Some Contributions to Filtering, Modeling and Forecasting of Heteroscedastic Time Series

Pär Stockhammar

Stockholm University

Doctoral Dissertation
Department of Statistics
Stockholm University
S-106 91 Stockholm
Sweden

Abstract

Heteroscedasticity (or time-dependent volatility) in economic and financial time series has been recognized for decades. Still, heteroscedasticity is surprisingly often neglected by practitioners and researchers. This may lead to inefficient procedures.

Much of the work in this thesis is about finding more effective ways to deal with heteroscedasticity in economic and financial data. Paper I suggest a filter that, unlike the Box-Cox transformation, does not assume that the heteroscedasticity is a power of the expected level of the series. This is achieved by dividing the time series by a moving average of its standard deviations smoothed by a Hodrick-Prescott filter. It is shown that the filter does not colour white noise.

An appropriate removal of heteroscedasticity allows more effective analyses of heteroscedastic time series. A few examples are presented in Paper II, III and IV of this thesis. Removing the heteroscedasticity using the proposed filter enables efficient estimation of the underlying probability distribution of economic growth. It is shown that the mixed Normal - Asymmetric Laplace (NAL) distributional fit is superior to the alternatives. This distribution represents a Schumpeterian model of growth, the driving mechanism of which is Poisson (Aghion and Howitt, 1992) distributed innovations. This distribution is flexible and has not been used before in this context. Another way of circumventing strong heteroscedasticity in the Dow Jones stock index is to divide the data into volatility groups using the procedure described in Paper III. For each such group, the most accurate probability distribution is searched for and is used in density forecasting. Interestingly, the NAL distribution fits best also here. This could hint at a new analogy between the financial sphere and the real economy, further investigated in Paper IV. These series are typically heteroscedastic, making standard detrending procedures, such as Hodrick-Prescott or Baxter-King, inadequate. Prior to this comovement study, the univariate and bivariate frequency domain results from these filters are compared to the filter proposed in Paper I. The effect of often neglected heteroscedasticity may thus be studied.

Keywords: Heteroscedasticity, variance stabilizing filters, the mixed Normal - Asymmetric Laplace distribution, density forecasting, detrending filters, spectral analysis, the connection between financial data and economic growth.

ISBN 978-91-7447-048-2

Printed in Sweden by US-AB, Stockholm 2010
Distributor: Department of Statistics, Stockholm University

To Alma

List of Included Papers

I A Simple Heteroscedasticity Removing Filter

Research Report 2007:1, Department of Statistics, Stockholm University. Submitted

II On the Probability Distribution of Economic Growth

Research Report 2008:5, Department of Statistics, Stockholm University.

Accepted for publication (with minor revisions) by *Journal of Applied Statistics*.

III Density Forecasting of the Dow Jones Stock Index

Research Report 2010:1, Department of Statistics, Stockholm University. Submitted

IV Comovements of the Dow Jones Stock Index and US GDP

Research Report 2010:2, Department of Statistics, Stockholm University. Submitted

Acknowledgments

First and foremost I offer my sincerest gratitude to my supervisor, Professor Lars-Erik Öller, who from the very first meeting sparked my interest in time series analysis. Your enthusiasm, humour, deep knowledge and ability to describe things clearly have been invaluable for the progress of this thesis. I can not thank you enough for showing such genuine interest in my work and my life. I have always looked forward to our entertaining and stimulating meetings, discussing not only the progress of my work but also other important things such as travels and good wines. I wish I could tell stories as vividly as you do.

I am also heartily indebted to my assistant supervisor, Professor Daniel Thorburn. I have always been amazed by your statistical intuition and your profound knowledge on literally all fields. Ventilating thoughts with you have been a great source of comfort to me. Thanks also to Mattias Villani for the interest in my work and for many valuable comments. I gratefully acknowledge Roy Batchelor for suggestions on Paper II.

I am obliged to all my friends and colleagues, former and present, at the department. Thank you for making each day at the department a joyful one. Special thanks go to my fellow doctoral students and in particular to my friend Bertil. Thanks also to Håkan for all the help with computers and software.

The Papers in this thesis have been presented at numerous occasions. I gratefully acknowledge the many suggestions from seminar participants.

Without financial support, this thesis would never have been written. I am grateful for the support from the Department of Statistics, Stockholm University. Financial support from the Royal Swedish Academy of Sciences, which facilitated my visiting scholarship in London, is highly appreciated. I am also thankful for the support from the International Institute of Forecasters (IIF).

Gratitude goes to all my friends outside the department. To Robert for being the best friend one could wish for. To Kalle and Jeanette for all the fun we have had and will have. To the regrettably occasional, but very appreciated acquaintances, such as Mattias, Nisse, Andreas etc. Regards to our new friends, Anna and Ben. I am also obliged to Emma, Sakari and Annalena for good times and for the constant willingness to help baby-sitting, watching the dog, building fences, etc.

Finally, I could not have done this without the love from my family. My parents, Inger and Gunnar, I can not thank you enough for believing in me, and for the endless support and encouragement throughout my life. Anna, you are not only my big sister and a very good marathon runner, you are also a real friend whom I can talk with about anything. My final gratitude goes to Lisa. You are truly exceptional and have the rare ability of being the best listener and talker at the same time. It is a privilege to spend time with you, our daughter Alma and dog Spassky (yes, named after the 10th chess world champion). Alma, thank you for making me happy every day - to you I dedicate this thesis.

Contents

1. Introduction **1**

2. Filters in the frequency domain **2**

2.1 Spectral analysis

3. Stationarity and heteroscedasticity **7**

3.1 Stationarity and heteroscedasticity tests
3.2 The Box-Cox transformation
3.3 Modeling volatility
3.4 Paper I: A Simple Heteroscedasticity Removing Filter

4. On finding and applying the most adequate probability distributions for heteroscedastic time series **16**

4.1 Paper II: On the Probability Distribution of Economic Growth
4.2 Density forecasting
4.3 Paper III: Density Forecasting of the Dow Jones Stock Index

5. Completing the circle **21**

5.1 Paper IV: Comovements of the Dow Jones Stock Index and US GDP

6. Conclusions and ideas for further developments **22**

References **24**

Included papers

1. Introduction

The stylized facts of economic and especially financial time series are that their variance, or volatility, changes over time. This characteristic is often referred to as heteroscedasticity (or volatility clustering) and was first recognized by Mandelbrot (1963):

> "...large changes tend to be followed by large changes - of either sign - and small changes by small changes..."

Still, the heteroscedasticity is surprisingly often neglected by practitioners and researchers. Much of the work in this thesis is about finding more effective ways to deal with heteroscedasticity in economic and financial data. This also enables measuring the effect of neglected heteroscedasticity.

The thesis is structured as follows. Potential readers of this thesis might be more or less familiar with econometrics or time series analysis. This is the reason why I in Section 2 and 3 present a toolbox of fundamentals, aimed to ease understanding of the material in Papers I-IV. These introductory sections should give the reader a presentation of the typical problems in this field, and of how the ideas and solutions presented in the thesis have evolved.
.

Section 2 presents some essentials about detrending filters and their properties. Filtering is about emphasizing or eliminating a chosen characteristic or interval of frequencies in the series. Thus, filtering is closely related to frequency domain analysis and is considered in Section 2.1. Section 3 contains some essentials about stationarity and unit root testing. The consequences of neglecting heteroscedasticity in unit root tests are discussed in Section 3.1. In the fortunate case of observing a variance that changes proportionally to the level of the series, it may be stabilized using the Box-Cox transformation described in Section 3.2. More often than not, the variance is "level invariant", and might be modelled using the techniques described in Section 3.3, or removed using the proposed procedure summarized in Section 3.4, and discussed in greater depth in Paper I.

An appropriate removal of heteroscedasticity allows more effective analysis of heteroscedastic time series. A few examples are presented in this thesis. Accounting for heteroscedasticity enables a efficient study of the underlying probability distribution of economic growth as summarized in Section 4.1.

A closely related topic is density forecasting as described in Section 4.2 and applied on Dow Jones stock index returns in Section 4.3. It is shown that the mixed Normal - Asymmetric Laplace (NAL) distribution is particularly suitable for fitting both GDP growth and stock index returns, thus hinting at an observable analogy between economic growth and financial data. Paper IV (summarized in Section 5) makes use of the proposed filter in Paper I prior to an investigation of the presumed analogy indicated in Papers II and III. Thus - Paper IV, in a sense, completes the circle.

Some concluding remarks and some ideas for future work are presented in Section 6, followed by Papers I-IV.

2. Filters in the frequency domain

Separating trends and cycles of seasonally adjusted data is essential to much macroeconomic analysis. The research to find proper methods to decompose time series was accelerated after the influential paper by Nelson and Plosser (1982), who argued that macroeconomic time series are characterized by stochastic trends rather than linear trends. This decomposition might be done using so called low-pass, high-pass or band-pass filters. Low-pass filters are used to pick out the trend in a time series (or the low frequency movements when viewed in the frequency domain). On the contrary, the high-pass filter eliminates the trend. The intermediate band-pass filter is designed to isolate midrange frequencies, often associated with business cycle fluctuations. An ideal filter completely eliminates the frequencies outside the prespecified interval, while passing the remaining ones unchanged. The exact filter would be a moving average of infinite order, impossible to design for a finite sample. A central issue in detrending time series involves finding good, hopefully optimal, approximations to the ideal filter. Perhaps the most popular (also frequently used in this thesis) approximation is the detrending filter proposed by Hodrick-Prescott (HP) (1997). The HP filter is an example of a low-pass filter. Baxter-King (BK) (1999) proposed a moving average type approximation of the business cycle band defined by Burns and Mitchell (1946). The BK filter is thus of band-pass type designed to pass through time series components with frequencies between 6 and 32 quarters, while dampening higher and lower frequencies.

Following Baxter-King (1999), a useful detrending method should satisfy six requirements. First, the filter should extract a cyclical component within a

specified range of periodicities, and leave the characteristics of this component as undistorted as possible. Secondly, the filter should not change the timing of the turning points in the series under analysis (thus, there should be no phase shift). Thirdly, the filter should be an optimal approximation to the ideal filter, according to some predesigned loss function measuring the discrepancies between the approximate and exact filters. Fourth, the filter should produce a stationary series. Fifth, the filter should yield business cycle components unrelated to the length of the observation period and finally the method must be operational. The first difference for instance, sometimes used to detrend a time series, has the drawback of being asymmetrical and thus induces phase shifts. Also, the first difference filter reweights the densities towards higher frequencies as indicated in Paper IV.

Working with filters, it is thus hard not to cross the paths of spectral analysis. The effect of any linear filter, $h(B) = \sum_{-\infty}^{\infty} h_j B^j$, where h_j, $j = 0, \pm 1, \pm 2,...$ are fixed weights and B is the lag operator such as $B^j y_t = y_{t-j}$, can be obtained from the frequency response function (or transfer function) found by replacing B by $\exp(-iw)$, where $0 \leq w \leq \pi$. Assuming that the series is stationary, the gain (defined as the modulus of the frequency response function) shows how the amplitude at each frequency is affected. Studying the gain thus provides information about whether the filter is of the low-pass, high-pass or the band-pass type. The accuracy of the approximation of the ideal filter might also be studied using the gain. The squared gain is the factor by which the original spectrum must be multiplied to yield the filtered spectrum. Other important spectral functions frequently used in this thesis include the phase shift and coherency functions. The formulae are given in the next section.

2.1 Spectral analysis

In the frequency domain, the variance of a time series is decomposed according to periodicity. This may reveal important features of univariate or bivariate time series, not apparent in the time domain. The estimation of spectral densities in the frequency domain raises some issues not encountered in the time domain.

If y_t is a real-valued stationary process with absolutely summable autocovariances, $\gamma(j)$, then the Fourier transform, $f(w)$, of $\gamma(j)$ exists and

$$f(w) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \gamma(j) e^{-iwj} = \frac{1}{2\pi} \left(\gamma(0) + 2 \sum_{j=1}^{\infty} \gamma(j) \cos wj \right). \quad (2.1)$$

This is the *spectral density function* defined in the range $[-\pi, \pi]$. Based on sample time series of n observations, it is logical to estimate $f(w)$ by replacing the theoretical autocovariances $\gamma(j)$ by the sample counterpart $\widehat{\gamma}(j)$. The spectrum is hence estimated as

$$\widehat{f}(w) = \frac{1}{2\pi} \sum_{j=-(n-1)}^{n-1} \widehat{\gamma}(j) e^{-iwj} = \frac{1}{2\pi} \left(\widehat{\gamma}(0) + 2 \sum_{j=1}^{n-1} \widehat{\gamma}(j) \cos wj \right).$$

The sample autocovariance function $\widehat{\gamma}(j)$ is asymptotically unbiased and

$$\lim_{n \to \infty} E\left(\widehat{f}(w) \right) = f(w).$$

Thus, $\widehat{f}(w)$ is also asymptotically unbiased. But the variance of $\widehat{f}(w)$ does not decrease as n increases, and so $\widehat{f}(w)$ is not a consistent estimator. It is clear that the precision of $\widehat{\gamma}(j)$ decreases as j increases, because the coefficients will be based on fewer and fewer observations. An intuitive way of reasoning would be to give less weight to $\widehat{\gamma}(j)$ as j increases. An estimator with this property is

$$\widehat{f}(w) = \frac{1}{2\pi} \left(\widehat{\gamma}(0) \upsilon_0 + 2 \sum_{j=1}^{M} \widehat{\gamma}(j) \upsilon_j \cos wj \right),$$

where $\{\upsilon_j\}$ is a set of weights called the *lag window*, and M $(< n)$ is called the truncation point. Several lag windows exist which all lead to consistent estimates of $f(w)$. Throughout the entire thesis, a Parzen window with truncation point $M = 20$ has been used to smooth the sample spectrum. This window has the advantage of not producing negative estimates. Where applied in this thesis, the chosen truncation point falls right between the two "rule of thumb" values, $M = \sqrt{n}$ and $M = 2\sqrt{n}$, see e.g. the discussion in Percival and Walden 1993, pp. 277-280.

