KOF Konjunkturforschungsstelle
Swiss Institute for
Business Cycle Research

Arbeitspapiere/ Working Papers

Marc Wildi, Bernd Schips

Signal Extraction: How (In)efficient Are Model-Based Approaches? An Empirical Study Based on TRAMO/SEATS and Census X-12-ARIMA

No. 96, December 2004

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Signal Extraction: How (In)efficient are Model-Based Approaches? An Empirical Study Based on TRAMO/SEATS and Census X-12-ARIMA

November 21, 2004

Marc Wildi

University of Technical Sciences, Zurich

Technopark, Jaegerstrasse 2

CH-8401 Winterthur, Switzerland

email: wia@zhwin.ch

phone: 032 345 12 71

Bernd Schips

Swiss Institute for Business Cycle Research

Swiss Federal Institute of Technology Zurich

Weinbergstrasse 35

CH-8092 Zurich, Switzerland

email: schips@kof.gess.ethz.ch

Abstract

Estimation of signals at the current boundary of time series is an important task in many practical applications. In order to apply the symmetric filter at current time, model-based approaches typically rely on forecasts generated from a time series model in order to extend (stretch) the time series into the future. In this paper we analyze performances of concurrent filters based on TRAMO and X-12-ARIMA for business survey data and compare the results to a new efficient estimation method which does not rely on forecasts. It is shown that both model-based procedures are subject to heavy model misspecification related to false unit root identification at frequency zero and at seasonal frequencies. Our results strongly suggest that the traditional model-based approach should not be used for problems involving multi-step ahead forecasts such as e.g. the determination of concurrent filters.

Keywords: Signalextraction, concurrent filter, unit root, amplitude and time delay.

Marc Wildi is Professor for econometrics and time series analysis at the University of Technical Sciences, Zurich. His recent work is devoted to statistical problems in the health care sector (forecasting the Swiss per capita annual health care expenditures, measuring hospital inefficiency) and in the financial area (robust regression and time series models). In his latest publication, the author reexamines the signal extraction problem and proposes a new efficient approach for solving the difficult estimation problem at the current time point of a time series.

Bernd Schips (1939) is Director of the Swiss Institute for Business Cycle Research at the Federal Institute of Technology of Zurich (KOF ETHZ). After his studies in Karlsruhe, he obtained his doctorate in economics and completed his education by habilitating in Bochum. He has been head of the KOF ETHZ and professor of econometrics since 1992 after a stay at the University of St. Gall. He is a member of various scientific committees and is President of the group of experts "Business Statistics", Neuchtel/Switzerland. Since 2000, he has been President of CIRET (Centre for International Research on Economic (Tendency) Surveys).

1 Introduction

A policy-oriented business cycle research needs reliable indicators of current economic activity (real time data). In general, these indicators are affected by seasonal patterns and noise. Therefore, the data have to be filtered accordingly. Unfortunately, symmetric filters cannot be used towards the boundaries of a sample. The actual (current) signal - which is often the most interesting in practical applications - cannot be computed directly, but has to be estimated instead. Moreover, the series considered here are part of a *leading* indicator of economic activity so that multivariate approaches are not really helpful in predicting outcomes of such an indicator. Therefore, signals or equivalently symmetric filters must be approximated by suitable (efficient) *univariate* asymmetric filters. The filter for the most actual (current) time point is called *concurrent* filter. As argued in [7], p.2 "As the first-published adjustment for month T, this is the adjustment (the output of the concurrent filter) that receives the most attention ... Thus it is especially important to consider properties of the concurrent filters".

In this article we analyze properties of concurrent filters based on various estimation methods. Model-based approaches (MBA) such as TRAMO/SEATS or Census X-12-ARIMA are widely used for signal extraction. For solving the above boundary problem - approximation of symmetric by asymmetric filters - MBA rely on forecasts generated by a model to 'stretch' (extend) the time series in order to apply the symmetric filter towards the boundaries of a sample. As shown in [2], [12] and [15] this procedure results in a particular asymmetric filter whose coefficients (filter weights) are optimized with respect to *one-step ahead* forecasting performances of the time series model (typically an ARIMA-model). However, if the weights of the symmetric filter decay slowly (which is typical for seasonal-adjustment or trend extraction, see the example below) then not only one-step but also multi-step-ahead forecasts of longer horizons are needed. Since it is known that one- and multi-step ahead forecasting performances of time series models are generally conflicting because of model misspecification (see, for example, [3] and [4]) it follows that the optimization criterion underlying MBA is not optimally designed for solving the boundary signal estimation problem i.e. for computing the *concurrent* filter.

In this article, we compare the performances of TRAMO and X-12-ARIMA with those of an efficient estimation method presented in [15] - the so called Direct Filter Approach (DFA) - using a representative sample of 36 monthly time series (business survey data collected at the Institute for Business Cycle Research at the Swiss Institute of Technology, Zurich. The data as well as the filters used can be downloaded from the sites www.zhwin.ch/˜wia/signalextraction or www.kof.ethz.ch/signalextraction. In sections 3 and 4 *specific* DFA filters - optimized for each time series - are used and comparisons 'in' and 'out of sample' are reported. In section 5 'similar' time series are clustered together into three different clusters and only three different filters are optimized for the DFA (one for each cluster, the same for all series in a cluster) which are compared to the 36 specific MBA-filters. As shown below, these results clearly contradict the implicitly assumed efficiency of model-based concurrent filters.

2 Experimental Design

In order to illustrate inefficiency issues related to model-based approaches we here consider concurrent filters for a trend component whose transfer function is defined by

$$\tilde{\Gamma}(\omega) := \begin{cases} 1 & 0 \leq |\omega| \leq \pi/14 \\ \dfrac{\pi/7 - |\omega|}{\pi/7 - \pi/14} & \pi/14 \leq |\omega| \leq \pi/7 \\ 0 & \pi/7 \leq |\omega| \leq \pi \end{cases} \tag{1}$$

The (symmetric) filter weights are

$$\tilde{\gamma}_k = \begin{cases} -\dfrac{1}{\pi(\pi/7 - \pi/14)} \left[\dfrac{\cos(k\pi/7) - \cos(k\pi/14)}{k^2} \right] & k \neq 0 \\ \dfrac{1}{2}\left(\dfrac{1}{7} + \dfrac{1}{14}\right) & \text{k=0} \end{cases} \tag{2}$$

The filter doesn't affect components with frequencies smaller than $\pi/14$ and eliminates completely components corresponding to frequencies greater than $\pi/7$. Therefore, it is suited for defining a particular 'trend' for monthly time series. We chose a trend component here, because it is relevant for many users of statistical data: as argued in [7] p.7 "A substantial number (of users of seasonally adjusted data) would also prefer that higher frequency components be suppressed in order to obtain a smoother adjusted series". As can be seen by direct computation, the filter weights $\tilde{\gamma}_k$ decay 'slowly'. Therefore good one- and multi-step ahead forecasts are required. In all our experiments, the filter weights have been truncated and we use the same symmetric MA(121)-filter $\Gamma(\cdot)$ defined by weights $\gamma_k = \begin{cases} C\tilde{\gamma}_k, & |k| \leq 60 \\ 0, & \text{otherwise} \end{cases}$ where $C := 1/\sum_{|k|\leq 60} \tilde{\gamma}_k$. As a result we can compute signals - so called 'final' estimates - for $t = 61, 62, ..., T - 60$.

We chose a constant trend definition for all 36 time series considered in order to make the comparisons more reliable[1]. Moreover, we avoid a signal specific to one particular model-based approach (such as for example a canonical trend or a Henderson filter with seasonal dips) in order not to favor one particular method. Although not explicitly reported here, the results do not dramatically differ if other trend definitions or seasonal adjustment filters are used. The reason may be imputed to the fact that the rate of decay of the weights based on alternative filter definitions is often slow too. Therefore, the corresponding estimation problem requires good one- and multi-step ahead forecasting performances (which is generally a conflicting requirement for MBA).

