hat{\beta}_p|/\hat{\sigma}$ would lead to BLIP end filters with improved performance over those using a global value for $|\hat{\beta}_p|/\hat{\sigma}$.

The following empirical comparison of the competing end filters was undertaken. First, for each local window we calculated the trend for the central point of that window using the central filter (6). Next, for each q and 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 (32), where the u_s are the weights of the end filter determined by Corollary 6 and 9 respectively. For the X-11 case the standard X-11 end filters were applied. Recall that these are given by Corollary 9 using the central X-11 Henderson filter, $\hat{\beta}_p^2/\hat{\sigma}^2 = 4/(\pi(3.5)^2)$ and the local linear model with $\lambda = 0$. For the ARIMA forecast extension case this meant predicting the trend as in (31) with y_{t+s} now predicted using the forecast function of a global ARIMA model restricted to data up to and including the local window. To obtain appropriate ARIMA model parameters we fitted global ARIMA models to all the data. This was done for reasons of simplicity and to provide stable estimates. Using standard diagnostics we selected ARIMA (0,1,1) models for Building Permits and Permanent Mi-

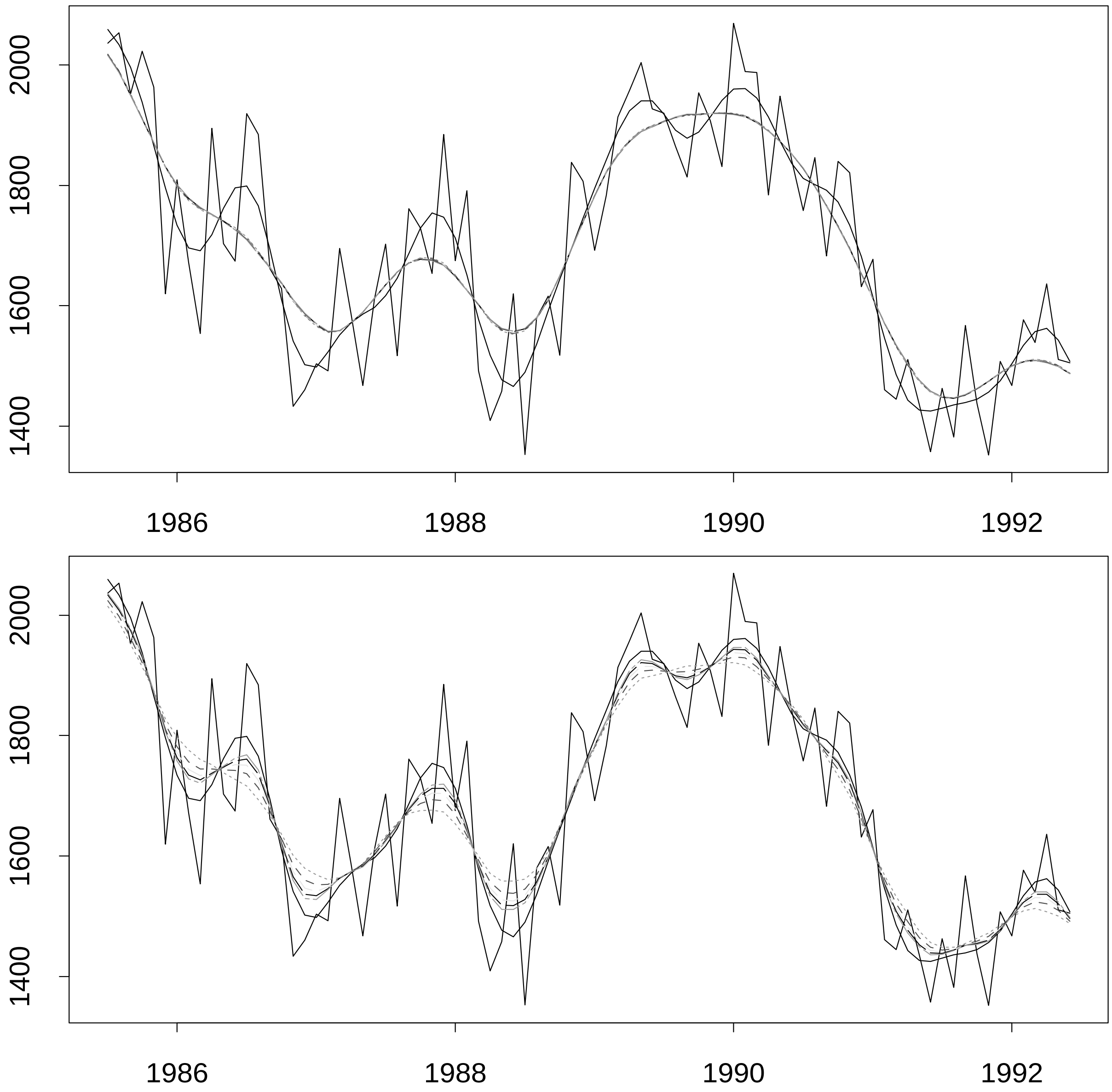


Figure 31: The Building Permits data is characterised by many local turning points, with a general downward drift. In the top graph, the data is filtered by the 13 point X-11 Henderson filter and the 13 point filters given by Theorem 1 for the local linear model and $\theta = 0$. In the bottom graph, the data is filtered by the 13 point X-11 Henderson filter and the 13 point filters given by Theorem 1 for the local linear model and $\theta = .5$. In both graphs, the solid line corresponds to the 13 point X-11 Henderson filter, the dotted and dashed lines to the 13 point filters given by Theorem 1 with $\lambda = .1, .2, .3, .4, .5$ respectively.

