547 LECTURE NOTES IN ECONOMICS AND MATHEMATICAL SYSTEMS

Marc Wildi

Signal Extraction

Efficient Estimation,
'Unit-Root'-Tests
and Early Detection
of Turning Points

Springer

Lecture Notes in Economics and Mathematical Systems 547

Marc Wildi

Signal Extraction

Efficient Estimation, 'Unit Root'-Tests and Early Detection of Turning Points

With 80 Figures and 15 Tables

Springer

Author

PD Dr. Marc Wildi
Institute for Data Analysis and Processdesign (IDP)
Technopark / Jägerstrasse 2
Postfach 805
8401 Winterthur, Switzerland

Library of Congress Control Number: 2004111007

ISSN 0075-8442
ISBN 3-540-22935-3 Springer Berlin Heidelberg New York

Springer is a part of Springer Science+Business Media

springeronline.com

Printed in Germany

Typesetting: Camera ready by author
Cover design: *Erich Kirchner*, Heidelberg

Printed on acid-free paper 42/3130Di 5 4 3 2 1 0

Herzlichen Dank Herr Stier und liebe Grüsse an 2*A+B+V Wildi

Foreword

The material contained in this book originated in interrogations about modern practice in time series analysis.

- Why do we use models optimized with respect to one-step ahead forecasting performances for applications involving multi-step ahead forecasts?
- Why do we infer 'long-term' properties (unit-roots) of an unknown process from statistics essentially based on short-term one-step ahead forecasting performances of particular time series models?
- Are we able to detect turning-points of trend components earlier than with traditional signal extraction procedures?

The link between 'signal extraction' and the first two questions above is not immediate at first sight. Signal extraction problems are often solved by suitably designed symmetric filters. Towards the boundaries ($t = 1$ or $t = N$) of a time series a particular symmetric filter must be approximated by asymmetric filters. The time series literature proposes an intuitively straightforward solution for solving this problem:

- Stretch the observed time series by forecasts generated by a model.
- Apply the symmetric filter to the extended time series.

This approach is called '*model-based*'. Obviously, the forecast-horizon grows with the length of the symmetric filter. Model-identification and estimation of unknown parameters are then related to the above first two questions.

One may further ask, if this approximation problem and the way it is solved by model-based approaches are important topics for practical purposes? Consider some 'prominent' estimation problems:

- The determination of the seasonally adjusted actual unemployment rate.
- An assessment of the 'trend' of the actual GDP movement.
- Inferences about the 'global heating' in recently observed climatologic changes.

These problems all suggest that there is some kind of 'signal' which is overlapped by undesirable perturbations which mask the actual state of an interesting phenomenon. Formally, *actuality* of the estimates translates into *boundary* signal estimation. Signals often have a prospective component towards the boundary $t = N$: the detection of a turning-point of a trend component is informative about the future of the time series. So the corresponding estimation problem is highly relevant for many applications. Furthermore, Since model-based approaches like TRAMO/SEATS or Census X-12-ARIMA[1] are widely

[1]Although X-12-ARIMA is not a 'pure' model-based approach, see chapter 2, the procedure nevertheless relies on forecasts for computing boundary estimates.

used for 'signal extraction' one may then ask if the resulting method is *efficient*[2]?

The empirical results obtained in chapter 7 and more recently in Wildi, Schips[99][3] demonstrate that 'traditional' model-based boundary signal estimates are far from being efficient. The examples demonstrate that the relative mean-square error (between outputs of symmetric and asymmetric filters) can be reduced substantially (more than 30% in the mean over all time series considered) when using the efficient estimation method presented in this book. Moreover, the new method outperforms model-based approaches for *all* 41 time series in Wildi/Schips[99]. Optimal filter designs and properties of important statistics involved in the estimation problem are presented in chapters 3 and 4. The consistency, the efficiency and the asymptotic distribution of the resulting filter parameter estimates are derived in chapter 5 for a wide class of input signals (processes). An extension of this method which enables a *faster detection of turning points* for 'smooth' trend components is also presented in chapter 5. Chapter 6 presents finite sample issues and empirical examples are to be found in chapters 7 and 8.

As shown in chapter 7 as well as in Wildi/Schips[99] the observed inefficiency of model-based approaches is partly due to wrongly inferred unit-roots. The business survey data analyzed in Wildi/Schips[99] cannot be integrated because the time series are bounded. However, traditional unit-root tests such as (augmented) Dickey-Fuller or Phillips-Perron are often unable to reject the null hypothesis (integration) for such time series.
It is in fact strange that 'long-term' dynamics (unit-roots) are often inferred from statistics based on 'short term' one-step ahead forecasting performances of particular time series models. Experience suggests that short term forecasting performances generally do not allow for sufficiently strong rejection of the null hypothesis : *'Traditional' ADF- or PP-test-statistics may be well-suited for short-term (one-step ahead) forecasting but they are often misleading for problems requiring good multi-step ahead forecasting performances.*
In the general context of 'signal extraction', unit-roots are important because they are related to particular restrictions of the asymmetric filters, see chapter 5. Therefore, great attention has been devoted to 'unit-roots' in this particular context and new solutions - which 'fit' specifically the signal extraction problem - are presented in chapter 5.

[2]It is known that one- and multi-step-ahead forecasting performances may be conflicting, see chapter 1. Therefore it is surprising that few attention has been deserved to efficiency issues in signal extraction problems.

[3]The authors analyze the performance of trend boundary estimates for a representative sample of 41 business survey indicators

Contents

Part I Theory

Part I

Theory

1

Introduction

1.1 Overview

For many applications a well known problem is to 'extract' or equivalently to estimate some predefined 'signal' or component from a time series contaminated by 'noise' (which is not necessarily a white noise process). Consider

$$X_t = Y_t + \nu_t \tag{1.1}$$

where X_t is observed, Y_t is the interesting signal and ν_t overlaps and 'contaminates' the signal. Let $t \in \mathbb{Z}$ (discrete time) and assume $X_1, X_2, ..., X_N$ have been observed. The problem is to 'compute' values for the unknown $Y_1, Y_2, ..., Y_N$. The following figures illustrate some practically relevant signals for monthly economic time series.

