BWI-paper
D.P. van Beusekom
November 2003

FORECASTING & DETRENDING
OF
TIME SERIES MODELS

Free University of Amsterdam
Faculty of Sciences
Study Bedrijfswiskunde & Informatica
De Boelelaan 1081A
1187 HV Amsterdam

Table of contents

Preface

One of the last subjects of the study Business Mathematics & Computer Science is writing a paper about a subject that is related to the study. The study is a combination of three fields, which are Economics, Mathematics and Computer Science.
The reason I chose this subject has a couple of reasons.
First, I wanted to extend my knowledge of the course Mathematical Systems Theory and apply it to the field of economics.
Secondly, I wanted to find a subject where I could apply the three fields of my study.

The final result is the subject Forecasting & Detrending of time series models.
The subject looked interesting, since it has a link to optimizing business processes, a field I am particularly interested in, and I could learn something from the statistical theories that are used, since it is one of my weakest points throughout my study.

I would like to thank my supervisor Dr. A. Ran from the Free University (Amsterdam, the Netherlands) for his time, advice and critics, which have been a great help writing this paper. I would also like to thank A. Cofino of the TESO boat company for supplying me with data which has been used in the paper.

Chapter 1 Introduction

The development of new techniques and ideas in econometrics has been rapid in recent years and these developments are now being applied to a wide range of areas and markets.

Especially the area of forecasting and control is a hot issue these days since a lot of companies try to optimize their business processes and want to have a good estimate of production planning throughout a large time period. Therefore, better ways of data analysis are being developed to ensure promising forecasting methods.

There are a lot of theories known that can make good trend estimations, however there are still a lot of problems that cannot be fully resolved when it comes to trends. It is often unclear where the trend ends and the fluctuations begin, and the desiderata for separating the two, if possible, have remained in dispute. Secondly, it is still hard to extract the trend, even when it is a clearly defined entity.

The purpose of this paper therefore is to make a review of some methods which are available for obtaining estimates of the trend and of the detrended series. Using an example some techniques will be discussed and compared to each other.

In order to accomplish these ends several topics will be discusses. First, the general meaning of forecasting and trend will be discussed. Secondly, the effects of one of the principal tools of time series modeling, the difference operator, will be discussed. Next, a few model-based methods of trend extraction shall be discussed within the context of ARIMA models and signal extraction. Finally, some enhanced en currently used methods of trend extraction shall be discussed which are independent of any model.

In general the paper will use peaces of the book *System dynamics in economic and financial models* as basis in which parts of the theory will be explained. The figures throughout the paper are programmed using a data sequence of the TESO boat company and a data sequence of the course Mathematical Systems Theory. The programmed parts can be found in Appendix B.

Chapter 2 Time series analysis

2.1 History

Macroeconomic and microeconomic time series often have an upward drift or trend which makes them non-stationary. Since many statistical procedures assume stationarity, it is often necessary to transform data before beginning analysis. There are a number of familiar transformations, including deterministic detrending, stochastic detrending and differencing. In recent years, methods for stochastic detrending have received much attention.
Trying to predict the market has been a hot issue for most companies. For many years forecasts were made using data of previous years. Yet it seemed that even though data of previous years resulted in quite adequate production plans a lot of plans were still not optimal and new workloads appeared that were not calculated in the plans. So trying to forecast the future became even more important and a better analysis on available data had to be done. These are the times that time series analysis became more and more a standard in company forecasting and new methodologies were invented to improve forecasting for companies and eventually try to optimize profits.

Though time series analysis is a broad area of research it is mostly used to optimize planning and consists of two primary goals: identifying the nature of the phenomenon represented by the sequence of observations and forecasting (predicting future values of the time series variables). Both of these goals require that the pattern of observed time series data is identified and more or less formally described. Once that pattern is established, it can be interpreted and integrated into other data. Regardless of the depth of our understanding and the validity of our interpretation of the phenomenon, one can extrapolate the identified pattern to predict future events.

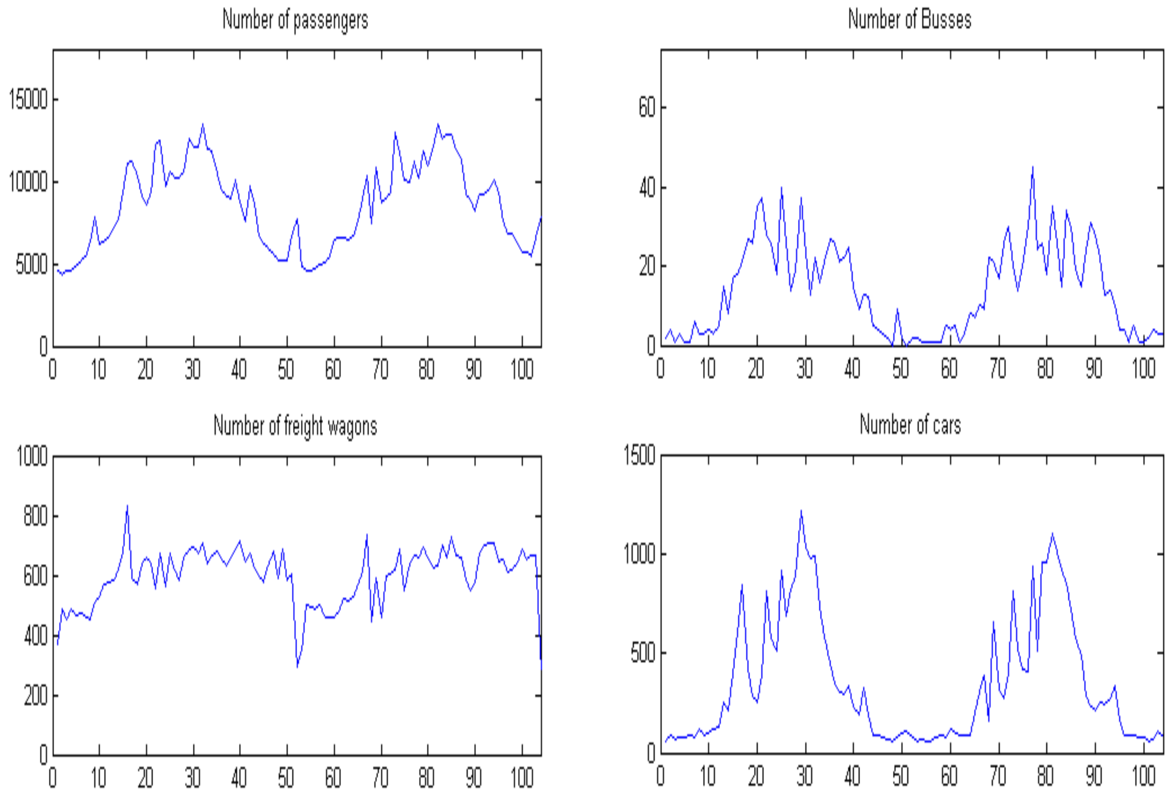
2.2 Enters the trend

As in most other analysis, in time series analysis it is assumed that the data consist of a systematic pattern and random noise which makes a pattern difficult to identify. Therefore, most time series techniques involve some form of filtering out noise in order to make the pattern more salient.

When conducting an analysis of a data sequence one of the first things that can be done is to find out if a trend is present. The first way to check is to look if the data sequence shows a repeating periodical entity which shows growth or decay. Looking at figures 2.1 through 2.4[1] one can see that all variables show a periodical entity. However, in this case nothing can be said about periodical growth or decay since the difference between each

[1] See Appendix B.1 for the programming of the figures

week of two following years may point out growth, but more years are required to make a firm assumption about growth or decay. Accordingly, the dataset of the course *Mathematical Systems Theory* will be used which has a sequence of 128 monthly observations.



Figures 2.1- 2.4 *A series of 104 weekly observations on 4 important variables in forecasting the number of ferry's that are required throughout a year of the TESO boat company*

Most time series patterns can be described in terms of two basic classes of components: trend and seasonality. In economic time series the trend is often the dominant feature. A trend resembles the trajectory of a massive, slow moving body which is barely disturbed by collisions with other, smaller bodies which cross its path.

In the economic time series that will be discussed a trend can be defined as a general systematic linear or nonlinear component, which can contain cycles and must be less volatile than the fluctuations that surround it.

This definition of a trend is flexible enough to allow for a motion which is a fluctuation in one perspective to be regarded as a trend in another. The justification of this assumption rests upon a distinction which shall be drawn between trend estimation and data smoothing. Data smoothing is a justifiable activity even when a meaningful distinction cannot be drawn between the trend and the fluctuations.

2.3 Detrending a trend

We now know basically how a trend can be described. But what is the reason that we want to extract the trend when trying to forecast a time series?
There are a couple of reasons that can be called upon. First, economists are often far more interested in the patterns of fluctuations which are superimposed upon the trends than they are in the trends themselves. In that case it is useful to remove the trend in order to see the patterns more clearly.
Another reason, and one of the most important ones when forecasting a time series, to most criteria of statistical estimation, the object in modeling the trajectory of a variable is to explain the variance as much as possible. If there is a trend present, however smooth and monotone the trend may be, it contributes a large proportion of the explained variance. So if a trend is not removed, the parameters of a model, which is supposed to explain the patterns of the fluctuations, will only be explaining the trend.
Now the only thing left to explain is how to achieve this goal. There have been made numerous methodologies that try, or even succeed to a certain level, in removing the trend of a time series. Amongst others there are the signal extraction method which will be discussed in chapter 5 and the Hodrick-Prescott filter which will be discussed in chapter 6.

Chapter 3 Use of the differencing operator

A big reason for using a stationary data sequence instead of a non-stationary sequence is that non-stationary sequences, usually, are more complex and take more calculations when forecasting is applied to a data series.
One of the methodologies that can be used to make a non-stationary time series stationary is to apply a difference operator to a data series.

3.1 The difference operator

When faced with a time series that shows irregular growth, differencing can be seen as predicting the change that occurs from one period to the next in a time series *Y(t)*. In other words, it may be helpful to look at the first difference of the series, to see if a predictable pattern can be discerned there. For practical purposes, it is just as good to predict the next change as to predict the next level of the series, since the predicted change can always be added to the current level to yield a predicted level.

Within forecasting, backward differencing is normally used. Now, given the data series *Y(t)* we can create the new series:

$$Z(t) = Y(t) - Y(t-1) \tag{3.1}$$

The differenced data will contain one point less than the original data. This imposes a time lag which shall be discussed later. Although one can difference the data more than once, one difference is usually sufficient and recommended. The more times one differences the data the bigger the chance will be that important parts of the data without trend are thrown away, which explain the data so a reasonable forecasting can be made.

3.2 Tintners way of differencing

There are many ways differencing can be applied to data series, but not all ways of differencing are favorable for the end result one is searching for. In general the first objective of differencing is to make the time series stationary, yet some differencing methods remove more then the non-stationary part of a time series and thus remove important information that is vital to the forecasting. One of the first econometricians that worked with differencing was Tintner (1940).
In his vision differencing could make sure that a sequence of ordinates of a polynomial of degree *m*, corresponding to equally spaced values of the argument, can be reduced to a constant by taking m differences. Next to that, if the trend also described a polynomial function the effect would be that, when taking a finite number of differences, a great deal

of the systematic part in the data could be eliminated. This sounds like a good way to remove the trend, and in fact it is, but the total effect will be much larger then actually intended. By taking differences and "throwing the original data away" other information that could be interesting to the economist might be thrown away as well.

What actually happens when applying a difference operator can be seen in figure 3.1[2].