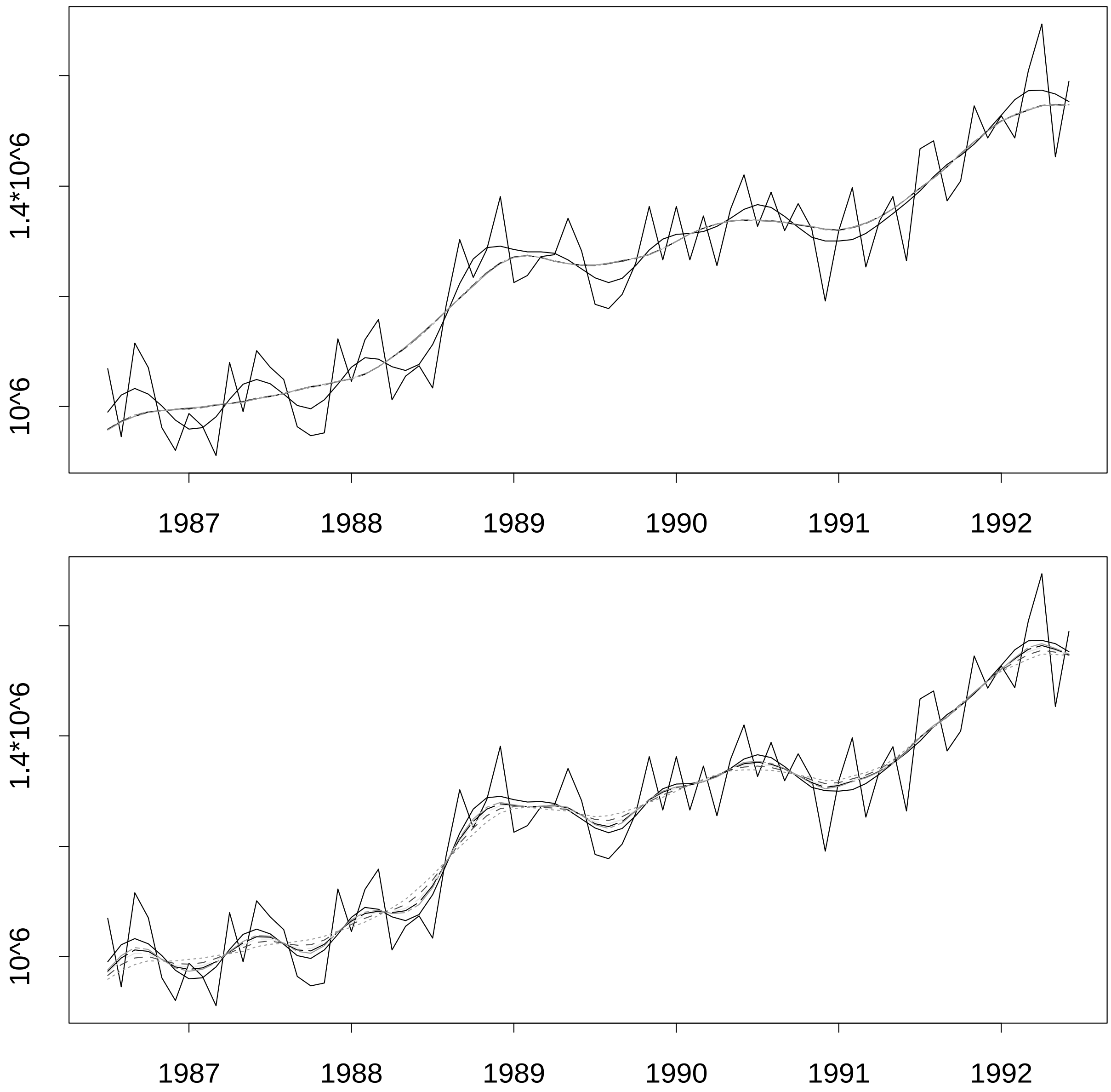


Figure 32: The Merchandise Exports data is characterised by a few weak local turning points, with a generally large upward drift. In the top graph, the data is filtered by the 13 point X-11 Henderson filter and the 13 point filters given by Theorem 1 for the local linear model and $\theta = 0$. In the bottom graph, the data is filtered by the 13 point X-11 Henderson filter and the 13 point filters given by Theorem 1 for the local linear model and $\theta = .5$. In both graphs, the solid line corresponds to the 13 point X-11 Henderson filter, the dotted and dashed lines to the 13 point filters given by Theorem 1 with $\lambda = .1, .2, .3, .4, .5$ respectively.

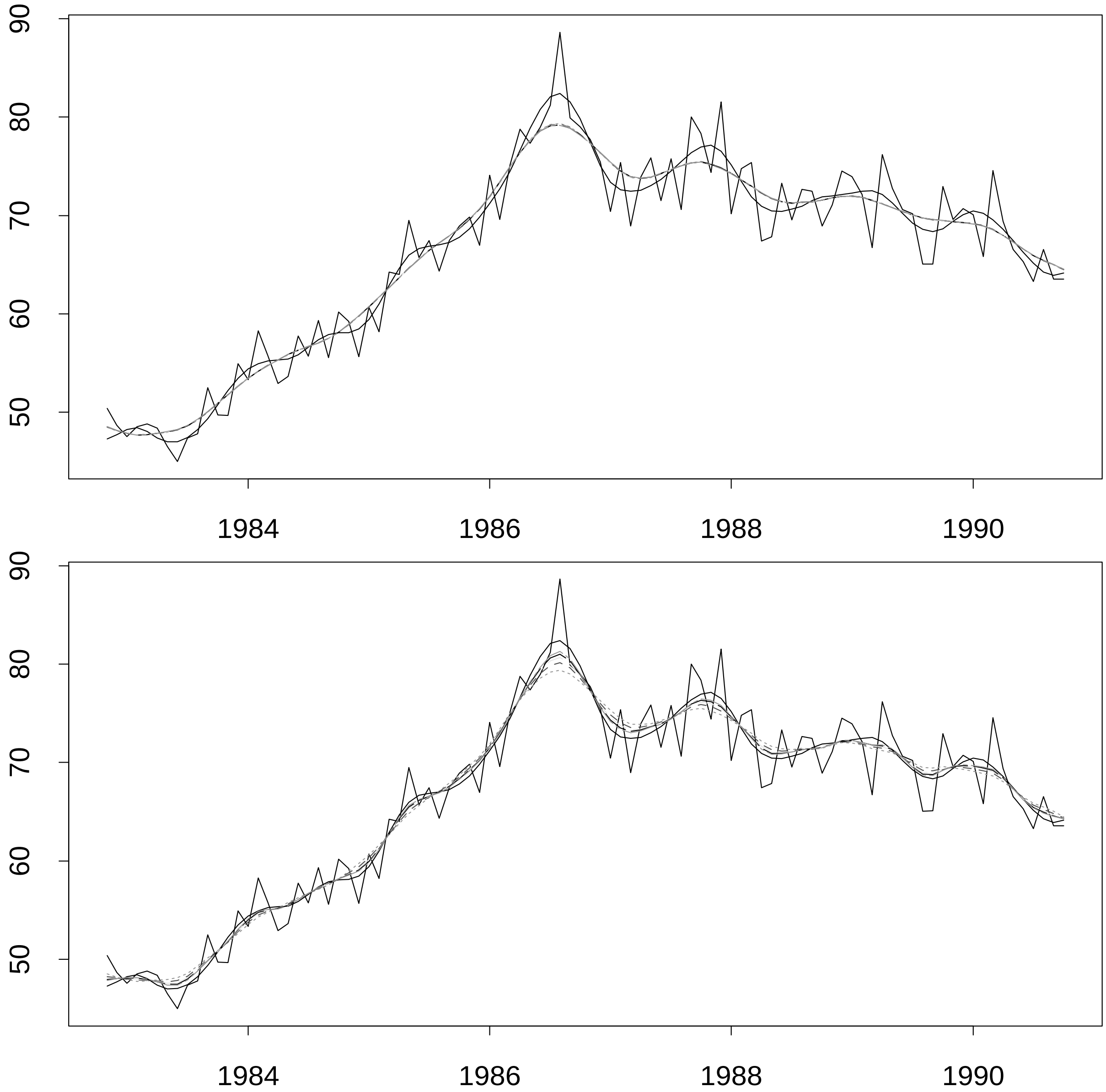


Figure 33: The NZ Furniture Sales data is characterised by a strong turning point with a large downward drift still continuing. In the top graph, the data is filtered by the 13 point X-11 Henderson filter and the 13 point filters given by Theorem 1 for the local linear model and $\theta = 0$. In the bottom graph, the data is filtered by the 13 point X-11 Henderson filter and the 13 point filters given by Theorem 1 for the local linear model and $\theta = .5$. In both graphs, the solid line corresponds to the 13 point X-11 Henderson filter, the dotted and dashed lines to the 13 point filters given by Theorem 1 with $\lambda = .1, .2, .3, .4, .5$ respectively.