A natural tool for examining the comovements of two stationary series in the

time domain is the cross-correlation function $r_{1,2}(j) = c_{1,2}(j)/s_1 s_2$, where $c_{1,2}(j)$ is the sample cross-covariance function on lag j, and s_1 and s_2 are the sample standard deviations for the two time series $y_{1,t}$ and $y_{2,t}$. In this study we mainly use a frequency domain approach with focus on the *cross-spectrum*. Frequency domain techniques allow for studying correlation differentiated by frequency. In practice, several cross-spectral functions are necessary to describe the comovements of two time series in the frequency domain. The cross-spectrum is most easily studied through the so called phase, the gain and the coherency functions. They are all derived from the cross-spectrum defined as the Fourier transform of the cross-covariance function $\gamma_{1,2}$, namely

$$f_{1,2}(w) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \gamma_{1,2}(j) e^{-iwj}.$$

Note that the cross-covariance function $\gamma_{1,2}(j)$ is real for real series $y_{1,t}$ and $y_{2,t}$, but $f_{1,2}(w)$ is complex because $\gamma_{1,2}(j) \neq \gamma_{1,2}(-j)$, but the cross-spectrum can be divided into one real and one imaginary part

$$f_{1,2}(w) = c_{1,2}(w) - iq_{1,2}(w),$$

where $c_{1,2}(w)$ and $q_{1,2}(w)$ are defined as

$$c_{1,2}(w) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \gamma_{1,2}(j) \cos wj$$

and

$$q_{1,2}(w) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \gamma_{1,2}(j) \sin wj.$$

The function $c_{1,2}(w)$ is called the *co-spectrum* and $q_{1,2}(w)$ the *quadrature spectrum* of the series $y_{1,t}$ and $y_{2,t}$. These functions are, however, difficult to interpret. An alternative way to express the cross-spectrum is in the form

$$f_{1,2}(w) = A_{1,2}(w) e^{i\phi_{1,2}(w)},$$

where

$$A_{1,2}(w) = \sqrt{c_{1,2}^2(w) + q_{1,2}^2(w)}$$

is real and is called the *cross-amplitude spectrum* between $y_{1,t}$ and $y_{2,t}$. The *phase spectrum*, is defined as

$$\phi_{1,2}(w) = \tan^{-1}\left[\frac{-q_{1,2}(w)}{c_{1,2}(w)}\right],$$

expressing the shift between the oscillations of the two variables. Note that $\phi_{1,2}(w)$ is discontinuous at frequency multiples of $\frac{\pi}{2}$. Another useful cross-spectral function is the *gain function* which is the ratio of the cross-amplitude spectrum to the input spectrum, i.e.

$$G_{1,2}(w) = \frac{A_{1,2}(w)}{f_1(w)},$$

the analogue of the regression coefficient in the time domain. Finally, the (squared) *coherency function* may be derived from the cross-spectrum as

$$K^2_{1,2}(w) = \frac{A^2_{1,2}(w)}{f_1(w)f_2(w)},$$

where $f_1(w)$ and $f_2(w)$ are the spectra of the individual series $y_{1,t}$ and $y_{2,t}$. The coherency is essentially the standardized cross-amplitude function and is analogous to the coefficient of determination, R^2, in the time domain. Cross-spectral analysis thus decomposes the series into individual cyclical components. The coherency is the squared correlation coefficient between $y_{1,t}$ and $y_{2,t}$ at frequency w. Clearly, $0 \leq K^2_{1,2}(w) \leq 1$. A value of $K^2_{1,2}(w)$ close to one implies a strong linear relationship of the two components at frequency w. The corresponding phase indicates at what lag this correlation occurs. It is only of interest to study the phase at frequencies where the coherency is large. Trends in the phase spectrum reveal information of the lead or lag relationship. If the trend is linear, the slope is the length of the lead or the lag. A nonlinear phase spectrum indicates varying lead or lag lengths.

Consider the linear filter $Z_t = \sum_{j=-\infty}^{\infty} h_j B^j Y_t = h(B)Y_t$, where $\sum_{j=-\infty}^{\infty} |h_j| < \infty$. It can be shown (see e.g. Priestley (1981), chapter 4.12) that the spectrum of the filtered series Z_t is given by

$$f_Z(w) = \left|h\left(e^{jw}\right)\right|^2 f_Y(w),$$

where $f_Y(w)$ is the spectral density function (2.1). The function $|h(e^{jw})|^2$ is the squared gain function often called the transfer function or the frequency response function, used to measure the effect of applying a linear filter on a series. As an example, the first difference filter

$$\begin{aligned} Z_t &= \Delta Y_t \\ &= h(B)Y_t, \end{aligned}$$

where $h(B)$ is the difference operator $\Delta = (1 - B)$ can, using standard trigonometrics, be expressed as

$$\begin{aligned} \left|h\left(e^{jw}\right)\right|^2 &= \left(1 - e^{jw}\right)\left(1 - e^{-jw}\right) \\ &= 2(1 - \cos w), \end{aligned}$$

which is a continuously increasing function for $0 \leq w \leq \pi$, see Paper IV. The transfer functions for other filters used in this thesis are found analogously.

3. Stationarity and heteroscedasticity

In time series analysis one does not usually have the luxury of obtaining an *ensemble*. That is, one typically observes only one observation at each measurement point for a specific variable, which adds up to just one *realization* of the same. Fortunately, if the series of interest, y_t, is *stationary*, the mean, variance and autocorrelations can be estimated by averaging across the single realizations. It is therefore desirable that the series is stationary and most time series models are based on the assumption that the time series of interest are approximately stationary, have been stationarized or are cointegrated with some other variables. There are various types of stationarity, see e.g. the classic work of Doob (1953, chapters 10 and 11) for a thorough treatment on the subject.

The joint distribution function of the finite set of random variables $\{Y_{t_1}, Y_{t_2}, ..., Y_{t_n}\}$ from the stochastic process $\{Y_t : t = 0, \pm 1, \pm 2, ...\}$ is defined by

$$F_{Y_{t_1}, Y_{t_2}, ..., Y_{t_n}}(y_{t_1}, y_{t_2}, ..., y_{t_n}) = P\{Y_{t_1} \leq y_{t_1}, ..., Y_{t_n} \leq y_{t_n}\},$$

where y_i, $i = 1, 2, ..., n$ are any real numbers. A time series is called *strictly* (or *strongly*) *stationary* if

$$F_{Y_{t_1},Y_{t_2},...,Y_{t_n}}\left(y_{t_1},y_{t_2},...,y_{t_n}\right)=F_{Y_{t_{1+h}},Y_{t_{2+h}},...,Y_{t_{n+h}}}\left(y_{t_1},y_{t_2},...,y_{t_n}\right),$$

for any n and h. If the series is strictly stationary, the joint distribution function is the same at each time point and depends only (if at all) on the distance between the elements in the index set. For the process Y_t, $\mu_t = E(Y_t)$ and $\sigma_t^2 = E\left(Y_t - \mu\right)^2$. The covariance and correlation functions are defined as

$$\gamma\left(t_1,t_2\right)=E\left(Y_{t_1}-\mu_{t_1}\right)\left(Y_{t_2}-\mu_{t_2}\right)$$

and

$$\rho\left(t_1,t_2\right)=\frac{\gamma\left(t_1,t_2\right)}{\sigma_{t_1}\sigma_{t_2}}.$$

Since the distribution function is the same for all t, the mean and variance functions for a strictly stationary process is constant provided that $E\left(|Y_t|\right) < \infty$ and $E\left(|Y_t^2|\right) < \infty$. Furthermore

$$\gamma\left(t_1,t_2\right)=\gamma\left(t_1+h,t_2+h\right)$$

and

$$\rho\left(t_1,t_2\right)=\rho\left(t_1+h,t_2+h\right),$$

for any t_1, t_2 and h. Thus the autocorrelation between Y_t and Y_{t+h} in a strictly stationary process with finite mean and variance, depends only on the time difference h.

A weaker form of stationarity (*weak* or *covariance stationarity*) is often used in empirical time series analysis. A weakly stationary process has constant (time invariant) joint moments up to order n. That is, a second order weakly stationary process has constant mean and variance and the covariance and autocorrelation functions being functions of the time difference alone. A strictly stationary process (with finite mean and variance) is also weakly stationary, but not so if the mean and/or the variance are infinite.

Usually, economic time series are not stationary and even after seasonal adjustment or deflation they will typically still exhibit trends, fairly regular cycles and other non-stationary behaviours. If the series has a long-run linear trend and tends to revert to the trend line following a disturbance, e.g.

$$y_t=\alpha+\beta_0 t+\varepsilon_t, \qquad (3.1)$$

where ε_t is stationary, it may be possible to stationarize the series by detrending. In this case it is done by fitting and subtracting a linear trend line prior to fitting a model. The result, $y_t = \varepsilon_t$, is stationary by definition. Such a time series is said to be *trend-stationary*. This concept can be generalized to more complicated types of trends. In practice, detrending is seldom sufficient to make the series stationary, in which case it is worthwhile to try to transform it into a series of differences, especially because many time series do not seem to follow a model of type (3.1), see Nelson and Plosser (1982). If the mean, variance, and autocorrelations of the original series vary over time, even after detrending, calculating changes (or differences) of the series between periods or between seasons is often a better stationarization method. If this results in a stationary series, it is said to be *difference-stationary*. The best example of a difference stationary process is the random walk, defined as

$$y_t = y_{t-1} + \varepsilon_t.$$

Clearly, $\Delta y_t = \varepsilon_t$. Sometimes it can be hard to tell the difference between a series that is trend-stationary and one that is difference-stationary. Using a difference to try to stationarize (3.1) yields

$$\Delta y_t = \beta_0 + \varepsilon_t - \varepsilon_{t-1}.$$

Thus, the first order MA coefficient is on the unit circle and Δy_t is noninvertible. Of course, the same problem of inducing noninvertible unit root processes may arise using models with other types of trend. In the same sense, it is inappropriate to subtract a deterministic trend from a difference-stationary process. It should also be noted that there are other more elaborated ways to detrend a time series, see the discussion in Paper I and IV.

In business cycle research, macroeconomic variables are usually decomposed into a trend and a (stationary) cyclical component. Still, in the 1970s it was widely believed that the long-run trend in macroeconomic variables is constant, i.e. trend-stationary. As already mentioned, Nelson and Plosser (1982) questioned this traditional view and argued that important macroeconomic variables (such as GDP) are instead difference-stationary.

3.1 Stationarity and homoscedasticity tests

The unit root test may be used to make statistical inference about a time series being difference-stationary or not. The two main unit root tests used in this thesis, are the augmented Dickey-Fuller (ADF) test and the Phillips-Perron (PP) test. The latter is a modification of the previous one, but unlike the ADF test, the PP test makes a non-parametric correction to the t-test statistic, see e.g. Wei (2006) chapter 9 for details.

There are a number ways to test for heteroscedasticity. In this thesis I have used the common ARCH-LM and the Breusch-Pagan tests. The former was introduced in Engle (1982) and starts by fitting the most adequate AR(q) model

$$y_t = \alpha + \sum_{i=1}^{q} \phi_i y_{t-i} + \varepsilon_t, \qquad t = 1, 2, ...$$

After that, the squared residuals $\widehat{\varepsilon}_t^2$ are regressed on a constant and q lagged values:

$$\widehat{\varepsilon}_t^2 = \widehat{\alpha}_0 + \sum_{i=1}^{q} \widehat{\alpha}_i \widehat{\varepsilon}_{t-i}^2. \qquad (3.2)$$

The null hypothesis is homoscedasticity in which case we would expect all $\widehat{\alpha}_i$ to be close to zero. The Lagrange-Multiplier (LM) test statistic nR^2, where n is the sample size and R^2 is the coefficient of determination in regression (3.2), asymptotically follows the $\chi^2(q)$ distribution. An even simpler test is obtained by regressing the squared residuals directly on the independent variables, which is the Breusch-Pagan test.

On several occasions in this thesis, unit root tests report stationarity for a series for which heteroscedasticity tests reject the null hypothesis of homoscedasticity. This contradiction reveals a weakness in the ADF and PP tests in that they fail to capture the heteroscedasticity in the series. It should also be noted that the null hypothesis of a unit root in Dickey-Fuller tests tend to be rejected too often in the presence of conditional heteroscedasticity, see e.g. Kim and Schmidt (1993). Heteroscedasticity affects estimates of parameters. The observations are unequally weighted and hence sample information is not optimally exploited, which results in inefficient estimates. So a mechanical use and interpretation of the results of unit root tests might lead to a statistically correct, but inefficient use of models which require stationary time series (such as ARIMA or ARFIMA models).

3.2 The Box-Cox transformation

In case the series is positive and where the standard deviation is changing proportionally to the level of the series, the power transformation

$$T(Y_t) = \frac{Y_t^{\rho} - 1}{\rho}, \qquad (3.3)$$

introduced by Box and Cox (1964), can be used to stabilize the variance. The Box-Cox transformation contains some commonly used transformations as special cases, for example:

ρ	Transformation
-1	$1/Y_t$
-0.5	$1/\sqrt{Y_t}$
0	$\ln Y_t$
0.5	$\sqrt{Y_t}$
1	Y_t

Ibid. showed how to estimate the transformation parameter, ρ, using maximum likelihood. The variance of economic and financial time series may change over time, not only as a function of the series, but also in other ways. What to do then?

The problem essentially has two[1] solutions which are presented in the subsequent Section 3.3 and 3.4, respectively. The first one is to model the (conditional) variance. The other is to remove the heteroscedasticity prior to model fitting. The second approach saves on parameters and enables an application of simple (second-order stationarity) models. The first approach is by far most used in practice. The reason for this is mainly due to powerful and proven tools to handle heteroscedasticity in both regression and time series data. A short survey is presented in Section 3.3.

[1] Yet another way to treat heteroscedasticity primarily in density forecasting is presented in Paper III in this thesis.

3.3 Modeling volatility

In regression it is well known that OLS estimates are not efficient in the presence of autocorrelated and/or heteroscedastic (nonspherical) disturbances. Given the model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \qquad (3.4)$$

where $E(\mathbf{u}) = 0$ and $E(\mathbf{uu}') = \sigma^2\boldsymbol{\Omega}$ (with $\boldsymbol{\Omega} \neq \mathbf{I}$), the OLS estimator of $\boldsymbol{\beta}$ will be unbiased and

$$\begin{aligned} Var(\mathbf{b}) &= E\left[(\mathbf{b}-\boldsymbol{\beta})(\mathbf{b}-\boldsymbol{\beta})'\right] = E\left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}\mathbf{u}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\right] \\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}, \qquad (3.5) \end{aligned}$$

which is obviously different from the OLS variance, $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$. Applications of the OLS estimate would lead to inefficient estimates of $\boldsymbol{\beta}$, invalid confidence intervals, t-tests and F-tests etc.

The generalized least squares (GLS) estimator multiplies (3.4) by a $n \times n$ nonsingular matrix $\mathbf{T}$

$$\mathbf{Ty} = (\mathbf{TX})\boldsymbol{\beta} + \mathbf{Tu}. \qquad (3.6)$$

Standard GLS theory (see for instance Hamilton (1994, chapter 8)) applies OLS to the transformed variables in (3.6) resulting in best linear unbiased estimators (BLUE´s) for $\boldsymbol{\beta}$ and $Var(\boldsymbol{\beta})$ in the model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$, with nonspherical disturbances.