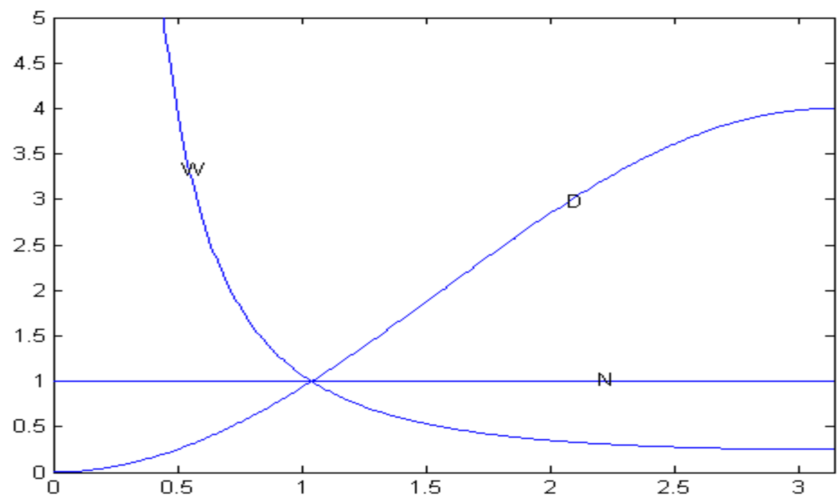


Figure 3.1 *The frequency response function D of the second order difference operator* $(I-L)^2$, *together with the power spectrum W of a first order random walk* $y(t)=(I-L)^{-1}\varepsilon(t)$, *where* $\varepsilon(t)$ *is a white-noise process. The power spectrum of* $\varepsilon(t)$ *is represented by the horizontal line N*

In this figure the curve labeled D represents the frequency-response function of the second-difference operator

$$\delta(L)=(I-L)^2=1+\delta_1 L+\delta_2 L^2 \qquad (3.2)$$

This function indicates the factors by which the operator attenuates or amplifies the amplitudes of the sinusoidal components of a time series to which it is applied.
To explain this, imagine that all stationary stochastic processes, and other processes besides, can be regarded as combinations of an indefinite number of sinusoidal components whose frequencies, denoted by ω, lie in the interval $[0,\pi]$, which is the range of the horizontal axis of the diagram.
Now take the frequency response function of a linear operator or filter $\psi(L)$, which can be defined by:

[2] See Appendix B.2 to see how the figure can be created

$$|\psi(z)| = \sqrt{\psi(z)\psi(z^{-1})}\,, \tag{3.3}$$

where z, a complex exponential whose locus is on the unit circle, can be denoted by $z = \exp\{i\omega\}$. Or, in other words, $|\,\psi(z)\,|$ denotes the modulus of the complex function $\psi(z)$.

When rewriting the second difference operator in the form of equation (3.3) one gets the following equation:

$$\begin{aligned} &\delta(L) = (I - L)^2 \Rightarrow \delta(z) = (1-z)^2 \\ &|\delta(z)| = \sqrt{\delta(z)\delta(z^{-1})} = \sqrt{(1-z)^2(1-z^{-1})^2} = \\ &|\delta(z)| = 2 - (z + z^{-1}) \end{aligned} \tag{3.4}$$

On setting $z = \exp\{i\omega\}$here as well, and using the identity that $1/2\,(\exp\{i\omega\} + \{-i\omega\}) = \cos(\omega)$ equation (3.4) becomes

$$\delta(e^{i\omega}) = 2 - 2\cos(\omega) \tag{3.5}$$

Now, when one looks a bit closer to the frequency response function in figure 3.1 a few more interesting points can be seen. Apparently the (second) difference operator nullifies a time series at zero frequency, which might be called a linear or a quadratic trend.
To see this, first take the first difference operator which can be defined by:

$$((I - L)y)(t) = y(t) - y(t-1) \tag{3.6}$$

Furthermore suppose that

$$y(t) = \gamma t^2 + \alpha t + \beta + \varepsilon(t) \qquad \text{with} \qquad \varepsilon(t) \to N(0,\mu) \tag{3.7}$$

Now if formulae (3.6) and (3.7) are combined one gets:

$$\begin{aligned} y(t) &= \quad \gamma t^2 + \alpha t + \beta + \varepsilon(t) \\ y(t-1) &= \gamma(t-1)^2 + \alpha(t-1) + \beta + \varepsilon(t-1) \\ &\text{-- -} \\ z(t) &= \gamma(2t-1) + \alpha + \varepsilon(t) - \varepsilon(t-1) \end{aligned} \tag{3.8}$$

Applying the second difference, that is, computing $z(t) - z(t-1)$, gives

$$\kappa(t) = \varepsilon(t) - \varepsilon(t-2) + 2\gamma\,, \tag{3.9}$$

which indeed proves the statement above.
Furthermore, on a range of $[0, \pi]$, the difference operator applies a higher and higher frequency to the data series until point π, where the data series is magnified by a factor 4. This already shows that Tintners method of differencing has a few flaws, since it is not desired that the data is scrambled when applying a difference operator.

To illustrate how inappropriate the method of differencing can be, when trying to remove a trend, first a way that works better then the differencing method of Tintner will be described. After that, the two methodologies will be applied to the data series of the course *Mathematical Systems Theory,* so the difference between the two can be seen.

3.3 Fitting a polynomial function to a time series

When trying to remove the non-stationary components from a polynomial data sequence one of the possible methodologies is to fit a polynomial function. In this case the residuals of the polynomial function and the original data sequence can be plotted to obtain a differenced sequence, but preserving important data.
To see this, first a subset of 128 observations is plotted in figure 3.2[3] together with a seventh degree polynomial that has been fitted onto the data sequence. Figure 3.3 shows the residuals after the polynomial time trend of degree 7 has been extracted. Figure 3.4 shows the effect of applying the difference operator to the series.

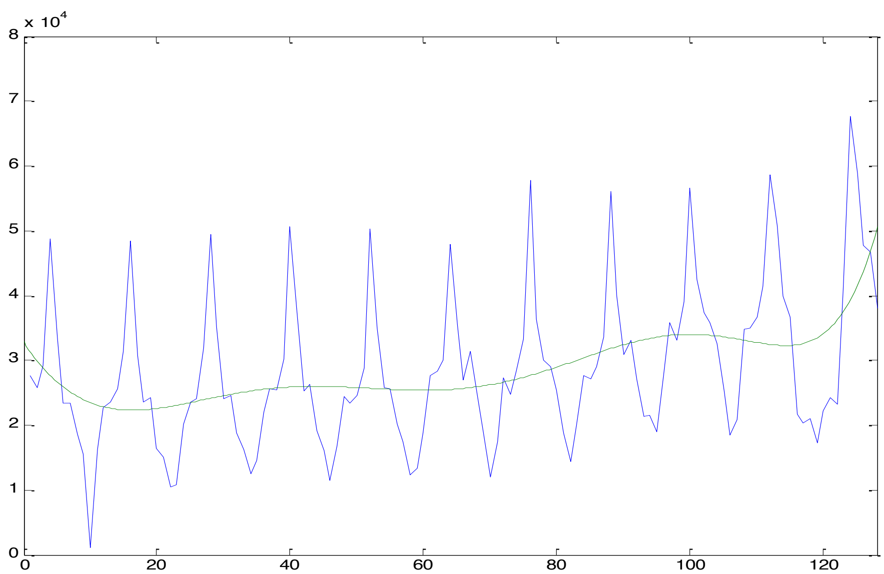


Figure 3.2 A *series of 128 monthly observations on the number of passengers that boarded a ferry*

[3] See Appendix B.3 to see how figures 3.2 through 3.4 are created

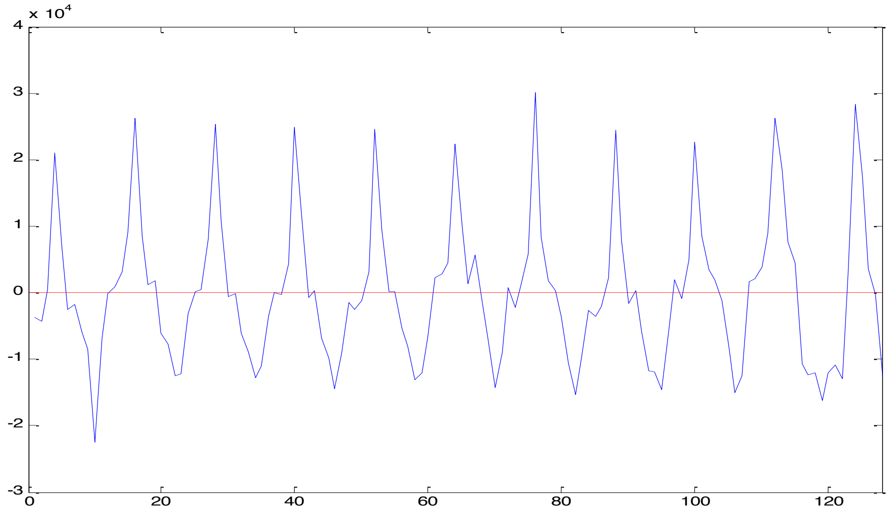


Figure 3.3 The *residuals from fitting a seventh-degree polynomial to the data on boarding passengers*

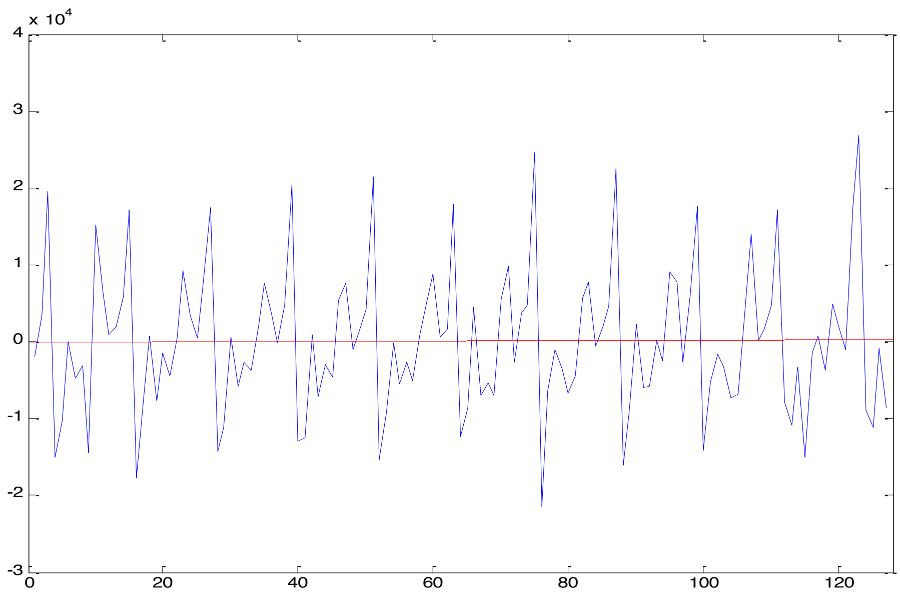


Figure 3.4 A *series generated by applying the difference operator to the data on boarding passengers*

Here it can be seen that when applying the difference operator not only the relative sizes of the peaks are misrepresented, but also another effect is evident. By applying the difference operator a time lag has been induced which scrambles the data even more, while the polynomial fit does not induce the time shift.

3.4 Removing the time lag of differenced data

As mentioned in paragraph 3.3 differencing has, next to scrambling the data, another undesired effect which can easily be avoided. This is the so-called phase effect whereby the transformed series suffers a time delay. This delay simply occurs since differences of two points are taken each time, thus the differenced data will have one data point less then the original series and when two successive differencing operations are applied to a series of weekly observations, a time lag of one week is induced.

To correct this phase effect another time shift will have to be applied to get the series in sync with the original time line again. The desired result can be achieved by simply shifting the affected series forward in time. This can be done without harming the data series, since the difference operator imposes the same time lag on all components regardless of their frequencies.