- In fig.1.1, a particular time series (described in chapter 7) and a 'trend' defined by the canonical decomposition (see section 2.3) can be seen.
- In fig.1.2, the same time series and the 'seasonally adjusted' component (signal) defined by the canonical decomposition (see section 2.3) can be seen.
- Finally, both signals are compared in fig.1.3.

These examples are treated in detail in chapters 7 and 8. The signals are documented in chapter 2.

A general approach for estimating Y_t given X_t in 1.1 relies on stochastic processes. The observable process X_t is then called the *input process* or the *input signal* and Y_t is called the *output signal* (this is because Y_t can often be estimated by the output of a particular filter, see section 1.2 below). It is intuitively reasonable to allow a signal estimation method to depend on the particular stochastic 'properties' of the input process X_t in 1.1. As an example, assume

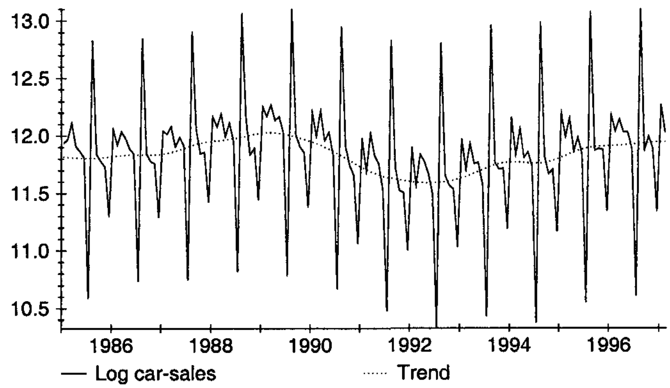

Fig. 1.1. Original series and trend

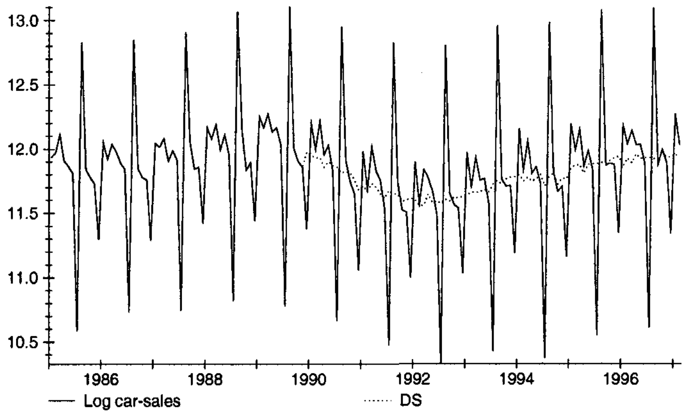

Fig. 1.2. Original and seasonally adjusted series

$$X_{1t} := Y_t + A_1 \cos(t\omega_1 + \Phi_1)$$
$$X_{2t} := Y_t + A_1 \cos(t\omega_1 + \Phi_1) + A_2 \cos(t\omega_2 + \Phi_2)$$

where $\omega_1 \neq \omega_2$ and Φ_1 and Φ_2 are independent random variables uniformly distributed in $[-\pi, \pi]$. Suppose the interesting signal is given by $Y_t = \cos(t\omega + \Phi)$, where Φ is uniformly distributed in the interval $[-\pi, \pi]$. X_{it}, $i = 1, 2$ and Y_t

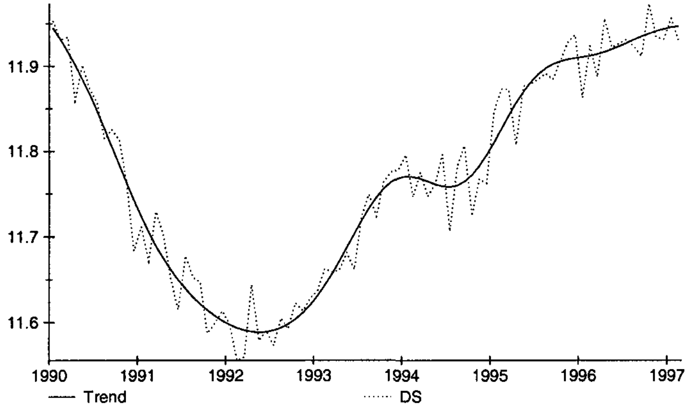

Fig. 1.3. Trend and seasonally adjusted series

are particular harmonic processes. The latter (Y_t) can be extracted from X_{2t} by eliminating $A_1\cos(t\omega_1+\Phi_1)$ and $A_2\cos(t\omega_2+\Phi_2)$. This may be achieved by a suitable 'filter' (see chapter 3). If the input process is X_{1t} instead, then the same filter could be used for extracting Y_t in principle. However, it is readily seen that the resulting estimation method would be unnecessarily complicated. In fact, a simpler filter eliminating $A_1\cos(t\omega_1+\Phi_1)$ 'only' could be used. For processes which are not deterministic (as the harmonic processes above) too complicated devices are generally inefficient: eliminating additional components involves a 'cost' which is quantified in chapter 5. Therefore, knowledge of particular stochastic properties of the DGP (Data Generating Process) of X_t is necessary for computing *efficient* signal estimates. If the relevant properties of X_t are unknown, then they must be inferred from the sample $X_1, ..., X_N$. Model-based approaches (MBA) are widely used for solving signal extraction problems because they *try* to infer the DGP of X_t from a finite sample $X_1, ..., X_N$. Resulting signal estimates can account for stochastic properties of the input signal X_t but *the efficiency cannot be asserted in general* (see section 1.2).

A new method, called *direct filter approach* (DFA) is presented here for solving the signal estimation problem. The main advantages of this approach are *efficiency* and *flexibility*. Filters can be optimized with respect to the traditional mean square error criterion or with respect to another practically important objective, namely the *'fast detection of turning-points'*. Often, signal estimates are subject to significant time delays towards the end point $t = N$ of a finite sample. Therefore, 'turning-points' of the signal cannot be detected 'in time'. The DFA enables to constrain filters such that the time delay becomes smaller. These issues are analyzed in chapters 3 and 5. Model-based approaches do not allow for time delay constraints.