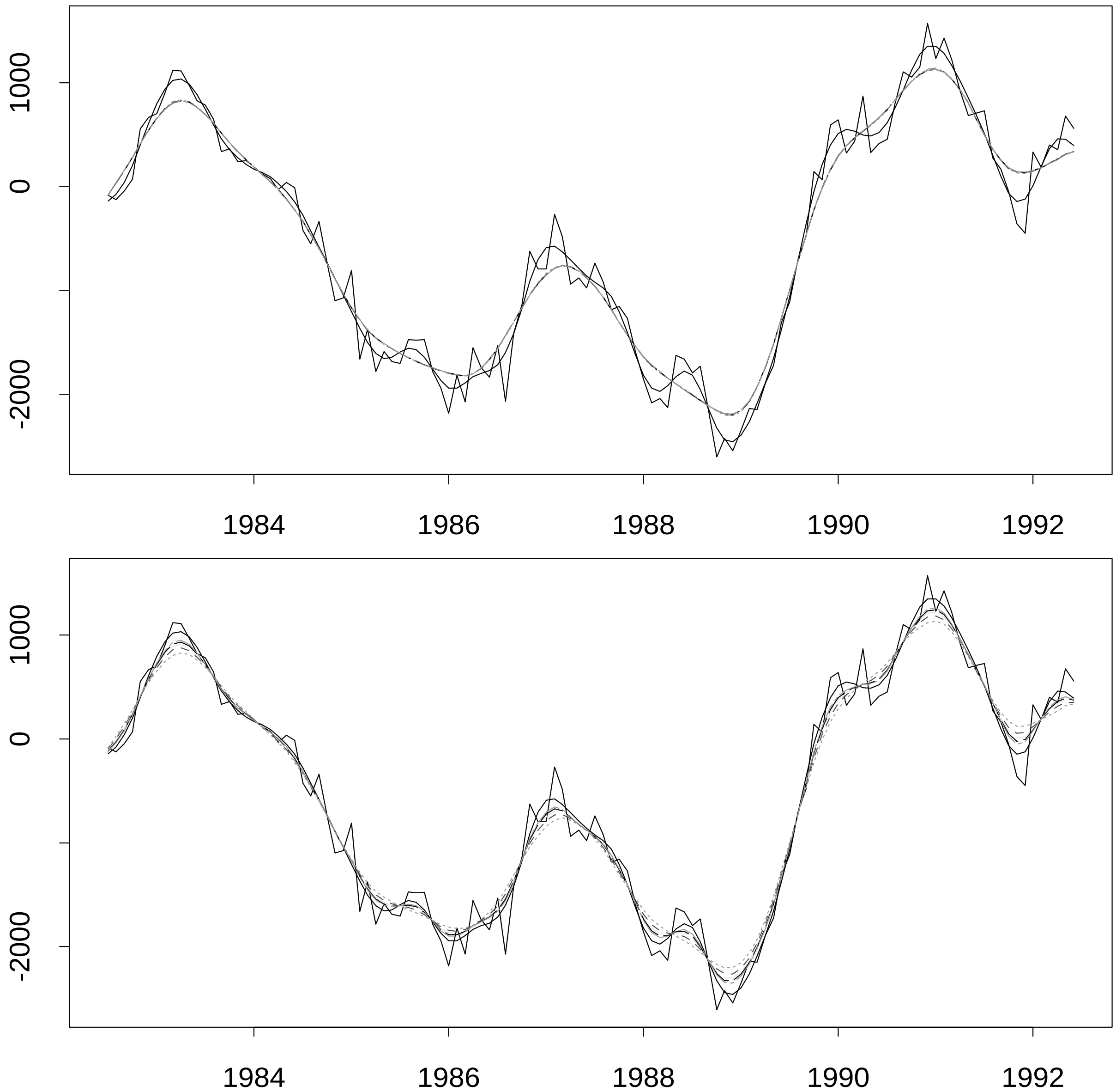


Figure 34: The Permanent Migration data is characterised by many local turning points, with almost zero drift long term. In the top graph, the data is filtered by the 13 point X-11 Henderson filter and the 13 point filters given by Theorem 1 for the local linear model and $\theta = 0$. In the bottom graph, the data is filtered by the 13 point X-11 Henderson filter and the 13 point filters given by Theorem 1 for the local linear model and $\theta = .5$. In both graphs, the solid line corresponds to the 13 point X-11 Henderson filter, the dotted and dashed lines to the 13 point filters given by Theorem 1 with $\lambda = .1, .2, .3, .4, .5$ respectively.

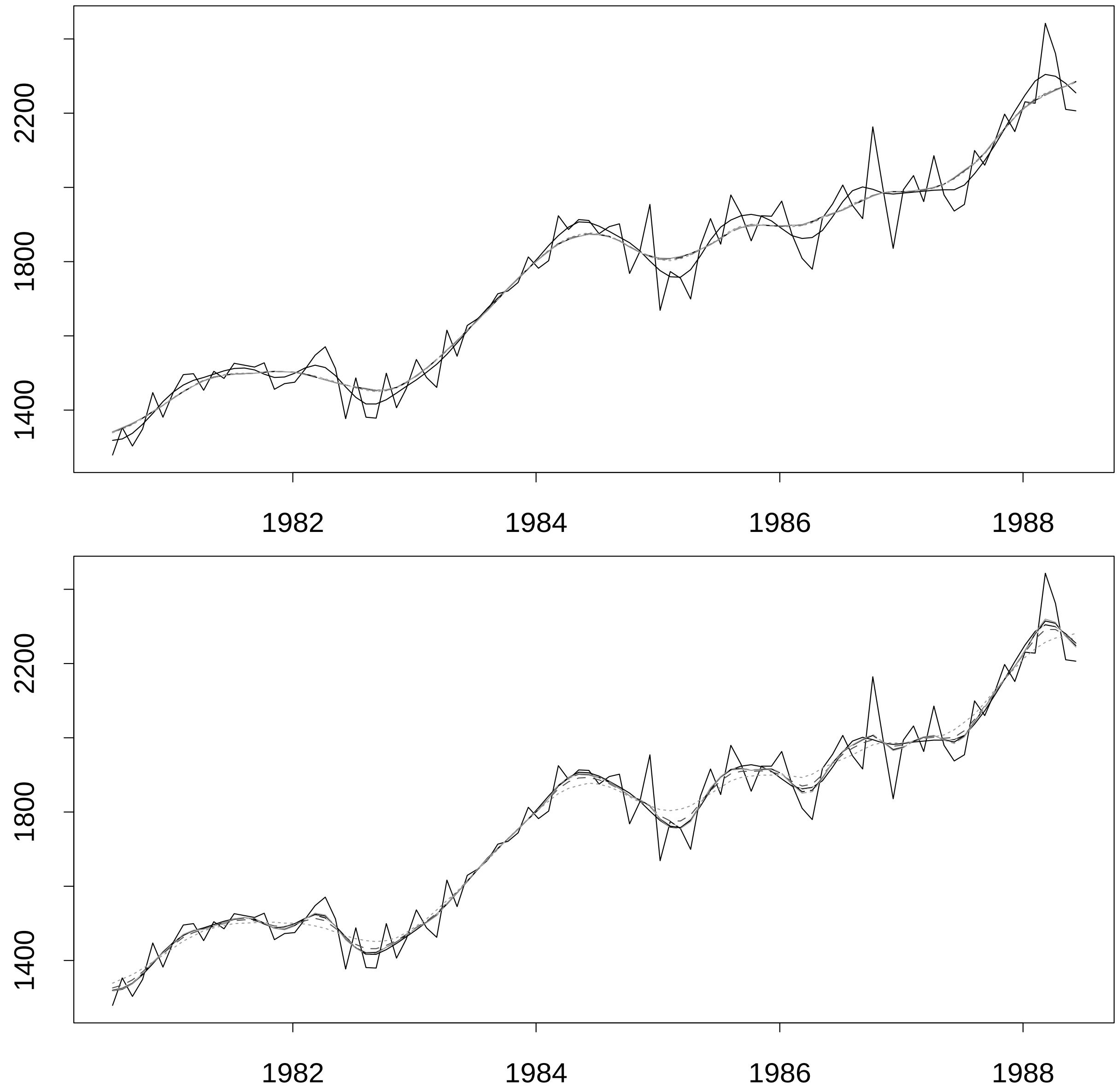


Figure 35: The US Furniture Sales data is characterised by many local turning points, with a general upward drift. In the top graph, the data is filtered by the 13 point X-11 Henderson filter and the 13 point filters given by Theorem 1 for the local linear model and $\theta = 0$. In the bottom graph, the data is filtered by the 13 point X-11 Henderson filter and the 13 point filters given by Theorem 1 for the local linear model and $\theta = .5$. In both graphs, the solid line corresponds to the 13 point X-11 Henderson filter, the dotted and dashed lines to the 13 point filters given by Theorem 1 with $\lambda = .1, .2, .3, .4, .5$ respectively.

gration, and an ARIMA (2,1,0) model for Exports. Means were fitted in each case. The difference between the central filter trend estimates and the end filter trend estimates give the revisions which occur as predictions of missing data are replaced by their actual values. Finally we made boxplots of the absolute value of the revisions and these are given in Figures 36–38.

In accord with the discussion of the theoretical properties of these filters, the greatest revisions and spread of revisions occurred when $q = 0$ followed by $q = 1$. It is not clear in making comparisons whether one should focus on the median or the lower quartile of the distribution of the revisions, or the upper quartile or the interquartile range, or even the outliers. Depending on the cost function associated with revisions, a case can be made for each of these. Here we focus on the median because that provides a robust estimate of the square root of the mean squared revisions (up to a scale factor). The mean squared revisions criterion was used to evaluate the theoretical performance of the filters in Section 3. To further assist comparison, Table 1 provides the values of the medians for each of the series for the various values of q.