It is possible that more complicated operators have been used, such as those which are formed from the ratios of lag-operator polynomials, which have different time lags at different frequencies. In this case the phase effect can be removed by applying the operator again, but in reversed time.
To explain this, let $F = L^{-1}$ denote the inverse of the lag operator which can be described as the forward shift operator. If one then takes a sequence *z(t)* the effect of this operator on *z(t)* will be $Fz(t) = z(t+1)$.
Now if one assumes that the operator $\varphi(L) = \delta(L)/\gamma(L)$ has been used upon the data sequence *y(t)* and thus invokes the time shift, then this will be removed by applying the reversed time operator $\varphi(F)$ to the transformed series $z(t) = \varphi(L)y(t)$. Thus *x(t)* will suffer no phase effect if it is formed from *y(t)* in two stages:

$$(i) \quad \gamma(L)z(t) = \delta(L)y(t) \tag{3.10}$$
$$(ii) \quad \gamma(F)x(t) = \delta(F)z(t)$$

Combining these two equations gives the expression for x(t):

$$x(t) = \varphi(F)\varphi(L)y(t) = \frac{\delta(F)\delta(L)}{\gamma(F)\gamma(L)}y(t) \tag{3.11}$$

Here the combined operator $\psi(L) = \varphi(F)\varphi(L)$ may be described as a bidirectional filter.

Chapter 4 ARIMA and the differencing operator

Although differencing isn't always an optimal way of achieving a nice fit, it is still widely used as data transformation because it plays a central role in a lot of methodologies by which ARIMA models are fitted to non-stationary time series.
To explain this, the subjects that will be discussed in this chapter will give a better idea at what ARIMA models actually are and give a review of how ARIMA models and random walk models can be combined. After that the Box & Jenkins method will be described, which uses ARIMA and differencing as basis.

4.1 The ARIMA model

ARIMA models are, in theory, the most general class of models for forecasting a time series which can be stationarized by transformations such as differencing and logging. In fact, the easiest way to think of ARIMA models is as fine-tuned versions of random walk and random-trend models. The fine-tuning consists of adding lags of the differenced series and/or lags of the forecast errors to the prediction equation, as needed to remove any last traces of autocorrelation from the forecast errors.

The acronym ARIMA stands for Auto-Regressive Integrated Moving Average. Lags of the differenced series appearing in the forecasting equation are called auto-regressive terms, lags of the forecast errors are called moving average terms, and a time series which needs to be differenced to be made stationary is said to be an integrated version of a stationary series.

A non-seasonal ARIMA model is classified as an ARIMA(p,d,q) model where:

- **p** is the number of autoregressive terms,
- **d** is the number of non-seasonal differences, and
- **q** is the number of lagged forecast errors in the prediction equation.

To identify the appropriate ARIMA model for a time series, one begins by identifying the order(s) of differencing needed to stationarize the series and remove the gross features of seasonality, perhaps in conjunction with a variance-stabilizing transformation such as logging or deflating. Different approaches and datasets will give different type of ARIMA processes.

4.2 Random walk models and ARIMA

The simplest way to explain the meaning of a random walk model is by first taking the equation

$$\hat{Y}(t) - Y(t-1) = \pi \qquad (4.1)$$

where π is the mean of the first difference, i.e. the average change of one period to the next, which follows from the white noise process $\varepsilon(t) \to N(\pi, \mu)$. If one rearranges this equation to put $\hat{Y}(t)$ by itself on the left, one gets:

$$\hat{Y}(t) = Y(t-1) + \pi \qquad (4.2)$$

In other words, the prediction holds that this period's value will equal last period's value plus a constant representing the average change between periods. This is the so-called "random walk" model. It assumes that, from one period to the next, the original time series merely takes a random "step" away from its last recorded position.
If the constant term is zero one can talk about a random walk without drift.

If the time series being fitted by a random walk model has an average upward (or downward) trend that is expected to continue in the future, one should include a non-zero constant term in the model. In that case it is said that the random walk undergoes drift. Since it includes (only) a non-seasonal difference and a constant term, it is classified as an ARIMA(0,1,0) model with constant term.

When modeling, an advantage of using the ARIMA model option to fit a random walk model is that it easily allows adding terms to correct the model for autocorrelation in the residuals, if this should be necessary. In particular, if the random walk model has significant positive autocorrelation in the residuals at lag 1, one should try and use a so called ARIMA(1,1,0) model, which can be defined as a differenced first-order autoregressive model with equation:

$$\widehat{Y}(t) = \pi + Y(t-1) + \phi(Y(t-1) - Y(t-2)) \qquad (4.3)$$

On the other hand, if the random walk model has significant negative autocorrelation in the residuals at lag 1, one should try and fit a ARIMA(0,1,1) model, which can be defined as a simple exponential smoothing model with growth with equation:

$$\widehat{Y}(t) = \pi + Y(t-1) - \theta e(t-1) \qquad (4.4)$$

4.3 The Box & Jenkins Method

One of the methodologies that use ARIMA processes and differencing as basis is the Box & Jenkins (B&J) method. According to the prescription of Box & Jenkins a few steps have to be taken when building an ARIMA model.

4.3.1 Determining the order of differencing

The first step is to determine if the series is stationary and if there is any significant seasonality that needs to be modelled. To achieve this one has to take as many differences of the original series as are needed to reduce it to stationarity.

Normally, the correct amount of differencing is the lowest order of differencing that yields a time series which fluctuates around a well-defined mean value and whose auto-correlation function plot decays rapidly to zero, either from above or below. If the series still exhibits a log-term trend, or otherwise lacks a tendency to return to its mean value, or if its autocorrelations are positive out to a high number of lags (10 or more), then it needs a higher order of differencing.
Next to this, differencing tends to introduce negative correlation. If the series initially shows strong positive autocorrelation, then a non-seasonal difference will reduce the autocorrelation and perhaps even the drive the lag-1 autocorrelation to a negative value. If a second non-seasonal difference is then applied (which is occasionally necessary), the lag-1 autocorrelation will be driven even further in the negative direction.

Now, if the lag-1 autocorrelation is zero or even negative, then the series does not need further differencing. In this case it is not recommended to difference another time, because this will only result in "overdifferencing" the series and end up adding extra AR or MA terms to undo the damage. If the lag-1 autocorrelation is more negative then –0.5 the series probably has been overdifferenced already.

Another symptom of possible overdifferencing is an increase in the standard deviation, rather than a reduction, when the order of differencing is increased.

Fourthly, a model with no orders of differencing assumes that the original series is stationary. A model with one order of differencing assumes that the original series has a constant average trend or random walk. A model with two orders of total differencing assumes that the original series has a time-varying trend.

A last consideration in determining the order of differencing is the role played by the constant term in the model, if one is included. The constant represents the mean of the series if no differencing is performed, it represents the average trend in the series if one order of differencing is used, and it represents that average trend in the trend if there are two orders of differencing. Normally it is assumed that a trend in the trend does not exist and the constant is removed from the model.

4.3.2 Assuming stationarity after differencing

When the differencing of the series is done successfully it is proposed that the differenced series can be treated in the same way as a stationary series, which has had no need of differencing.

To clarify this step of the B&J prescription assume that many trends in economic time series can be represented by random walk processes of the sort which can be depicted by:

$$y(t) = (I - L)^{-d} v(t), \tag{4.5}$$

wherein *v(t)* represents a stationary stochastic process of the ARMA variety. In simplest instance, *v(t)* is a white-noise process which generates a sequence of independently and identically distributed random variables.

The effect of the difference operator upon *y(t)* is no less drastic in the context of ARIMA modelling than in any other context. However, if the process generating *y(t)* is indeed a kind of random walk, which imposes that the data series *y(t)* is non-stationary, then the use of the difference operator to reduce it to stationarity is undoubtedly called for. Though this is a sound method, the question still arises if random walk models are appropriate analogies for the types of time series models which are likely to be encountered when forecasting a time series.
One way to find out whether random walks are indeed appropriate analogies is to compare the power spectra of economic time series with the spectra which are generated by random walks. If a big difference can be detected between the two and are consistent on a selected range, then doubt may be cast upon the appropriateness of random-walk models.

In figure 3.1, the function labelled W represents the spectral density function, or power[4] spectrum, of a first-order random walk y(t) which is described by:

$$y(t) = (I - L)^{-1} \varepsilon(t), \tag{4.6}$$

wherein $\varepsilon(t)$ represents a white-noise process with a variance of $V\{\varepsilon(t)\} = 2\pi$.
The function indicates the power which is attributable to the sinusoidal components of which the random walk is composed. An evident feature of this spectrum is that there is infinite power at zero-frequency as can be seen in figure 3.1. This phenomenon corresponds to the theoretical condition that the values generated by a random walk defined on an indefinite set of integers in unbounded.

The notion of power is synonymous with the notion of variance and, for a sinusoidal function, the variance is half the square of the amplitude. The white-noise process $\varepsilon(t)$,

[4] See Appendix A.1 for more detailed information about a power spectrum

which is the motive force that drives the random walk, has a uniform distribution of power over the frequency interval $[0,\pi]$. Therefore the power spectrum of *y(t)* is, in effect, the square of the frequency-response function of the operator $(I-L)^{-1}$ scaled by the variance, or power, of the process $\varepsilon(t)$.

4.3.3 Comparing the response function and the random walk model

To further understand the assumption that is stated above, two comparisons can be made to clarify the matter. The first comparison that can be made is between the functions *W* and *D* in figure 3.1. The frequency response operator *D*, which represents the second-difference operator $(I-L)^2$, can now also be seen as the square of the frequency response function of the operator $(I-L)$ which is effective in reducing the first-order random walk y(t) to the white-noise sequence $\varepsilon(t)$. When multiplying curve *W* by curve *D* the horizontal line *N* is created, which represents the power spectrum of the white-noise process $\varepsilon(t)$.
The second comparison that can be made is between the generated power spectrum of a typical economic time series and that of the first-order random walk.