Unit-roots of the DGP are important properties of the input signal which affect the performance of the estimation procedure if they are ignored, see chapter 7. It is shown in chapter 5 that unit-roots of the DGP 'translate' into particular constraints for the optimal asymmetric filter. In principle these constraints allow for more general non-stationarities than 'unit-roots' of the DGP only. A formal procedure for testing these hypotheses (constraints) is presented in chapter 5. The advantage of such a test is that it is specifically designed for the signal estimation problem whereas 'traditional' unit-root tests (such as Dickey-Fuller or Phillips-Perron for example) are derived from one-step ahead forecasting performances (of a model for the DGP) only. Therefore, the power of 'traditional' tests against stationary alternatives with roots close to the unit-circle is typically low (this situation is common for a lot of applications including many economic time series) because a 'long-term' property (a unit-root at frequency zero) is inferred from a statistic based on 'short-term' performances. Cochrane [18], p.914, argues "These models (ARIMA) ... draw inferences about the long-run dynamics from a model fit to the short-run dynamics ... However, if the long-run dynamics cannot be captured in the model used to study the short-run, these identification procedures bias conclusions about long-run behavior". The new test implicitly accounts for one- *and* multi-step ahead forecasting performances and it is explicitly designed for the signal estimation problem.

For the proposed DFA, particular attention is accorded to *finite sample issues* (overfitting problem, see chapter 6). 'Parsimony' in the sense of 'cautiously' parameterized models (see Box and Jenkins [9]) is a relevant concept. Feldstein [31], p.829, argues: "A useful model is not one that is 'true' or 'realistic' but one that is parsimonious, plausible and informative". The proposed direct filter approach is based on a new filter class, so called *Zero-Pole-Combination* (ZPC-) filters. ZPC-filters are obtained by a parsimonious parameterization of ARMA-filters for which each parameter (degree of freedom) becomes straightforwardly interpretable, see chapter 3.
Although the principle of parsimony may help in alleviating the overfitting problem, it is not a 'panacea'. Therefore, new solutions are proposed for the DFA in order to avoid specific overfitting problems, see chapter 6. Empirical evidences listed in chapters 7 and 8 confirm the effectiveness of the proposed method. Simulated and 'real-world' time series are analyzed and the performances of the DFA and the MBA are compared both 'in' and 'out of sample'.

A signal estimation method which relies on an explicit model for the DGP of X_t is called a MBA. Different methods have been proposed which are characterized by various assumptions and/or model structures. Chapter 2 provides an (necessarily limited) overview on the topic. Model-based approaches are often referenced as '*the* MBA' here and in the following chapters (despite methodological differences of various approaches) by opposition to '*the* DFA' which does not rely on an explicit model for the DGP of X_t. A brief description of the MBA is proposed in the following section. It is suggested that the optimization criterion underlying the MBA does not 'match' the signal estimation problem for misspecified models (which is the rule in practice). Therefore, model-based estimates may be inefficient. Empirical results in chapter 7 as well as in Wildi/Schips[99] confirm this statement.

1.2 A General Model-Based-Approach

For 'general' (stationary or non-stationary integrated) linear stochastic processes, the signal estimation problem is solved by *linear filters*. A (linear) filter is a sequence γ_k, $k \in \mathbb{Z}$ of square summable (in our context real) numbers: $\sum_{k=-\infty}^{\infty} |\gamma_k|^2 < \infty$. MA-, AR- and ARMA-filters are characterized by particular finite sets of parameters generating γ_k. If the sequences $\hat{Y}_t$ and X_t are related by

$$\hat{Y}_t = \sum_{k=-\infty}^{\infty} \gamma_k X_{t-k} \tag{1.2}$$

then $\hat{Y}_t, X_t$ are called the *output* and the *input signals* of the filter γ_k respectively. If $X_t = Y_t + \nu_t$ where X_t, Y_t and ν_t are linear stochastic processes, then it has been shown that the best estimate $\hat{Y}_t$ (in the mean square sense) of Y_t is the output of a particular linear filter if some 'mild' assumptions are satisfied (see Whittle [95] for stationary X_t and Bell [4] for non-stationary integrated X_t; results for non-linear processes are presented in Gihman and Skorohod [39], p.273-274).

For a realization of infinite length $(..., X_{-2}, X_{-1}, X_0, X_1, X_2, ...)$ (infinite sample), the best extraction filter is generally *symmetric* ($\gamma_k = \gamma_{-k}, k > 0$) and of *infinite order* (i.e. there does not exist a n_0 such that $\gamma_k = 0$ for all $k > n_0$). The symmetry ensures that the phase or equivalently the time shift of the filter vanishes, see chapter 3. The following example illustrates these properties for a particular signal estimation problem: X_t is given by 1.1, where ν_t is a white noise process and Y_t is a random walk (so called Muth-model, see for example Mills [67] p.69 ff.):

$$\begin{aligned} X_t &= Y_t + \nu_t \\ Y_t &= Y_{t-1} + \epsilon_t \end{aligned}$$

where ϵ_t, ν_t are independent iid sequences. The best mean square estimate of the signal (the random-walk) is then given by :

$$\hat{Y}_t = \frac{(1-\theta)^2}{1-\theta^2} \sum_{k=-\infty}^{\infty} \theta^{|k|} X_{t-k}$$

where θ depends on the signal to noise ratio (the ratio of the variances of ν_t and ϵ_t). The optimal filter coefficients $\frac{(1-\theta)^2}{1-\theta^2}\theta^{|k|}$ are symmetric and decay exponentially fast but they never vanish if $\theta \neq 0$.