The business survey data collected by the Institute for Business Cycle Research at the Swiss Federal Institute of Technology consists of 36 monthly time series beginning in June 1979 and ending in August 2004 ($T = 303$). The specific dynamics of the time

[1] Signals for X-12-ARIMA are not explicitly defined so that direct comparisons across methods are not possible. Model-based signal definitions resulting for example from canonical decompositions (SEATS) depend on misspecified models (see below), the characteristics of the time series and eventually the (latest) version of the software used. Therefore we chose the above unifying framework for our comparisons.

series involved are very different which makes the considered sample in some sense 'representative' for many practical applications: noise amount and spectral peaks of seasonal components vary in height (intensity) and/or width (stability of the season). Another important fact is that the time series are *bounded*. Therefore, the corresponding processes *cannot be integrated*. Identification of an integrated process by the MBA would be a formal misspecification suggesting that one- and multi-step ahead forecasting performances may be conflicting. Note that boundedness of time series is quite frequent in practical applications e.g. 'rates' such as unemployment-rates for example.

If the data-generating process (DGP) of a time series were known, then the MBA would be optimal. In practice the DGP is unknown and MBA attempt to identify it from data. However, the empirical results presented here strongly suggest that an identification strategy based on one-step ahead forecasting errors (as in TRAMO or X-12-ARIMA for example) is not optimal in general. An alternative to the 'traditional' MBA would be to find an asymmetric filter which minimizes the *revision error variance*

$$\min_{\hat{\Gamma}} E[(\hat{Y}_t - Y_t)^2] \tag{3}$$

where

- $\hat{Y}_t$ denotes the estimate, i.e. the output of the asymmetric concurrent filter with transfer function $\hat{\Gamma}$ (the argument of the above minimization procedure);
- Y_t is the (unknown) output of the symmetric filter;
- The optimization is operated with respect to the unknown (asymmetric) filter coefficients.

An optimization criterion based on (3) would not primarily contribute to the identification of the DGP but it would *implicitly* account for one- *and* multi-step ahead forecasts as well as for the 'shape' of the symmetric filter to be approximated, something that MBA such as TRAMO or X-12-ARIMA cannot do. Although DGP-identification and the optimisation problem (3) are closely related, at least asymptotically and under suitable model constraints both problems differ due to finite sample effects and model misspecifications.

Unfortunately, the expectation in (3) is unknown (in fact even Y_t is generally unknown). The idea of the direct filter approach (DFA) as presented in [13], [14] and [15] is to approximate the unknown expectation by an efficient estimate and to optimize unknown (asymmetric) filter coefficients by minimizing this estimate. Consider a minimization of

$$\min_{\hat{\Gamma}} \frac{2\pi}{N} \sum_{k=-[N/2]}^{[N/2]} |\Gamma(\omega_k) - \hat{\Gamma}(\omega_k)|^2 I_{NX}(\omega_k) \tag{4}$$

where $\Gamma(\cdot)$ and $\hat{\Gamma}(\cdot)$ are the transfer functions of symmetric and asymmetric filters, $I_{NX}(\omega_k)$ is the periodogram of the input process and $\omega_k = k2\pi/N$ are equidistant discrete frequencies in $[-\pi, \pi]$. It is shown in [15] that the output of the resulting asymmetric filter is asymptotically efficient if the input signal (original time series) is stationary. The intuition behind this assertion may be based on the following sequence of approximations

$$\begin{aligned}\frac{2\pi}{N}\sum_{k=-[N/2]}^{[N/2]}|\Gamma(\omega_k)-\hat{\Gamma}(\omega_k)|^2 I_{NX}(\omega_k) &\simeq \frac{2\pi}{N}\sum_{k=-[N/2]}^{[N/2]} I_{N\Delta Y}(\omega_k) \qquad (5)\\ &= \frac{1}{N}\sum_{t=1}^{N}(Y_t-\hat{Y}_t)^2\\ &\simeq E[(\hat{Y}_t-Y_t)^2] \qquad (6)\end{aligned}$$

where $I_{N\Delta Y}(\omega_k)$ is the (unknown) periodogram of the filter error $\Delta Y_t := Y_t - \hat{Y}_t$. More formally, it is shown in [15] that the approximation error in (5) is of order $o(N^{-1/2})$ (superconsistency) and that the error in (6) is the 'smallest' possible (asymptotically) from which the asserted efficiency can be derived. The distribution of the estimated filter parameters obtained in (4) is derived in [15] under some general assumptions about the input signal. Generalizations to integrated processes as well as generalizations of 'traditional' information criteria and of unit-root tests - which match the structure of the boundary signal estimation problem - are presented too in [15]. However, in deriving the empirical results in this article we rely deliberately on an 'agnostic' approach for the DFA and do not account neither for 'identification' nor for 'hypothesis testing': *the same filter design is used for all 36 time series considered and unknown parameters are optimized with respect to (4) using allways the same set of initial values for the unknwon coefficients.*

In the following, both MBA are compared to the above efficient DFA. The empirical *in sample* results are based on a comparison of concurrent filter outputs (with a 'rolling' boundary $T' = 61, ..., 303 - 60 = 243$) for the trend-signal defined by the (truncated and renormalized) MA(121)-filter based on (2). Therefore, squared errors of the asymmetric filters are computed for $t = 61, ..., T - 60 = 243$ (183 observations) and sample means (of squared errors) are taken accordingly. DFA- and MBA-filters are obtained by using information in the whole sample (303 observations). However, the 'in sample' experiment is not fully relevant because future observations are not available for estimating the filter coefficients in practice. Therefore, 'out of sample' results are also reported. For the latter, only $T - 60 = 243$ observations can be used for estimation. Furthermore, 40 observations[2] are retained at the end of the remaining sample for assessing the various methods. All in all, the number of available observations for identification (MBA) and estimation (MBA and DFA) is reduced by 100 for the 'out of sample' experiment: from $N = 303$ to $N = 243 - 40 = 203$ observations.

MBA asymmetric filters are obtained by forecasting the time series at the rolling endpoints and then applying the truncated symmetric MA(121) to the stretched time series. DFA asymmetric filters solve (4). For the DFA, the filter design is based on zero-pole pairs located at the frequencies $k\pi/6$, $k = 0, 1, ..., 6$. A corresponding zero-pole pair is characterized by 2 degrees of freedom. However, the particular zero-pole pair at frequency

[2] Filter errors are autocorrelated so that a sufficiently long sample is needed for assessing out of sample performances.

0 aims at small time delays of the filter and is characterized by one degree of freedom only, see [15] for formal definitions and results. An additional unconstrained zero-pole pair (three degrees of freedom since the argument is optimized too) and a normalizing constant are used to match the filter properties in the important passband of the filter. This amounts to $5 * 2 + 1 + 2 + 3 + 1 = 17$ degrees of freedom. The stable and invertible minimum-phase DFA-filters used here are of the type ARMA(15,15) with parameters satisfying a set of 13 restrictions (filter parameters can be downloaded from www.zhwin.ch/~wia/signalextraction or www.kof.ethz.ch/signalextraction). The two degrees of freedom of zero-pole pairs at $k\pi/6$, $k = 1, ..., 6$ account for *height* and *width* of potential spectral peaks (seasonal fundamental and harmonics). If there is no apparent spectral peak at a particular frequency, corresponding zeroes and poles of the ARMA-filter are allowed to cancel each other. Therefore we renounce to 'identify' an optimal filter design as mentioned earlier. As a result, some of the AR- and/or MA-coefficients are almost vanishing and the (concurrent) filter may seem to be unnecessarily complicate. However, as shown below the performances (especially 'out of sample') are not affected by this circumstance.

Filter errors are computed for MBA (TRAMO and X-12-ARIMA) and the DFA and relative sample error variances

$$\frac{\sum_{t=61}^{T'}(\hat{Y}_t^{DFA} - Y_t)^2 - \sum_{t=61}^{T'}(\hat{Y}_t^{MBA} - Y_t)^2}{\sum_{t=61}^{T'}(\hat{Y}_t^{DFA} - Y_t)^2} \tag{7}$$

are recorded for all series, where Y_t is the output of the symmetric MA(121) filter (T' varies depending on the analysis being in- or out-of-sample). We use TRAMO and X-12-ARIMA as implemented in DEMETRA, version 2 (SP1). This is a user-friendly interface running under Windows which gives access to TRAMO/SEATS and X-12-ARIMA (release version 0.2.8). It can be downloaded from http://forum.europa.eu.int/Public/irc/dsis/eurosam/library[3]. Since MBA generally rely on adjustments - outliers and calendar effects such as trading-day or Easter-effect adjustments - and/or log-transforms we used original (unadjusted) as well as *linearized* (adjusted and transformed) series for our comparisons. Note that the linearized time series generally depend on the MBA used, since testing procedures may differ.