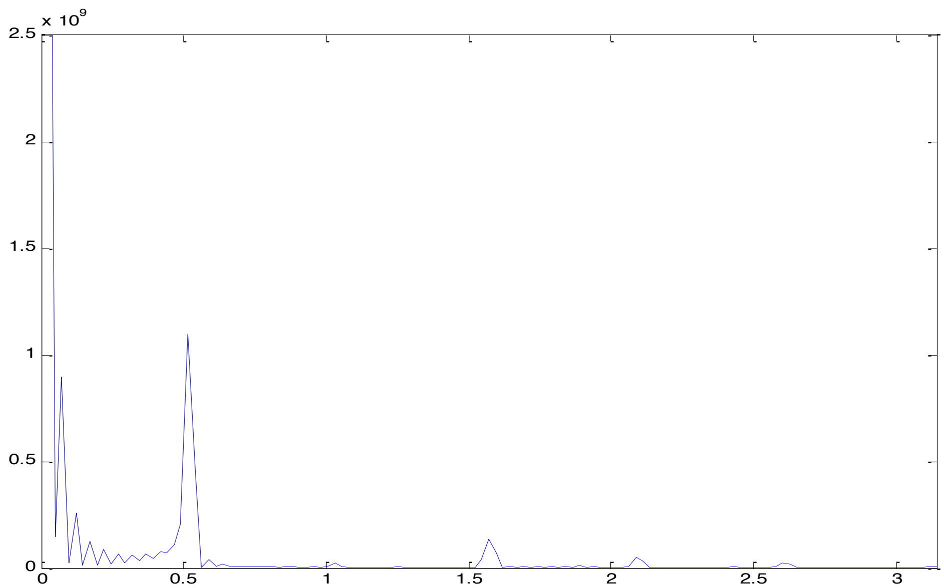


Figure 4.1 *The periodogram of 128 observations on passengers that go with a ferry*

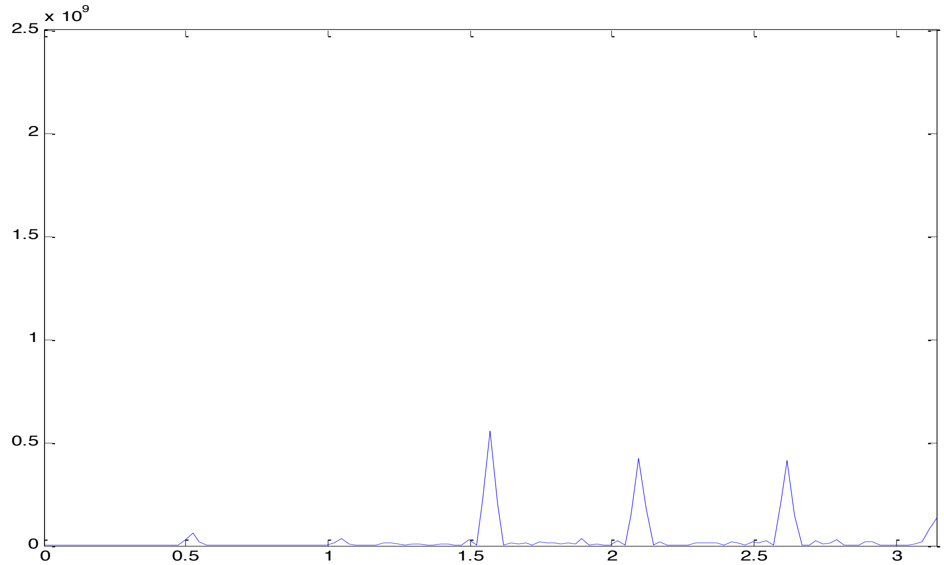


Figure 4.2 *The periodogram of a sequence obtained by applying the second difference operator to the data on passengers that go with a ferry*

The reason that periodograms[5] are used is that the periodograms of empirical time series rarely show anything like the slew of power across the range of frequencies which is characteristic for a simple random walk. As one can see in figure 4.1[6] a trend is present since there is a strange peak at the beginning of the series which is characteristic for a trend. This is exactly the reason why random walk models are said to be inappropriate models for non-stationary economic time series. Figure 4.2 shows that when applying the difference operator the trend has indeed been removed, but also the peak that was present at $\omega = 1/6\pi$ has almost vanished. But why is it then that random walk models feature so largely in econometric time series modelling?

To explain this one needs to take a closer look at the ARMA part of an ARIMA model. The ARMA part of an ARIMA model provides a model for the *v(t)* process in equation (4.5). Now, define $v(t) = (I - \theta L)^d \varepsilon(t)$, where $\varepsilon(t)$ is white noise and θ is close to unity, so equation (4.5) can be rewritten into:

$$(I - L)^d y(t) = (I - \theta L)^d \varepsilon(t) \qquad (4.7)$$

Here one can see that the effect of the operator $(I - L)^d$, which induces the random walk, will be counteracted by that of the operator $(I - \theta L)^d$. The effect of this counter action will take the power of the process to the neighbourhood of $\omega = 0$.

[5] See Appendix A.2 for more detailed information about periodograms
[6] See Appendix B.4 to see how figures 4.1 and 4.2 are created

Seeing that this works one also needs to change the formula that removes the trend. Looking back at equation (3.10), one can add the parameter θ here as well, which changes the difference operator into:

$$\delta(L) = \left\{ \frac{(1+\theta)^2}{4} \right\}^d \frac{(I-F)^d (I-L)^d}{(I-\theta F)^d (I-\theta L)^d} \tag{4.8}$$

This new defined operator has a few new features, which can be quite useful. For instance, the parameter θ now serves to limit the effects of the difference operator. Next to that, the filter has a bidirectional filter which eliminates the phase effect.
There has been added a factor on the right side of equation (4.8) to ensure that the frequency response function attains the value of unity when $\omega = \pi$, which is when $z = -1$. To actually see the effect of this new operator figure 4.3[7] shows the frequency response function for the operator in the case where $d = 2$ and for various values of θ. The reason for taking $d = 2$ is that equation (4.8) is very similar to the Hodrick-Prescott detrending filter which shall be examined later. Two things can be learned from figure 4.3. Even though the major effects of the modified filter are confined to the lower reaches of the frequency range, there is still a gradual transition between the effect of nullifying a frequency component, as happens at $\omega = 0$, and that of preserving it, as happens at $\omega = \pi$

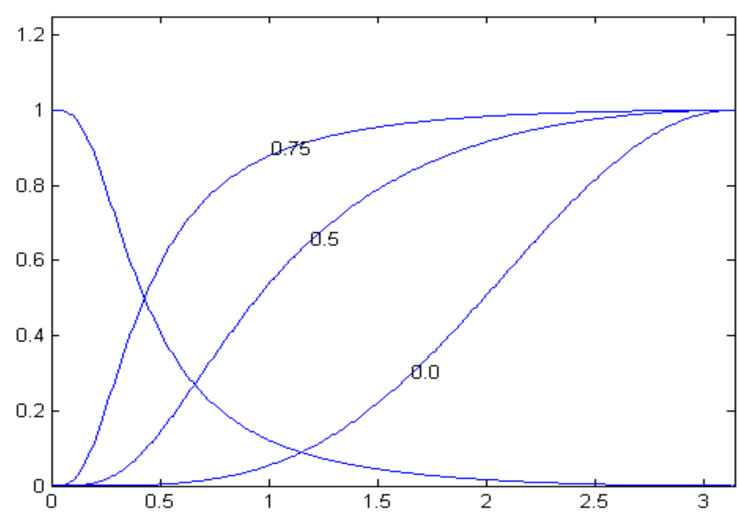


Figure 4.3 *The frequency response function of the modified difference operator of equation (4.8) when d = 2 for the values of* $\theta = 0.0$, $\theta = 0.5$, $\theta = 0.75$, *together with the complementary trend estimation filter for the case of* $\theta = 0.75$

[7] See Appendix B.5 to see how figure 4.3 is created

These two things can be a serious problem when examining time series models. When a periodogram shows a sharp distinction between those components which belong to the trend and those which belong to the fluctuations some problems may be encountered.
It is arguable that such a distinction can be found in the periodogram of figure 4.3 where the power is virtually zero in the vicinity of $\omega = \pi / 8$.

This is the point at which we should propose separating the trend from the fluctuations. The frequency of $\omega = \pi / 6$, which corresponds to the fundamental frequency, lies beyond the range of the frequencies which belongs to the trend.

The next few chapters will be devoted to discovering the appropriate means of separating the trend and the fluctuations when a clear distinction is evident in the periodogram of the data.

Chapter 5 Trend estimation by signal extraction

A trend has only a tenuous existence within the context of an ARIMA model. A trend represents nothing more than the accumulation of fluctuations, which are created by applying a filter to a white noise sequence $\varepsilon(t)$ of independently and identically distributed random variables. If the trend and the fluctuations are due to the same motive force, which is the white noise process, then it is meaningless to draw a distinction between them.

If the trend and fluctuations of a data series indeed are due to separate sources a method will have to be used to define them as such. An obvious resource is to attribute a separate ARIMA model to each of them. In that case, the model which is generated by adding together the component ARIMA models is an ARIMA model itself.
Given that the ARIMA models for the structural components of a time series combine so seamlessly to form a reduced-form ARIMA model, it might seem like wasted effort to attempt to separate the combined series into its constituent parts. Thus in the absence of any seams to show where the joints are to be found and where the cut is to be made, it is to be expected that whatever separation is achieved is liable to be a doubtful one.

Though this has been a good argument not to separate the trend and fluctuations, econometricians have not been deterred by these difficulties and they have developed a sophisticated methodology for separating an ARIMA model into its putative components. The methodology, which can separate the trend from the fluctuations, is based on the signal extraction technique.

5.1 Signal extraction basics

To clarify this matter, one may begin by considering a general model of the processes which have generated the data. Assume that the trend, or signal sequence, $\xi(t)$ is generated by a non-stationary ARIMA process and that the residual component, or noise process $\eta(t)$, which is its complement, is generated by an ordinary stationary autoregressive process. Furthermore assume that both processes are unobservable[8] series which are assumed to be statistically independent.

Thus the data sequence $y(t)$, which is an observable time series and a function mapping from the set of integers $\Gamma = \{t = 0, \pm1, \pm2, ...\}$ onto the real line, may be represented by

$$y(t) = \xi(t) + \eta(t) \qquad t = 0, \pm1, \pm2, ... \qquad (5.1)$$

[8] See Appendix A.3 for more detailed information about unobservable series

The object is to infer the trajectory of the trend from the observations on $y(t)$ given a knowledge of the covariance structures of $\xi(t)$ and $\eta(t)$. To achieve this one can use the signal extraction method in which one needs to find the best estimate, or minimum mean squared error, of the signal $\xi(t)$, for any fixed t given the observed data and taking $\eta(t)$ as (white) noise. However, the analogy may be misleading since, in the case of trend extraction, the residue contains information which is of primary interest whereas, in the usual case of signal extraction, the noise sequence is liable to be regarded as a nuisance which is to be discarded.

5.1.1 A general lag operator

If an estimate of $\xi(t)$ has to be defined, one way to obtain this by filtering the data series $y(t)$, which can be denoted by

$$x(t) = \psi(L)y(t) \tag{5.2}$$

Here $\psi(L) = \sum_j \psi_j L^j$ is a function of the lag operator L which may be a polynomial or a rational function. A single element of the sequence $x(t)$ at time t can be denoted by $x_t = \sum_j \psi_j y_{t-j}$. When both succeeding and preceding values of $y(t)$ are available for the purpose of estimating the current value of $\xi(t)$, then $\psi(L)$ is liable to be a two-sided function containing both positive and negative powers L. This fact is true since t is chosen between time points 0 and a last observed point T.

The coefficients of the filter $\psi(L)$ are estimated by invoking the minimum mean square error criterion. The errors in question are the elements of the sequence $e(t) = \xi(t) - x(t)$. The principle of orthogonality, by which the criterion is fulfilled, indicates that the errors must be uncorrelated with the elements in the information set $Y_t = \{y_{t-k}; k = 0, \pm 1, \pm 2, ...\}$. Thus

$$\begin{aligned} 0 &= E\{y_{t-k}(\xi_t - x_t)\} \\ &= E(y_{t-k}\xi_t) - \sum_j \psi_j E(y_{t-k}y_{t-j}) \\ &= \gamma_k^{y\xi} - \sum_j \psi_j \gamma_{k-j}^{yy} \end{aligned} \tag{5.3}$$

for all k = 0. The equation may be expressed, in terms of the z-transform, as

$$\gamma^{y\xi}(z) = \psi(z)\gamma^{yy}(z) \tag{5.4}$$

where

$$\gamma^{yy}(z) = \gamma^{\xi\xi}(z) + \gamma^{\eta\eta}(z) \qquad \text{and} \qquad \gamma^{y\xi}(z) = \gamma^{\xi\xi}(z) \tag{5.5}$$

are respectively the auto covariance generating function of $y(t)$ and the cross covariance generating function of $y(t)$ and $\xi(t)$. It follows from equations (5.4) and (5.5) that

$$\psi(z) = \frac{\gamma^{y\xi}(z)}{\gamma^{yy}(z)} = \frac{\gamma^{\xi\xi}(z)}{\gamma^{\xi\xi}(z) + \gamma^{\eta\eta}(z)} \tag{5.6}$$

Now by setting $z = \exp\{i\omega\}$, one can derive the frequency response function of the filter, which is used in estimating signal $\xi(t)$. The effect of the filter is to multiply each of the frequency components of $y(t)$ by the fraction of its variance which is attributable to the signal. The same principle applies to the estimation of the residual detrended component. The trend elimination filter is just the complementary filter

$$1 - \psi(z) = \frac{\gamma^{\eta\eta}(z)}{\gamma^{\xi\xi}(z) + \gamma^{\eta\eta}(z)} \tag{5.7}$$

It has been shown by Cleveland and Tiao (1976) and, more recently, by Bell (1984) that these formulae apply equally to stationary ARMA processes and to non-stationary ARIMA processes using the Wiener Kolmogorov theory. This however, shall not be discussed in this paper.