As expected from Theorem 10, the BLIP end filters where $|\hat{\beta}_p|/\hat{\sigma}$ is determined from the data generally have smaller revisions than BLUP end filters and smaller revisions than the Musgrave end filters used in X-11. (These are BLIP end filters based on the central X-11 Henderson filter, a local linear model with $\lambda = 0$ and $\hat{\beta}_p^2/\hat{\sigma}^2$ set to $4/(\pi(3.5)^2)$.) For the Exports and Permanent Migration series, although the choice of $|\hat{\beta}_p|/\hat{\sigma}$ is similar to the X-11 value, the value of λ used in the X-11 end filters is inappropriate and so here, in the main, the X-11 end filters have larger revisions than the BLUP end filters. For all cases when $q = 0$ and for Building permits and Permanent Migration when $q = 1$, the BLIP end filter with $|\hat{\beta}_p|/\hat{\sigma}$ determined from the data performed at least as well as if not better than ARIMA forecast extension. One might hope that that these encouraging results will be further improved by using local (adaptive) estimates of $|\hat{\beta}_p|/\hat{\sigma}$.

		Building Permits		Exports		Permanent Migration	
		median	% difference	median	% difference	median	% difference
$q = 5$	L	1.8	-7.8	1503	21.0	3.7	-44.6
	X	1.9	-	1242	-	6.6	-
	H	1.9	-0.6	1562	25.7	4.3	-35.7
	A	1.6	-15.0	1247	0.4	3.7	-43.9
$q = 4$	L	4.2	-0.6	2804	2.5	10.4	-37.2
	X	4.2	-	2737	-	16.5	-
	H	4.5	6.7	2852	4.2	10.4	-37.0
	A	3.3	-21.1	2827	3.3	10.4	-37.4
$q = 3$	L	5.7	-4.3	2818	1.6	14.4	-26.3
	X	6.0	-	2775	-	19.5	-
	H	5.8	-2.1	2889	4.1	15.0	-23.1
	A	4.7	-20.8	3299	18.9	13.9	-28.7
$q = 2$	L	5.9	-0.5	3931	6.6	13.6	-8.2
	X	5.9	-	3688	-	14.8	-
	H	5.9	-0.7	3866	4.8	13.3	-10.6
	A	5.0	-16.1	5557	50.7	13.6	-8.5
$q = 1$	L	13.8	-3.1	13459	8.0	29.2	-34.8
	X	14.2	-	12570	-	44.8	-
	H	13.9	-2.5	13534	7.7	30.7	-31.5
	A	14.1	-1.1	11777	-6.3	29.3	-34.5
$q = 0$	L	27.5	-9.3	19187	-27.4	71.3	-34.6
	X	30.3	-	25472	-	108.9	-
	H	32.1	5.7	21201	-16.8	73.2	-32.8
	A	27.5	-9.2	20584	-19.2	74.9	-31.3

Table 1: Using the central X-11 Henderson filter in the body and the local linear model, comparisons are made between the various end filters and for the various values of q. The medians of the absolute value of the revisions are shown, as are the percentage differences (on unrounded values) between the median for the standard X-11 end filter and the medians for the other end filters. Here L refers to the BLIP end filter based on the local linear model and the choices of λ and $\hat{\beta}_p^2/\hat{\sigma}^2$ given in Section 4.1. Likewise, X refers to the standard X-11 end filter and H refers to the BLUP end filter based on the local linear model. Note that the X-11 end filters are BLIP end filters based on the central X-11 Henderson filter, a local linear model with $\lambda = 0$ and $\hat{\beta}_p^2/\hat{\sigma}^2 = 4/(\pi(3.5)^2)$ or .104. Finally A refers to the filter obtained by using the central filter with unknown observations predicted by a global ARIMA model.

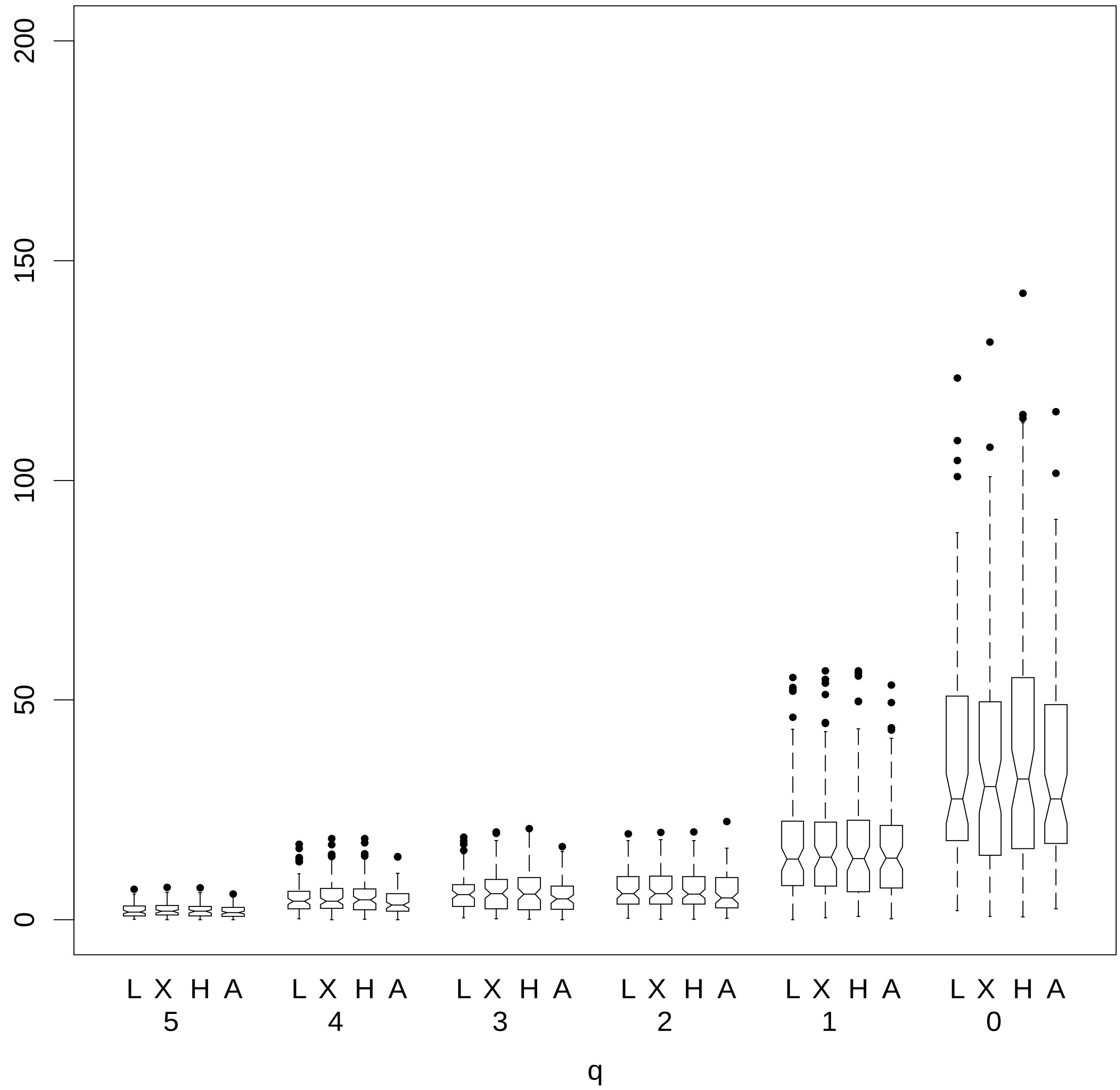


Figure 36: Building Permits data. Using the central X-11 Henderson filter in the body and the local linear model, comparisons are made between the various end filters and for the various values of q. The choice of $\lambda = 0.046$ is determined by searching for values which improve the revisions for the BLUP end filter for the $q = 0$ case. The choice of $\hat{\beta}_p^2/\hat{\sigma}^2 = 0.22$ is determined by searching for values which improve the revisions given λ. Here L refers to the BLIP end filter based on the local linear model and the choices of λ and $\hat{\beta}_p^2/\hat{\sigma}^2$ given above. Likewise, X refers to the standard X-11 end filter and H refers to the BLUP end filter based on the local linear model. Note that the X-11 end filters are BLIP end filters based on the central X-11 Henderson filter, a local linear model with $\lambda = 0$ and $\hat{\beta}_p^2/\hat{\sigma}^2 = 4/(\pi(3.5)^2)$ or .104. Finally A refers to the filter obtained by using the central filter with unknown observations predicted by a global ARIMA model.