3 Empirical Results 'in sample'

In this section, comparisons are based on *three distinct experimental 'set-ups'*. As filter inputs we use original data in the first two and linearized data (adjusted for outliers and/or

[3]Recent and intended future developments of X-12-ARIMA (X-12-ARIMA/SEATS, see [10]) seem to point towards TRAMO/SEATS since for example it is claimed in [10] that "It (the new development) is a prototype of a merged version of the two programs" and "The version of X-12-ARIMA used in this prototype has an updated model selection procedure based on the procedure found in the TRAMO time series modelling program". For our purposes, however, we prefer that the versions of X-12-ARIMA and TRAMO used here still differ in particular with respect to model identification. Since TRAMO does not perform better than the 'older' X-12-ARIMA version it is reasonable to expect that in the light of the former comments the newer version(s) of X-12-ARIMA won't do neither.

calendar effects and/or log-transformed) in the last experiment. In the first experiment, identification is based on the adjusted data (it is well known that model identification may be heavily affected by outliers, see for example [11]) and estimation is based on the original unadjusted time series. For that we make use of functionalities in the 'detailed analysis' module in DEMETRA: in estimating parameters for the original data we simply constrain model orders to those obtained for the linearized data. In the second experiment, we estimate model orders and parameters using the original data. Hereby, we consider the fact that the DFA does not 'benefit' from this features neither. Again we use the 'detailed analysis' module of DEMETRA in which we exclude adjustments and transformation facilities. In the last experiment, identification and estimation are both based on the linearized time series (which are then also the input series for our concurrent filters). For the DFA, identification is not necessary and estimation is based on the original unadjusted time series only: therefore the same concurrent DFA filters are used for all three experimental designs. As a result, the last experiment favors MBA since linearized time series are used as input signals.

3.1 In-sample results for all series

The model orders obtained when allowing for series adjustments are summarized in tables 1 (TRAMO) and 2 (X-12-ARIMA). For X-12-ARIMA only those models were listed in table 2 which differed from TRAMO: thus the two identification procedures agreed for exactly half of the time series only which is a first indication of model misspecification. It is well known that different ARIMA-models may lead to very similar one-step ahead forecasting performances although the same models may considerably differ with respect to multi-step ahead performances especially if integration orders do not agree: see for example time series 1 and 4 which are identified as I(1) by TRAMO and as I(2) by X-12-ARIMA. As shown in [7] (section 5) such indeterminacy may have substantial (severe) impacts on the resulting filters. The 'd'-columns correspond to the estimated integration orders of the models. As already mentioned, the data generating process (DGP) of the business survey data used cannot be integrated. However, both MBA identified integrated processes for *all* series except series number 33 which is identified as a stationary process by TRAMO (note that it is I(2) according to X-12-ARIMA). All in all, 18 (TRAMO) and 32 (X-12-ARIMA) time series are identified as I(2)-processes and X-12-ARIMA selects the airline-model for 27 time series (75% of the cases). Note that usage of unit-root tests did not lead to substantial improvements of estimated integration orders, see section 3.2 (this is not really surprising because traditional tests based on one-step ahead forecasting performances cannot sufficiently discriminate series with 'longer' (bounded) swings from realizations of integrated processes). Moreover, relaxing the imposed unit-root constraints may lead to overfitting.

Empirical results for the comparison between the DFA and the MBA based on the first experimental 'set up' above are to be found in table 3: in the column entitled "X-12-T" model orders are based on TRAMO and parameter estimates are based on X-12-ARIMA[4] whereas in the column "X-12-A" the comparison is fully based on X-12-ARIMA. The

[4]Therefore the effect due to different estimation routines can be isolated.

numbers in the columns are based on expression (7): thus negative values indicate that the DFA performs better. As can be seen, the DFA outperforms TRAMO as well as X-12-ARIMA for *all* time series and the mean relative gain (reduction in error variance) of the DFA is close to 40% at the rolling boundary of the time series. Empirical results not reported here show that a linear combination of MBA and DFA filters does not outperform the DFA, suggesting that the MBA is 'encompassed' by the DFA. The difference between the estimation procedures of X-12-ARIMA and TRAMO is negligible (see the columns entitled "TRAMO" and "X-12-T") at least with respect to concurrent filter characteristics and the different model identification procedures of TRAMO and X-12-ARIMA do not lead to statistically significant differences neither although X-12-ARIMA performs slightly better (compare columns "X-12-T" and "X-12-A").

For the second experimental 'set-up' we only report the aggregate mean performances (in order to save space) which are −35% (when comparing the DFA to TRAMO) and −46% (when comparing the DFA to X-12-ARIMA). TRAMO performed slightly (though not significantly) better when the model identification was based on the unadjusted time series and conversely for X-12-ARIMA. Again, the DFA outperformed both MBA for all time series.

In the third experiment, we use the linearized (adjusted and/or log-transformed) time series as input signals and compare the performance of the DFA - *still using the filters optimized for the original series* - with 'optimal' MBA concurrent filters. Note that the linearized series depend on the method used (in their introduction to [9] the authors argue "The procedure of X-12-ARIMA differs from that of TRAMO in several ways, related mainly to parameter and likelihood calculation and to outlier identification). This particular experimental design clearly favors the MBA since filter coefficients for the DFA were not re-optimized and linearized series should be 'optimally' transformed to account for model assumptions. Again, we only present the mean performance for all series which was −30% (TRAMO) and −28% (X-12-ARIMA).

Since the DFA relies on a much larger set of estimated parameters (17 parameters are estimated), one may suspect that the above results may be due at least partially to overfitting. In [15] it is shown that out of sample performances of the DFA do not significantly differ from in sample properties, because the particular design of so called Zero-Pole-Combination (ZPC)-filters specifically accounts for the 'salient features' of a time series - height and width of spectral peaks at $k\pi/6$, $k = 0, ..., 6$ -, see for example section 3.3 in [15] for formal definitions and results. The empirical results in sections 4 and 5 below confirm that overfitting (by the DFA) is not an issue, but inefficiency (of MBA) seems to be one.

Although we now know that the loss in performance of MBA is important, we still do not know why the DFA does better. A detailed analysis for two particular time series in the next section offers more insights into the relevant topics.

3.2 A detailed analysis based on two particular time series

Due to methodological differences the performances of DFA and MBA vary. Two particular time series - series number 30 and 31 - are selected here for revealing these differences in more detail. Although the conclusions drawn do not markedly differ for the other series of the sample we have chosen number 30 and 31 because they are 'representative' for what happens in general and are therefore illustrative. The models identified by TRAMO and X-12-ARIMA for the time series $X_{t,30}$ and $X_{t,31}$ (using all observations) are

$$\begin{aligned} Y_{t,30} &:= (1-B)X_{t,30} \\ Y_{t,30} &= (1-b_1)\epsilon_{t,30} \end{aligned} \quad (8)$$

where $b_1^{TRAMO} = -0.1765$ and $b_1^{X12} = -0.1315$ and

$$\begin{aligned} Y_{t,31} &:= (1-B)(1-B^{12})X_{t,31} \\ Y_{t,31} &= (1-b_1B)(1-\beta_1B^{12})\epsilon_{t,31} \end{aligned} \quad (9)$$

with $b_1^{TRAMO} = -0.6879, \beta_1^{TRAMO} = -0.7791$ and $b_1^{X12} = -0.6916, \beta_1^{X12} = -0.8361$. Since both approaches selected identical models, results were (almost) identical too (insignificant variations are due to differences between estimation routines). The MBA selected I(1)- and I(2)-DGP's respectively. For series 30, 'traditional' unit-root tests (Zivot-Andrews, Elliot-Rothenberg-Stock and Phillips Perron) do reject the I(1)-unit-root hypothesis[5]. For series 31 Zivot-Andrews does not reject the null hypothesis, Elliot-Rothenberg-Stock rejects it on the 5%- but not on the 1%-level and Phillips- Perron rejects the null hypothesis that the seasonal differences are I(1) on the 1%-level. The asymmetric model-based concurrent filter is generated by applying the truncated MA(121) based on (2) to the time series stretched by forecasts generated by models (8) and (9): we here only report results for estimates based on TRAMO (the differences between both MBA are negligible).