In a regression with k explanatory variables the heteroscedasticity might take the form

$$\sigma_t^2 = \sigma^2 x_{jt}^2, \qquad t = 1, 2, ...$$

where x_j^2 is the explanatory variable that can be thought of as the source of heteroscedasticity. The original model can then be transformed to (for details, see again Hamilton (1994, chapter 8))

$$\frac{y_t}{x_{jt}} = \beta_1(\frac{1}{x_{jt}}) + \beta_2(\frac{x_{2t}}{x_{jt}}) + ... + \beta_j + ... + \beta_k(\frac{x_{kt}}{x_{jt}}) + (\frac{u_t}{x_{jt}}). \qquad (3.7)$$

The standard inference procedures are valid for the transformed variables in (3.7). Equation 3.7 can in matrix notation be generalized to

$$\frac{y_t}{\sigma_t} = \frac{\mathbf{x}_t'\boldsymbol{\beta}}{\sigma_t} + \frac{u_t}{\sigma_t}, \qquad (3.8)$$

where the variance of $u_t^* = \frac{u_t}{\sigma_t}$ is constant:

$$E\left[u_t^{*2}\right] = E\left[(\frac{u_t}{\sigma_t})^2\right] = \frac{1}{\sigma_t^2}E\left[u_t^2\right] = \frac{\sigma_t^2}{\sigma_t^2} = 1.$$

The procedure to divide each observation by the standard deviation of the disturbances is for obvious reasons often called weighted least squares. The estimates, $\widehat{\boldsymbol{\beta}}$, are found by minimizing

$$\sum_{t=1}^{n}\left(\frac{u_t}{\sigma_t}\right)^2 = (\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'\boldsymbol{\Phi}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}), \qquad \text{where } \boldsymbol{\Phi}=\text{diag}(\sigma_1^2,\sigma_2^2,....,\sigma_n^2)$$

In other words, observations with low σ_t are considered more reliable and are weighted more heavily. The observations with high σ_t however, have a smaller influence on the estimate of $\boldsymbol{\beta}$.

In the univariate case one likes to preserve the dynamic structure (autocorrelation) while making the series homoscedastic. Then (3.7) with index

$$k = \begin{cases} 0 & \text{if no intercept} \\ 1 & \text{intercept} \end{cases}$$

would be appropriate. But σ_t is unknown and must be estimated. With just one realization of the series this can not be done. A way out is to estimate σ_t recursively using a window of observations. As in the GLS case, appropriate weights would produce estimates that are close to being BLUE.

For heteroscedastic time series data, ARCH-type models are considered as benchmarks. They were first introduced in the seminal article by Engle (1982), who was awarded with the price in Economic Sciences in Memory of Alfred Nobel, 2003. Engle´s original Autoregressive Conditional Heteroscedasticity (ARCH) model has afterwards been developed into many directions, see e.g. Bollerslev et al. (1992) for a exhaustive exposition. Below follows a short survey of the models used in this thesis.

Consider the first-order autoregressive model, AR(1):

$$y_t = \phi y_{t-1} + \varepsilon_t,$$

where ε_t is i.i.d.$(0, \sigma^2)$. This model can be too restrictive in applications. A more general model allows for time varying variance. As in heteroscedastic regression (see above), the standard approach to handle heteroscedastic data is to use an exogenous variable to predict the variance. Engle (1982) proposed the model

$$\begin{aligned} \varepsilon_t &= v_t \sqrt{h_t} \\ h_t &= w + \alpha_1 \varepsilon_{t-1}^2, \end{aligned}$$

where v_t is i.i.d.$(0, 1)$ and h_t is the conditional variance. This is essentially the ARCH(1) model and may be generalized to include q lags of ε_t :

$$h_t = w + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2, \qquad (3.9)$$

which is the ARCH(q) model. The ARCH model captures the tendency of volatility clustering. In order to ensure declining weights of the shocks and to reduce the number of parameters, a linearly declining lag structure was proposed in ibid. A simple scoring algorithm for the likelihood function was also provided.

It was soon recognized that the ARCH(q) model was too restrictive in many cases. Bollerslev (1986) presented an improvement of ARCH by adding lagged values of h_t in equation (3.9):

$$h_t = w + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i h_{t-i}. \qquad (3.10)$$

This model was called the generalized ARCH, or GARCH(p, q). To ensure a well-defined process, all the infinite order AR parameters must be positive. The GARCH is able to describe the persistence in the conditional volatility. By rearranging terms, (3.10) is interpreted as an ARMA model for ε_t^2 with auroregressive parameters $\sum_{i=1}^{q} \alpha_i + \sum_{i=1}^{p} \beta_i$, and moving average parameters, $-\sum_{i=1}^{p} \beta_i$. This idea can be used to find the proper orders of p and q following Bollerslev (1988).

The simple structure of equation (3.10) induces some important limitations on the GARCH models. As first noted by Black (1976), stock returns are negatively correlated with volatility changes in stock returns. That is, the volatility tends to decline in the response to "good news" and vice versa. This phenomenon is sometimes called the leverage effect. ARCH and GARCH are examples of models unable to capture such asymmetric effects of positive and negative shocks. "Symmetric" models are usually classified as "*linear volatility models*". Next to the ARCH and GARCH models, the best known members of this class of models include the GARCH-M, FIGARCH and IGARCH models. In order to capture possible leverage effects, various nonlinear extensions of the GARCH model have been developed over the years. The earliest, and also the most commonly used one, is the exponential GARCH, or EGARCH, model introduced by Nelson (1991). The EGARCH model describes the relationship between past shocks and the *logarithm* of the conditional variance:

$$\ln(h_t) = w + \sum_{i=1}^{q} \alpha_i g(v_{t-i}) + \sum_{i=1}^{p} \beta_i \ln h_{t-i}, \qquad (3.11)$$

where $g(v_t) = \theta v_t + \delta \left[|v_t| - E\left|v_t\right| \right]$ and $v_t = \varepsilon_t \, / \sqrt{h_t}$. Because of the loglinear form of (3.11) there are, unlike the GARCH model, no restrictions on the parameters α_i and β_i to ensure nonnegativity of the conditional variance. As with the class of linear GARCH models, there are numerous nonlinear parameterizations with exotic names such as the GJR-GARCH, TGARCH, STGARCH, MSW-GARCH and QGARCH model.

3.4 Paper I: A Simple Heteroscedasticity Removing Filter

Paper I of this thesis suggests another way to handle heteroscedastic time series namely by simply removing it. This is achieved by dividing the time series by a moving average of its standard deviations (STDs), smoothed by a Hodrick-Prescott filter (HP). The suggested filter is applied on the logarithmic, quarterly and seasonally adjusted US, UK and Australian GDP series. The unfiltered (Diff ln) GDP series were all found to be stationary according to the ADF test, but significantly heteroscedastic according to the ARCH-LM test. Moreover, they are all characterized by decreasing volatility over time. Consequently, parameter estimates are strongly based on an obsolete

structure. After filtering no heteroscedasticity remains. Moreover, it was shown that the filter does not colour white noise when applied on 10 000 simulated realizations of white noise with 200 observations each. That is an important property - we do not want the filter to induce spurious characteristics into the series.

Following the discussion in Section 3.3, the most straightforward way to remove heteroscedasticity in the GDP series could be to divide the heteroscedastic series by the conditional volatility estimated from ARCH/GARCH models or from any of their many generalizations. Besides being more cumbersome, it is shown to be significantly less effective than the proposed filtering procedure.

After applying the proposed filter, an adequate ARIMA-model is estimated for the filtered GDP series, and the parameter estimates are then used in point forecasting the unfiltered time series. The forecasts are compared to those from ARIMA, ARFIMA and GARCH models estimated from unfiltered data. It is demonstrated that estimating ARIMA models from the filtered series generates significantly more accurate forecasts when pooling across all horizons, according to the Diebold-Mariano test of equal forecasting performance. Much as seasonality is suppressed by seasonal adjustment filters, this simple filter could be used as a standard method to remove heteroscedasticity prior to model fitting or just to get a glimpse of the underlying structure, not corrupted by heteroscedasticity.

4. On finding and applying the most adequate probability distributions for heteroscedastic time series

Paper I is actually the product of an idea to empirically test a reduced form of the Aghion-Howitt (AH, 1992) model. The AH model is based on the Schumpeterian idea of creative destruction, i.e. the economy is driven by welfare augmenting better products (innovations, or shocks) and temporary declines (Schumpeter, 1942, Chapter 8). AH further assumes that innovations arrive according to a Poisson process with arrival rate $c\lambda$, where c is the amount of labour used in research and $\lambda > 0$ is the parameter indicating the productivity of the research technology.

Aggregating a Poisson number of shocks (as assumed in the AH model) will lead to asymmetric distributions. This is true no matter the impact of the shocks. It is not possible to test this AH hypothesis by trying to generate realizations from some distribution and compare them to, say the US GDP series. The filter proposed in Paper I enables us to work with mean and variance stationary time series, and thus to make a fair comparison between the frequency distributions of the GDP growth series and various probability distributions (notably some asymmetric ones related to the Poisson distribution). Suitable Kernel functions of these distributions can then be compared to the Kernel distributions of the frequency distributions of the filtered series. In this thesis the Gaussian Kernel function is used together with the bandwidth proposed in Silverman (1986). This combination is considered to be optimal when data are close to normal as they are here (see the next section).

4.1 Paper II: On the Probability Distribution of Economic Growth

The distribution closest to represent the reduced form of the AH model is the exponential distribution which is the distribution of the time between innovations in the Poisson process. To also allow for negative growth, the double exponential (Laplace) distribution obtained as the difference between two exponentially distributed variables with the same value on the parameter λ is examined. The Laplace distribution is symmetric around its mean where the left tail describes below average shocks and vice versa. Due to the expected asymmetries in these series the AH representative is further modified. Allowing the exponential distribution to take different λs in the two tails leads to the asymmetric Laplace (AL) distribution which is the main model candidate.

The series studied here are the US, UK and the compound G7 GDP quarterly series. It is first recognized that data lend some support to the AH hypothesis. Significant skewness was found in the unfiltered (Diff ln) UK and G7 GDP series. As expected, the mean and standard deviation in these series are stabilized using the filter in Paper I. Also, the skewness and kurtosis are more stable to the ones estimated on unfiltered data. This indicates that the moment estimates are more accurate for the complete filtered series, an important property, especially as the parameters are here estimated using the method of moments (MM). It was also found that the excess kurtosis in

the AL distribution is too large for the filtered (and unfiltered) growth series. The AL could therefore not be the only source of innovations, so Gaussian noise is added, leading to the weighted mixed Normal-AL (NAL) distribution. This distribution is capable of generating a wide range of skewness and kurtosis, making the model very flexible. A convolution of the N and AL distributions (called c-NAL) and a Normal Mixture (NM) distribution was also considered. The parameters are estimated using MM by equating the first four noncentral sample moments with the theoretical ones and then solving those equations for the quantities to be estimated. Thus, the theoretical central and noncentral moments are provided for the NM, NAL and c-NAL distributions.

After estimating the NAL parameters it is found that the Gaussian noise component dominates. The N, NM, NAL and c-NAL distributions are compared to the empirical distributions at 1 000 equidistant point of the Kernel distribution in the interval $(\widehat{\mu} - 4\widehat{\sigma}, \widehat{\mu} + 4\widehat{\sigma})$. The accuracy is measured using four measures (RMSE, MdAPE, sMdAPE and MASE). It is found that the NAL distribution is superior to the N, NM and the c-NAL distribution according to every measure, except RMSE for the US. Kernel estimation is sometimes criticized to be based on too subjective choices both of function and of bandwidth. But so are goodness of fit tests and it is well known that tests based on both approaches have low power. To be on safer ground, χ^2 tests using three different numbers of bins are performed. The results of this test point in the same direction as before, the NAL distribution fits growth best. Thus, the US, UK and G7 GDP series could be looked upon as samples from a NAL distribution. According to the AH model, λ measures the intensity of only positive shocks. The technique presented in Paper II provides a way to estimate related quantities (though buried in Gaussian noise), and perhaps to compare different economies.

4.2 Density forecasting

A point forecast of some variable by itself contains no description of the associated uncertainty. This stand in contrast to the density forecast, which is an estimate of the probability distribution of the possible future values of that variable. It thus provides complete information of the uncertainty associated with a prediction. Between these two extremes is the interval forecast, i.e. the probability that the outcome will fall within a stated interval. The density

forecast provides information on all possible intervals.

Density forecasting is rapidly becoming a very active and important area among both researchers and practitioners of economic and financial time series. E.g. density forecasts of inflation in the UK are published each quarter both by the Bank of England in its "fan" chart and the National Institute of Economic and Social Research (NIESR) in its quarterly forecast. .

The need to consider the full density of a time series rather than, say, its conditional mean or variance has for long been recognized among decision makers. If the loss function depends asymmetrically on the outcome of future values of possibly non-Gaussian variables it is important to have information not only about the first two moments, but also full conditional density of the variables.

The issue of density forecasting heteroscedastic time series has been treated in some studies. The logical idea to use GARCH-type models have been used by e.g. Diebold et al. (1998) and Granger and Sin (2000). Weigend and Shi (2000) instead suggested hidden Markov experts for predicting the conditional probability distributions. Paper III of this thesis present yet another way to handle heteroscedasticity in density forecasting.

4.3 Paper III: Density Forecasting of the Dow Jones Stock Index

Instead of modeling the conditional variance using the above suggestions, the data (the daily Dow Jones Industrial Average, DJIA, 1928-2009), are here divided into three parts of volatility (denoted high, medium and low). Each part is being roughly homoscedastic which enables the use of simple distributions to describe each part. For each part, the most accurate density forecast distribution is searched for and the result is used to provide easy guidelines for the intervening situations of local volatility. The density forecasting ability of the NM distribution (as used by e.g. the Bank of England when calculating density forecasts of macroeconomic variables in the UK, albeit using a different parameterization, Wallis (1999)), is here compared to the N and NAL distributions. In Paper II, the latter distribution (then originated from the AH model) was found to accurately fit GDP series and

it is interesting to see if the same applies to stock index returns. To further improve user-friendliness, simplified versions of the NM and NAL (using two fixed parameters) are also considered. The density forecast ability is evaluated using the probability integral transform (PIT). Standard tests signal no autocorrelation in mean corrected powers of the PIT scores, and finding the most suitable distribution for density forecasts is a matter of finding the distribution with the most uniform PIT histogram. This is done using goodness of fit tests for the different parts of volatility, separately.

It is found that the fitted NAL distributions are superior to the N and NM on average. Also, there is no great loss of information by using the simplified NM and NAL distributions, in fact the fit is slightly improved for the NM. The NM fit is nevertheless inferior to both the NAL and the N distributions.
.

This proposed procedure of circumventing strong heteroscedasticity in the entire series involves taking decisions on how to react to different degrees of local volatility. This could be made either by constantly reestimating the parameters using the MM method and the new, local set of moment estimates. Using the simplified NAL distribution also facilitates a strict judgmental estimation of the parameters using the estimated distributions for the high, medium and low volatility parts as guidelines.