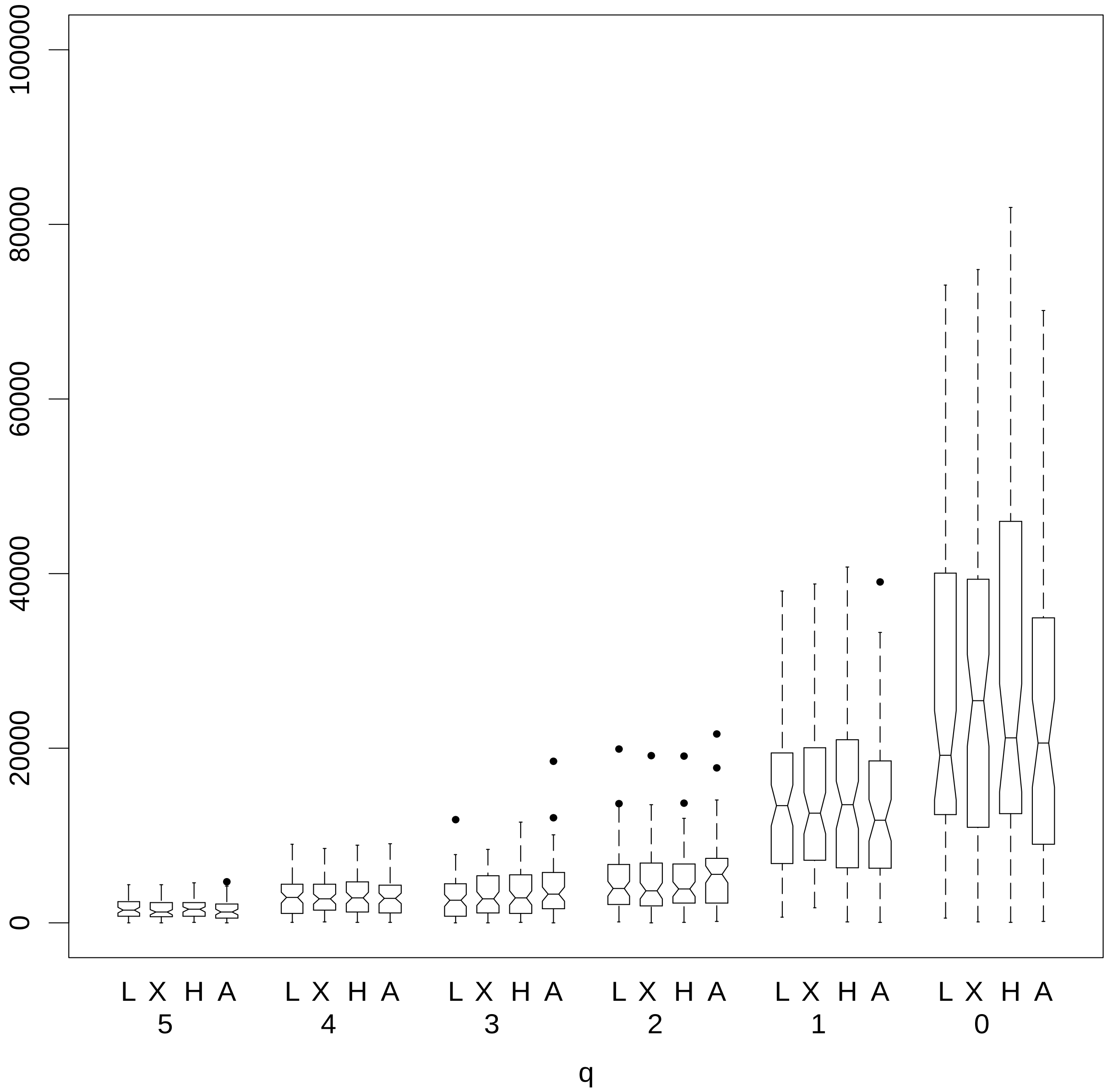


Figure 37: Merchandise Exports data. Using the central X-11 Henderson filter in the body and the local linear model, comparisons are made between the various end filters and for the various values of q. The choice of $\lambda = 0.52$ is determined by searching for values which improve the revisions for the BLUP end filter for the $q = 0$ case. The choice of $\hat{\beta}_p^2/\hat{\sigma}^2 = 0.30$ is determined by searching for values which improve the revisions given λ. Here L refers to the BLIP end filter based on the local linear model and the choices of λ and $\hat{\beta}_p^2/\hat{\sigma}^2$ given above. Likewise, X refers to the standard X-11 end filter and H refers to the BLUP end filter based on the local linear model. Note that the X-11 end filters are BLIP end filters based on the central X-11 Henderson filter, a local linear model with $\lambda = 0$ and $\hat{\beta}_p^2/\hat{\sigma}^2 = 4/(\pi(3.5)^2)$ or .104. Finally A refers to the filter obtained by using the central filter with unknown observations predicted by a global ARIMA model.