Clearly, filters of infinite order cannot be used if the available input sample $X_1, ..., X_N$ is finite. But the symmetry property leads to problems even for filters of finite orders. Difficulties arise if t is 'close' to the boundaries $t = 1$ or $t = N$ of the sample. Therefore, the filter output $\hat{Y}_t$ of the symmetric filter (which solves the so called *signal extraction problem*) must be estimated too, say by $\hat{\hat{Y}}_t$. The latter is called a solution of the *finite sample signal estimation problem*. Model-based approaches provide solutions for both $\hat{Y}_t$ and $\hat{\hat{Y}}_t$. The latter problem is solved as follows (see Stier and Wildi [87] and Wildi [98]) :

- replace unknown X_i ($i < 1$ or $i > N$) in 1.2 by fore- and/or backcasts $\hat{X}_i$ generated from a model of the DGP (for example an ARIMA or a RegARIMA-model, see Findley et al.[32] or EUROSTAT [30])
- apply the symmetric filter $(\gamma_k)_{k\in\mathbb{Z}}$ to the 'extended' sample $X_t^e := \begin{cases} \hat{X}_t & t \notin \{1, ..., N\} \\ X_t & \text{else} \end{cases}$

One obtains :

$$\begin{aligned}
\hat{\hat{Y}}_t &= \sum_{k=-\infty}^{\infty} \gamma_k X_{t-k}^e && (1.3)\\
&= \sum_{k=t-N}^{t-1} \gamma_k X_{t-k} + \sum_{k=-\infty}^{t-N-1} \gamma_k \hat{X}_{t-k} + \sum_{k=t}^{\infty} \gamma_k \hat{X}_{t-k}\\
&= \sum_{k=t-N}^{t-1} \gamma_k X_{t-k} + \sum_{k=-\infty}^{t-N-1} \gamma_k \sum_{j=1}^{N} a_{t-k,j} X_j + \sum_{k=t}^{\infty} \gamma_k \sum_{j=1}^{N} a_{t-k,j} X_j\\
&= \sum_{j=1}^{N} \gamma_{t-j} X_j + \sum_{j=1}^{N} \left(\sum_{k=-\infty}^{t-N-1} \gamma_k a_{t-k,j} \right) X_j + \sum_{j=1}^{N} \left(\sum_{k=t}^{\infty} \gamma_k a_{t-k,j} \right) X_j\\
&= \sum_{j=1}^{N} \hat{\gamma}_{t-j} X_j && (1.4)
\end{aligned}$$

where $a_{t-k,j}$ are the coefficients of X_j, $j = 1, ..., N$, in the (linear) forecasting function of $\hat{X}_{t-k}$ if $t-k \notin \{1, ..., N\}$ and

$$\hat{\gamma}_{t-j} := \begin{cases} \gamma_{t-j} + \sum_{k=-\infty}^{t-N-1} \gamma_k a_{t-k,j} + \sum_{k=t}^{\infty} \gamma_k a_{t-k,j} & j = 1, ..., N \\ 0 & \text{else} \end{cases} \qquad (1.5)$$

Note that $(\hat{\gamma}_{t-j})_{j=1,...,N}$ depends on t and that it is an *asymmetric* filter in general.

If the DGP of X_t is known, then the above estimate $\hat{\hat{Y}}_t$ satisfies a mean square optimality criterion (see for example Cleveland [15], Bell [3], Bell [4], Huot and all [55] and Bobbitt and Otto [7]). The 'true model' (DGP) can be used for

1. linearizing the sample (identify 'outliers' or 'shifts' and remove them from the original series)
2. supplying missing values
3. defining components and corresponding symmetric signal extraction filters for realizations of infinite length (see chapter 2)
4. supplying fore- and backcasts in order to compute signal estimates for finite samples.

In the following, the last point i.e. the determination of an efficient signal estimate for finite samples is analyzed. This is an important problem for many applications (an example is given in section 1.5) because in practice only finitely many observations of an input process X_t are available. It is now suggested that the MBA does not efficiently solve this problem if the DGP is unknown.

If the DGP is unknown, then a 'suitable' model must first be identified. In this case, 'misspecification' is the rule for most applications, see for example Box [8]. Therefore, it is generally impossible to assert optimality properties for the proposed MBA. Also, in case of misspecification the minimization of the one-step ahead mean-square forecasting error does not necessarily 'match' the signal estimation problem (for finite samples) because 1.3 involves one- *and multi-step-ahead* forecasts. Clements and Hendry [14], p.244, argue : "as it is not possible to prove that 1-step estimation is optimal when models are misspecified, dynamic estimation could improve multi-period forecast accuracy" (dynamic estimation means that parameters of forecasting functions are estimated separately for each forecasting step, by minimizing directly the corresponding forecasting error) and p.282 "Indeed the 'best' model on 1-step forecasts need not dominate at longer horizons". However, dynamic estimation is cumbersome and it is not a 'panacea', as shown by the same authors. With regards to the model selection procedure, Clements and Hendry p.281. claim "we find that the usual criteria based on t- and F-tests are not applicable when models are to be chosen on the basis of their ability to multi-step forecast". As a result, inferences based on 'traditional' tests do not straightforwardly extend to estimation problems involving multi-step ahead forecasts (such as the signal estimation problem). But even if the right model has been selected (for example in an artificial simulation context), Clements and Hendry are warning against careless use p.292 "... a poor forecast could result from the estimated DGP relative to the false autoregressive model" (in their study, the

'false' model is a pure random-walk model whereas the true DGP is a stationary process with an AR-root close to one). The authors show that the relative performances of 'true' and 'false' models generally depend on the chosen forecast horizon.
Consequently, *the model-based optimization procedure does not 'match' the signal estimation problem for finite samples if the DGP is unknown because one-step and multi-step ahead forecasting performances are generally conflicting in the presence of misspecification.* In fact $E[(\hat{Y}_t - \hat{\hat{Y}}_t)^2]$ should be minimized instead of the mean square error of the residuum in the model equation for X_t. More generally, *optimizing with respect to 'short term' performances (one-step ahead forecasts) may be misleading when estimating 'long term' components (like a trend for example).*
The approximation of $\hat{Y}_t$ by $\hat{\hat{Y}}_t$ can be stated in terms of a filter approximation problem. For that purpose, a suitable 'distance' measure is needed. The DFA bases on the minimization of such a measure. It is shown in chapter 5 that the solution of the corresponding optimization criterion minimizes $E[(Y_t - \hat{\hat{Y}}_t)^2]$ up to an error term which is smallest among a general class of estimators. Also, the asymptotic distribution of the estimated filter parameters can be derived, see chapter 5. Therefore, inferences for the DFA are not based on one-step ahead performances only (as for the MBA) but implicitly account for one- and multi-step ahead performances simultaneously. This is particularly important when testing for unit-roots for example, see chapters 5 and 7, since unit-roots determine the 'long-term' dynamics of a process.