Note that the NAL distribution fit both GDP growth in Paper II and now stock index data. This could hint at a new analogy between the financial sphere and the real economy, further investigated in Paper IV.

5. Completing the circle

Applying the filter suggested in Paper I on heteroscedastic GDP growth series not only resulted in better point forecasts. It also enabled a proper study of their underlying probability distributions. In Paper II, the NAL distribution was found to be close to these probability distributions and, interestingly, also accurately fitted the DJIA in Paper III. This indicates common characteristics in GDP and financial data, further investigated in the concluding Paper IV. Their joint relationship is probably best explored in the frequency domain using the spectral tools described in Section 2.1. Both US GDP growth and Dow Jones contain a positive trend and are heteroscedastic. This must be eliminated before further investigation. The effectiveness in removing the trend and heteroscedasticity of the filter proposed in Paper I was shown there. It is logical to believe that the same filter conveniently fits in this application as well. Thus Paper IV makes use of the results in Paper I, II and III, and thus in a way, completes the circle.

5.1 Paper IV: Comovements of the Dow Jones Stock Index and US GDP

As first noted by Granger (1966), national product series such as GDP typically contain a unit root. As shown in Paper IV, the same applies to the Dow Jones stock index. In the frequency domain, this shows up as low or infinite frequency variation in the spectral density. Standard analysis requires stationarity and hence time series are detrended prior to further analysis. As mentioned in Section 2, given a finite time series it is impossible to design an ideal filter. Many approximations have been suggested. The most popular ones are the HP filter, the BK filter and the filter suggested by Beveridge and Nelson (BN) (1981). Also, simply the first difference and the centered moving average are frequently used for detrending purposes.

Surprisingly, none of the above filters takes the heteroscedasticy into account. Neglected heteroscedasticity distorts both time domain and frequency domain results. The filter proposed in Paper I not only removes heteroscedasticity, but also the trend in the series and consequently seems like a good alternative. The univariate and bivariate frequency domain results of this filter are compared to the results from the filters that do not take heteroscedasticity into account. Hence, the effect of neglected heteroscedasticity is measured.

No matter which filter is used, significant comovements exist between the DJIA and US GDP growth series. It is found that accounting for heteroscedasticity somewhat shortens the business cycles. The coherency seems quite robust across filters, but using the filter proposed in Paper I slightly shifts the coherency peak to the left and results in larger than average coherency values comparing to the other studied filters. The phase shift is less robust, especially for the BK filtered series. Most filters report that DJIA leads US GDP at peak coherency frequency (about two years), but also reveal a feedback from US GDP to DJIA at around half a year. This is also confirmed in the time domain using cross correlations and Granger-causality tests. Using the BK filter with frequency band 6 to 32 quarters by definition does not utilize this information. The same applies to the BN filter. It is therefore advisable to extend the frequency bands to 2 - 32 quarters in comovement studies like this one, provided that the series are homoscedastic. The filtered series using the suggested heteroscedasticity removing filter induce the longest lead shifts at the peak coherency frequency, and also above average feedback lag. When applied on subperiods in accordance with US GDP volatility, most filtered series showed scattered first order cross correlations, but less so in the homoscedastic series.

Thus, the choice of detrending filter affects both univariate and bivariate frequency domain results. More importantly, heteroscedasticity matters and must be eliminated prior to comovement studies like this one.

6. Conclusions and ideas for further development

This thesis provides simple, yet effective, ways to handle heteroscedasticity in economic and financial time series. The heteroscedasticity removing filter in Paper I allows new and more efficient analysis and applications. A few are presented in this thesis, such as improving point forecast accuracy of linear time series models (Paper I) or rendering a efficient study of the underlying probability distribution of economic growth possible (Paper II).

During the work with the thesis many ideas of further development have crossed my mind. Most of them were dismissed more or less immediately as simply bad ideas. But some have been stored to mature in my brain for

quite some time. Two of them even resulted in half-finished papers awaiting to be completed.

The first one involves making the heteroscedasticity removing filter in Paper I model-based inspired by the pioneering works on seasonal adjustments by Cleveland and Tiao (1976), Burman (1980) and Hillmer and Tiao (1982), and later by e.g. Maravall (1987 and 1993). These approaches typically employ ARIMA processes for the trend and seasonal components and white noise for the irregular component. Most detrending filters are ad hoc by nature, and a proper model-based approach, which jointly models the heteroscedasticity is called for. The maximum likelihood function quickly gets very complicated rendering maximum likelihood estimation of the parameters difficult.

The other half-finished project concerns using the proposed NAL as an error distribution in linear time series models such as ARIMA. In the theory of time series analysis it is common practice to assume that the noise series generating the process is normal. It is widely known that this is too restrictive in many applications, e.g. modeling financial or growth series as seen in this thesis. Even if the true process is non-normal and we mistakenly maximize a normal log likelihood for an autoregressive model of order p, the resulting estimates of the parameters are consistent but, as first mentioned by White (1982), the standard errors for the estimates need not be correct. Any linear time series model applied on skewed and leptokurtic data will produce skewed and leptokurtic residuals, so it is a straightforward idea to investigate the properties of linear models assuming different error distributions. This has been done in some studies, Tiku et al. (2000) and Damsleth and El-Shaarawi (1989) used a student´s t marginal and a double exponential (Laplace) marginal, respectively. Nielsen and Shepard (2003) investigated the case of exponential noise in the AR(1) model. However, none of the above examples accounts for the skewness, and this makes it tempting to use the NAL distribution for the noise. Very recently, Lanne and Lütkepohl (2010) used a normal mixture distribution for the noise in structural vector autoregressions. As before, the maximum likelihood functions get complicated, and numerical optimization to estimate the parameters is called for.

References

Aghion, P. and Howitt, P. (1992) A model of growth through creative destruction. *Econometrica,* 60, 323-351.

Baxter, M. and King, R. G. (1999) Measuring Business-cycles: Approximate band-pass filters for economic time series. *The Review of Economics and Statistics*, 81, 575-593.

Beveridge, S. and Nelson, C. R. (1981) A new approach to the decomposition of economic time series into permanent and transitory components with particular attention to measurement of the business cycle. *Journal of Monetary Economics,* 7, 151-174.

Black, F. (1976) Studies of stock price volatility changes. *Proceedings from the American Statistical Association, Business and Economic Statistics Section*, 177–181.

Bollerslev, T. (1986) Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics*, 31, 307-327.

Bollerslev, T. (1988) On the correlation structure for the generalized autoregressive conditional heteroskedastic process. *Journal of Time Series Analysis*, 9, 121-131.

Bollerslev, T. Chou, R. Y. and Kroner, K. F. (1992) ARCH modeling in finance: a review of the theory and empirical evidence. *Journal of Econometrics*, 52, 5-59.

Box, G. E. P. and Cox, D. R. (1964) An analysis of transformations. *Journal of Royal Statistical Society*, B, 26, 211–252.

Burman, J. P. (1980) Seasonal adjustment by signal extraction. *Journal of the Royal Statistical Society A,* 143, 321-337.

Burns, A. F. and Mitchell, W. C. (1946) Measuring business cycles. *National Bureau of Economic Research*. New York.

Cleveland, W. P. and Tiao, G. C. (1976) Decomposition of seasonal time series: A model for the Census X-11 program. *Journal of the American Statistical Association,* 71, 581-587.

Damsleth, E. and El-Shaarawi, A. H. (1989) ARMA models with double-exponentially distributed noise. *Journal of Royal Statistical Society B*, 55,

61-69.

Diebold, F. X., Gunther, T. A. and Tay. A. S. (1998) Evaluating density forecasts with applications to financial risk management. *International Economic Review*, 39, 863-883.

Doob, J. L. (1953) *Stochastic processes*. Wiley, New York.

Engle, R. F. (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50, 987-1007.

Granger, C. W. J. (1966) The typical spectral shape of an economic variable. *Econometrica*, 34, 150-161.

Granger, C. W. J. and Sin, C.-Y. (2000) Modelling of absolute returns of different stock indices: Exploring the forecastability of an alternative measure of risk. *Journal of Forecasting*, 19, 277-298.

Hamilton, J. D. (1994) *Time Series Analysis*, Princeton University Press, Princeton, New Jersey.

Hillmer, S. C. and Tiao, G. C. (1982) An ARIMA-model-based approach to seasonal adjustment. *Journal of the American Statistical Association,* 77, 63-70.

Hodrick, R. J. and Prescott, E. C. (1997) Postwar U.S. business cycles: An empirical investigation. *Journal of Money, Credit and Banking*, 29, 1-16.

Kim, K. and Schmidt, P. (1993) Unit root tests with conditional heteroskedasticity. *Journal of Econometrics*, 59, 287–300.

Lanne, M. and Lütkepohl, H. (2010) Structural vector autoregressions with nonnormal residuals. *Journal of Business and Economic Statistics*, 28, 159-168.

Mandelbrot, B. (1963) The variation of certain speculative prices. *The Journal of Business*, 36, 394-419.

Maravall, A. (1987) On minimum mean squared error estimation of the noise in unobserved component models. *Journal of Business and Economic Statistics*, 5, 115-120.

Maravall, A. (1993) Stochastic linear trends: Models and estimators. *Journal of Econometrics*, 56, 5-37.

Nelson, B. (1991) Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, 59, 347-370.

Nelson, C. and Plosser, C. (1982) Trends and random walks in macroeconomic time series: some evidence and implications. *Journal of Monetary Economics*, 10, 130-162.

Nielsen, B. and Shepard, N. (2003) Likelihood analysis of a first-order autoregressive model with exponential innovations. *Journal of Time Series Analysis*, 24, 337-344.

Percival, D. B. and Walden, A. T. (1993) *Spectral analysis for physical applications. Multitaper and conventional univariate techniques*, Cambridge University Press, Cambridge.

Priestley, M. B. (1981) *Spectral analysis and time series, vol. 1: univariate series.* Academic Press, London.

Schumpeter, J. A. (1942) *Capitalism, Socialism and Democracy.* Harper, New York.

Silverman, B. W. (1986) *Density estimation for statistics and data analysis.* Chapman and Hall, London.

Tiku, M. L., Wong, W. K., Vaughan, D. C. and Bian, G. (2000) Time series models in non-normal situations: symmetric innovations. *Journal of Time Series Analysis*, 21, 571-596.

Wallis, K. F. (1999) Asymmetric density forecasts of inflation and the Bank of England´s fan chart. *National Institute Economic Review*, 167, 106-112.

Wei, W. S. (2006) *Time Series Analysis: Univariate and Multivariate Methods (2nd ed.),* Addison-Wesley, Boston.

Weigend, A. S. and Shi, S. (2000) Predicting daily probability distributions of S&P500 returns. *Journal of Forecasting*, 19, 375-392.

White, H. (1982) Maximum likelihood of misspecified models. *Econometrica*, 50, 1-25.

A Simple Heteroscedasticity Removing Filter

Pär Stockhammar and Lars-Erik Öller

Department of Statistics, Stockholm University
S-106 91 Stockholm, Sweden
E-mail: par.stockhammar@stat.su.se

Abstract

In this paper variance stabilizing filters are discussed. A new filter with nice properties is proposed which makes use of moving averages and moving standard deviations, the latter smoothed with the Hodrick-Prescott filter. This filter is compared to a GARCH-type filter. An ARIMA model is estimated for the filtered GDP series, and the parameter estimates are used in forecasting the unfiltered series. These forecasts compare well with those of ARIMA, ARFIMA and GARCH models based on the unfiltered data. The filter does not colour white noise.

Keywords: Economic growth, heteroscedasticity, variance stabilizing filters, the Hodrick-Prescott filter.

1. Introduction

Data transformations are made in order to facilitate analysis of empirical time series. There are a number of reasons why one might want to remove heteroscedasticity before modeling. For one thing, it saves on parameters. Another reason is the fact that most time series models require stationarity. Constant mean and variance are necessary requirements for (weak) stationarity. In case the variance is proportional to the level of the series, a logarithmic transformation may make the series both homoscedastic and stationary in variance. But many time series do not have constant, or even stationary variance even after transformations. Often heteroscedasticity is simply ignored and there does not seem to exist heteroscedasticity removing filters in

the literature. Handling this problem is still a hot topic, see e.g. Baltagi et al. (2009). Giordani and Villani (2010) suggested the locally adaptive signal extraction and regression (LASER) model to capture the dynamics of the signal and noise. The mean and variance of these components are allowed to slowly or abruptly shift and the model (using a normal mixture distribution for each component) is then able to capture these movements.

In this paper we present a new filter designed to remove heteroscedasticity. This is achieved by dividing the time series with a moving average of its standard deviations (STDs), smoothed by a Hodrick-Prescott filter (HP). Here we apply the filter to the logarithmic, quarterly and seasonally adjusted US GDP series. The same series of the UK and Australia are also analyzed. All the above time series are found to be significantly heteroscedastic. After filtering no heteroscedasticity remains. Moreover, white noise is not coloured.

Within the model building framework, there exists numerous ways to handle heteroscedasticity such as weighted least squares (WLS) or generalized least squares (GLS) in regression models, autoregressive conditional heteroscedasticity (ARCH) and generalized ARCH (GARCH) models for time series data. The most straightforward way to remove heteroscedasticity in the GDP series above is to divide the heteroscedastic series by the conditional volatility estimated from ARCH/GARCH models or from any of their many generalizations. A comparison and a discussion of the two approaches will be pursued in this paper.

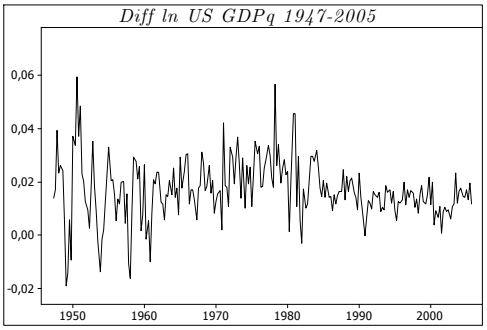
Despite the heteroscedasticity in the first differenced logarithmic (Diff ln) series, augmented Dickey-Fuller (ADF) tests do not signal any unit root. This can lead to a statistically correct, but inefficient use of e.g. ARIMA or ARFIMA models. Hess and Iwata (1997) conclude that there is no model that better replicates business cycle features than the simple ARIMA(1,1,0) model. This was contested by Candelon and Gil-Alana (2004) who considered fractionally integrated models, and showed that ARFIMA models even more accurately describe the business cycle characteristics in the US and the UK. After applying the proposed filter, an adequate ARIMA-model is estimated for the filtered GDP series, and the parameter estimates are then used in forecasting the unfiltered time series. The forecasts are compared with those from ARIMA, ARFIMA and GARCH models estimated from unfiltered data, showing that ARFIMA is not the best model. The effect of the inefficiency is measured.

The data are presented in Section 2 and the effects of heteroscedasticity are discussed in Section 3. Section 4 contains a discussion about the filters, where we also test for possible side effects. Section 5 contains the forecast comparison and Section 6 concludes.