5.2 Model-based methods of trend estimation

It has been argued that the best way of separating the trend from the fluctuations is to model both of them at the same time within the framework provided by a structural ARIMA model which assigns separate parameters to the components. There have been a lot of examples of detrending which follow this prescription.
Two of them that will be discussed are the method described by Hillmer and Tiao, which are concerned with extracting the hidden components from existing seasonal ARIMA models, and a methodology which is concerned with filtering short sequences using a rational operator.

5.2.1 Hillmer and Tiao

The model in question of Hillmer and Tiao is specified by the equation

$$(I - L^s)y(t) = (I - \theta L^s)\varepsilon(t) \tag{5.8}$$

where s is the number of observations within the span of an annual cycle.
When taking $\theta = 0$, $y(t)$ will be denoted by the same period of the previous year plus the amount of a white noise disturbance. The accumulation of such disturbances over the years will cause the seasonal value to follow a first-order random walk.
The elaboration of the model which introduces a non-zero value of θ has the same effect in limiting the random walk as it does in the analogous non-seasonal equation under (4.7). In the context of the seasonal model, it increases the concentration of the power of the process in the vicinities of the seasonal frequencies $\omega = 2\pi j / s; j = 1, \ldots.., s/2$, thereby reducing the drift and regularising the cycles.
The seasonal patterns which are typically exhibited by economic time series show a degree of persistence which is not found in the stochastic output generated by equation (5.8). There is no bound in the long run on the amplitude of the cycles generated by this equation. Also, there is a tendency for the phases of the cycles to drift without limit.

The virtue of equation (5.8) is not a model of the processes generating the seasonal time series but as a device for forecasting the series. The forecasting rule which is implied by the equation when $\theta = 0$ is that the most recent set of s observations, which represent an annual cycle, should be taken as the pattern for all future years. In the case where $\theta \neq 0$, the pattern of the annual cycle, which is to be extrapolated, should be formed from a weighted average of all previous cycles, with θ as a discount factor which is applied repeatedly to the cycles as the years recede.
The autocovariance generating function of the model may be factorized into two partial fractions whose denominators contain, respectively, the trend operator *1 – z* and the seasonal operator $S(z) = 1 + z + z^2 + \ldots. + z^{s-1}$:

$$\begin{aligned}\gamma^{yy}(z) &= \frac{(I - \theta z^s)(1 - \theta z^{-s})}{(1 - z^s)(1 - z^{-s})} \\ &= \frac{A(z)}{(1-z)(1-z^{-1})} + \frac{B(z)}{S(z)S(z^{-1})}\end{aligned} \tag{5.9}$$

To solve this equation, multiply the equation by the factor $(1-z)(1-z^{-1})$ and take *z = 1*: This will remove the *B(z)* component and leave the equation:

$$A(1) = \lim_{z \to 1} \frac{(1-z)}{(1-z^z)}(1 - \theta z^s)(1 - \theta z^{-s})\frac{(1-z^{-1})}{(1-z^{-s})} \tag{5.10}$$

Now using the identity that

$$\lim_{z \to 1} \frac{1-z}{1-z^s} = \frac{1}{s} \tag{5.11}$$

equation (5.10) can be solved, and *A(z)* can be defined as

$$A(z) = \frac{1}{s^2}(1-\theta)^2 \tag{5.12}$$

Since this holds true, the partial fraction that is associated with the trend can be defined by:

$$\varphi(z) = \frac{(1-\theta)^2}{s^2(1-z)(1-z^{-1})} \tag{5.13}$$

This is the autocovariance generating function of a random walk process. Setting $z = \exp\{i\omega\}$ in the function generates the pseudo spectral density function of the process. The function $\varphi(z)$ attains a minimum value of $(1-\theta)^2/(4s^2)$ at the Nyquist frequency of $\omega = \pi$.
Now consider a horizontal line drawn at this height over the interval $[0, \pi]$. The line represents the spectral density function of a white-noise component of variance $(1-\theta)^2/(4s^2)$ which is an integral part of the trend component as it is currently defined. According to the principle of canonical factorization this white noise component should be subtracted from the trend component and attributed to the irregular component which is currently part of the residue. The subtraction of the white noise component leads to a revised trend component of the form:

$$\begin{aligned} \gamma^{\xi\xi}(z) &= \frac{(1-\theta)^2}{s^2(1-z)(1-z^{-1})} - \frac{(1-\theta)^2}{4s^2} \\ &= \frac{(1-\theta)^2(1+z)(1+z^{-1})}{4s^2(1-z)(1-z^{-1})} \end{aligned} \tag{5.14}$$

It follows that the filter which is appropriate to extracting the trend takes the form

$$\psi(z) = \frac{\gamma^{\xi\xi}(z)}{\gamma^{yy}(z)} = \frac{(1-\theta)^2}{4s^2} \frac{(1+z)S(z)S(z^{-1})(1+z^{-1})}{(1-\theta z^s)(1-\theta z^{-s})} \tag{5.15}$$

The interesting thing about this formula is, when taking *z = 1*, the outcome of the equation is $\psi(1)=1$. This exactly the case when the filter should preserve the component of $y(t)$ at zero frequency. This is what should be expected of a filter designed to estimate the trend component.

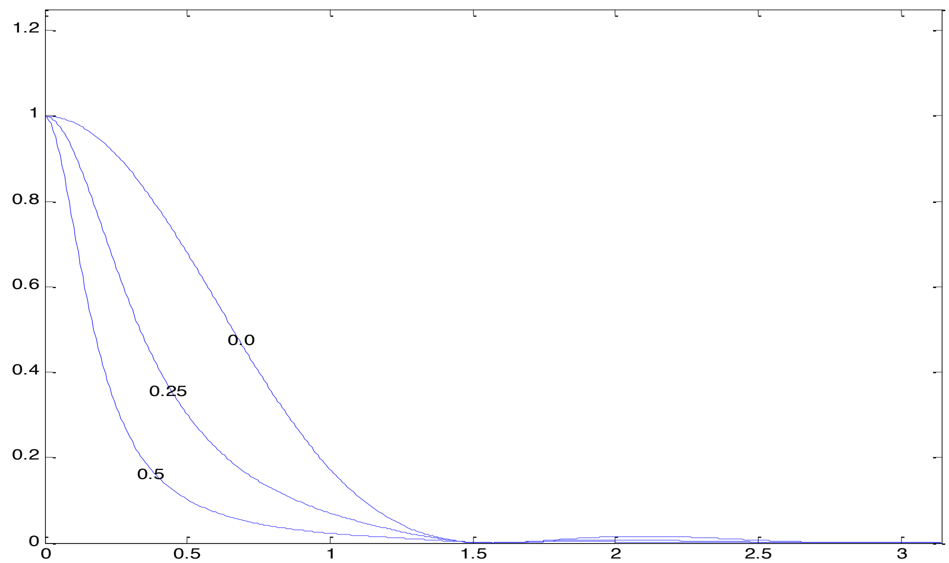


Figure 5.1 *The frequency response function of the model based trend estimation filter of equation (5.15) when s = 4 for various values of* θ

When looking at figure 5.1 it can be seen that the filter does not make a firm distinction between the components which belong to the trend and those which belong to the fluctuations. Its characteristics are attuned to those of an ARIMA model in which the frequency ranges of the two sets of components are bound to overlap substantially. Another remark that can be made is that the frequency response of the model based filter is barely distinguishable from that of the filter which is complementary to the modified differencing filter of equation (4.8). The latter could well be used in place of the model-based filter.

The recommendation that trend estimation should be conducted only within the context of a structural ARIMA model is now in doubt. If there is a manifest distinction between the frequency domains of the trend and the fluctuations, then sharper tools are needed for separating the two.

5.2.2 A rational lag operator

Now consider using a rational lag operator, so the sequence $y(t)$ can be defined by:

$$\begin{aligned} y(t) &= \xi(t) + \eta(t) \\ &= \frac{\psi_T(L)}{(I-L)^d}\nu(t) + \psi_R(L)\varepsilon(t), \end{aligned} \tag{5.16}$$

where $\nu(t)$ and $\varepsilon(t)$ are statistically independent sequences generated by normal white noise processes with $V\{\nu(t)\} = \sigma_\nu^2$ and $V\{\varepsilon(t)\} = \sigma_\varepsilon^2$ respectively.
Here $\psi_T(L)$ and $\psi_R(L)$ are proper rational functions of the lag operator L, which have all of their poles and zeros lying outside the unit circle. The operator $(I-L)^{-d}$, which is apt to be described as a summation or integration operator, is wholly responsible for the non-stationary character of $\xi(t)$.
Equation (5.16) may be multiplied throughout by the product of the denominators to give

$$g(t) = \delta_T(L)\nu(t) + \delta_R(L)\varepsilon(t) = \varsigma(t) + \kappa(t) \tag{5.17}$$

where $\delta_T(L)$ and $\delta_R(L)$ are polynomials of finite degree. From now on assume that both $\psi_T(L)$ and $\psi_R(L)$ are finite degree polynomials too. This assumption entails only a small loss of generality, but serves to simplify the exposition.
It follows that

$$g(t) = (I-L)^d y(t) \tag{5.18}$$

In the more general case, where $\psi_T(L)$ and $\psi_R(L)$ are rational functions, there would be an additional factor in the mapping from $y(t)$ to $g(t)$ compounded from the denominators of $\psi_T(L)$ and $\psi_R(L)$.
Now, if an estimate of the sequence $\varsigma(t)$ is to be made, the process needs to be filtered to remove the phase shift which has been induced. This gives

$$z(t) = \frac{\delta_T(F)\delta_T(L)}{\gamma(F)\gamma(L)} g(t) \tag{5.19}$$

where

$$\gamma(F)\gamma(L) = \delta_T(F)\delta_T(L) + \lambda\delta_R(F)\delta_R(L) \tag{5.20}$$

with

$$\lambda = \frac{\sigma_\varepsilon^2}{\sigma_\nu^2} \tag{5.21}$$

An estimate k(t) of the sequence $\kappa(t)$ can be obtained from $g(t)$ via similar operations which are summarised by the equation

$$k(t) = \lambda \frac{\delta_R(F)\delta_R(L)}{\gamma(F)\gamma(L)} g(t) \tag{5.22}$$

The filter of equation (5.19) will generate the minimum mean-square-error estimate of the stationary sequence $\varsigma(t)$ provided that the smoothing parameter has the value of $\lambda = \sigma_\varepsilon^2 / \sigma_\nu^2$. Since the Wiener Kolmogorov theory applies equally to non-stationary processes the filter of equation X also generates the mean-square-error of $\xi(t)$.
Though this approach seems solid it is beset by two problems. On the one hand, there is the difficulty posed by the initial conditions. On the other hand, there is the danger that the unbounded nature of the data sequence $y(t)$ and the disparity of the values within it will lead to problems of numerical representation. Therefore, in pursuit of an alternative approach, consider the equation

$$\begin{aligned} \xi(t) &= y(t) - \eta(t) \\ &= y(t) - \frac{\kappa(t)}{(I-L)^d} \end{aligned} \tag{5.23}$$

The equation suggests that one may begin by estimating the stationary sequence $\kappa(t)$ by applying the filter of 5.22 to $g(t)$ which is the differenced version of the data sequence. Thereafter, an estimate of the stationary sequence $\eta(t)$ can be obtained by a d-fold process of accumulation. Finally, the estimate of $\xi(t)$ can be obtained by a simple subtraction.