Before introducing the DFA, a well known identification problem is stated in the following section.

1.3 An Identification Problem

Let

$$X_t = T_t + C_t + S_t + I_t \tag{1.6}$$

Then there are $4N$ unknowns or unobservable variables for N equations only. Without additional (strong) assumptions the components on the right hand side are unidentified. To simplify, suppose one is interested in estimating the trend T_t given $X_1, ..., X_N$. If it is assumed that the trend evolves according to a predefined deterministic time pattern (for example a polynomial in t) then 'ad hoc' filters can be used (for example a Spencer filter, see Brockwell and Davis [10] and Kendall and Stuart [57] or a Henderson filter, see Gray and Thomson [41]). However, components such as the trend are often assumed to be stochastic. In this case various identifying assumptions exist like for example:

- impose perfect dependence of the components so that knowledge of a particular one determines the others (see Beveridge and Nelson [6] and section 2.2 below);
- impose independence of the components and regularity or smoothness of trend and seasonal components (see the canonical decomposition in Hillmer and Tiao [52], Burman [11] and section 2.3 below).
- specify the individual component-models a priori (see the structural models approach in Grether and al. [70], Harvey [47] and section 2.4 below);

The above methods rely on explicit components (see section 2): the components are then estimated by the output of a particular 'extraction' filter. Alternatively, components could be defined implicitly by the output of a filter satisfying a particular criterion. As an example, the output T_t of a Hodrick-Prescott filter minimizes

$$\sum_{t=1}^{N}(X_t - T_t)^2 + \lambda \sum_{t=2}^{N-1}((T_{t+1} - T_t) - (T_t - T_{t-1}))^2 \tag{1.7}$$

where λ is given a priori. Larger λ lead to increased 'smoothness' of the filter output, see Hodrick and Prescott [54]. The first term penalizes deviations of T_t from the original time series and the second one penalizes 'roughness' (as defined by the mean of the squared second order differences). A similar approach underlies the Henderson filter, see Henderson [51] and section 2.5 below. Many of these methods were introduced by Whittaker [93] and [94]. At first sight, the identification problem seems to be 'circumvented' by implicit component definitions. However, criteria such as 1.7 are often difficult to interpret. For the Henderson filter, Wallis [92] p.164 argues: "... nor any later author has asked whether the symmetric Henderson filter produces a good estimate of the trend, however: for this purpose the trend is simply defined as the Henderson output". Moreover, the identification problem is often shifted towards the more or less arbitrary choice of a particular parameter of the filter (for example λ in 1.7).
The following fig.1.4 plots the Hodrick-Prescott 'growth component' T_t (solid line, $\lambda := 1600$ is a 'default' setting for many applications) and the canonical trend from TRAMO/SEATS (dotted line, see section 2.3) for a particular time series (UK-car-sales series, see chapter 7).

To summarize, the signal *identification* problem can be stated as follows

- different particular signal definitions generally lead to different components, see for example figs.1.4 and 2.11;
- a priori knowledge is always necessary for a unique identification of the components in 1.6, due to the 'large' number of unobservable variables (which define the so called 'structural form' of the process X_t). Therefore, a 'universal' definition of unobservable components of a time series is impossible. At last, implicit subjective convictions based on individual experience seem to motivate particular definitions.

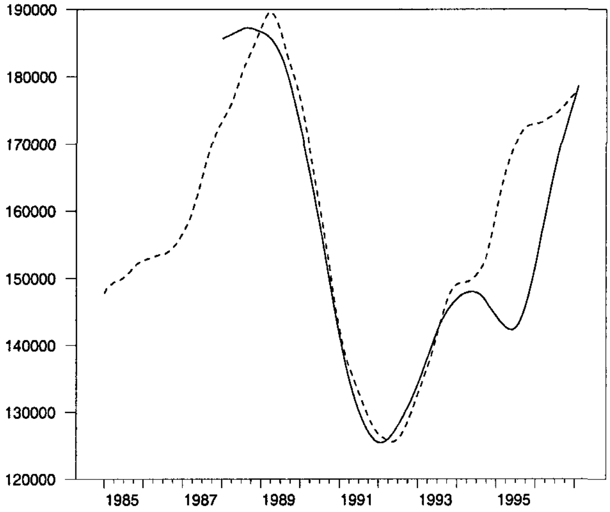

Fig. 1.4. HP- and Tramo/SEATS-Trend

Bell [5] p. 176 argues that "in seasonal adjustment the components are really artificial constructs presumed useful to estimate; there is no objective 'truth"'. In agreement with this comment, neither a new component definition nor a corresponding symmetric extraction filter are proposed here. Instead, *the signal estimation problem for finite samples ($\hat{\hat{Y}}_t$) is stressed: given $t \in \{1, ..., N\}$ and $\hat{Y}_t$ the output of a symmetric signal extraction (or 'smoothing') filter of possibly infinite order, find $\hat{\hat{Y}}_t$ which approximates $\hat{Y}_t$ given $X_1, ..., X_N$.*

In the next section, the DFA is briefly introduced. This is a new signal estimation method for finite samples. The presentation is informal. 'Technical' issues are postponed to following chapters.