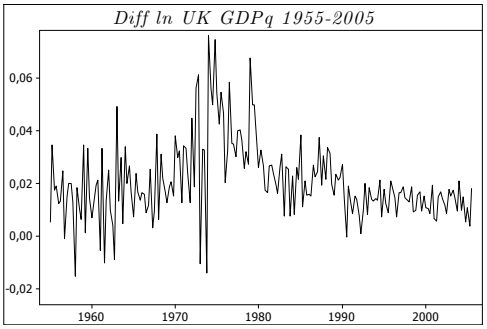
2. The data

The seasonally adjusted US GDPq (quarterly) series 1947-2005 ($n = 236$ observations) can be found on the website of *Bureau of Economic Analysis*, www.bea.gov. The seasonally adjusted UK GDPq 1955-2005 ($n = 204$), and the Australian GDPq series 1959-2005 ($n = 188$) were copied from the websites *National Statistics* (www.statistics.gov.uk) and of *OECD* (www.oecd.org), respectively. The Diff ln GDP quarterly time series together with its first four estimated moments are shown in Figure 2.1. The results of the ADF test for a unit root and the ARCH-Lagrange multiplier (ARCH-LM) test for homoscedasticity are also included.

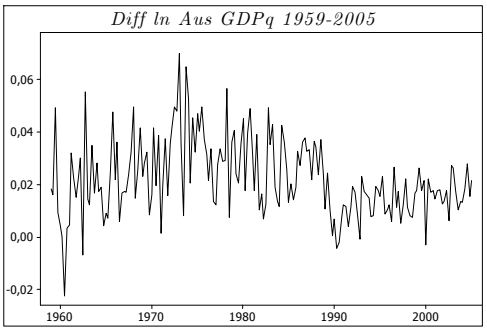
Figure 2.1: The Diff ln GDP series and their properties



Mean ($\widehat{\mu}$)	0.017
STD($\widehat{\sigma}$)	0.011
Skewness ($\widehat{\tau}$)	0.22
Kurtosis ($\widehat{\kappa}$)	2.05
ADF (p-value)	0.000
ARCH-LM (p-value)	0.000



Mean ($\widehat{\mu}$)	0.021
STD($\widehat{\sigma}$)	0.015
Skewness ($\widehat{\tau}$)	0.99
Kurtosis ($\widehat{\kappa}$)	2.07
ADF (p-value)	0.005
ARCH-LM (p-value)	0.000



Mean ($\widehat{\mu}$)	0.022
STD($\widehat{\sigma}$)	0.015
Skewness ($\widehat{\tau}$)	0.49
Kurtosis ($\widehat{\kappa}$)	0.45
ADF (p-value)	0.000
ARCH-LM (p-value)	0.005

The skewness is significantly nonzero in the UK and Aus series and significant leptokurtocity appears in the US and UK series. Unit root tests are highly sensitive to heteroscedasticity. Hamori and Tokihisa (1997) showed that a permanent STD shift strongly affects the size of Dickey-Fuller type tests. The effect of a single break in STD on the ADF test has been analyzed by Kim et al. (2002), who reported risks of over-rejection of the null hypothesis in the presence of a negative break. This builds on the study by Kim and Schmidt (1993) who showed that Dickey-Fuller tests tend to reject too often in the presence of conditional heteroscedasticity. The discussion was extended by Cavaliere (2004), who came to the same conclusion for other commonly used unit root tests. In our study, the null hypothesis of a unit root could not be rejected in all original series, but rejected for the Diff ln series, signalling stationarity. However the persistence in the series in Figure 2.1 may hint at an integrating order between zero and one, which may be hard to detect due to heteroscedasticity.

3. Effects of heteroscedasticity

Often nonstationary economic time series can be made stationary simply by differencing. But this can remedy only nonstationarity in mean - nonstationarity in variance must be handled in other ways.

Let the variance of a nonstationary process change with the level, $Var(y_t) = cf(\mu_t)$, for any positive proportionality constant c, and where f is an increasing function of the time varying level μ_t. Then it is possible to find a transformation T so that $T(y_t)$ has constant variance by approximating the function by a first order Taylor series around μ_t

$$T(y_t) \approx T(\mu_t) + T'(\mu_t)(y_t - \mu_t).$$

Now

$$Var\left[T(y_t)\right] \approx \left[T'(\mu_t)\right]^2 Var(y_t) = cf(\mu_t)\left[T'(\mu_t)\right]^2.$$

For the variance of $T(y_t)$ to be constant the transformation must be chosen so that

$$T'(\mu_t) = \frac{1}{\sqrt{f(\mu_t)}},$$

that is

$$T(\mu_t) = \int \frac{1}{\sqrt{f(\mu_t)}} d\mu_t.$$

If, for example, $Var(y_t) = c^2\mu_t^2$, then $T(\mu_t) = \int \frac{1}{\sqrt{\mu_t^2}} d\mu_t = \ln \mu_t$. Hence, a logarithmic transformation of the series will have constant variance.

If instead the variance of the series is linearly proportional to the level so that $Var(y_t) = c\mu_t$, then the square root transformation $\sqrt{y_t}$ will produce a constant variance. More generally, the Box-Cox transformation (Box and Cox, 1964) that includes the logarithmic and the square root transformations as special cases, is often used to stabilize the variance. However the transformations mentioned above are only defined for positive series, and more importantly, what to do when the standard deviation is not a function of the series?

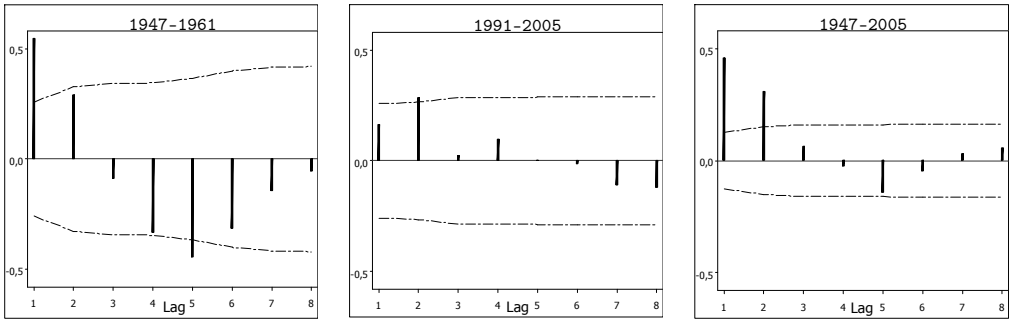
In regression it is well known that OLS estimates are not efficient in the

presence of heteroscedastic disturbances. The problem can be solved using the standard theory of GLS by simply dividing each observation and each explaining variable by an estimate of the varying standard deviation of the disturbances. OLS estimates of the transformed model would then be efficient.

In the univariate case one likes to preserve the dynamic structure (autocorrelation) while making the series homoscedastic. But, here too, the non-constant standard deviation σ_t is unknown and must be estimated. With just one realization of the series this can not be done. A way out is to estimate σ_t recursively using a window of observations[2]. As in the GLS case, dividing by appropriate observation weights would produce estimates that are BLUE.

Consider Figure 3.1 showing the autocorrelation function (ACF) of three time periods of the US GDP series, the first and last 15 years, and the entire series.

Figure 3.1 ACF of Diff ln US GDPq



There are obvious similarities in autocorrelation structure between the first and most volatile period (1947-1961), and the entire series (1947-2005). The most recent period (1991-2005), which is crucial for forecasts, does not seem to have much influence on the ACF estimates for the entire period. There has been a change in the autocorrelation structure, but with equal observation weights the model estimates will mainly be based on the old structure, because of the high volatility in the beginning of the time series.

[2]The variance is defined using a moving window of length $2\eta + 1$, with η even: $Var(y_t) = \frac{1}{2\eta+1} \sum_{i=-\eta}^{\eta} (y_t - \mu_t)^2$, where $\mu_t = \frac{1}{2\eta+1} \sum_{i=-\eta}^{\eta} y_t$

4. Heteroscedasticity filters

Prior to heteroscedasticity filtering, the series is stationarized and normalized to vary around zero. Otherwise the filtered GDP series would contain a trend. The Diff ln series in Figure 2.1 all have a nonzero mean and despite the rejection of a unit root, they exhibit slowly changing levels. This hints at the possibility of a double root, or the true integrating order could lie somewhere between $I(1)$ and $I(2)$, suggesting fractionally integrated models, such as ARFIMA. This is also supported by Candelon and Gil-Alana (2004) who concluded that the US and UK GDP series are integrated of order around $I(1.4)$. In the nine cases in the subsequent Section 5, the root is (along with the other parameters) estimated to lie between 1.27 and 1.47. Taking just one difference is not enough to extract all trend whereas taking another difference over-differences the data. One solution to the problem is simply to take a fractional difference of degree $d = 1.4$ prior to heteroscedasticity filtering. The local trends in the series will then be close to eliminated. In general, yet another parameter (d) needs to be estimated and results based on fractionally integrated time series are difficult to interpret.

4.1 A simple filter

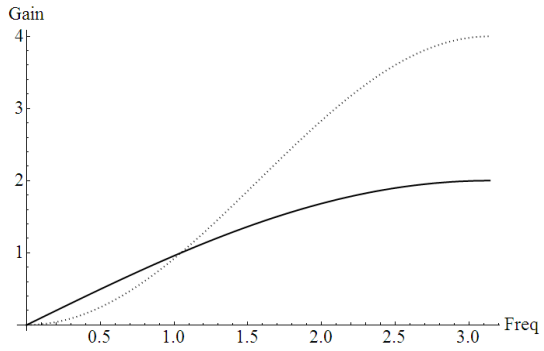
Transforming the series by subtracting from each (Diff ln) observation a local mean is a tempting alternative. It is possible to capture integration orders between $I(1)$ and $I(2)$ using one of the following two operations:

$$\begin{aligned} (a) \qquad z_t^{(a)} &= \Delta y_t - \sum_{\tau=t-\eta}^{t+\eta} \Delta y_\tau \,/k \,, \qquad t = \eta+1, \eta+2, ..., n-\eta \\ (b) \qquad z_t^{(b)} &= \Delta y_t - \sum_{\tau=t-\eta}^{t+\eta} \Delta y_{\tau-1} \,/k \,, \qquad t = \eta+2, \eta+3, ..., n-\eta+1 \end{aligned} \tag{4.1}$$

where $\Delta y_t = y_t - y_{t-1}$, y_t is the ln GDP series at time t, k (odd) is the window length, $\eta = (k-1)/2$ and even. Using different values on η in (4.1), different degrees of integration are achieved. There are two extremes. For $\eta = (n-1)/2$, the term $\sum_{\tau=t-\eta}^{t+\eta} \Delta y_\tau \,/k$ equals $\overline{\Delta y}$. The other extreme appears when k equals one, that is $\eta = 0$. Operation (4.1b) is used only in

the latter case and is equivalent to a the second difference operation, $\Delta^2 y_t$. The choice of η depends on the series studied. If it is close to $I(1)$ then you should just choose η close to $(n-1)/2$, and if the series is close to $I(2)$ then choose $\eta = 0$ in (b) or a small value on η in (a). No matter what choice of η, (4.1) is a high-pass filter in that it removes the low frequency movements of the series. The gain is the change in the output when a step change of unit size hits the input. Figure 4.1 shows the gain for the two extremes.

Figure 4.1: The gain functions of the special cases one difference, $\eta = (n-1)/2$ (solid line) and two differences, $\eta = 0$ (dotted line)



This transformation can be generalized by raising $z_t^{(i)}$ to the power d, $\left(z_t^{(i)}\right)^d$ where $i = a, b$. This enables us to handle integrating orders below one and above two. However, this is not needed for the series studied in this paper.

Figure 4.2 shows the estimated spectral densities of the (4.1) for the Diff ln US, UK and Aus data, using window length $k \approx n/2$, $k = 35$, $k = 15$ and $k = 1$. In all cases a Parzen window was used to smooth the spectral densities.

Figure 4.2: The spectrum of the ln US, UK and Aus data

For the extreme of just one difference one gets a spectrum dominated by low frequency variations, due to the persistence in these series. On the contrary, using (4.1b) with $k = 1$ ($\Delta^2 y_t$) removes all variation at the zero frequency in these series. Neither of the two extremes is very attractive to use here, the dominant low or high frequency properties overshadow frequencies in between. The remaining alternatives all have very similar high frequency properties. As expected when the windows get shorter the low frequencies are transferred to higher frequencies. The high frequency variations resulting from even shorter windows are too dominating. It seems that window length $k = 15$ is a good middle course which succeeds in detrending the data without removing business cycle features. Using $k = 15$, all three filtered spectral densities are bimodal with peaks at around three and ten quarters. The first four moments of $z_t^{(a)}$ from (4.1), using $k = 15$, are reported in Table 4.1.

Table 4.1: The moments of $z_{t,US}^{(a)}$, $z_{t,UK}^{(a)}$ and $z_{t,Aus}^{(a)}$ *using* $k = 15$

	US	UK	Aus
Mean ($\widehat{\mu}$)	0.000	0.000	0.000
STD($\widehat{\sigma}$)	0.011	0.011	0.012
Skewness ($\widehat{\tau}$)	0.034	−0.459	0.123
Kurtosis ($\widehat{\kappa}$)	2.362	4.681	0.932

The skewness is significantly nonzero in the UK, and significant leptokurtocity appears in all series. Moreover the series are still significantly heteroscedastic but stationary.

Removing the heteroscedasticity is a matter of dividing (4.1) by estimates of the changing volatility. It does not seem reasonable to assume that the volatility changes abruptly all the time - smoothing is necessary as a compromise between contrafactual constant STD and unrealistically large volatility changes from one quarter to the next. Hence we assume that the variance is

slowly evolving over time. Thus, a logical estimate of the volatility at time t is

$$HP^{(\gamma)}\left\{\sqrt{\frac{\sum_{\tau=t-\nu}^{t+\nu}\left(z_\tau^{(i)}\right)^{2d}}{2\nu}}\right\}, \qquad i=a,b \qquad (4.2)$$

where $\nu=(l-1)/2$ and l is the window length which might not be equal to the window length, k, in (4.1). In this study however, $k=l$. $HP^{(\gamma)}$ is the Hodrick-Prescott (1997) filter designed to decompose a macroeconomic time series into a nonstationary trend component and a stationary cyclical residual.