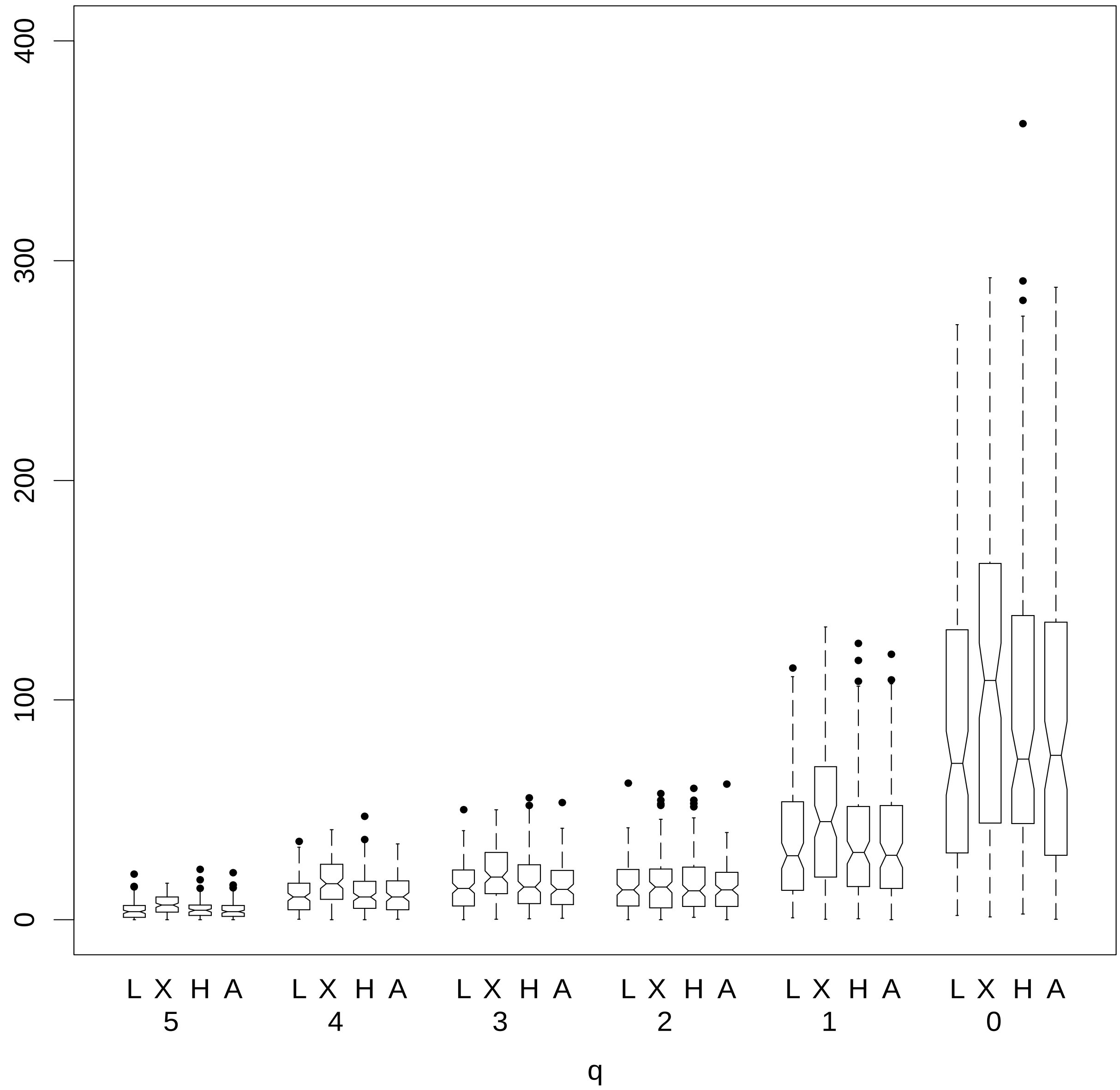


Figure 38: Permanent Migration data. Using the central X-11 Henderson filter in the body and the local linear model, comparisons are made between the various end filters and for the various values of q. The choice of $\lambda = 3.8$ is determined by searching for values which improve the revisions for the BLUP end filter for the $q = 0$ case. The choice of $\hat{\beta}_p^2/\hat{\sigma}^2 = 0.32$ is determined by searching for values which improve the revisions given λ. Here L refers to the BLIP end filter based on the local linear model and the choices of λ and $\hat{\beta}_p^2/\hat{\sigma}^2$ given above. Likewise, X refers to the standard X-11 end filter and H refers to the BLUP end filter based on the local linear model. Note that the X-11 end filters are BLIP end filters based on the central X-11 Henderson filter, a local linear model with $\lambda = 0$ and $\hat{\beta}_p^2/\hat{\sigma}^2 = 4/(\pi(3.5)^2)$ or .104. Finally A refers to the filter obtained by using the central filter with unknown observations predicted by a global ARIMA model.

5 Conclusions

A flexible family of semi-parametric, finite, moving-average filters has been developed, both in the body and at the ends of series, based on suitably chosen local dynamic models operating within the span of the central filter.

For the body of the series, the central filters are constructed using fidelity and smoothness criteria with θ, the balance between fidelity and smoothness, specified by the user. Through θ the analyst can (subjectively) impose an estimate of trend on the data whose smoothness conforms to other criteria such as the overall objectives of the analysis. Of the two criteria, it is smoothness that plays the dominant role in terms of its variation with θ.

For the ends of the series, the end filters are constructed using a minimum revisions criterion and optimal linear (biased) prediction. It is shown that a suitable choice of a bias term related to X-11's I/C ratio can lead to minimum revisions that are competitive and sometimes better than those achieved using ARIMA forecast extension. These gains are dependent on the choice of central filter adopted. In particular, the choice of a central filter with an appropriate balance of smoothness and fidelity may well result in lower mean squared revisions. It would appear that whereas smoothness plays the influential role in the body of the series, it is fidelity that plays the dominant role in minimising expected revisions at the ends.

This work constitutes a beginning and much remains to be done. The latter include more case studies, a more thorough analysis of ways to fit local dynamic models to time series, the incorporation of these non-seasonal trend filters within seasonal adjustment procedures such as X-11, and the extension of these filters to handle outliers and structural changes.

A Appendix

A.1 Minimising Q at the ends

The following result considers the minimisation of Q at the ends of the series where now Q is defined by (8), (9) with $\hat{g}_t$ replaced by $\tilde{g}_t$ and $\tilde{g}_t$ is given by (32). In particular it shows that the minimum revisions end filters specified by Corollary 6 will not be the same as the end filters obtained from minimising the criterion Q. Thus the minimum revisions end filters will not preserve the same balance of smoothness and fidelity at the ends as in the body; conversely filters that maintain the same balance of fidelity and smoothness at the ends as in the body will not minimise the revisions. An important exception to the latter is the case of the Macaulay filters where $\theta = 1$. In this case the minimum revisions end filters are exactly the same as the optimum fidelity end filters as established in Theorem 7.

Theorem A1 *Consider a time point t at the end of the series where $q = T - t$ and $0 \le q < r$. The value of $\mathbf{u} = (u_{-r}, \ldots, u_q)^T$ that minimises Q subject to (34) is given by*

$$\mathbf{u}(t) = \mathbf{L}_1^T(\mathbf{I} - \mathbf{G}\mathbf{L}_2(\mathbf{L}_2^T\mathbf{G}\mathbf{L}_2)^{-1}\mathbf{L}_2^T)\mathbf{u}^{(0)}(t)$$

where

$$\mathbf{u}^{(0)}(t) = \mathbf{E}_1^{-1}\mathbf{C}(\mathbf{C}^T\mathbf{E}_1^{-1}\mathbf{C})^{-1}\mathbf{c} + (1-\theta)\mathbf{G}\mathbf{\Phi}\mathbf{v}(t)$$

with $\mathbf{E}_1$, $\mathbf{C}$, $\mathbf{c}$ and $\mathbf{G}$ given by Theorems 1 and 5 respectively. Moreover the $(2r+1)\times$ $(2r+p+2)$ matrix $\mathbf{\Phi}$ has typical element $\mathbf{E}_{1,ij}$ with $-r \le i \le r, -r-p-1 \le j \le r$. and $\mathbf{v}(t)$ has typical element

$$\sum_{k=1}^{p}(-1)^{k+1}\binom{p}{k}u_{s+k}(t-k) \qquad (-r-p-1 \le s \le r).$$

Here $u_s(t-k)$ is the typical element of $\mathbf{u}(t-k)$ and $\mathbf{v}(t)$ is defined to be zero when $t \le T-r$.

A.2 Limiting forms of the filters

From Theorem 1 the optimal central filter that minimises Q has weights given by

$$\mathbf{w} = \mathbf{E}_\theta^{-1}\mathbf{C}(\mathbf{C}^T\mathbf{E}_\theta^{-1}\mathbf{C})^{-1}\mathbf{c}$$

where

$$\mathbf{E}_\theta = \theta(\sigma^2\mathbf{I} + \mathbf{\Omega}) + (1-\theta)(\sigma^2\mathbf{B}_{p+1} + \mathbf{\Gamma}).$$

We consider the special case where

$$\mathbf{E}_\theta = \sigma^2(\mathbf{A}_\theta + \lambda\mathbf{B}_\theta) \tag{67}$$

with

$$\begin{aligned} \mathbf{A}_\theta &= \theta\mathbf{I} + (1-\theta)\mathbf{B}_{p+1} \\ \mathbf{B}_\theta &= \theta\mathbf{\Omega}_0 + (1-\theta)\mathbf{\Gamma}_0 \end{aligned} \tag{68}$$

and $\mathbf{\Omega} = \lambda\mathbf{\Omega}_0$, $\mathbf{\Gamma} = \lambda\mathbf{\Gamma}_0$. Note that this family includes the local dynamic models (55), (56) where $\lambda = \sigma_\eta^2/\sigma^2$ and the cases for ξ_t given by (24). It would be a straightforward exercise to extend the results that follow to handle more general forms for ξ_t. However this has not been done, partly for simplicity of presentation and partly because the cases considered should cover most situations met in practice.

With these qualifications we now consider the limiting forms of the optimal central and end filters as λ increases. In the context of (55) and (56) this limit corresponds to the case where $\sigma^2 = 0$ or, in practice, where σ_η^2 is very much larger than σ^2. In such cases the model has no noise component so that y_t is just the trend g_t.

Theorem A2 *The optimal impulse response function given by Theorem 1 with $\mathbf{E}_\theta$ given by (67) converges as $\lambda \to \infty$ to*

$$\bar{\mathbf{w}} = \begin{cases} \mathbf{B}_\theta^{-1}\mathbf{C}(\mathbf{C}^T\mathbf{B}_\theta^{-1}\mathbf{C})^{-1}\mathbf{c} & (0 \le \theta < 1) \\ \delta_0 & (\theta = 1) \end{cases}. \tag{69}$$

where $\mathbf{B}_\theta$ is given by (68) and δ_0 has zero elements with the exception of the central element which is unity.