As can be seen from the periodogram in the left panel in the middle of figure 3, series 30 is characterized by 'local' trends and a weak seasonal component located at the fundamental $\pi/6$ (which cannot be accounted for by the non-seasonal ARIMA(0,1,1)(0,0,0)-model (8). The corresponding panel in figure 4 shows that series 31 is characterized by stronger seasonal harmonics as well as noise (the panels on the right in the middle of these figures are the periodograms after the filtering process). Outputs of the asymmetric TRAMO- and DFA-filters as well as corresponding revision errors (differences between outputs of symmetric and asymmetric filters) are depicted in figure 1 for both time series. The spectral decomposition of the revision errors in figure 2 reveals a pattern which - as suggested by experience - seems to be typical when comparing DFA and MBA: the former generally strongly outperforms the latter in the lower frequency portion of the spectrum as well as around or at seasonal peaks (look, for example, at the remaining peak at the fundamental $\pi/6$ in the revision errors produced by TRAMO for series 30 in the left panel of figure 2). Towards frequency zero the worse performance of MBA may be explained by *model-misspecifications due to erroneous unit root identification* which imply unnecessarily severe

[5]The tests were performed using the R-routines ur.pp, ur.za and ur.ers implemented in the package urca. A constant model (no trend) is used with lag length 4 (ERS,ZA) or lags="long" (PP).

restrictions for the amplitude function of the corresponding concurrent filter and larger time delays. Special ‘care’ of the lower frequency components is needed because they are often strong and because they belong to the passband of the filter. Therefore, any (unnecessary) distortion may result in a substantial loss of performance. For the DFA, five filter parameters are allocated to this important frequency band in order to match relevant time series characteristics: see more about interpreting them below. The worse performance of MBA-filters towards seasonal frequencies may be imputed to the fact that often a single parameter - for example an SMA(12) - must account for all seasonal components (fundamental and harmonics) simultaneously. Experience suggests that model-based filters are often too sparsely parameterized to account for the complex dynamics of practical time series. The richer parameterized ZPC-filters of the DFA which define the constrained ARMA-filters are specifically designed to adapt for *strength* (height) and *stability* (width) of *each* of the spectral peaks located at (or in the vicinity of) $k\pi/6$, $k = 0, ..., 6$. Ultimately, MBA rest on (too) parsimonious models - i.e. filters - because their parameters are not immanently constrained to adapt for the ‘salient features’ only. Similar arguments are put forward in [8] to invoke a state-space model approach to signal extraction[6]. Relaxing the parsimony constraint may therefore result in overfitting and even worse filter performances. The problem is that even severe misspecifications cannot be detected because one-step ahead forecast errors are not informative enough.

Ideally, amplitude and time delay (phase divided by frequency) functions of the asymmetric filter should ‘mimic’ the corresponding functions of the symmetric filter. Unfortunately, both requirements are generally conflicting. As can be seen from the lower two panels in figures 3 and 4 , the ‘fit’ of the amplitude and the time delay functions (units on the vertical axis correspond to time units i.e. months) of the concurrent DFA-filter depends on the spectrum of the input process, getting better for dominant components which is a direct consequence of the optimization criterion (4). The time delay, for example, is small towards low frequencies where the bulk of the spectrum is located. From the shape of the amplitude function in the lower left panel in figure 4 one can see that the asymmetric DFA-filter generally damps seasonal components without removing them completely (as would be the case for a MBA-filter derived from an ARIMA-model with ‘crude’ seasonal differences, see [2], section 4 and [15] section 5.3). Note that the finite MA(121)-filter does not remove them completely neither but the infinite one in 2 would. The periodogram of the output signal in the right-hand panel in the middle of figure 4 reveals that the overall damping of the filter in the stopband is ‘optimal’ for removing the seasonal components of the time series.
broader trough of the

As suggested above, the shape of amplitude and time delay functions of the concurrent filter in the passband is important. As a result of the optimization (4) the ‘free’ zero-pole pair (whose argument is not constrained a priori) of the DFA concurrent filters is always located in the passband: together with the zero-pole pair at frequency zero and

[6]However, as for the above MBA these models are optimized with respect to one-step ahead forecasting performances only so that model misspecification will lead to inefficient concurrent filters too. Moreover, the ‘traditional’ basic structural model (BSM) corresponds to an I(2)-process which would again result in unnecessarily severe restrictions for the concurrent filters. The I(2)-hypothesis for the BSM is derived from the requirement that the slope of the trend (first differences) should be ‘adaptive’.

the normalizing constant five (3+2) degrees of freedom are used to match the series characteristics in the passband of the filter (whereas the widely used airline-model relies on two parameters only for the whole spectrum). One may note first that $\hat{A}_{DFA}(0) \neq 1$ (i.e. the MA(∞)-weights of the ARMA-filter do not add to one) in contrast to the MBA-filter which must satisfy the constraint $\hat{A}_{MBA}(0) = \Gamma(0) = 1$ because of the misspecified unit-roots at frequency zero, see [15] for a formal treatment of this problem. Intuitively, the level restriction $\hat{A}(0) = \Gamma(0)$ is necessary for (asymptotically) unbounded integrated time series, because otherwise the revision error variance would become infinite (asymptotically). Evidently, for bounded time series this is no more true. Therefore, the unnecessarily severe 'level' constraint is relaxed for the DFA. It is important to emphasize that *the relaxation of this restriction has not only incidences on the normalizing constant of the filter (as one might expect at first sight) but also on its zeroes and poles, thus affecting the transfer function of the filter in a more fundamental manner*. Whereas the real zero-pole pair located at frequency zero aims at a small time delay of the filter in the pass band (see for example figures 3 and 4 and [15] section 5.4 for formal details), the additional 'free' zero-pole pair shifts undesirable filter characteristics - overshooting of the amplitude function or 'large' time delays - to regions in the passband where the spectrum of the input series is weak, see figures 3 and 4. As a result, effects due to undesirable filter characteristics are less pronounced, see for example the spectra of the revision errors in figure 2.

The 'invisible hand' at work is the optimization criterion (4) which moves zeroes and poles of the ARMA-filter in order to exploit the individual spectral shape of a time series optimally. It would be a difficult task to search for forecasting-models whose concurrent filters would behave 'similarly'. Formally, the structure of such a model could be obtained from the concurrent DFA-filter but it is very unlike that the same model could be obtained using statistics based on one-step ahead forecasting errors only. This is because one-step ahead forecasting is not directly related to signal extraction at the (current) boundary of a time series. More fundamentally, time series characteristics which are relevant for boundary signal extraction do not necessarily import in the context of one-step ahead forecasting: a large spectral peak in the vicinity of (but not necessarily at) frequency zero may be accounted for by first differences when computing a one-step ahead forecast but for the signal extraction problem additionally the *location* of the peak is important because transfer functions of symmetric and concurrent filters must match there (equivalently: undesirable filter characteristics must be shifted away from the peak). In this sense, *computation of concurrent filters requires more flexibility and is more 'information demanding' than one-step ahead forecasting which is often erroneously confounded with DGP-identification in the signal extraction literature*.

As the above examples demonstrate, misspecification of the integration order seems to have severe impacts on the efficiency of the resulting MBA-filters. Maravall in [6], p.156 argues that "moderate overdifferencing causes, in practice, little damage". While this statement may eventually be verified for one-step ahead forecasting applications the above findings show that it should be revised for signal extraction in general.