Given a time series x_t (in this case the time-dependent variance), the decomposition into unobserved components is

$$x_t = g_t + c_t,$$

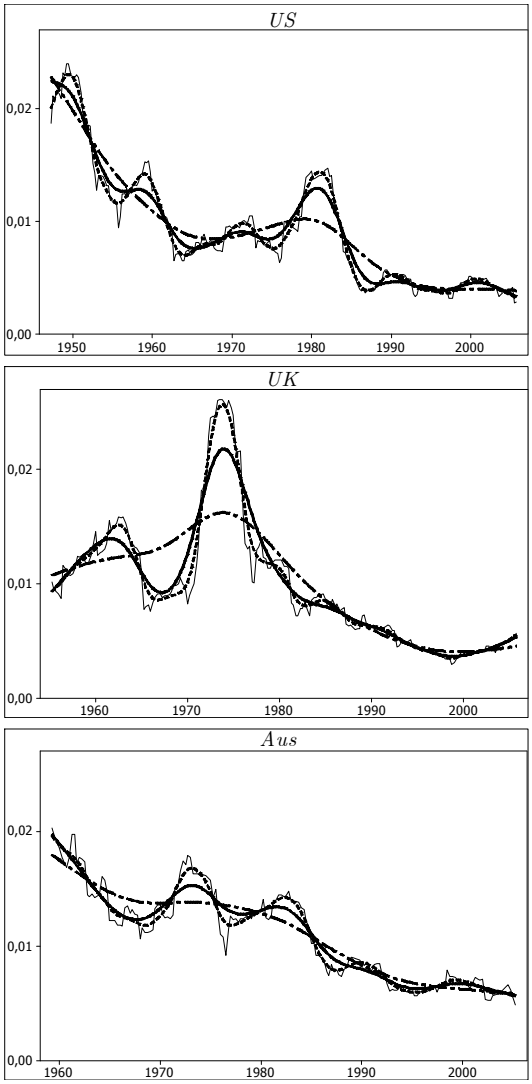
where g_t denotes the unobserved trend component at time t, and c_t the unobserved cyclical residual at time t. Estimates of the trend and cyclical components are obtained as the solution to the following minimization problem

$$\min_{[g_t]_{t=1}^n}\left\{\sum_{t=1}^{n} c_t^2+\gamma\sum_{t=3}^{n}(\triangle^2 g_t)^2\right\}, \qquad (4.3)$$

where $\triangle g_t = g_t - g_{t-1}$ and $g_{\min}$ is the HP-filter. The first sum of (4.3) expresses the closeness between the HP trend and the original series, while the second sum represents the smoothness of the trend. The positive smoothing parameter γ controls the weight between the two components and is thereby a measure of the signal-to-noise variance ratio. As γ increases, the HP trend becomes smoother and vice versa. Note that the second sum, $(\triangle^2 g_t)$, is an approximation to the second derivative of g at time t.

Figure 4.3 shows the moving STDs (using $l=15$) and illustrates the effect of the HP-filter for different values of γ.

Figure 4.3: The moving STDs (thin solid line) and the HP trend using γ=100 (dashed), γ=1 600 (thick solid) and γ=50 000 (dashed/dotted)



If we would not have made the mean correction in (4.1) a positive trend would have resulted in $\widetilde{z}_t$ below, cf. Figures 2.1 and 4.3. For quarterly data a commonly used value is $\gamma = 1\ 600$, originally proposed in Hodrick and Prescott (1997)[3]. In this study, $\gamma = 1\ 600$ accords well with the principle

[3] " ...a 5 percent cyclical component is moderately large, as is a one-eighth of 1 percent

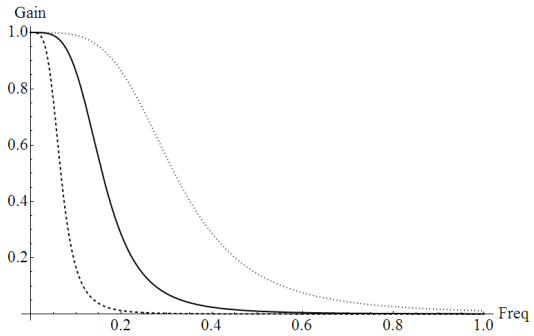
"not too rugged - not too smooth".

Dividing (4.1) by (4.2) and multiplying by the overall STD of the entire period, s_y, and adding the corresponding arithmetic average $\overline{y}$, we get the proposed heteroscedasticity removing filter

$$\widetilde{z}_t = s_y \left[\frac{\left(z_t^{(i)}\right)^d}{HP^{(\gamma)}\left(\sqrt{\sum_{\tau=t-\nu}^{t+\nu}\left(z_\tau^{(i)}\right)^{2d}/2\nu}\right)} \right] + \overline{y}, \qquad (4.4)$$

where $\widetilde{z}_t$ is the filtered series, $i = a, b$ and $t = \max[k-\eta, l-\nu], \max[k-\eta+1, l-\nu+1], ...$ Dividing (4.1) by the moving STDs directly remains an alternative hypothesis to be considered later on in this section. Whatever the choice of γ in (4.4), the trend component, g_t, of the HP filter is a low-pass filter, see Figure 4.4.

Figure 4.4: The gain functions of the HP filter using γ=100 (dotted line), γ=1 600 (solid line) and γ=50 000 (dashed line)[4]



This means that for any parameter values, (4.4) is a high-pass filter. The filter is ad hoc in the same sense as the X11 seasonal adjustment. It is well known that seasonal adjustment filters may introduce spurious autocorrelation (as

change in the growth rate in a quarter", yielding $\sqrt{\gamma} = 5/(1/8)$, $\gamma = 1\ 600$.

[4]The gain function of the HP filter is $G(w,\gamma) = \frac{1}{1+4\gamma(1-\cos w)^2}$, where w is the frequency.

does X11), see e.g. Wallis (1974). This would be a serious disadvantage of the filter and can be tested by feeding white noise into (4.4). This filter (using three different values of γ) was applied on 10 000 simulated i.i.d $N(0,1)$ series with 200 observations each, the approximate length of the GDP series in this study. The results are shown in Table 4.2.

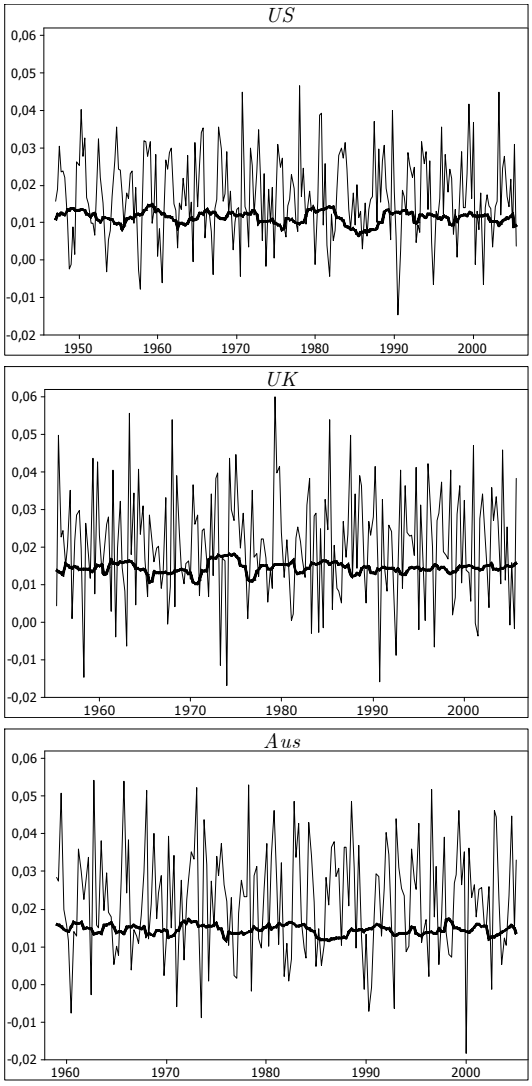
Table 4.2: Filter effects on white noise

	White noise	$\gamma = 1\ 600$	$\gamma = 10\ 000$	$\gamma = 100\ 000$
Mean	0.00005	0.00001	0.00002	0.00001
STD	0.99968	0.99244	0.99468	0.99725
Skewness (τ)	0.00002	0.00046	0.00035	0.00021
Kurtosis (κ)	2.99983	2.94535	2.96130	2.98420
Jarque-Bera	491/10000	488/10000	490/10000	504/10000
$Q^*_{0.05}(12)$	486/10000	651/10000	589/10000	507/10000
$Q^*_{0.05}(24)$	489/10000	691/10000	612/10000	521/10000

The first four moments in Table 4.2 are close to identical for the white noise series, unfiltered and filtered. The size of the Jarque-Bera test for normality is also correct at confidence level 0.05, the size of the Ljung-Box autocorrelation tests $Q^*_{0.05}(12)$ and $Q^*_{0.05}(24)$ is slightly too large, but no major distortion appears. Still, for $\gamma = 1\ 600$, the right hand tail of the χ^2-distribution is slightly too thin. Thus, for $\gamma = 1\ 600$ a threshold value of 0.065 ($\chi^2_{0.065}(12) = 19.85$) is more appropriate for the 0.05 significance level, for 12 lags, and 0.07 ($\chi^2_{0.07}(24) = 35.24$) for 24 lags. When $\gamma \to \infty$, the HP-filter becomes linear, exactly preserving the distribution.

Figure 4.5 shows the filtered series $\widetilde{z}_t$ with $k = l = 15$, $d = 1$ and $\gamma = 1\ 600$. Series of moving STDs (thick solid lines) are also included (using window length 15) in order to elucidate the effect of the filtering.

Figure 4.5: The filtered series



Filtered series using (4.4) but without HP smoothing of the STDs

$$\widetilde{z}_t^{(\text{no HP})} = s_y \left[\frac{\left(z_t^{(i)} \right)^d}{\sqrt{\sum_{\tau=t-\nu}^{t+\nu} \left(z_\tau^{(i)} \right)^{2d} / 2\nu}} \right] + \overline{y}$$

are very similar. For the sake of completeness, the two alternatives will be compared. Figure 4.5 shows that both the mean and the STD look stable in the filtered series. As expected these series pass stationarity and homoscedasticity tests, see Table 4.3. Because of the low power of the ADF and ARCH-LM test in possible concurrent effects of levelshifts and heteroscedasticity, these tests are supplemented by the results of the Phillip-Perron (P-P) test for a unit root and the Breusch-Pagan (B-P) test of homoscedasticity.

Table 4.3: Testing for a unit root and for homoscedasticity in the filtered series (p-values) when (4.1) is divided by unfiltered $\widehat{\sigma}_t$ (no HP) and according to (4.4) (HP)

	US		UK		Aus	
	no HP	HP	no HP	HP	no HP	HP
ADF	0.00	0.00	0.00	0.00	0.00	0.00
P-P	0.00	0.00	0.00	0.00	0.00	0.00
ARCH-LM	0.20	0.71	0.66	0.90	0.81	0.63
B-P	**0.03**	0.31	0.43	0.89	0.84	0.93

As expected, the null hypothesis of a unit root is rejected in all cases. Also, filter (4.4) successfully removes heteroscedasticity from the US, UK and Aus GDP series, without seriously affecting the dynamics of the series. Note that the B-P test signals heteroscedasticity (at significance level 0.05) in the filtered US series when no HP filter has been used to smooth the volatility.

4.2 A GARCH-type filter

It is interesting to compare these results to the other approach mentioned earlier, namely to divide (4.1) by the conditional volatility estimated from GARCH-type models. Specifically, in the presence of asymmetries and leptokurtocity, nonlinear GARCH models must be considered which have the ability to capture asymmetric effects. Several such models exist in the literature, most notably the EGARCH model of Nelson (1991), the TGARCH model of Zakoian (1994) and the GJR-GARCH model of Glosten et al. (1993). These models were originally introduced to capture the leverage

effect of stock returns. The models that are found adequate and minimize Akaike´s criterion are:

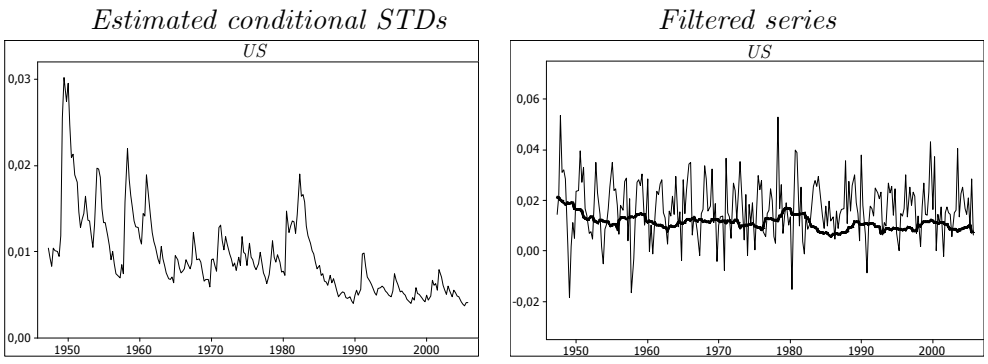
$$
\begin{aligned}
z^{(a)}_{t,US} &= \varepsilon_t \\
\ln(h_{t,US}) &= \underset{(0.26)}{-0.66} - \underset{(0.14)}{0.43}\left(\frac{\varepsilon_{t-1}}{\sqrt{h_{t-1}}}\right) + \underset{(0.11)}{0.42}\left(\left|\frac{\varepsilon_{t-1}}{\sqrt{h_{t-1}}}\right| - E\left(\left|\frac{\varepsilon_{t-1}}{\sqrt{h_{t-1}}}\right|\right)\right) + \underset{(0.03)}{0.93}\ln(h_{t-1})
\end{aligned}
$$

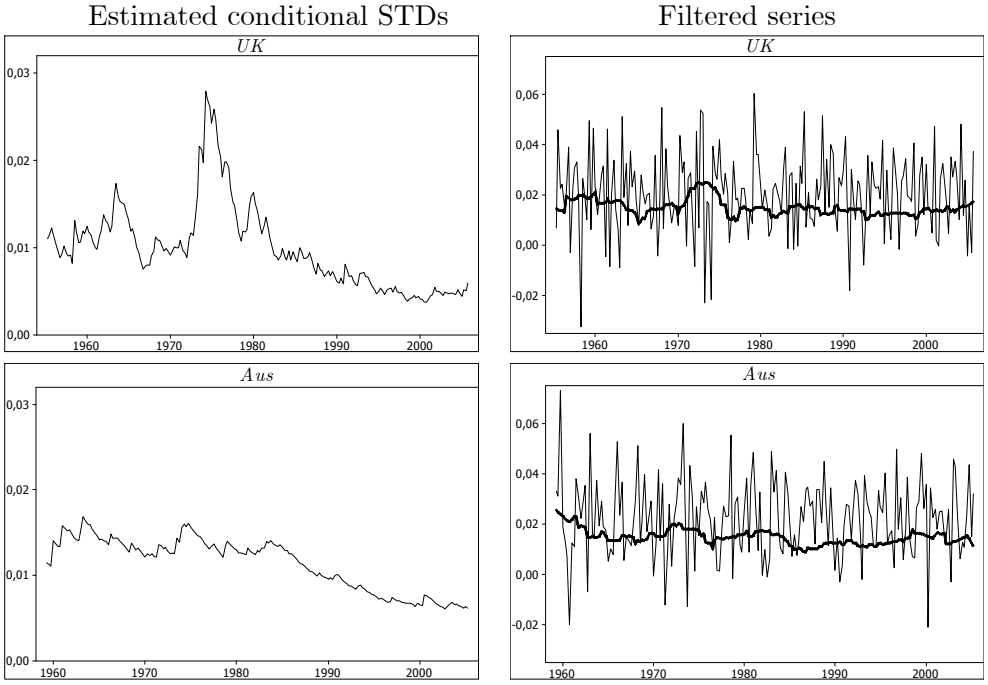
$$
\begin{aligned}
z^{(a)}_{t,UK} &= \underset{(0.07)}{0.22}\, z^{(a)}_{t-1,UK} + \varepsilon_t \\
\ln(h_{t,UK}) &= \underset{(0.08)}{-0.15} - \underset{(0.02)}{0.11}\left(\frac{\varepsilon_{t-1}}{\sqrt{h_{t-1}}}\right) + \underset{(0.09)}{0.30}\left(\left|\frac{\varepsilon_{t-1}}{\sqrt{h_{t-1}}}\right| - E\left(\left|\frac{\varepsilon_{t-1}}{\sqrt{h_{t-1}}}\right|\right)\right) + \underset{(0.02)}{0.98}\ln(h_{t-1})
\end{aligned}
$$

$$
\begin{aligned}
z^{(a)}_{t,Aus} &= \varepsilon_t \\
h_{t,Aus} &= \underset{(0.03)}{0.06}\,\varepsilon^2_{t-1} + \underset{(0.04)}{0.93}\,h_{t-1}
\end{aligned}
$$

Note that the coefficient estimates of h_{t-1} are close to one indicating a strong persistence in the conditional variance, or outright nonstationarity. This is also seen in Figure 4.6 showing the conditional volatility estimated using the above models (left panel) and the corresponding filtered series (right panel).