Proof

For $0 \le \theta < 1$

$$\lim_{\lambda\to\infty} \mathbf{E}_\theta/\lambda = \sigma^2\mathbf{B}_\theta$$

so that

$$\bar{\mathbf{w}} = \mathbf{B}_\theta^{-1}\mathbf{C}(\mathbf{C}^T\mathbf{B}_\theta^{-1}\mathbf{C})^{-1}\mathbf{c}.$$

For $\theta = 1$ note that $\mathbf{B}_\theta$ is singular. In this case

$$\mathbf{E}_1 = \sigma^2(\mathbf{I} + \lambda\mathbf{\Omega}_0)$$

and $\mathbf{\Omega}_0$ can be partitioned as

$$\mathbf{\Omega}_0 = \begin{pmatrix} \mathbf{\Omega}_- & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \mathbf{\Omega}_+ \end{pmatrix}$$

where the $r \times r$ dimensional matrices $\mathbf{\Omega}_-$, $\mathbf{\Omega}_+$ are both positive definite. Partition $\mathbf{C}$ accordingly as

$$\mathbf{C} = \begin{pmatrix} \mathbf{C}_- \\ \mathbf{c}^T \\ \mathbf{C}_+ \end{pmatrix}$$

where $\mathbf{C}_-$, $\mathbf{C}_+$ are $r \times (p+1)$ matrices and $\mathbf{c}$ is given by (17). Then

$$\mathbf{C}^T\mathbf{E}_1^{-1}\mathbf{C} = (\mathbf{D}_\lambda + \mathbf{c}\mathbf{c}^T)/\sigma^2$$

where

$$\mathbf{D}_\lambda = \mathbf{C}_-^T(\mathbf{I} + \lambda\mathbf{\Omega}_-)^{-1}\mathbf{C}_- + \mathbf{C}_+^T(\mathbf{I} + \lambda\mathbf{\Omega}_+)^{-1}\mathbf{C}_+$$

and

$$\lim_{\lambda\to\infty} \lambda\mathbf{D}_\lambda = \bar{\mathbf{D}} = \mathbf{C}_-^T\mathbf{\Omega}_-^{-1}\mathbf{C}_- + \mathbf{C}_+^T\mathbf{\Omega}_+^{-1}\mathbf{C}_+. \tag{70}$$

In this case the optimal impulse response function is given by

$$\mathbf{w} = \mathbf{E}_1^{-1}\mathbf{C}(\mathbf{C}^T\mathbf{E}_1^{-1}\mathbf{C})^{-1}\mathbf{c} = \begin{pmatrix} (\mathbf{I} + \lambda\mathbf{\Omega}_-)^{-1}\mathbf{C}_-\mathbf{D}_\lambda^{-1}\mathbf{c} \\ \mathbf{c}^T\mathbf{D}_\lambda^{-1}\mathbf{c} \\ (\mathbf{I} + \lambda\mathbf{\Omega}_+)^{-1}\mathbf{C}_+\mathbf{D}_\lambda^{-1}\mathbf{c} \end{pmatrix} /(1 + \mathbf{c}^T\mathbf{D}_\lambda^{-1}\mathbf{c})$$

and so

$$\lim_{\lambda\to\infty} \mathbf{w} = \delta_0$$

as required.

□

Now consider the BLUP end filters based on Corollary 6 and specified central filter weights given by $\mathbf{w}$. Define the $n \times n$ dimensional partitioned matrix $\bar{\mathbf{G}}$ as

$$\bar{\mathbf{G}} = \begin{pmatrix} \bar{\mathbf{G}}_{11} & \bar{\mathbf{g}}_1 & \bar{\mathbf{G}}_{12} \\ \bar{\mathbf{g}}_1^T & \bar{\mathbf{g}}_0 & \bar{\mathbf{g}}_2^T \\ \bar{\mathbf{G}}_{12}^T & \bar{\mathbf{g}}_2 & \bar{\mathbf{G}}_{22} \end{pmatrix} \tag{71}$$

where the $r \times r$ dimensional matrices $\bar{\mathbf{G}}_{11}$, $\bar{\mathbf{G}}_{12}$, $\bar{\mathbf{G}}_{22}$ are given by

$$\begin{array}{rcl} \bar{\mathbf{G}}_{11} & = & \mathbf{\Omega}_-^{-1} - \mathbf{\Omega}_-^{-1}\mathbf{C}_-\bar{\mathbf{E}}\mathbf{C}_-^T\mathbf{\Omega}_-^{-1} \\ \bar{\mathbf{G}}_{12} & = & \phantom{\mathbf{\Omega}_-^{-1}} - \mathbf{\Omega}_-^{-1}\mathbf{C}_-\bar{\mathbf{E}}\mathbf{C}_+^T\mathbf{\Omega}_+^{-1} \\ \bar{\mathbf{G}}_{22} & = & \mathbf{\Omega}_+^{-1} - \mathbf{\Omega}_+^{-1}\mathbf{C}_+\bar{\mathbf{E}}\mathbf{C}_+^T\mathbf{\Omega}_+^{-1} \end{array}$$

with

$$\bar{\mathbf{E}} = \bar{\mathbf{D}}^{-1} - (\bar{\mathbf{D}}^{-1}\mathbf{c})(\bar{\mathbf{D}}^{-1}\mathbf{c})^T/\mathbf{c}^T\bar{\mathbf{D}}^{-1}\mathbf{c}$$

and $\bar{\mathbf{D}}$ is defined by (70). Moreover the r dimensional vectors $\bar{\mathbf{g}}_1$, $\bar{\mathbf{g}}_2$ are given by

$$\begin{aligned}
\bar{\mathbf{g}}_1 &= -\boldsymbol{\Omega}_-^{-1}\mathbf{C}_-\bar{\mathbf{D}}^{-1}\mathbf{c}/\mathbf{c}^T\bar{\mathbf{D}}^{-1}\mathbf{c} \\
\bar{\mathbf{g}}_2 &= -\boldsymbol{\Omega}_+^{-1}\mathbf{C}_+\bar{\mathbf{D}}^{-1}\mathbf{c}/\mathbf{c}^T\bar{\mathbf{D}}^{-1}\mathbf{c}
\end{aligned}$$

and the scalar $\bar{\mathbf{g}}_0$ is given by $\bar{\mathbf{g}}_0 = 1/\mathbf{c}^T\bar{\mathbf{D}}^{-1}\mathbf{c}$. With this notation in place we now establish the following result.

Theorem A3 *The BLUP end filter given by Corollary 6 with central filter weights specified by* $\mathbf{w}$ *and* $\mathbf{E}_\theta$ *given by (67) converges as* $\lambda \to \infty$ *to a limit specified by Corollary 6 with* $\mathbf{G}$ *replaced by* $\bar{\mathbf{G}}$ *where* $\lim_{\lambda\to\infty} \lambda\mathbf{G} = \bar{\mathbf{G}}$ *is given by (71).*

Proof

Observe that the BLUP end filter specified in Corollary 6 and Theorem 5 does not depend on the scale parameter σ^2. Thus, for the purposes of determining these end filters, it is sufficient to consider the case $\sigma^2 = 1$.

Assuming $\sigma^2 = 1$ and the notation of Theorem A2 we further note that

$$\lambda\mathbf{E}_1^{-1} = \begin{pmatrix} (\lambda^{-1}\mathbf{I}+\boldsymbol{\Omega}_-)^{-1} & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & (\lambda^{-1}\mathbf{I}+\boldsymbol{\Omega}_+)^{-1} \end{pmatrix}.$$

Furthermore $\lambda\mathbf{E}_1^{-1}\mathbf{C}(\mathbf{C}^T\mathbf{E}_1^{-1}\mathbf{C})^{-1}\mathbf{C}^T\mathbf{E}_1^{-1}$ is given by

$$\lambda^{-1}\begin{pmatrix} (\lambda^{-1}\mathbf{I}+\boldsymbol{\Omega}_-)^{-1}\mathbf{C}_- \\ \lambda\mathbf{c}^T \\ (\lambda^{-1}\mathbf{I}+\boldsymbol{\Omega}_+)^{-1}\mathbf{C}_+ \end{pmatrix} (\mathbf{D}_\lambda + \mathbf{c}\mathbf{c}^T)^{-1} \begin{pmatrix} (\lambda^{-1}\mathbf{I}+\boldsymbol{\Omega}_-)^{-1}\mathbf{C}_- \\ \lambda\mathbf{c}^T \\ (\lambda^{-1}\mathbf{I}+\boldsymbol{\Omega}_+)^{-1}\mathbf{C}_+ \end{pmatrix}^T$$

and

$$(\mathbf{D}_\lambda + \mathbf{c}\mathbf{c}^T)^{-1} = \mathbf{D}_\lambda^{-1} - (\mathbf{D}_\lambda^{-1}\mathbf{c})(\mathbf{D}_\lambda^{-1}\mathbf{c})^T/(1+\mathbf{c}^T\mathbf{D}_\lambda^{-1}\mathbf{c}).$$

Thus, for example,

$$(\lambda^{-1}\mathbf{I}+\boldsymbol{\Omega}_-)^{-1}\mathbf{C}_-(\mathbf{D}_\lambda + \mathbf{c}\mathbf{c}^T)^{-1}(\lambda^{-1}\mathbf{I}+\boldsymbol{\Omega}_+)^{-1}\mathbf{C}_+$$

and

$$\lambda - \lambda\mathbf{c}^T(\mathbf{D}_\lambda + \mathbf{c}\mathbf{c}^T)^{-1}\mathbf{c} = 1/(\lambda^{-1} + \mathbf{c}^T(\lambda\mathbf{D}_\lambda)^{-1}\mathbf{c})$$

converge to $\bar{\mathbf{G}}_{12}$ and $\bar{\mathbf{g}}_0$ respectively as λ increases. The other elements of $\bar{\mathbf{G}}$ follow similarly.