1.4 The Direct Filter Approach

The following section relies on Wildi [98]. Suppose (the output of) some symmetric filter with transfer function $\Gamma(\omega)$, $-\pi \leq \omega \leq \pi$ must be approximated by (the output of) an asymmetric filter with transfer function $\hat{\Gamma}(\omega)$. As seen in the preceding section, the asymmetry results from the 'truncation' of realizations of infinite length. For notational convenience one 'hat' of the estimates $\hat{Y}_t$ and $\hat{\hat{Y}}_t$ in section 1.2 is eliminated. Therefore, $\hat{Y}_t$ becomes Y_t (the output of the symmetric filter) and $\hat{\hat{Y}}_t$ becomes $\hat{Y}_t$ (the output of the asymmetric filter). Assume $\Omega_N := \{\omega_k | \omega_k = k2\pi/N, |k| = 0, ..., [N/2]\}$ where $[N/2]$ is the greatest integer smaller or equal to $N/2$ and N is the sample size. Assume for simplicity of exposition that X_t is stationary (generalizations for non-stationary integrated processes are provided in chapter 5) and let

$$I_{NX}(\omega_k) := \frac{1}{2\pi N}\left|\sum_{t=1}^{N} X_t \exp(-i\omega_k t)\right|^2 \tag{1.8}$$

denote the *periodogram* of the input process computed for $\omega_k \in \Omega_N$, see chapter 4. Then, under suitable regularity assumptions (see chapter 5) the solution $\hat{\Gamma}_0(\cdot)$ of

$$\min_{\hat{\Gamma}} \frac{2\pi}{N} \sum_{k=-[N/2]}^{[N/2]} |\Gamma(\omega_k) - \hat{\Gamma}(\omega_k)|^2 I_{NX}(\omega_k) \tag{1.9}$$

generates an output $\hat{Y}_{t0}$ which minimizes $E[(Y_t - \hat{Y}_t)^2]$ up to an asymptotically negligible error term which is smallest possible (for a given class of estimators, see below). This result was first stated in Wildi [96]. The solution of 1.9 is attained within a general class of filters described in chapter 3. An intuitive explanation of the preceding statement can be given by considering the following approximation:

$$\frac{2\pi}{N} \sum_{k=-[N/2]}^{[N/2]} |\Gamma(\omega_k) - \hat{\Gamma}(\omega_k)|^2 I_{NX}(\omega_k) \simeq \frac{2\pi}{N} \sum_{k=-[N/2]}^{[N/2]} I_{N\Delta Y}(\omega_k) \tag{1.10}$$

$$= \frac{1}{N}\sum_{t=1}^{N}(Y_t - \hat{Y}_t)^2 \tag{1.11}$$

where $\Delta Y_t := Y_t - \hat{Y}_t$. The approximation 1.10 corresponds to a *finite sample convolution* and 1.11 corresponds to *a finite sample spectral decomposition* of the mean square filter approximation error (see chapter 5). Under suitable regularity assumptions, 1.11 is a *best linear unbiased estimate* (BLUE) of the theoretical mean square error $E[(Y_t - \hat{Y}_t)^2]$, see chapter 5. Efficiency of the DFA then depends on the error term in the approximation 1.10: it is shown that the expression on the left hand side is a *superconsistent estimate* of 1.11, see chapter 5. Therefore, the DFA 'inherits' the efficiency property (BLUE) of 1.11, i.e. $\hat{Y}_{t0}$ minimizes $E[(Y_t - \hat{Y}_t)^2]$ up to an error term which is smallest possible among the class of linear estimators (of $E[(Y_t - \hat{Y}_t)^2]$). Note that in general Y_t and therefore 1.11 and $E[(Y_t - \hat{Y}_t)^2]$ are unknown for finite samples whereas the left hand side of 1.10 can be computed.

In order to derive the consistency and the efficiency as well as the distribution of the estimated filter parameters for a large class of input signals (including non-stationary integrated processes) technical results involving the periodogram 1.8 are needed. These are reported in chapter 4 and in the appendix. It is shown in chapter 5 that 1.9 can be generalized so that the time delay of the resulting filter is 'smaller'. It is then possible to detect 'turning-points' of a particular component earlier, see chapter 8.

The proposed signal estimation method for finite samples is called a *direct* filter approach because the coefficients $\hat{\gamma}_k$ of the resulting asymmetric filter are computed 'directly' from the minimization of (an efficient estimate of) the mean square error $E[(Y_t - \hat{Y}_t)^2]$. In comparison, the filter coefficients of model-based approaches are derived indirectly from the equivalence between 1.3 and 1.4. They rely on the minimization of the mean square *one*-step ahead forecasting error of the model (whereas the signal estimation problem requires good one- and multi-step ahead forecasting performances). Moreover, time constraints (for the resulting asymmetric filter) cannot be 'build' into 1.3 for the MBA so that turning-points of trend components cannot be detected 'earlier'.

In the following section, a typical application for an efficient finite sample signal estimation method is provided. Also, the content of the following chapters is briefly summarized.

1.5 Summary

For economic time series, interesting signals are often seasonally adjusted components or trends, see chapter 2 (recall that component definitions depend on strong a priori assumptions, see section 1.3). An efficient and general signal estimation method is needed for these important applications because economic time series are characterized by randomness (the DGP is not deterministic) and complex dynamics. Moreover, 'typical' users are often interested in signal estimates for time points near the upper boundary $t = N$[1]. Consequently, filters are heavily asymmetric so that efficient estimation methods are required.

A new method, the DFA, is presented here. The book is organized as follows:

- In chapter 2, model-based approaches are presented. The aim is not to provide an exhaustive list of existent methods but to describe established procedures which are implemented in 'widely used' software packages. The objective is to compare the DFA to established MBA.
- The main concepts needed for the description of filters in the frequency domain (such as transfer functions, amplitude functions or phase functions) are proposed in chapter 3. A new filter class (ZPC-filters) is derived whose characteristics 'match' the signal estimation problem.
- For the DFA, an eminent role is awarded to the periodogram (or to statistics directly related to the periodogram). It 'collects' and transforms the

[1] For assessing the actual state of the 'business cycle' for example

information of the sample $X_1, ..., X_N$ into a form suitable for the signal estimation problem. Therefore, properties of the periodogram and technical details related to the DFA are analyzed in chapter 4. In particular, the statistic is analyzed for integrated processes. Stochastic properties of *squared* periodogram ordinates are analyzed in the appendix. Both kind of results are omitted in the 'traditional' time series literature and are needed here for proving theoretical results in chapter 5. An explorative instrument for assessing possible 'unit-root misspecification' of the filter design for the DFA is proposed also.