Figure 4.6: EGARCH/GARCH filtered series



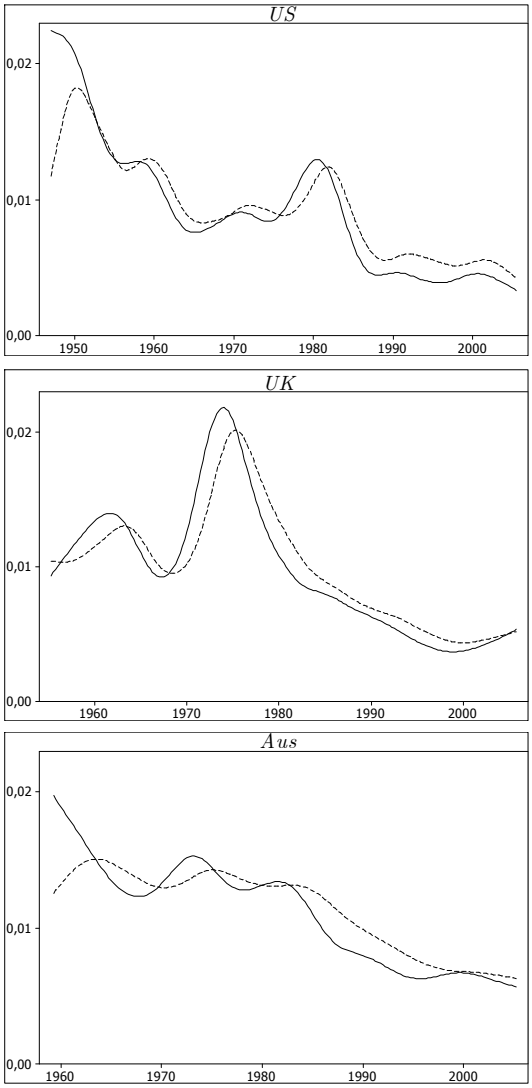


Note the rugged shape of the estimated volatility. It seems that HP-filtering of the volatility series would be even more needed here than in filter (4.4), cf. Figure 4.3. This is performed in the following subsection showing a comparison of the above filters.

4.3 Comparing the filters

Figure 4.7 shows a comparison of the HP trends (with $\gamma = 1\ 600$) of the estimated volatility using filter (4.4), see cf. Figure 4.3, and of the EGARCH/ GARCH filter.

Figure 4.7: The HP trend of moving STDs of z_t in filter (4.4) (solid lines) and of EGARCH/GARCH estimated STDs (dashed lines)



The amplitude of the HP trend is generally higher using filter (4.4) than the alternatives, rendering a heavier heteroscedasticity filtering. Also there are phase shifts between the HP trends, which are due to the fact that filter (4.4) is centered and the EGARCH/GARCH is a causal filter. This can be corrected for by lagging the HP trend according to the calculated phase shift or constructing a two-sided filter, as in model-based seasonal adjustment,

see Hillmer and Tiao (1982). At peak coherency frequency between the HP trends, the estimated phase shifts are 12.3, 10.7 and 10.9 quarters for the US, UK and Aus respectively. After correcting the series, almost identical filters are obtained. Note however that this procedure results in losing observations in the end of each series.

Despite the rather unstable looking moving STDs of the EGARCH/GARCH filtered series, the ARCH-LM tests fail to reject the null hypotheses of homoscedasticity, see Table 4.4

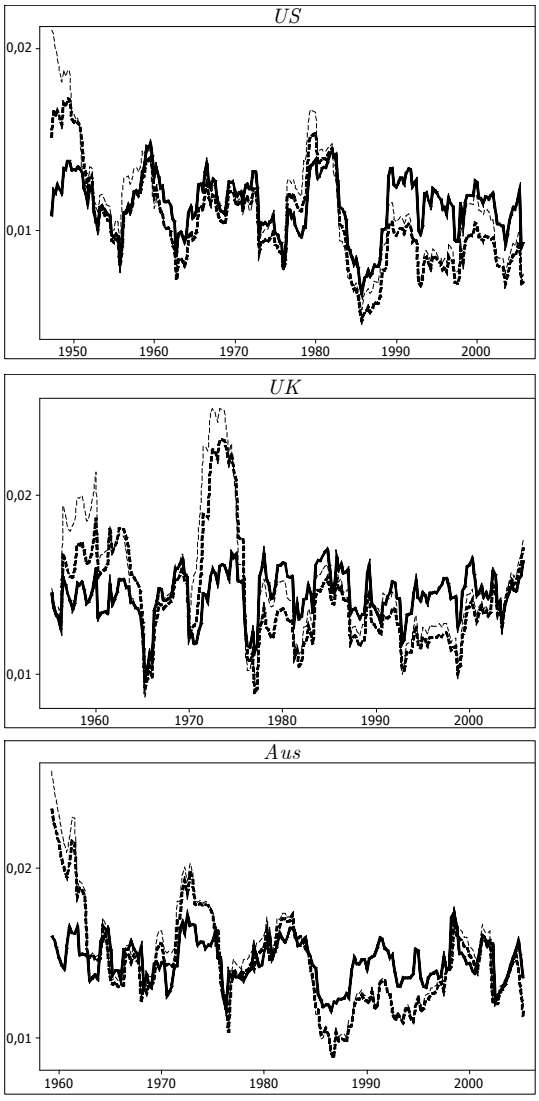
Table 4.4: Test for a unit root and homoscedasticity in the EGARCH/GARCH filtered series (p-values)

	US		UK		Aus	
	no HP	HP	no HP	HP	no HP	HP
ADF	0.00	0.00	0.00	0.00	0.00	0.00
P-P	0.00	0.00	0.00	0.00	0.00	0.00
ARCH-LM	0.97	0.31	0.39	0.68	0.65	0.80
B-P	0.90	0.11	0.07	0.32	0.44	0.83

Whether the HP filter renders the heteroscedasticity filtering more effective or not is still not clear. The p-values for the UK and Aus are higher using the HP filter than without, but it is the other way around for the US. Note, however, that the p-values for the ARCH-LM and B-P tests of the filtered series are considerably smaller on average compared to the ones shown in Table 4.3.

To elucidate the differences, Figure 4.8 presents the moving STDs (again using $k = 15$) of each filtered series. The $\widetilde{z}_t$ series without the HP filter have been omitted in order to simplify comparisons. Also, to enable graphical comparisons in both ends, filter (4.4) has been modified there.

Figure 4.8: The moving STDs using filter (4.4) (solid line), and the EGARCH/ GARCH filter (dashed line), with HP (thick line) and without HP(thin line)

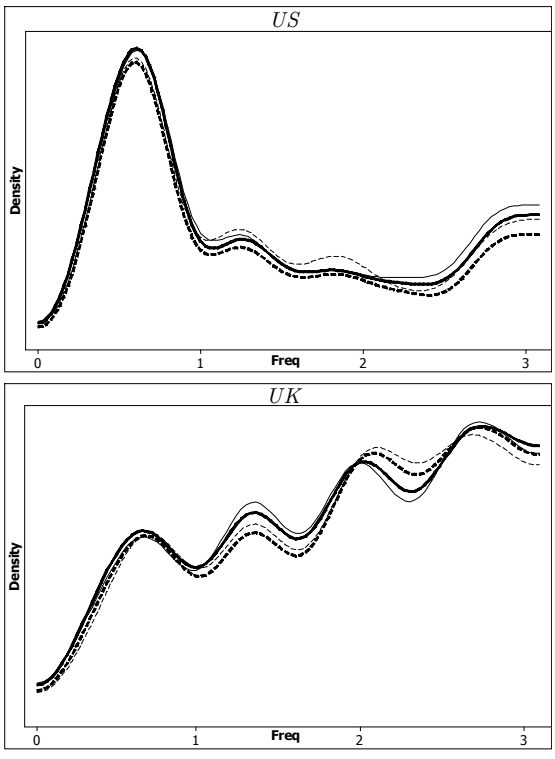


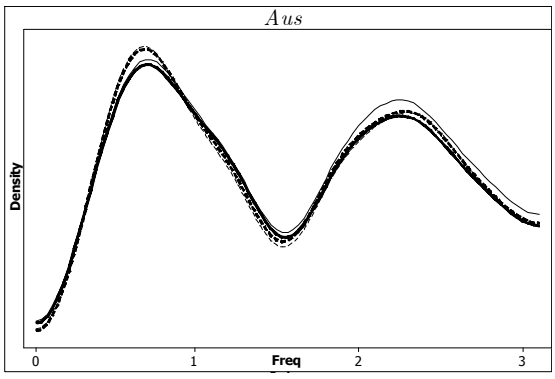
For all three countries it is again evident that filter (4.4) more effectively removes heteroscedasticity in these series. The moving STDs of the EGARCH/ GARCH filtered series all contain a decreasing local trend in the beginning, and after that the amplitude of the swings seem to be larger. This is especially true when no HP filter was used in the filtering. This result is somewhat

surprising, particularly for the US for which both the ARCH-LM and B-P tests in Table 4.4 report very high p-values compared to the other filters. According to e.g. the sign test, the median absolute deviations from the overall STD is significantly smaller using filter (4.4) than the EGARCH/GARCH filter. Applying the HP filter on the STDs or not does not change this result. Our recommendations to apply it are based on logical considerations, see the discussion in Section 4.1. Also, this way of reasoning is supported by the test results in Table 4.3, 4.4, the sign test and Figure 4.8.

Figure 4.9 shows that the spectral densities of the filtered series using (4.4) or the EGARCH/GARCH filter, with and without the HP filter, are almost identical.

Figure 4.9: Spectral densities of the filtered series using (4.4) (solid line), and the EGARCH/GARCH filter (dashed line), with HP (thick line) and without (thin line)





The close similarities of the spectral densities are to some extent due to the Parzen window and truncation point used, $M = 20$, which for each country is between the existing rules-of-thumb values $M = \sqrt{n}$ and $M = 2\sqrt{n}$ suggested in the literature, see e.g. Percival and Walden 1993, pp. 277-280. By reducing M, larger differences will appear in the spectral densities, but the fundamental features will still be similar. An application of the simple filter (4.4) is the subject of the next section.

5. A forecast competition

In this section we generate 1-20 quarters ex ante point forecasts, using adequate ARIMA models estimated for the proposed *filtered* series $\widetilde{z}_{t,US}$, $\widetilde{z}_{t,UK}$ and $\widetilde{z}_{t,Aus}$ (with $k = l = 15$, $\gamma = 1\ 600$ and $d = 1$). The parameter estimates are then applied in forecasting the *unfiltered* (Diff ln) data. In table 5.1 the results of this proposed procedure are marked with an asterisk *. We compare the accuracy of these with forecasts from ARIMA, ARFIMA and GARCH models estimated for the *unfiltered* data. The accuracy of the out of sample forecasts is measured by the root mean square forecast error (RMSFE) divided by the STD, and multiplied by 100. Hence, RMSFE = 100 would signal that the forecast error is of the same size as that of a naive forecast. The fact that the values are considerably smaller, even for the longest horizons, is another symptom of the decreasing variance of the time series studied. For instance, the STD of Diff ln US GDP of the latest 15 years is just 41 percent of the STD of the entire period.

Table 5.1: GDP forecast comparisons[5] (RMSFE)/STD

Country	ARIMA	ARIMA*	ARFIMA	GARCH	Horizon (q)
US	25.5	**25.3**	37.7	25.6	4
UK	51.2	**40.8**	55.7	61.2	4
Aus	27.8	**19.3**	42.7	36.7	4
Country	ARIMA	ARIMA*	ARFIMA	GARCH	Horizon (q)
US	**24.6**	36.4	27.2	36.0	8
UK	47.4	**38.3**	49.1	46.5	8
Aus	50.2	42.1	**40.0**	46.7	8
Country	ARIMA	ARIMA*	ARFIMA	GARCH	Horizon (q)
US	46.3	45.9	48.6	**41.7**	12
UK	39.8	**36.7**	40.2	38.2	12
Aus	53.7	50.8	**44.9**	50.2	12
Country	ARIMA	ARIMA*	ARFIMA	GARCH	Horizon (q)
US	80.5	**69.8**	71.5	97.9	20
UK	36.9	**36.3**	46.2	98.2	20
Aus	50.6	45.5	**42.3**	95.4	20

Bold figures mark the lowest RMSFE of each row. Table 5.1 offers little support to Candelon and Gil-Alana (2004); ARFIMA models produce the least accurate forecasts at the horizon of four quarters. For the longer horizons ARFIMA has the lowest RMSFE only for Australian GDP. Estimating ARIMA models from the filtered series, rather than from the unfiltered ones, generates significantly more accurate forecasts when pooling across all horizons above, according to the Diebold-Mariano test of equal forecasting performance (p-value $= 0.02$). When comparing our method with ARFIMA and GARCH-type models (also estimated on the unfiltered series) the p-values of the Diebold-Mariano test is 0.04 and 0.01, respectively. The superiority of models based on filtered data is due to the inefficiency of estimates from heteroscedastic data. The series in this study can be seen as a worst case scenario, because the volatility of the series decreases over time. Consequently,

[5]The forecasts with a horizon of four quarters are made for 2005, using data up to quarter 4 of year 2004 for estimation. Consequently, four point forecasts are compared to four actual outcomes. Similarly, the outcomes of 2004-2005, 2003-2005 and 2001-2005, are omitted in the estimation step but used when comparing the forecast accuracy on the horizons 8, 12 and 20 quarters, respectively.

parameter estimates are strongly based on an obsolete structure.