□

Finally we consider the BLIP end filters based on Theorem 10, a given value for $\hat{\beta}_p^2/\hat{\sigma}^2$ and specified central filter weights $\mathbf{w}$. Once again further notation is necessary. Define

$$\bar{\gamma} = \begin{pmatrix} \boldsymbol{\Omega}_-^{-1}\mathbf{C}_-\bar{\mathbf{E}}\mathbf{d} \\ \mathbf{c}^T\bar{\mathbf{D}}^{-1}\mathbf{d}/\mathbf{c}^T\bar{\mathbf{D}}^{-1}\mathbf{c} \\ \boldsymbol{\Omega}_+^{-1}\mathbf{C}_+\bar{\mathbf{E}}\mathbf{d} \end{pmatrix}, \qquad \bar{\phi} = -\frac{\bar{\gamma}^T\bar{\mathbf{H}}\mathbf{w}}{\mathbf{d}^T\bar{\mathbf{E}}\mathbf{d} + \bar{\gamma}^T\bar{\mathbf{H}}\bar{\gamma}} \tag{72}$$

where $\bar{\mathbf{H}} = \mathbf{L}_2(\mathbf{L}_2^T\bar{\mathbf{G}}\mathbf{L}_2)^{-1}\mathbf{L}_2^T$ and $\bar{\mathbf{E}}$, $\bar{\mathbf{G}}$ are given by (71). Then we have the following result.

Theorem A4 *The BLIP end filter given by Theorem 10, a given value for $\hat{\beta}_p^2/\hat{\sigma}^2$, specified central filter weights $\mathbf{w}$ and $\mathbf{E}_\theta$ given by (67) converges as $\lambda \to \infty$ to a limit specified by Theorem 10 with $\mathbf{G}$, γ, ϕ, replaced by $\bar{\mathbf{G}}$, $\bar{\gamma}$, $\bar{\phi}$ respectively. Here $\lim_{\lambda\to\infty} \lambda\mathbf{G} = \bar{\mathbf{G}}$ is given by (71) and $\bar{\gamma}$, $\bar{\phi}$ are given by (72).*

The proof of this result is established in essentially the same way as that for Theorem A3. Note, in particular, that the limit does not depend on the value chosen for $\hat{\beta}_p^2/\hat{\sigma}^2$.

A.3 Evaluating the I/C ratio

We consider evaluating ratios of the form

$$\frac{I_p}{C_p} = \frac{E|\Delta^p \epsilon_t|}{E|\Delta^p g_t|}$$

where $y_t = g_t + \epsilon_t$ follows the local dynamic model (2) of order p and it is further assumed that ϵ_t and g_t are independent Gaussian processes. Note that when $p = 1$ this gives the population analogue of X-11's $\bar{I}/\bar{C}$ ratio.

Given these assumptions it is evident that $\Delta^p \epsilon_t$ and $\Delta^p g_t$ are independent Gaussian random variables with means zero and $p!\beta_p$ respectively. Their variances are given by

$$\text{Var}\{\Delta^p \epsilon_t\} = \, ^{2p}C_p \sigma^2 \qquad \text{Var}\{\Delta^p g_t\} = E\{(\Delta^p \xi_t)^2\}$$

where the latter quantity depends on the particular form of the local dynamic model adopted. In the case of the local linear model (55), $E\{(\Delta \xi_t)^2\} = \lambda\sigma^2$ and in the case of the local quadratic model (56), $E\{(\Delta^2 \xi_t)^2\} = 2\lambda\sigma^2$. Using the following result the forms (61) and (62) can now be determined.

Theorem A5 *If X is a Gaussian random variable with mean μ and unit variance then*

$$E|X| = |\mu|(\Phi(|\mu|) - \Phi(-|\mu|)) + \sqrt{\frac{2}{\pi}}\exp(-\frac{1}{2}\mu^2)$$

where $\Phi(x)$ is the standard normal cumulative distribution function.

Proof

Here

$$\begin{aligned}
E|X| &= \int_0^\infty x\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(x-\mu)^2}dx - \int_{-\infty}^0 x\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(x-\mu)^2}dx \\
&= \int_{-\mu}^\infty (x+\mu)\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}dx + \int_{\mu}^\infty (x-\mu)\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}dx \\
&= \mu(\Phi(\mu) - \Phi(-\mu)) + 2\int_{|\mu|}^\infty x\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}dx \\
&= |\mu|(\Phi(|\mu|) - \Phi(-|\mu|)) + \sqrt{\frac{2}{\pi}}\exp(-\frac{1}{2}\mu^2)
\end{aligned}$$

as required. □

References

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

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.

Bell, W. R. and Hillmer, S. C. (1984). Issues involved with the seasonal adjustment of economic time series. *Journal of Business and Economic Statistics*, 2.

Brillinger, D. R. (1965). A property of low pass filters. *SIAM Review*, 7:65–67.

Cholette, P.-A. (1981). A comparison of various trend-cycle estimators. In Anderson, O. D. and Perryman, M. R., editors, *Time Series Analysis*, pages 77–87. North-Holland, New York.

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.

Cleveland, W. S. (1979). Robust locally-weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, 74:829–836.

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.

Cleveland, W. S., Dunn, D. M., and Terpenning, I. J. (1978). SABL - A resistant seasonal adjustment procedure with graphical methods for interpretation and diagnosis. In Zellner, A., editor, *Seasonal Analysis of Economic Time Series*, pages 201–231. U.S. Department of Commerce, U.S. Bureau of the Census, Washington, DC.

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

Dagum, E. B. (1996). A new method to reduce unwanted ripples and revisions in trend-cycle estimates from X-11-ARIMA. *Survey Methodology*, 22(1):77–83.

Dagum, E. B. and Laniel, N. (1987). Revision of trend-cycle estimators of moving average seasonal adjustment methods. *Journal of Business and Economic Statistics*, 5(2):177–189.

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.

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

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.

Gray, A. G. and Thomson, P. J. (1990). Commentary on STL: A seasonal-trend decomposition procedure based on Loess. *Journal of Official Statistics*, 6:47–54.

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.

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

Henderson, R. (1916). Note on graduation by adjusted average. *Transactions of the Actuarial Society of America*, 17:43–48.

Henderson, R. (1924). A new method of graduation. *Transactions of the Actuarial Society of America*, 25:29–40.

Kendall, M. G. (1973). *Time-Series.* Hafner Press, New York.

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.

Leser, C. E. V. (1963). Estimation of quasi-linear trend and seasonal variation. *Journal of the American Statistical Association*, 58:1033–1043.

London, D. (1985). *Graduation: The Revision of Estimates.* ACTEX, Abington, Connecticut.

Macaulay, F. (1931). *The Smoothing of Time Series.* National Bureau of Economic Research, New York.

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.

Ramsay, C. M. (1991). Minimum variance moving-weighted-average graduation. *Transactions of the Society of Actuaries*, XLIII:305–333.

Schlicht, E. (1981). A seasonal adjustment principle and a seasonal adjustment method derived from this principle. *Journal of the American Statistical Association*, 76:374–378.

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, Bureau of the Census, U.S. Department of Commerce, Washington, D.C.

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.

Whittaker, E. T. (1923). On a new method of graduation. *Proceedings of the Edinburgh Mathematical Society*, 41:63–75.