- The main theoretical results for the DFA are reported in chapter 5: the consistency, the efficiency, the generalization to non-stationary integrated input processes, the generalized conditional optimization (resulting in asymmetric filters with smaller time delays) and the asymptotic distribution of the estimated filter parameters (which enables hypothesis testing). In particular, a generalized unit-root test is proposed which is designed for the signal estimation problem.
- In order to prove the results in chapter 5, regularity assumptions are needed. One of these assumptions is directly related to finite sample issues (overfitting problem). Therefore, the overfitting problem is analyzed in chapter 6. Overparameterization and overfitting are distinguished and new procedures are proposed for 'tackling' their various aspects. An estimation of the order of the asymmetric filter is presented (which avoids more specifically overparameterization), founding on the asymptotic distribution of the parameter estimates. The proposed method does not rely on 'traditional' information criteria, because the DGP of X_t is not of immediate concern. However, it is shown in the appendix that 'traditional' information criteria (like AIC for example) may be considered as special cases of the proposed method. Also, new procedures ensuring regularity of the DFA solution are proposed which solve specific overfitting problems. The key idea behind these new methods is to modify the original optimization criterion such that overfitting becomes 'measurable'. It is felt that these ideas may be useful also when modelling the DGP for the MBA.
- Empirical results which are based on the simulation of artificial processes (I(2), I(1) and stationary processes) and on a 'real-world' time series are presented in chapter 7. The DFA is compared with the MBA with respect to mean square performances. It is shown that the DFA performs as well as maximum likelihood estimates for artificial times series. If the DGP is unknown, as is the case for the 'real-world' time series, the DFA outperforms two established MBA, namely TRAMO/SEATS and CENSUS X-12-ARIMA (see chapter 2 for a definition). The increased performance is achieved with respect to various signal definitions (two different trend signals and a particular seasonal adjustment) both 'in' and 'out of sample'. It is also suggested that statistics relying on the one-step ahead forecasts, like 'traditional' unit-root tests (augmented Dickey-Fuller and Phillips-Perron tests) or diagnostic tests (like for example Ljung-Box tests) may

be misleading for the signal estimation problem if the true DGP is unknown. Instead, specific instruments derived in chapters 4, 5 and 6 are used for determining the optimal filter design for the DFA. These instruments, which are based on estimated filter errors (rather than one-step ahead forecasting errors of the model), indicate smaller integration orders for the analyzed time series (I(1)- instead of I(2)-processes as 'proposed' by the majority of the unit-root tests). A possible explanation for these differences may be seen in the fact that filter errors implicitly account for one- *and* multi-step ahead forecasts simultaneously. A further analysis of the revision errors (filter approximation errors) suggests that the I(2)-hypothesis should be rejected indeed.
- Finally, an empirical comparison of the DFA and the MBA with respect to their ability of detecting 'turning-points' (of two different trend components) is conducted in chapter 8. The MBA is compared with the 'original' DFA and with the result of a generalized constrained optimization (whose filter solution has a smaller time delay). As in the preceding chapter, the DFA generally outperforms the MBA with respect to the proposed criterion.

In the following chapter 2, well established model-based approaches are presented. Two of them are used as 'benchmarks' in chapters 7 and 8.

2

Model-Based Approaches

2.1 Introduction

Model-based approaches attempt to identify the DGP of the input process and to estimate its parameters. They provide

1. Definitions of the theoretical components Y_{tj} (identification), where $j = 1, ..., n$ and n is the number of components.
2. Estimates $\hat{Y}_{tj}$ of the components for realizations of infinite length.
3. Estimates $\hat{\hat{Y}}_{tj}$ of the components for finite samples.

The general identification problem analyzed in section 1.3 led us to examine the last estimation problem only. Therefore, we here use the terminology 'model-based approach' whenever a method relies on back- or forecasts generated by a model for approximating $\hat{Y}_{tj}$ by $\hat{\hat{Y}}_{tj}$. From this perspective, the well-known X-11-ARIMA and X-12-ARIMA procedures can be considered as 'model-based' although the definitions of the signals at the first stage are 'implicit' (not model-based), see for example Dagum [22], Findley et al. [32] and section 2.5 below.

Most of the approaches to be presented here are based on the following two decompositions of X_t

$$X_t = T_t + C_t + S_t + I_t \tag{2.1}$$

$$X_t = T_t C_t S_t I_t \tag{2.2}$$

where T_t, C_t, S_t and I_t are the 'trend', the 'cyclical', the 'seasonal' and the 'irregular' components respectively (see Nerlove, Grether and Carvalho [70] for an interpretation of these components). The number of four components is not to be seen as a limitation. More (or less) components may be considered too. The multiplicative decomposition (2.2) can be justified by the observation that seasonal or irregular variations often grow with the 'level' T_t of a series.

Methods based on the multiplicative decomposition can be defined explicitly (see section 2.5) or they can be derived from the additive representation (2.1) by using a preliminary log-transform of X_t. Besides the additive and the multiplicative decompositions, some methods allow for additional representations of X_t, see for example section 2.5.