4 Out-of sample results for all series

The results in this section are based on the original (unadjusted) time series and correspond to the first experimental 'set up' in section 3.1. As mentioned before in section 2, the filters in the following experiment are based on information from $t = 1$ to $t = 203$ (which corresponds to 17 years of observations). The symmetric MA(121) is computed up to $t = 243$ and therefore *relative* out of sample performances (based on (7)) are computed for the latest 40 observations from $t = 204$ to $t = 243$, see the first column of table 4. Neither model parameters nor DFA coefficients were re-estimated as new information became available. In fact, re-estimation has only minor impact on the results as already shown in [5], p.7. Negative signs indicate that the DFA outperforms TRAMO. Similar results are obtained for X-12-ARIMA, so that we do not report them here explicitly: the mean relative gain of the DFA is -37% (reduction of revision error variance) 'out of sample'. Additionnally, out- and in-sample performances of each method are compared and corresponding results are reported in the second and third columns of table 4: here negative signs indicate that out of sample performances are better (than 'in-sample').

Note that 'in sample' results in the present section may differ from 'whole sample' results obtained in the preceding section 3 because models and DFA-filters are not computed on the whole sample (303 observations) but on the shorter subsample $t = 1, ..., 203$ (so model orders and/or parameter estimates may vary). For the DFA, a direct comparison of 'out of sample' and 'whole sample' (analyzed in section 3.2) concurrent filters in figure 5 shows that the DFA filter characteristics 'in' and 'out of sample' are very similar which explains the results in table 4. Note that for series 30 the weak fundamental at $\pi/6$ is still accounted for by the DFA (see the corresponding dip in the amplitude function in figure 5) whereas it is ignored by the MBA which selects the same non-seasonal ARIMA(0,1,1)(0,0,0)-model as for the whole sample. For series 31 the weak fundamental at $\pi/6$ is slightly damped by the 'whole sample' DFA-filter whereas the 'out of sample' DFA-filter ignores it (an effect due to the additional information in the whole sample). As seen in the last two columns of table 4 neither estimation method seems to be affected by overfitting (mean performances 'out of sample' are not significantly different from zero). It is therefore not surprising that the mean gain of the DFA 'out of sample' (approximately 38%) is close to the mean gain 'in sample' reported in table 1. Despite a relatively large number of parameters being estimated for the DFA, overfitting is obviously avoided by constraining parameters to fit the 'salient features' of the time series only.

As expected, the identification procedure of TRAMO selects different models depending on the available sample length. Therefore, we briefly investigate the additional gain obtained by the automatic model selection procedure. For that we compare the performances of TRAMO with automatic model selection set on (as above) with concurrent filters based on the airline model only (no model selection). The mean relative decrease of the filter error by using the automatic identification procedure was -8% (stand. dev. 3%) when comparing only those time series for which the airline model was *not* selected and -5%(stand. dev. 2%) in the mean over *all* time series on the whole sample (N=303).

5 Frequency-Clustering: 3 DFA-filters outperform 36 specific MBA-filters

Despite evident advantages of the DFA from a statistical point of view, the method suffers from difficult numerical optimization because of strong non-linearities and multidimensionality (17-dimensional parameter space). Numerical solutions of the criterion (4) have been found through combination of genetic algorithms (slow convergence towards global extremum) and methods based on Nelder-Mead and BFGS (fast convergence to local extremum) as implemented in R[7]. In order to reduce the computational effort we decided to simplify the estimation problem for the whole sample by defining clusters of 'similar' time series for which a single concurrent filter is computed. For the DFA, two time series are claimed to be 'similar' if the normalized periodograms (the normalization is obtained by standardizing the time series) look 'similar' whereby a formal measure of 'similarity' is provided by cluster analysis based on normalized periodograms, using complete linkage and the Euclidean distance (other clustering methods do not lead to substantially different clusters). Based on a dendrogram-analysis the time series were partitioned into three clusters as can be seen from table 5. The arithmetic means of the periodograms of the series in each of the clusters are shown in figure 6 (the original unadjusted series are used here). The *mean periodogram statistics* are obtained by averaging the periodograms of series in identical clusters. They are used in (4) in order to compute *three distinct concurrent filters, one for each cluster: the same filter is used for all series in a cluster.* Therefore, three concurrent DFA-filters only compete with 36 specific MBA-filters. Note that the various processes identified by TRAMO for the time series in a given cluster (for example cluster 2) suggest much more heterogeneity among time series, for example series 13 (cluster 2) is identified as I(1) without seasonals whereas for series 7 (also in cluster 2) an airline-model has been identified. The relative performances of both approaches are summarized in table 5. Negative numbers indicate that the particular DFA approach chosen here (three filters only) performs better. The three DFA-filters outperform the MBA for 30 (out of 36) time series and the mean gain in performance obtained is approximately -14% which is strongly significant (stand. dev. 3%). Note that the loss in performance of the three DFA filters with respect to *specific* DFA filters is quite important (about 22% larger revision error variance) which indicates that a lot of information has been lost by aggregating the periodograms. However, this information can obviously not be recovered by MBA.

These results confirm that the MBA is an inefficient estimation method. Moreover, as already shown by the out of sample results in section 4, overfitting cannot be an issue for the DFA since the method used here performs very well although the statistics used (the mean periodograms) do no more directly depend on one particular time series. Finally, these findings may also suggest that the automatic identification method of the above MBA may not be well-suited for signal extraction since very different process classes (different integration orders) are identified for time series with 'similar' characteristics.

[7] The Comprehensive Archive R Network, http://cran.r-project.org.

6 Summary and Conclusion

The above results provide strong evidence against the generally assumed efficiency of MBA in signal extraction problems. It is shown that the dynamic structure of many practical time series is too rich to be accounted for by simple models such as, for example, airline-models. In an attempt to generalize the airline-model in [1] it is argued in the conclusion: "The one-step-ahead forecast error diagnostic does not suggest strong forecast performance gains for the new models ... Our experience with these models strengthens our confidence in the robustness and flexibility of the airline model". Our experience confirms that statistics based on one-step ahead forecasting performances are not well suited for problems involving multi-step ahead forecasts but we do not agree with the authors when they claim that the airline model is "flexible".

Integrated processes are often identified by MBA for time series which cannot be integrated (for example bounded time series) thus resulting in inefficient concurrent filters. Although the unit-root constraints imposed by 'traditional' models (like for example the airline-model) may be useful for short-term one-step-ahead forecasting they may be severely misleading when computing multi-step ahead forecasts of longer horizons from the same model. Therefore, model-based concurrent filters are inefficient if the weights of the symmetric filter decay slowly. This situation is common to many applications including well-known Henderson (13- or higher order) or Hodrick-Prescott (with $\lambda = 1600$) as well as model-based (canonical) filters for example. Very often, the above MBA select models for I(2)-processes which is a misspecification for (economic) time series that are not extremely trending. This evident misspecification (at least for the above bounded time series) may be at least partially due to the 'crude' seasonal operator in $(1-B^{12}) = (1-B)(1+B+...+B^{11})$ which induces a spurious unit root towards frequency zero: the additional $(1-B)$ of the I(2)-model compensates the distortion induced by the seasonal operator $(1+B+...+B^{11})$. As a result, MBA-filters generally perform worse in the passband - as suggested in figure 2 - because they satisfy unnecessarily severe restrictions due to erroneous unit root identification. Furthermore, the traditional $(1-B^s)$-operator ($s = 12$ for monthly data) generates filters which must (often unnecessarily) remove spectral power at *all* seasonal harmonics (see for example [2], section 4) at the expense of worse approximations at other frequencies implying poorer performances of the filter. A detailed analysis for two particular time series reveals that *MBA do not match the 'salient features' of the above data which exhibit a dynamic structure that cannot be accounted for by 'too parsimonious' models*. Out-of-sample results confirm that more flexible (richer parameterized) ZPC-ARMA-filters are able to account for the complex dynamics of practical time series - location, height and width of spectral peaks - without being affected by overfitting (these results straighforwardly extend to integrated processes, see [15]). The 'invisible hand' at work is the optimization criterion (4) which accounts for the *relevant* information much better than statistics based on (one-step ahead) model residuals.

Note that the shape of the symmetric filter is explicitly accounted for by the DFA in (4) whereas classical model identification and estimation does not care about the signal to be approximated. If misspecification is to be expected, then this issue becomes an additional topic with regard to efficiency: for a symmetric MA(3) filter the concurrent filter is based on one-step ahead forecasts only but for a symmetric MA(5) one- and two-steps ahead forecasts are needed. Therefore the best model should perform well for both forecasting-steps.