5.3 Filtering short data sequences

Though the signal extraction methodology, which was described in the previous paragraph, wasn't flawless it has been used in combination with the HP filter to obtain a new methodology which filters short data sequences.
To explain this, imagine that instead of long data sequences, there are only T observations of the process $y(t)$ of equation (5.18) which run from $t = 0$ to $t = T - 1$.

These are gathered in a vector

$$y = \xi + \eta = x + h \tag{5.24}$$

where ξ is the trend vector and η is the residual vector which is generated by a stationary process with

$$E(\eta) = 0 \qquad \text{and} \qquad D(\eta) = \sigma_\varepsilon^2 \Sigma \tag{5.25}$$

The estimates of these vectors are denoted by x and h respectively. To find the finite sample the counterpart of equation (5.19), the *d*-th difference operator of $(1-L)^d = 1 + \delta_1 L + ... + \delta_d L^d$ needs to defined in the form of a matrix. Therefore, let the identity matrix of order T be denoted by

$$I_T = [e_0, e_1, ..., e_{T-1}], \tag{5.26}$$

where e_j represents a column vector in the position j and with zeros elsewhere. Then the finite sample lag operator is the matrix

$$L_T = [e_1, ..., e_{T-1}, 0], \tag{5.27}$$

which has units on the first sub diagonal and zeros elsewhere. This matrix is formed by deleting the leading vector of the identity matrix and by appending a zero vector to the end of the array. The lag operator polynomials can then be converted to matrix operators of order T simply by replacing the lag operator L by the matrix L_T. Thus, the matrix, which takes the *d*-th difference of a vector of order T, is given by $\Delta = (I - L_T)^d$. Taking differences within a vector entails a loss of information. Thus, if $\Delta = [Q_*, Q]'$, where Q_*' has *d* rows, then the *d*-th differences of the vector $y = [y_0, ..., y_{T-1}]'$ are the elements of the vector $g = [g_d, ..., g_{T-1}]'$ which is found in the equation

$$\begin{bmatrix} g_* \\ g \end{bmatrix} = \begin{bmatrix} Q_*' \\ Q' \end{bmatrix} y. \tag{5.28}$$

The vector $g_* = \Delta_*' y_*$ in this equation, which is a transform of the vector $y_* = [y_0, .., y_{d-1}]'$ of the initial elements, is liable to be discarded. The matrix of the transformations is the operator $\Delta_* = (I - L_T)^d$.
Premultiplying equation (5.24) by Q' gives

$$g = Q' y = \varsigma + \kappa = z + k,$$

where $\varsigma = Q'\xi$ and $\kappa = Q'\eta$ and where $z = Q'x$ and $k = Q'h$ are the corresponding estimates. The first and second moments of the vector ς may be denoted by

$$E(\varsigma) = 0 \qquad \text{and} \qquad D(\varsigma) = \sigma_\nu^2 \Omega_T, \tag{5.29}$$

and those of κ by

$$\begin{aligned} &E(\kappa) = 0 \qquad \text{and} \\ &D(\kappa) = Q' D(\eta) Q = \sigma_\varepsilon^2 Q' \Sigma Q = \sigma_\varepsilon^2 \Omega_R, \end{aligned} \tag{5.30}$$

where both Ω_T and Ω_R are symmetric Toeplitz matrices with a limited number of nonzero diagonal bands. The generating functions for the coefficients of these matrices are, respectively, $\delta_T(z)\delta_T(z^{-1})$ and $\delta_R(z)\delta_R(z^{-1})$, where δ_T and δ_R are defined by equation (5.20).
The optimal predictor z of the vector $\varsigma = Q'\xi$ is given by the following conditional expectation:

$$\begin{aligned} E(\varsigma \mid g) &= E(\varsigma) + C(\varsigma, g) D^{-1}(g)\{g - E(g)\} \\ &= \Omega_T(\Omega_T + \lambda\Omega_R)^{-1} g = z \end{aligned} \tag{5.31}$$

The optimal predictor k of $\kappa = Q'\eta$ is given, likewise, by

$$\begin{aligned} E(\kappa \mid g) &= E(\kappa) + C(\kappa, g) D^{-1}(g)\{g - E(g)\} \\ &= \lambda\Omega_R(\Omega_T + \lambda\Omega_R)^{-1} g = k \end{aligned} \tag{5.32}$$

It may be confirmed that $z + k = g$.
The estimates are calculated, first, by solving the equation

$$(\Omega_T + \lambda\Omega_R) b = g \tag{5.33}$$

for the value of b and, thereafter, by finding

$$z = \Omega_T b \qquad \text{and} \qquad k = \lambda\Omega_R b. \tag{5.34}$$

The solution of equation (5.34) is found via a Cholesky factorisation which sets $\Omega_T + \lambda\Omega_R = GG'$, where G is a lower triangular matrix. The system $GG'b = g$ may be cast in the form of $Gp = g$ and solved for p. Then $G'b = p$ can be solved for b.

Observe that the generating function for the matrix $GG^{'}$ is the polynomial $\gamma(z)\gamma(z^{-1})$ defined in (5.19). The next step in this process would be to recover the trended sequence. This can be done by recovering from *z* or from *k* an estimate *x* of the trend vector ξ.
To obtain the estimate *x* first recover *h*, which is the estimate of η, from $k = Q^{'}h$, which is the estimate of $\kappa = Q^{'}\eta$. After that *x* can be easily found by the simple subtraction $x = y - h$. This, however, shall not be discussed in this paper.

Chapter 6 Trend estimation by heuristic methods.

The Wiener-Kolmogorov theory of signal extraction depends upon the availability of a statistical model which can represent the processes generating the data. The task of specifying the model can be approached in various ways.
Econometricians have often favoured a structural approach. This has two objectives. First, it is intended that the output of the model should mimic the data series as closely as possible. Secondly, it is proposed that the model should contain as many separate elements as there are discernible components in the data. Motivating this approach is the notion that the quality of the signal extraction filter will be a function of the degree of realism in the underlying model.
A heuristic approach is one in which the model is determined solely with a view to ensuring that the resulting signal extraction filter has certain preconceived properties. A common objective is to derive a low pass filter with a designated cut-off frequency for which the transition from the pass band to the stop band is as rapid as possible, given the constraints of the filter order and the need to maintain numerical stability.
To illustrate this heuristic approach the Hodrick-Prescott (HP) filter is introduced.

6.1 The Hodrick-Prescott and Reinsch filter

The HP filter has a lot of similarities with the Reinsch smoothing spline (see below), which is used a lot in industrial surroundings and can be derived from the following model:

$$\begin{aligned} y(t) &= \xi(t) + \eta(t) \\ &= \frac{1}{(1-L)^2} v(t) + \eta(t) \end{aligned} \tag{6.1}$$

The common goal of both filters is to pursue a criterion of curve fitting which balances the conflicting objectives of smoothness and goodness of fit, which can be seen as the filter and the smoothing spline. A parameter is introduced, which is used to regulate the trade-off between the two and thus can point out if the balance between filter and spline is good enough.
The real difference between the HP filter and Reinsch smoothing spline follows from the use of the parameter. If the parameter is chosen correctly, the HP filter can be seen as an optimal predictor of the trajectory of a discrete-time second order random walk observed with error.
The Reinsch spline, on the other hand, represents the optimal predictor of the trajectory of an integrated Wiener process of which the periodic observations are obscured by white-noise errors. This result has been used as basis in deriving an algorithm for fitting the spline.

6.1.1 The Reinsch smoothing spline

The equation (6.1) represents nothing more then a second order random walk $\xi(t)$ which is obscured by disturbances or errors of observation which form a white noise sequence $\eta(t)$.
Now to take a look at how close the HP filter resembles the Reinsch smoothing spline take an integrated Wiener process which can be defined as a discrete time integrated moving average IMA(2,1) process described by

$$(I-L)^2\xi(t)=(1+\mu L)v(t) \tag{6.2}$$

Here $\mu=2-\sqrt{3}$ and $V\{v(t)\}=\sigma_v^2$ are solutions of the equations

$$\frac{2}{3}\kappa=\sigma_v^2(1+\mu^2), \qquad \text{and} \qquad \frac{\kappa}{6}=\sigma_v^2\mu^2 \tag{6.3}$$

where κ is a scale parameter which affects only σ_v^2.
The autocovariance generating function of the IMA(2,1) signal process $\xi(t)$ is

$$\gamma^{\xi\xi}(z)=\sigma_v^2\frac{(1+\mu z)(1+\mu z^{-1})}{(1-z)^2(1-z^{-1})^2} \tag{6.4}$$

whilst that of the observable noise corrupted process $y(t)=\xi(t)+\eta(t)$ is

$$\gamma^{yy}(z)=\frac{\sigma_v^2(1+\mu z)(1+\mu z^{-1})}{(1-z)^2(1-z^{-1})^2}+\sigma_\eta^2 \tag{6.5}$$

Applying this tot the signal extraction equation (5.15) gives

$$\psi(z)=\frac{\gamma^{\xi\xi}(z)}{\gamma^{yy}(z)}=\frac{\sigma_v^2(1+\mu z)(1+\mu z^{-1})}{\sigma_\eta^2(1-z)^2(1-z^{-1})^2+\sigma_v^2(1+\mu z)(1+\mu z^{-1})} \tag{6.6}$$

whereas the formula for the complementary "detrending" filter is

$$1-\psi(z)=\frac{\sigma_\eta^2(1-z)(1-z^{-1})}{\sigma_\eta^2(1-z)^2(1-z^{-1})^2+\sigma_v^2(1+\mu z)(1+\mu z^{-1})} \tag{6.7}$$

The smoothing parameter is the ratio $\lambda=\sigma_\eta^2/\sigma_v^2$ of the variance of the error process $\eta(t)$ which obscures the observations and the variance of the process $v(t)$ which is the

motive power of the signal. Note that the ratio is the same one that has been used already in equation (5.19).
Since equation (6.6) is optimal for the IMA(2,1) process, it is also optimal for extracting the signal $\xi(t) = (1+\mu L)v(t)$ from the sequence

$$y(t) = (1+\mu L)v(t) + (I-L)^2 \eta(t), \qquad (6.8)$$

where the equation has been multiplied throughout by $(I-L)^2$.
One can now obtain the HP filter by setting the smoothing parameter $\mu = 0$. Figure 6.1 shows the frequency response function of the Reinsch smoothing spline, which in general, does not differ that much from the HP filters frequency response function.