Additive or multiplicative component *models* are defined by supplying specific stochastic assumptions. Model-based approaches generally differ with respect to these assumptions. For the MBA in the following section, components are assumed to be *dependent.*

2.2 The Beveridge-Nelson Decomposition

The Beveridge-Nelson decomposition is a so called 'ARIMA'-model-based-approach. Let

$$X_t = T_t + C_t$$

where it is assumed that

$$X_t = X_{t-1} + \mu + \Xi(B)\epsilon_t \tag{2.3}$$

where $\Xi(B) := \sum_{k=0}^{\infty} \xi_k B^k = \frac{\sum_{k=0}^{p} \alpha_k B^k}{\sum_{j=0}^{q} \beta_j B^j}$ is a stable ARMA operator ($\alpha_0 = \beta_0 = 1$), see Beveridge and Nelson [6]. Consider a forecast $\hat{X}_{t+k}|X_t, X_{t-1}, ...$ of X_{t+k} for k 'large':

$$\begin{aligned}
\hat{X}_{t+k} &= k\mu + X_t + \left(\sum_{j=1}^{k} \xi_j\right)\epsilon_t + \left(\sum_{j=2}^{k+1} \xi_j\right)\epsilon_{t-1} + ... \\
&\simeq k\mu + X_t + \left(\sum_{j=1}^{\infty} \xi_j\right)\epsilon_t + \left(\sum_{j=2}^{\infty} \xi_j\right)\epsilon_{t-1} + ... \\
&=: k\mu + T_t
\end{aligned}$$

where $\left|\sum_{j=1}^{\infty} \xi_k\right| \leq \sum_{j=1}^{\infty} |\xi_k| < \infty$ because of the ARMA-structure (which induces an exponential decay of the coefficients). The slope of the forecast is given by μ and its 'level' is defined by T_t which is a stochastic process. In fact

$$T_t - T_{t-1} = \mu + \left(\sum_{j=0}^{\infty} \xi_j\right)\epsilon_t \tag{2.4}$$

so that T_t is a random walk with drift μ. Beveridge and Nelson call T_t the *permanent component*: "the value the original series would have if it were on the long-run path (as defined by the long run forecast) in the current time period. The permanent component is then the long-run forecast of the series adjusted for its mean rate of change...", see [6], p.156.

Remarks

- The permanent component can be interpreted as a 'trend'. From (2.4) the successive trend increments increase by $\mu + (\sum_{k=0}^{\infty} \xi_k)\, \epsilon_t$. If $|\sum_{k=0}^{\infty} \xi_k| > 1$ then the trend is more 'erratic' than the original series. Figure 2.1 illustrates the latter point : the solid line corresponds to a seasonally adjusted series (UK-car-sales series, see chapter 7) whereas the dotted line corresponds to the Beveridge-Nelson 'trend'. The permanent component T_t is estimated using the software-package 'RATS' (see below). AR- and MA-model orders were set to $p = 0$ and $q = 1$ so that $X_t - X_{t-1}$ is a MA(1) process. The estimated positive lag coefficient θ then implies $\sum_{j=0}^{\infty} \xi_j = 1 + \theta > 1$ in (2.4). Note that this phenomenon ('erratic' trend) has lead to criticism, see for example Metz [66] p.290. In fact, for many applications 'smooth' components are of interest (because it is felt that 'short term' variations should be 'smoothed out').
- As shown in equation 10 in Beveridge and Nelson [6] the 'cyclical' component $C_t := X_t - T_t$ is stationary and its innovation process is given by ϵ_t. Therefore, trend and cyclical components are *dependent* since they share the same innovation ϵ_t : the 'shocks' which generate the business cycle are the same as those which generate the growth process. Beveridge and Nelson interpret C_t as "a stationary process which represents the forecastable momentum present at each time period but which is expected to be dissipated as the series tends to its permanent level", see [6], p.158.
- Finite sample signal extraction problems do not exist here because T_t can be computed without knowledge of 'future' observations $X_{N+1}, X_{N+2}, ...$ as can be seen from (2.4).

An algorithm for computing the Beveridge-Nelson-decomposition has been proposed in Newbold [71]. This algorithm has been implemented in RATS. The corresponding procedure is called 'bndecomp.src'. The text-file can be downloaded from www.estima.com. The time series in figure 2.1 has been computed accordingly. Note that (2.3) does not allow for a seasonal component. Therefore, the input series has been previously seasonally adjusted. The corresponding seasonal adjustment procedure is presented in the following section.

2.3 The Canonical Decomposition

The following model-based approach is based on ARIMA-models too. However, the identifying assumptions for the components are 'at the opposite' of those in the preceding section (recall section 1.3). Indeed, it is assumed that T_t,

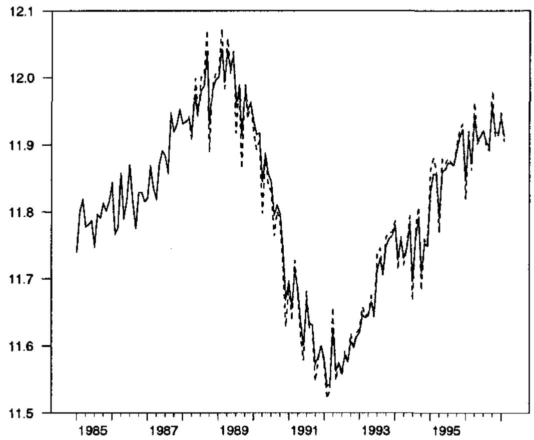

Fig. 2.1. UK-car sales (solid) and Permanent Component (dotted)

S_t and I_t are independent processes and that T_t and S_t are 'smooth'. Hillmer and Tiao [53] argue "To perform seasonal adjustment of the data, an *arbitrary* choice must be made. Considering that the seasonal and trend components should be *slowly* evolving, it seems reasonable to extract as much white noise as possible from the seasonal and trend components... Thus we seek to maximize the innovation variance of the noise component". The "slowly evolving" (smooth) trend and seasonal components or, more precisely, the maximization of the variance of the noise component characterizes the canonical decomposition.

Once an ARIMA-model for the DGP of X_t has been selected and (parameters) estimated, models for the individual DGP's of the components must be defined such that

- the resulting model is *admissible* i.e. the components sum up to X_t and are independent and
- the components may be interpreted as 'trend', 'seasonal' or 'irregular'.

Together with the above 'smoothness' property (see Box, Hillmer and Tiao [37] and Pierce [21]) these assumptions uniquely define the components. A good 'initiation' to the method is given in Maravall and Pierce [65] who consider a very simple ARIMA-process generating trend, seasonal and irregular variations. This is described in the following section.

2.3.1 An Illustrative Example

Let

$$(1 - B^2)X_t = (1 - B)(1 + B)X_t = \epsilon_t \tag{2.5}$$