It is remarkable that three DFA-filters clearly outperform 36 specific MBA filters, especially if the identification procedures of MBA suggest very different DGP's for series belonging to the same cluster (filtered by the same concurrent DFA-filter). This empirical fact provides further strong evidence against efficiency of MBA. Moreover, it also suggests that the 'traditional' identification procedures of MBA might not be well suited for the signal extraction problem at hand. For the 36 time series analyzed in this article, the gain in efficiency (of the resulting concurrent filter) obtained by the automatic model identification of TRAMO over constant usage of the airline model is only 5% (differences between estimation routines were irrelevant, see section 3.1). Additional 13% are obtained by using three DFA-filters which are constrained to be identical within a cluster of 'similar' time series. However, by far the largest improvement is obtained by using specific DFA-filters, revealing that a lot of information has been lost by the aggregation of the periodograms (clustering) which could not be retrieved by the MBA. The gain in using the DFA - approximately 40% (in and out of sample) in the first two experimental designs (using original time series) and approximately 30% when using linearized time series (the third design favors the MBA) - is huge and in effect much larger than that obtained by using automatic model identification procedures. Note that the DFA performs much better than the MBA for every experimental design and for *all* series and that no 'identification' procedure is necessary since the same ARMA(15,15)-filter design is used for all series.

From a methodological point of view, the optimization procedure underlying MBA solves a statistical problem (one-step ahead forecasting) which is only indirectly related to signal extraction. This fact partly explains the reported inefficiencies. Other issues may be seen in insufficient flexibility (too sparsely parameterized models) and unspecific model design. Typical characteristics such as the location of potential spectral peaks are often known in advance. Only the height and the width of the peaks must then be matched by the filter. Focussing on these features only by using suitable designs (ZPC-filters) avoids overfitting. Also, the 'legitimity' of model-based signal definitions (for example canonical components as defined in SEATS) cannot be asserted anymore, especially in the context of 'heavy' model misspecification (false integration order) as the above examples demonstrate, see also [7], section 5 for further evidence in this direction when 'weak' model misspecification (correctly identified unit roots but misspecified model orders) is at work. Finally, since efficiency cannot plead for MBA, corresponding signal extraction software should also provide additional informations such as amplitude and time delay functions of the concurrent (boundary) filters, as we did in figure 5. These important characteristics (especially the time delay in the passband of the filter) can aid users in deciding whether a given asymmetric filter is well suited for a particular application or not. As argued in [7] "In particular, the gain function of the infinite symmetric model-based filter provided by SEATS can fail to show significant features of the finite filter gain function, and it provides virtually no insight into the properties of the one-sided concurrent filter, whose gain and phase delay offer more relevant information for most users of seasonally adjusted data".

Briefly: the statement that properties of a model-based asymmetric concurrent filter are not important because it is ought to be 'optimal' (efficient) should be revised in the light of the above results.

References

[1] J. Aston, D.F. Findley, K.Wills, and D.E.K. Martin. Generalizations of the box-jenkins airline-model with frequency-specific seasonal coefficients, 2003.

[2] W.R. Bell and D.E. Martin. Computation of asymmetric signal extraction mean squared error for arima component models. Technical report, Census Bureau, 2002.

[3] Michael P. Clements and David F. Hendry. *Forecasting Economic Time Series*. Cambridge University Press, Cambridge, 1998.

[4] Michael P. Clements and David F. Hendry. *Forecasting Non-Stationary Economic Time Series*. The MIT Press, 2004.

[5] J Dosse and C. Planas. Revisions in seasonal adjustment methods : an empirical comparison of x12-arima and seats. Downloadable from : http://forum.europa.eu.int/Public/irc/dsis/eurosam/library (documents of methodological studies), 1996.

[6] D. F. Findley, B.C. Monsell, W.R. Bell, M. C. Otto, and B.C. Chen. New capabilities and methods of the X-12-arima seasonal-adjustment program. *Journal of Business and Economic Statistics*, 16(2), 1998.

[7] D.F. Findley and D.E.K. Martin. Frequency domain analyses of seats and x-11/12-arima seasonal adjustment filters for short and moderate-length time series. Technical report, Census Bureau, 2003.

[8] A.C. Harvey. *Forecasting Structural Time Series Models and the Kalman Filter*. Cambridge University Press, Cambridge, 1989.

[9] K. McDonald-Johnson, T.T. Ta Nguyen, C.C. Hood, and Brian C. Monsell. Improving the automatic regarima model selection procedures of x-12-arima version 0.3, 2002.

[10] B.C. Monsell, J.A.D. Aston, and S.J. Koopman. Toward x-13?, 2003.

[11] W. Stier. *Methoden der Zeitreihenanalyse*. Springer, Berlin-Heidelberg-New York, 2001.

[12] W. Stier and M. Wildi. About model-based times series procedures: Some remarks to tramo/seats and census x-12-arima. *Allgemeines Statistisches Archiv*, (86), 2002.

[13] M. Wildi. Forecasting non-stationary financial data with oiir-filters and composed threshold models. In *Decision Technologies for Computational Finance*, pages 391–402. Kluwer Academic, 1998.

[14] M. Wildi. Trend estimation and turning-point detection for economic time series. In *Theorisations du long terme et depassement des phases depressives*. L'Harmattan, 2000.

[15] M. Wildi. *Signalextraction: Efficient estimation, Unit Roots and Early Detection of Turning Points*, volume 547. Lecture Notes in Economics and Mathematical Systems, Berlin/Heidelberg/New York, 2004.

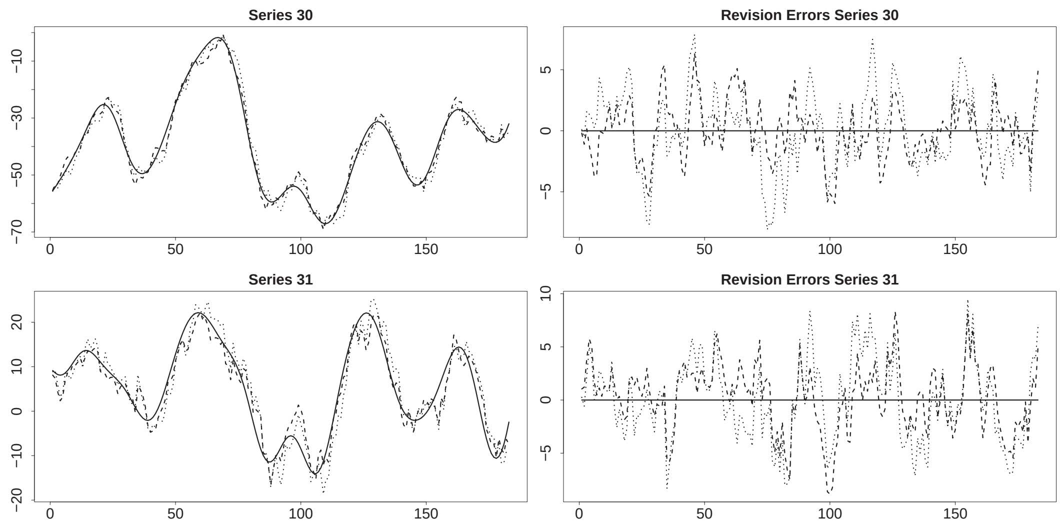

Figure 1: Boundary estimates (concurrent filter) and revision errors : MBA (dotted) and DFA (dashed)