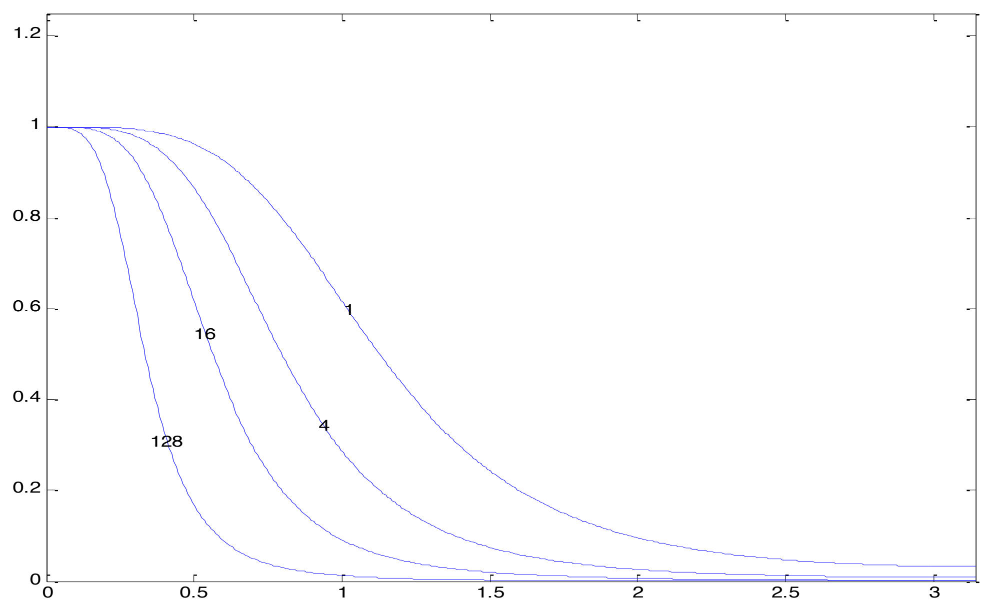


Figure 6.1 *The frequency response function of the Reinsch smoothing filter for various values of the smoothing parameter*

One of the advantages of the Reinsch filter is that it has a sharper transition than that of the filter of equation (5.8). Yet when equation (5.8) is multiplied by a factor $(I-L)$, which gives

$$(I-L)(I-L^s)y(t) = (I-\theta L^s)\varepsilon(t) \qquad (6.9)$$

the trend component of the model follows a second order random walk, which makes the differences in transition of the Reinsch filter and that of equation (6.9) a lot smaller. Though the Reinsch filter attenuates noisy and distracting detail a clear distinction should still be made when one wants to smooth the data and trend estimation.

When trend estimation is the main objective, the Reinsch filter alone is not enough and a different type of filter will have to be used, which is capable of selecting the frequencies needed to make a correct estimation.

6.1.2 Extending the Reinsch filter

Though the HP filter has almost the same effect as the Reinsch filter, it has one flaw which can occur. When choosing the smoothing parameter incorrectly, the HP filter can fail to remove motions in the data set that should have been removed together with the trend component. Also, when the transition rate in the frequency response function is not rapid enough for separating the trend component and fluctuations not all data which should be removed are in fact removed.
To solve this, a group of filters of type (6.8) may be used to increase the transition rate of the process. In this case we get the observable equation

$$q(t) = (I + \mu L)^n v(t) + (I - \theta L)^n \eta(t) \tag{6.10}$$

The optimal filter will then be denoted by

$$\psi(z) = \frac{\sigma_v^2 (1+\mu z)^n (1+\mu z^{-1})^n}{\sigma_\eta^2 (1-\theta z)^n (1-\theta z^{-1})^n + \sigma_v^2 (1+\mu z)^n (1+\mu z^{-1})^n} \tag{6.11}$$

Now by taking $\theta = \mu = 1$ the appropriate filter is obtained which increases the transition rate and should result in a satisfactory separation of trend and fluctuations.
The HP filter can be depicted as a special case when setting $n = 2$.
By raising the value of *n*, the sharpness of the transition can be increased. Now if the smoothing parameter $\lambda = \sigma_\eta^2 / \sigma_v^2$ is varied as well, the midpoint of the transition, or the nominal cut-off point, will create a new type of filter which may be described as a square-wave filter.

6.1.3 Square wave filters

A square wave filter is usually defined by a bidirectional form of equation (6.11):

$$\psi(z) = \frac{\delta(z)\delta(z^{-1})}{\gamma(z)\gamma(z^{-1})} = \frac{(1+z)^n (1+z^{-1})^n}{\gamma(z)\gamma(z^{-1})} \tag{6.12}$$

The filter is chosen in this form since expressions for the roots of the polynomial factors can be found more easily. In general, it is impossible to find analytic expressions for the roots of polynomials of a degree in excess of four. However, in the present case, it is possible to find analytic expressions. When taking $\mu = \theta = 1$ the expressions become reasonably tractable since one then gets the equation

$$\psi(z) = \frac{\sigma_v^2 (1+z)^n (1+z^{-1})^n}{\sigma_\eta^2 (1-z)^n (1-z^{-1})^n + \sigma_v^2 (1+z)^n (1+z^{-1})^n} \tag{6.13}$$

Though the square wave filter is very useful in extracting a trend, programming a square wave filter is much more complex. In order to construct the constituent filters $\varphi(z) = \delta(z)/\gamma(z)$ and $\varphi(z^{-1}) = \delta(z^{-1})/\gamma(z^{-1})$ which are used in the forwards and backwards passes, it is necessary to factorize the numerator and the denominator of the filter.
There is no difficulty in factorizing the numerator, yet factorizing the denominator is far more complex.
One should first find a suitable solution for $\gamma(z)\gamma(z^{-1})$ by solving the equation

$$\frac{(1+z)^n (1+z^{-1})^n}{\gamma(z)\gamma(z^{-1})} = \frac{\sigma_v^2 (1+z)^n (1+z^{-1})^n}{\sigma_\eta^2 (1-z)^n (1-z^{-1})^n + \sigma_v^2 (1+z)^n (1+z^{-1})^n} \tag{6.14}$$

which gives:

$$\gamma(z)\gamma(z^{-1}) = \frac{\sigma_\eta^2}{\sigma_v^2} (1-z)^n (1-z^{-1})^n + (1+z)^n (1+z^{-1})^n \tag{6.15}$$

And thus

$$\psi(z) = \frac{(1+z)^n (1+z^{-1})^n}{\lambda (1-z)^n (1-z^{-1})^n + (1+z)^n (1+z^{-1})^n} \tag{6.16}$$

To find the zeros of $\gamma(z)\gamma(z^{-1})$ knowledge of complex function theory is required. This shall not be discussed in this paper.
All in all the square wave filter poses to be a very useful extension for business cycle analysts who might wish to define a smoother trend in terms of a narrower range of frequencies.

6.2 Future research

Though all mentioned methodologies resulted in a pretty good method for trend extraction using the HP filter and Reinsch smoothing spline the research in the field of forecasting is still not at an end. The HP filter and Reinsch smoothing spline are used as

basis for new methodologies that are being created to further improve forecasting since however much science would like it the future can still only be estimated at best.
As a conclusion a few methodologies that have been researched and have the HP filter or Reinsch smoothing spline will be described to give an idea on which fields new improvements have risen.

6.2.1 Data smoothing using eighth order algebraic splines

A new type of algebraic spline is used to derive a filter for smoothing or interpolating discrete data points. Splines have many practical applications, including image processing and robot path planning. The spline is dependent on control parameters that specify the relative importance of data fitting and the derivatives of the spline. A general spline of arbitrary order is first formulated using matrix equations. After that the attention goes to eighth order splines because of the continuity of their first three derivatives (desirable for motor and robotics applications). The spline's matrix equations are rewritten to give a recursive filter that can be implemented in real time for lengthy data sequences. The filter is low pass with a bandwidth that is dependant on the spline's control parameters. Numerical results, including a simple image processing application, show the tradeoffs that can be achieved using the algebraic splines.

6.2.2 Spectral analysis using Fourier and wavelet techniques

This methodology tries to illustrate the properties of various measures of New Zealand's output gap. Measures of the output gap are estimated using a number of different methods: a Structural VAR model, a multivariate unobserved components model, the Hodrick-Prescott filter, a multivariate time series filter, and a linear time trend filter. In the research, spectral densities, calculated using the Fourier transform, highlight a number of important differences in the cyclical properties of the various output gap measures. However, the Fourier transform requires time series to be (weakly) stationary. Additionally, the research also uses time-dependant spectra, calculated using wavelet analysis, to further illustrate the cyclical characteristics of the different techniques used to estimate the output gap.

6.2.3 The Hodrick-Prescott and Baxter-King Filters

Recently, Baxter and King have found a new type of filter. As a result a test has been conducted which examines how well the HP and the band-pass filter proposed by Baxter and King (BK) extract the business-cycle component of macroeconomic time series. It is assessed that these filters use two different definitions of the business-cycle component. First, define that component to be fluctuations lasting no fewer than six and no more than thirty-two quarters; this is the definition of business-cycle frequencies used by Baxter and King. Second, define the business-cycle component on the basis of a decomposition of the series into permanent and transitory components. In both cases the conclusions are the same. The filters perform adequately when the spectrum of the original series has a peak at business-cycle frequencies. When the spectrum is dominated by low frequencies, the filters provide a distorted business cycle. Since most macroeconomic series have the typical Granger shape, the HP and BK filters perform poorly in terms of identifying the business cycles of these series. This is another example which has new outcomes that have positive properties and negative ones as well.

Appendix A Theoretical explanations

This appendix is devoted to explain some terms used in the paper, which are assumed to be known to the reader.

A.1 Power spectra

Let y be a stationary process with zero mean and covariances

$$R(k) = Ey(t)y^T(t-k), \qquad k \in Z \tag{A1}$$

The spectrum of the process can the be defined by the formal power series

$$S(z) = \frac{1}{2\pi}\sum_{k=-\infty}^{\infty} R(k)z^{-k}. \tag{A2}$$

The autocovariances of a process with spectrum S, which satisfies the condition $\int_{-\pi}^{\pi} \| S(e^{i\omega}) \| \, d\omega < \infty$, are given by

$$R(k) = \int_{-\pi}^{\pi} e^{ik\omega} S(e^{i\omega}) d\omega. \tag{A3}$$

Next consider the spectrum of a MA process

$$y(t) = \sum_{k=0}^{\infty} G_k \varepsilon(t-k), \tag{A4}$$

where ε is a white noise process.
In this case the spectrum of the MA process (A4) is given by

$$S(z) = \frac{1}{2\pi} G(z) G^T(z^{-1}), \tag{A5}$$

where $G(z) := \sum_{k=0}^{\infty} G_k z^{-k}$.

A.2 Periodograms

A cyclical process $y(t) = \sin(\omega_t + \theta)$ has covariance

$$R(k) = \frac{1}{2}\cos(\omega k) = \frac{1}{4}(e^{i\omega k} + e^{-i\omega k}) \tag{A6}$$

In this case the spectrum is given by

$$S(e^{i\omega}) = \frac{1}{4}(\delta(\omega) + \delta(-\omega)), \tag{A7}$$

where $\delta(\omega)$ is the Dirac distribution with the property that

$$\int_{-\pi}^{\pi} e^{i\lambda k}\delta(\omega)d\lambda = e^{i\omega k} \tag{A8}$$

So the frequency of a cyclical process is easily determined from the spectrum $S(e^{i\omega}) = \frac{1}{2\pi}\sum_{k=-\infty}^{\infty} R(k)e^{-i\omega k}$. A natural estimate of the spectrum is obtained by replacing the covariances by the sample autocovariances. This is called a periodogram.
The periodogram of the N observations $\{y(t), t = 1,..., N\}$, is defined by

$$\hat{S}_N(e^{i\omega}) = \frac{1}{2\pi}\sum_{k=-(N-1)}^{N-1}\hat{R}(k)e^{-i\omega k} \tag{A9}$$

where the autocovariances are estimated by $\hat{R}(k) = 1/N\sum_{t=k+1}^{N} y(t)y(t-k), k = 0,..., N-1$ and with $\hat{R}(k) = \hat{R}(-k)$ for k < 0.

A.3 Unobservable series

A state space system is called observable if the state vector can be reconstructed from the inputs and outputs. By $y(t; x_0, u)$ one can denote the output at time *t* generated by the input *u* and initial state $x(0) = x_0$ in the system

$$\begin{cases} x(t+1) = Ax(t) + Bu(t), & t \geq 0, \\ y(t) = Cx(t) + Du(t). & \end{cases} \tag{A10}$$

Next to this a state *x* in the state space R^n is said to be *unobservable* over the time interval $t = 0,..., k-1$ if $y(t; x, 0) = CA^t x = 0, t = 0,...., k-1$.

Appendix B Programming of all figures

This appendix is made to let the reader see how all figures can be made using Matlab.

B.1 Weekly observations of the TESO boat company

Using the dataset of 104 weekly observations the figures can be created by inputting the following code into Matlab:

```
pas = TESOweek(:,1);     vracht = TESOweek(:,2);
bus = TESOweek(:,3);     persAuto = TESOweek(:,4);

subplot(2,1,1);
plot(pas)
title('Number of passengers')
axis([0, 104, 0, 18e3])
subplot(2,1,2);
plot(vracht)
title('Number of freight wagons')
axis([0, 104, 0, 10e2])
subplot(2,1,1);
plot(bus)
title('Number of Busses')
axis([0,104,0,75])
subplot(2,1,2)
plot(persAuto)
title('Number of cars')
axis([0,104,0,1500])
```

B.2 Effect of the difference operator

To make a plot of the frequency response function and the power spectrum of a first order random walk the formulae $(I - L)^2$ and its complement, this can be seen as the power spectrum of the first order random walk, has to be rewritten. Using equation (3.4)
This will give no problem and one can input the following command lines:

```
t = 0:0.0025*pi:pi;

g = 0;

for x = 1:401
freqResp(x) = 2 - 2*cos(g*pi);
firstOrder(x) = 1/(2 - 2*cos(g*pi));
product(x) = freqResp(x)*firstOrder(x);
g = g + 0.0025;
end;
```

```
plot(t, freqResp); hold on;
plot(t, firstOrder); hold on;
plot(t, product);
axis([0 pi 0 5])
```

B.3 The difference operator and polynomial fit

First the data of the 128 monthly series has to be plotted. After that a polynomial function can be fitted to the series using a tool within the Matlab tool itself.

```
passagiers = Veerdata(:,1);

plot(passagiers);
axis([0 128 0 80000]);
```

After that the difference between the original series and the created polynomial function can be plotted.

```
verschil = passagiers - polyPas;
plot(verschil);
axis([0 128 -30000 40000]);
```

The difference operator can be implemented by simple using the command *diff()*.

```
diffOp = diff(passagiers);

plot(diffOp);
axis([0 128 -30000 40000]);
```

B.4 Creating periodograms

The periodograms of a time series can easily made using the command *ifft()* in Matlab. In this cased it has been applied to the original time series and the series after applying the second order difference operator.

```
perioPas = (abs(256*ifft([passagiers'
zeros(1,128)]))).^2/(2*pi*128);
w = [0:255]*pi/128;
plot(w, perioPas);

secDiff = diff(diffOp);
perioDif = (abs(252*ifft([secDiff' zeros(1,126)]))).^2/(2*pi*126);
q = [0:251]*pi/126;
plot(q, perioDif);
```

B.5 A modified difference operator

In this case the equation

$$\delta(L) = \left\{ \frac{(1+\theta)^2}{4} \right\}^d \frac{(I-F)^d (I-L)^d}{(I-\theta F)^d (I-\theta L)^d} \tag{B1}$$

had to implemented for the values of d = 2 and $\theta = 0.0$, $\theta = 0.5$, $\theta = 0.75$.
Using equation (3.4) equation (B1) can first be rewritten into:

$$\delta(L) = \left\{ \frac{(1+\theta)^2}{4} \right\}^2 \frac{(2-2\cos(\omega))^2}{(1+\theta^2-\theta\cos(\omega))^2} \tag{B2}$$

This can be easily be implemented into Matlab:

```
g = 0;

for x = 1:401
      filter1(x) = (1/16)*(((2 - 2*cos(g*pi))^2)/1);
      filter2(x) = (81/256)*(((2 - 2*cos(g*pi))^2)/((1 + (1/4) -
            cos(g*pi))^2));
      filter3(x) = ((49/64)^2)*(((2 - 2*cos(g*pi))^2)/((1 + (9/16)
            - 1.5*cos(g*pi))^2));
      filterC3(x) = 1 - ((49/64)^2)*(((2 - 2*cos(g*pi))^2)/((1 +
            (9/16) - 1.5*cos(g*pi))^2));
      g = g + 0.0025;
end;

plot(t, filter1); hold on;
plot(t, filter2); hold on;
plot(t, filter3); hold on;
plot(t, filterC3);
axis([0 pi 0 1.25]);
```

B.6 Signal extraction filter

The signal extraction filter is given by

$$\psi(z) = \frac{\gamma^{\xi\xi}(z)}{\gamma^{yy}(z)} = \frac{(1-\theta)^2}{4s^2} \frac{(1+z)S(z)S(z^{-1})(1+z^{-1})}{(1-\theta z^s)(1-\theta z^{-s})} \tag{B3}$$

One has to rewrite this equation so it can be plotted. Basically it can be done in the same way as in appendix B.5. In this case $s = 4$ and $S(z) = 1 + z + ... + z^{s-1}$ which gives the equation:

$$\psi(z) = \frac{(1-\theta)^2}{64} \frac{(1+z)(1+z^{-1})(1+z+z^2+z^3)(1+z^{-1}+z^{-2}+z^{-3})}{(1-\theta z^4)(1-\theta z^{-4})} \quad \text{(B4)}$$

Using the fact that $(1+z)(1+z^{-1})$ can be rewritten as $2-2\cos(\omega)$ gives:

$$\psi(z) = \frac{(1-\theta)^2}{64} \frac{(2+2\cos(\omega))(4+6\cos(\omega)+4\cos(2\omega)+2\cos(3\omega))}{(1+\theta^2-2\theta\cos(4\omega))} \quad \text{(B5)}$$

Now one can implement this into Matlab with various values for θ.
The coding for $\theta = 0, \theta = 1/4$ and $\theta = 1/2$ is:

```
g = 0;

for x  = 1:401
      newFilter(x) = (1/64)*(((2 + 2*cos(g*pi))*(4 + 6*cos(g*pi) +
            4*cos(2*g*pi) + 2*cos(3*g*pi)))/1);
      newFilter2(x) = (9/1024)*(((2 + 2*cos(g*pi))*(4 +
            6*cos(g*pi) + 4*cos(2*g*pi) + 2*cos(3*g*pi)))/(1 +
            (1/16) - (1/2)*cos(4*g*pi)));
      newFilter3(x) = (1/256)*(((2 + 2*cos(g*pi))*(4 + 6*cos(g*pi)
            + 4*cos(2*g*pi) + 2*cos(3*g*pi)))/(1 + (1/4) -
            cos(4*g*pi)));
      g = g + 0.0025;
end;

plot(t, newFilter); hold on;
plot(t, newFilter2); hold on;
plot(t, newFilter3);
axis([0 pi 0 1.25]);
```

B.7 The Reinsch smoothing filter

As given in chapter 6 the Reinsch smoothing filter can be denoted by

$$\psi(z) = \frac{\gamma^{\xi\xi}(z)}{\gamma^{yy}(z)} = \frac{\sigma_v^2(1+\mu z)(1+\mu z^{-1})}{\sigma_\eta^2(1-z)^2(1-z^{-1})^2 + \sigma_v^2(1+\mu z)(1+\mu z^{-1})} \quad \text{(B6)}$$

with $\mu = 2 - \sqrt{3}$ and $\lambda = \sigma_\eta^2 / \sigma_v^2$.

This equation can be rewritten into:

$$\psi(z) = \frac{\sigma_v^2(1+\mu^2+2\mu\cos(\omega))}{\sigma_\eta^2(2-2\cos(\omega))+\sigma_v^2(1+\mu^2+2\mu\cos(\omega))} \qquad \text{(B7)}$$

Implementing this into Matlab with $\lambda = 128, \lambda = 16, \lambda = 4, \lambda = 1$ gives the following code:

```
mu = 2 - sqrt(3);

g = 0;

for x = 1:401
      Reinsch(x) = ((1 + mu^2 + 2*mu*cos(g*pi)))/((128*(2 -
            2*cos(g*pi))^2) + (1 + mu^2 + 2*mu*cos(g*pi)));
      Reinsch2(x) = ((1 + mu^2 + 2*mu*cos(g*pi)))/((16*(2 -
            2*cos(g*pi))^2) + (1 + mu^2 + 2*mu*cos(g*pi)));
      Reinsch3(x) = ((1 + mu^2 + 2*mu*cos(g*pi)))/((4*(2 -
            2*cos(g*pi))^2) + (1 + mu^2 + 2*mu*cos(g*pi)));
      Reinsch4(x) = ((1 + mu^2 + 2*mu*cos(g*pi)))/(((2 -
            2*cos(g*pi))^2) + (1 + mu^2 + 2*mu*cos(g*pi)));
      g = g + 0.0025;
end;

plot(t, Reinsch); hold on;
plot(t, Reinsch2); hold on;
plot(t, Reinsch3); hold on;
plot(t, Reinsch4); hold on;
axis([0 pi 0 1.25]);
```

References

[1] Bell, W. (1984) Signal extraction for non-stationary time series. *The annals of statistics 12*, 646-664.

[2] Burman, J.P. (1980) Seasonal Adjustment by Signal Extraction. *Journal of the Royal Statistical Society 143,* 321-337.

[3] Cogley, T. and Nason, J.M. (1995) Effects of the Hodrick-Prescott filter on trend and difference stationary time series. Implications for business cycle research. *Journal of Economic Dynamics and Control 19,* 253-278.

[4] Harvey, A. and Koopman, S.J. (1999) Signal Extraction and the Formulation of Unobserved Components Models. *Center for Economic Research.*

[5] Heij, C., Ran, A.C.M. Schagen, F. van (2001) Mathematische Systeemtheorie. *Faculty of Sciences, Free University of Amsterdam.*

[6] Heij, C. and Schumacher, H. (1997) System dynamics in economic and financial models.

[7] Hillmer, S.C. and Tiao, G.C. (1982) An ARIMA-model-based approach to seasonal adjustment. *Journal of the American Statistical Association 7,* 63-70.

[8] Pierce, D.A. (1979) Signal Extraction Error in Nonstationary Time Series. *The Annals of Statistics,* 1303-1320.

[9] Tintner, G. (1940) The Variate Difference Method. John Wiley, Bloomington, IN.

[10] Vaart, A.W. van der (2001) Time Series Analysis. *Faculty of Sciences, Free University of Amsterdam.*

[11] Wecker, W.E. and Ansley, C.F. (1983) The Signal Extraction Approach to Nonlinear Regression and Spline Smoothing. *Journal of the American Statistical Association 78,* 81-89.

Weblinks

[1] Time Series Analysis, http://www.statsoftinc.com/textbook/sttimser.html.

[2] Time Series Analysis Signal Processing and Dynamics, http://www.qmw.ac.uk/~ugte133/book/protsdbk.htm.

[3] Data smoothing and interpolation using eighth order algebraic splines, http://academic.csuohio.edu/simond/splines/GeneralSpline.pdf.

[4] A spectral analysis of New Zealand output gaps using Fourier and wavelet techniques, http://www.rbnz.govt.nz/research/discusspapers/dp00_6.pdf.

[5] Do the Hodrick-Prescott and Baxter-King Filters Provide a Good Approximation of Business Cycles?, http://128.252.177.192/WoPEc/data/Papers/crecrefwp53.html.