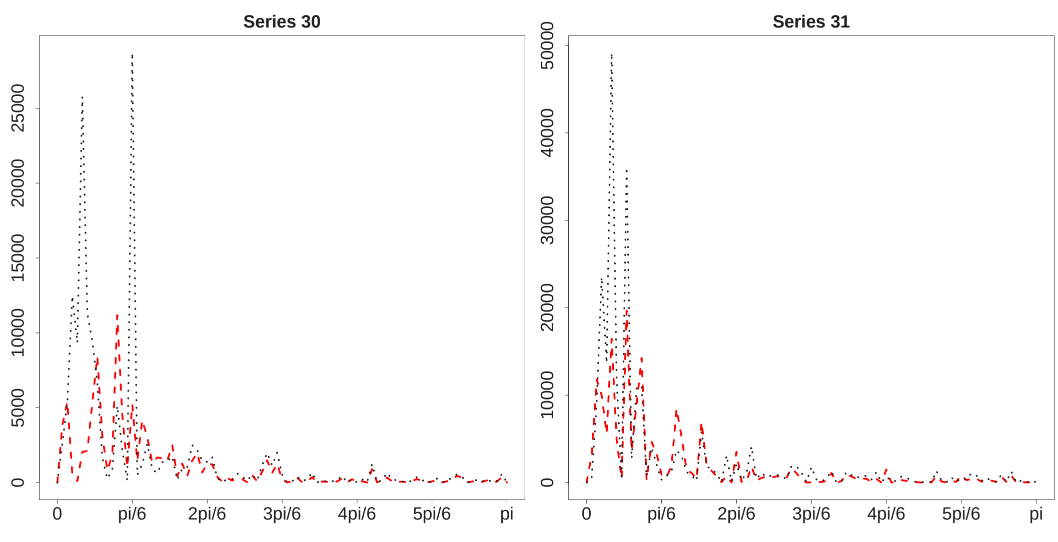

Figure 2: Periodograms of revision errors for MBA (dotted) and DFA (dashed)

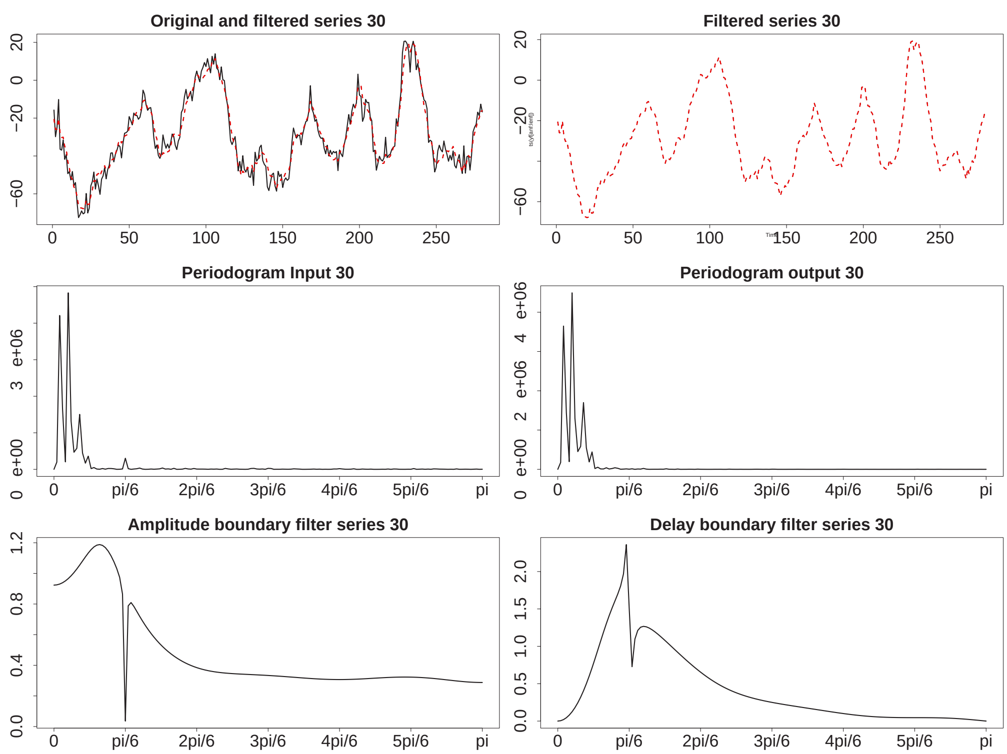

Figure 3: Characteristics of the asymmetric DFA-filter (series 30)

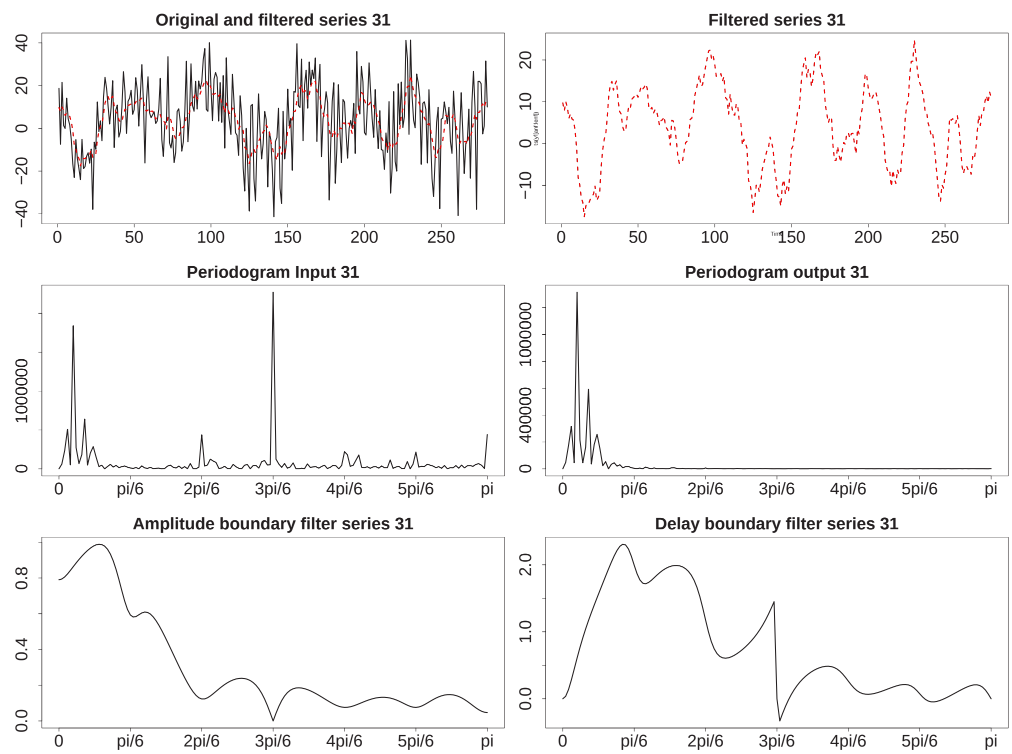

Figure 4: Characteristics of the asymmetric DFA-filter (series 31)

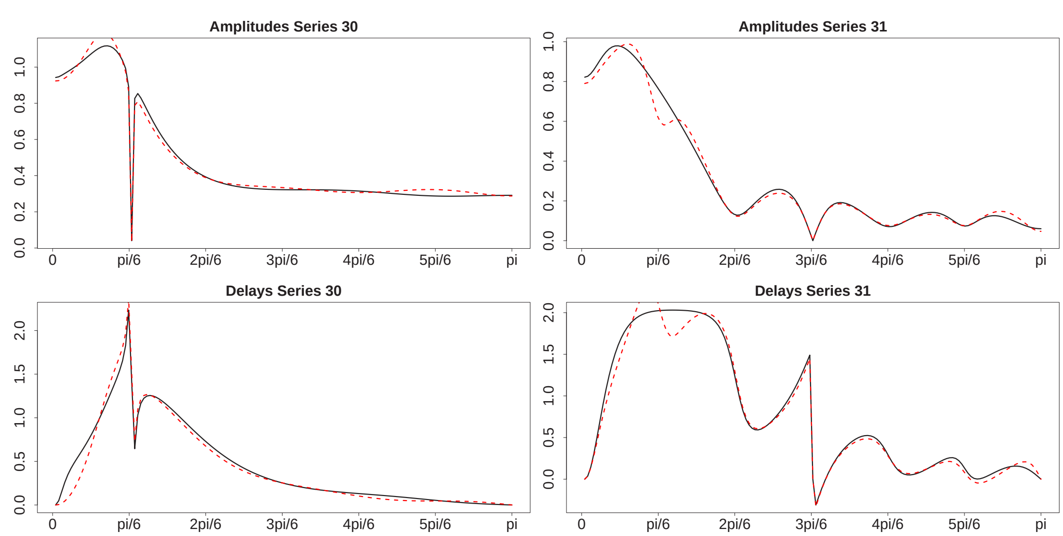

Figure 5: In- (shaded) and out-of sample (solid) filter characteristics

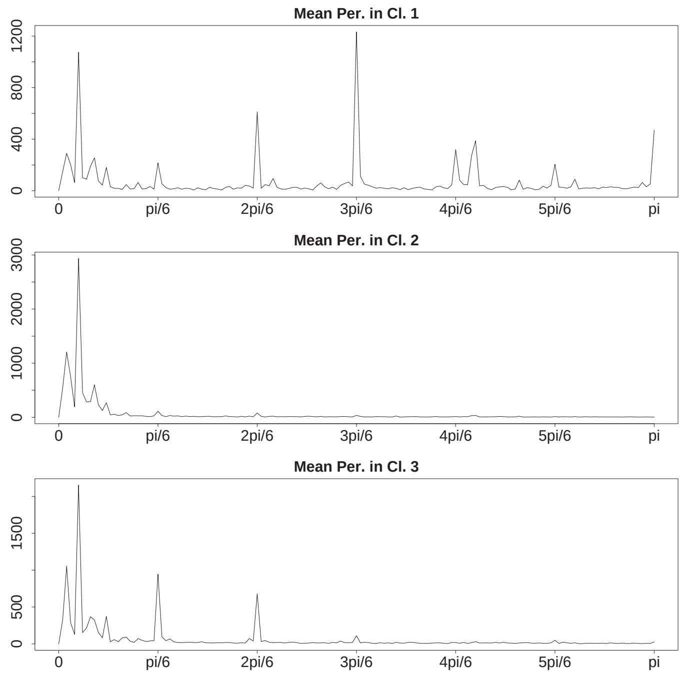

Figure 6: 'Mean periodograms'

Table 1: Models identified by TRAMO

Series	$(p, d, q) \times (P, D, Q)$	d	Series	$(p, d, q) \times (P, D, Q)$	d
1	(0,1,1)(1,0,1)	1	19	(0,1,1)(0,1,1)	2
2	(0,1,3)(0,0,1)	1	20	(3,1,0)(0,0,1)	1
3	(0,1,1)(0,1,1)	2	21	(0,1,1)(0,1,1)	2
4	(0,1,1)(1,0,0)	1	22	(0,1,1)(0,0,0)	1
5	(0,1,1)(0,1,1)	2	23	(0,1,1)(0,1,1)	2
6	(1,1,0)(0,0,0)	1	24	(0,1,1)(0,1,1)	2
7	(0,1,1)(0,1,1)	2	25	(0,1,1)(0,1,1)	2
8	(0,1,1)(0,1,1)	2	26	(3,1,0)(0,0,0)	1
9	(0,1,0)(0,1,1)	2	27	(0,1,1)(0,1,1)	2
10	(0,1,0)(0,1,1)	2	28	(2,1,0)(0,0,1)	1
11	(0,1,0)(0,1,1)	2	29	(0,1,1)(0,1,1)	2
12	(3,1,1)(0,0,0)	1	30	(0,1,1)(0,0,0)	1
13	(0,1,2)(0,0,0)	1	31	(0,1,1)(0,1,1)	2
14	(0,1,3)(0,0,0)	1	32	(2,1,0)(0,0,0)	1
15	(3,0,0)(0,1,1)	1	33	(1,0,1)(0,0,0)	0
16	(0,1,1)(1,0,0)	1	34	(0,1,1)(0,1,1)	2
17	(0,1,1)(0,1,1)	2	35	(0,1,3)(0,0,0)	1
18	(0,1,0)(0,0,1)	1	36	(0,1,1)(0,1,1)	2

Table 2: Models identified by X-12-ARIMA (only those which differ from TRAMO)

Series	$(p, d, q) \times (P, D, Q)$	d	Series	$(p, d, q) \times (P, D, Q)$	d
1	(0,1,1)(0,1,1)	2	16	(0,1,3)(0,1,1)	2
4	(0,1,1)(0,1,1)	2	18	(2,1,2)(0,1,1)	2
6	(0,1,3)(0,0,0)	1	20	(0,1,1)(0,1,1)	2
9	(0,1,1)(0,1,1)	2	22	(0,1,1)(0,1,1)	2
10	(0,1,1)(0,1,1)	2	26	(0,2,3)(0,0,0)	2
11	(0,1,1)(0,1,1)	2	28	(0,1,1)(0,0,1)	1
12	(3,1,1)(0,0,1)	1	32	(0,1,1)(0,1,1)	2
13	(0,1,1)(0,1,1)	2	33	(0,1,1)(0,1,1)	2
15	(0,1,1)(0,1,1)	2	35	(0,1,1)(0,1,1)	2

Table 3: Performance of DFA vs. MBA (‚in sample')

Series	TRAMO	X-12-T	X-12-A	Series	TRAMO	X-12-T	X-12-A
1	−71%	−64%	−46%	19	−42%	−40%	−40%
2	−25%	−22%	−22%	20	−43%	−44%	−51%
3	−21%	−21%	−21%	21	−27%	−33%	−33%
4	−74%	−70%	−68%	22	−14%	−12%	−7%
5	−47%	−39%	−39%	23	−14%	−16%	−16%
6	−67%	−67%	−45%	24	−26%	−28%	−28%
7	−33%	−33%	−33%	25	−22%	−20%	−20%
8	−15%	−16%	−16%	26	−54%	−54%	−56%
9	−26%	−27%	−23%	27	−42%	−35%	−35%
10	−17%	−16%	−16%	28	−27%	−27%	−30%
11	−19%	−19%	−18%	29	−29%	−29%	−29%
12	−37%	−40%	−40%	30	−76%	−75%	−75%
13	−35%	−34%	−32%	31	−42%	−38%	−38%
14	−38%	−37%	−37%	32	−32%	−31%	−34%
15	−20%	−22%	−50%	33	−20%	−17%	−43%
16	−52%	−51%	−33%	34	−28%	−27%	−27%
17	−28%	−28%	−28%	35	−15%	−15%	−44%
18	−94%	−111%	−2%	36	−35%	−34%	−34%
Mean					−36%	−36%	−34%

Table 4: Out of sample Performances of DFA and MBA

Series	DFA vs MBA	Out vs In (DFA)	Out vs In (MBA)
1	−57%	−16%	−8%
2	−57%	−13%	10%
3	−34%	−37%	−36%
4	−49%	20%	13%
5	−73%	−15%	10%
6	−32%	41%	32%
7	−26%	−67%	−77%
8	−56%	−90%	−32%
9	−6%	7%	−15%
10	−30%	23%	22%
11	−22%	17%	19%
12	−55%	6%	17%
13	−23%	43%	36%
14	−52%	1%	11%
15	−43%	−70%	4%
16	−28%	13%	27%
17	−35%	−72%	−44%
18	−88%	23%	22%
19	−61%	2%	36%
20	−18%	48%	49%
21	−40%	18%	27%
22	−4%	40%	41%
23	−45%	−86%	−37%
24	−53%	−42%	−10%
25	−3%	11%	17%
26	−25%	31%	18%
27	−26%	37%	30%
28	−70%	−56%	−7%
29	−27%	28%	23%
30	−81%	19%	21%
31	−37%	22%	8%
32	−45%	−2%	11%
33	−16%	−75%	−156%
34	−33%	−81%	−75%
35	3%	38%	21%
36	−24%	53%	51%
Mean	−38%	−5%	2%

Table 5: Performance of three DFA filters vs. MBA

Ser.nmb.	DFA vs. MBA	Cluster	Ser.nmb.	DFA vs. MBA	Cluster
1	−26%	1	19	−26%	1
2	−12%	2	20	−27%	2
3	−40%	3	21	26%	3
4	10%	2	22	−1%	2
5	−27%	1	23	−6%	3
6	−13%	2	24	−1%	3
7	0%	2	25	−3%	3
8	−1%	2	26	−11%	2
9	−19%	3	27	−28%	1
10	−5%	3	28	−33%	2
11	−5%	3	29	−27%	3
12	4%	2	30	−41%	2
13	−30%	2	31	−44%	1
14	−20%	2	32	−31%	2
15	−67%	1	33	28%	2
16	−15%	1	34	−23%	2
17	−11%	3	35	12%	2
18	−14%	2	36	−15%	2
Mean				−14%	