6. Conclusions

In this paper we discussed the issue of removing heteroscedasticity. We also propose a simple filter that successfully removes the heteroscedasticity in GDP series without significantly distorting the dynamics. Unlike the Box-Cox transformation, the filter does not assume that the heteroscedasticity is proportional to the level of the series. Dividing the heteroscedastic series by ARCH/GARCH estimated volatility is much more cumbersome and is significantly less effective than the proposed filter.

A mechanical estimation of e.g. ARIMA or ARFIMA models on heteroscedastic Diff ln GDP series is unbiased, but inefficient. Using filtered data resulted in better forecasts in a large majority of cases. When pooling across all horizons, the ARIMA models estimated for the filtered series generated significantly more accurate forecasts compared with ARIMA, ARFIMA and GARCH on the unfiltered data. The result is based on only three series, but they are the longest and the most important quarterly GDP series around. This simple filter could be used as a standard method to remove heteroscedasticity, much as seasonality is suppressed by seasonal adjustment filters.

Acknowledgments: We gratefully acknowledge helpful comments from Daniel Thorburn of Stockholm University, Mattias Villani of the Swedish Riksbank and from seminar participants at Cass Business School (London), Stockholm University, University of Helsinki, Åbo Akademi University (Turku) and the International Symposium on Forecasting in New York and in Nice. The research was supported by the Department of Statistics at Stockholm University, the Royal Swedish Academy of Sciences, the International Institute of Forecasters and by the Societas Scientiarum Fennica.

References

Baltagi, B. H., Jung, B. C. and Song, S. H. (2009) Testing for heteroskedasticity and serial correlation in a random effects panel data model. *Journal of Econometrics*, 154, 122-124.

Box, G. E. P. and Cox, D. R. (1964) An analysis of transformations. *Journal of Royal Statistical Society*, B, 26, 211-252.

Candelon, B. and Gil-Alana, L. A. (2004) Fractional integration and business cycle features. *Empirical Economics,* 60, 343-359.

Cavaliere, G. (2004) Unit root tests under time-varying variances. *Econometric Reviews*, 23, 259-292.

Giordani, P. and Villani, M. (2010) Forecasting macroeconomic time series with locally adaptive signal extraction. *International Journal of Forecasting*, 26, 312-325.

Glosten, L., Jagannathan, R. and Runkle, D. (1993) On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance*, 48(5), 1779-1801.

Hamori, S. and Tokihisa, A. (1997) Testing for a unit root in presence of a variance shift. *Economics Letters*, 57, 245-253.

Hess, G. D. and Iwata, S. (1997) Measuring and comparing business-cycle features. *Journal of Business and Economics and Statistics*, 15, 432-444.

Hillmer, S. C. and Tiao, G. C. (1982) An ARIMA-model-based approach to seasonal adjustment. *Journal of the American Statistical Association,* 77, 63-70.

Hodrick, R. J. and Prescott, E. C. (1997) Postwar U.S. business cycles: An empirical investigation. *Journal of Money, Credit and Banking*, 29, 1-16.

Kim, K. and Schmidt, P. (1993) Unit root tests with conditional heteroskedasticity. *Journal of Econometrics*, 59, 287-300.

Kim, T. H., Leybourne, S. and Newbold, P. (2002) Unit root tests with a break in innovation variance. *Journal of Econometric*s, 109, 365-387.

Nelson, B. (1991) Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, 59, 347-370.

Percival, D. B. and Walden, A. T. (1993) *Spectral analysis for physical applications. Multitaper and conventional univariate techniques*, Cambridge University Press, Cambridge.

Wallis, K. F. (1974) Seasonal adjustments and relations between variables. *Journal of the American Statistical Association*, 69, 18-31.

Zakoian, M. (1994) Threshold heteroscedastic models. *Journal of Economic Dynamics and Control*, 18, 931-955.

On the Probability Distribution of Economic Growth

Pär Stockhammar and Lars-Erik Öller

Department of Statistics, Stockholm University
S-106 91 Stockholm, Sweden
E-mail: par.stockhammar@stat.su.se

Abstract

Normality is often mechanically and without sufficient reason assumed in econometric models. In this paper three important and significantly heteroscedastic GDP series are studied. Heteroscedasticity is removed and the distributions of the filtered series are then compared to Normal, Normal Mixture and Normal - Asymmetric Laplace (NAL) distributions. NAL represents a skewed and leptokurtic distribution, which is in line with the Aghion and Howitt (1992) model for economic growth, based on Schumpeter's idea of creative destruction. Statistical properties of the NAL distributions are provided and it is shown that NAL competes well with the alternatives.

Keywords: The Aghion-Howitt model, asymmetric innovations, mixed Normal - Asymmetric Laplace distribution, Kernel density estimation, Method of Moments estimation.

1. Introduction

In the Schumpeterian world growth is driven endogenously by investments into R&D, leading to better products, which initially capture monopoly profits. The quality improvements occur randomly over time. The main contributions to endogenous growth are given by Romer (1986) and Lucas (1988).

They both argued that the underlying growth is determined by the accumulation of knowledge, with occasional setbacks. Other important papers on endogenous growth are: Segerstrom, Anant and Dinopoulos (1990), Grossman and Helpman (1991) and Aghion and Howitt (1992, henceforth AH).

The AH model is based on the Schumpeterian idea of creative destruction, i.e. the economy is driven by welfare augmenting better products (innovations, or shocks) and temporary declines (Schumpeter, 1942, Chapter 8). The expected rate of economic growth in AH is determined by the amount of research and its productivity. Innovations are assumed to arrive according to a Poisson process. To quote Aghion and Howitt (1998, p.54):

"When the amount n is used in research, innovations arrive randomly with a Poisson arrival rate λn, where $\lambda > 0$ is a parameter indicating the productivity of the research technology."

This was also assumed in e.g. Helpman and Trajtenberg (1994) and in Maliar and Maliar (2004). The latter study recognizes short waves, but unlike the present study neither accepts negative shocks. There are many real life examples that justify negative shocks, e.g. unsuccessful investments in physical or human capital, bad loans, losses when old investments become worthless and political conflicts. By negative (destructive) random shocks we try to mimic the setbacks in our reduced univariate approach. Moreover, all the models in the quoted works are specified in the time domain, while density distributions are the object of this study.

Aggregating a Poisson number of shocks (as assumed in the AH model) will lead to asymmetric distributions. It is found the growth series studied in this paper exhibit heteroscedasticity, which must be removed in order to enable a proper investigation of this hypothesis. This is done using the filter described in Stockhammar and Öller (2007). The filtered series are shown to be homoscedastic, and both skewness and leptokurtocity are found in the filtered (and unfiltered) series, rendering a hypothesis of normality dubious. Thus, data lend some support to the AH hypothesis.

In the Poisson process the time between each shock is exponentially distributed with intensity λ. Thus, the distribution closest to represent the AH model is the exponential distribution. In this study however, we have used the exponential distribution to describe the amplitude of the shocks. When

λ is small we expect infrequent but large shocks and vice versa. This intuitive way to describe the shocks accords well with modern economic theory. Specifically, to allow for negative or below average shocks, we have used the double exponential (Laplace) distribution obtained as the difference between two exponentially distributed variables with the same value on the parameter λ. The Laplace distribution is symmetric around its mean where the left tail describes below average shocks and vice versa. Due to the expected and empirically confirmed asymmetries in the GDP series we have allowed the exponential distribution to take different parameter values in the two tails, giving rise to the asymmetric Laplace (AL) distribution. This is one possible AH model representative. The asymmetric properties of the AL distribution have proved appealing for modeling currency exchange rates, stock prices, interest returns etc. see for instance Kozubowski and Podgorski (1999, 2000) and Linden (2001).

Another plausible explanation is that the long growth series have passed through alternating regimes over the years. Every such regime has its own normal distribution giving rise to a Normal Mixture (NM) distribution, which is our alternative hypothetical distribution[6]. The NM distribution, where skewness and leptokurtocity are introduced by varying the parameters, was used as early as the late nineteenth century by e.g. Pearson (1895).

It was found that the excess kurtosis in AL models is too large for the filtered (and unfiltered) growth series. The AL could therefore not be the only source of shocks, so Gaussian noise is added. AL distributed innovations are combined with normally distributed shocks leading to the weighted mixed Normal-AL (NAL) distribution. The NAL distribution is capable of generating a wide range of skewness and kurtosis, making the model very flexible. We also consider a convolution of the N and AL distributions. The parameters are estimated using the Method of Moments (MM).

This paper is organized as follows. The data are presented in Section 2 and the data preparation in Section 3. A model discussion together with the proposed model is the topic of Section 4. Section 5 contains the estimation set-up and a distributional accuracy comparison. Section 6 concludes.

[6]This could also be studied using regime switching models, but given the few observations, we did not pursue this idea.

2. The data

In this paper the important US GDPq (quarterly) 1947-2007, UK GDPq 1955-2007 and the compound GDP 1960-2007 series of the G7 countries[7] are studied as appearing on the websites of *Bureau of Economic Analysis* (www.bea.gov), *National Statistics* (www.statistics.gov.uk) and of *OECD* (www.oecd.org), respectively. All series are quarterly and seasonally adjusted, final figures, from which we form logarithmic differences, henceforth "growth" for short.

In order to accurately estimate the N, NM and NAL parameters, long series are required. The above series are the longest and most important quarterly GDP series available, and the G7 series is based on a large number of observations, albeit not as long as the US and UK ones. The first differenced logarithmic (Diff ln) series and their corresponding frequency distributions are shown in Figure 2.1. The frequency distributions are supplemented with the N distribution with the same mean and variance as those of the series, and an estimate of the Kernel density[8].

[7] Consists of Canada (1961-2007), France (1978-2007), Germany (1991-2007), Italy (1980-2007), Japan (1980-2007), UK (1960-2007) and US (1960-2007). The individual series have been scaled up/down to the price level year 2000, and the compound G7 series was calculated as their sum.

[8] The Kernel density estimate is defined as

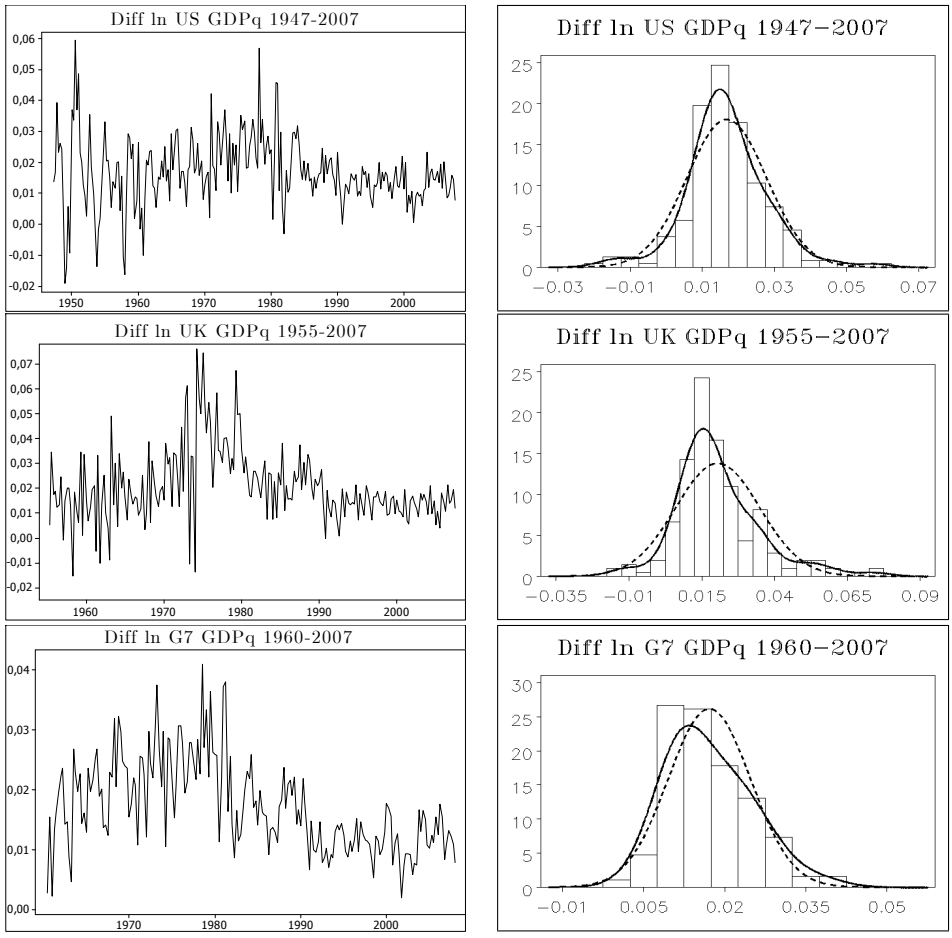
$$\widehat{f_h}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x_i - x}{h}\right)$$

where h is the bandwidth and $K(\cdot)$ is the Kernel function. In this study we have used the Gaussian Kernel, $K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}$, and the Silverman (1986) "Rule of Thumb" bandwidth

$$\widehat{h} = \left(\frac{4\widehat{\sigma}^5}{3n}\right)^{1/5} \approx 1.06\widehat{\sigma} n^{-1/5}$$

which is considered to be optimal when data are close to normal.

Figure 2.1: The Diff ln GDP series. The panels on the right hand show the frequency distributions of the data. The dashed line is the N distribution and the solid line is the Kernel density



The Diff ln series seem leptokurtic, especially the US and UK. This is also confirmed in Table 2.1 where the excess kurtosis in all cases exceeds zero. The results of the augmented Dickey-Fuller (ADF) test and the ARCH-LM test are also included.

Table 2.1: The moments and the results of unit root and homoscedasticity tests of the Diff ln US, UK and G7 GDP series

	US	UK	G7
Mean ($\widehat{\mu}$)	0.017	0.020	0.017
STD($\widehat{\sigma}$)	0.011	0.014	0.008
Skewness ($\widehat{\tau}$)	0.25	1.03	0.62
Kurtosis ($\widehat{\kappa}$)	2.14	2.26	0.06
ADF (p-value)	0.000	0.004	0.021
ARCH-LM (p-value)	0.000	0.000	0.001

The skewness is significantly nonzero in the UK and G7 series and thus lend some support to the AH hypothesis. Significant leptokurtocity appears in the US and UK series. The null hypothesis of a unit root using the ADF test is rejected for the Diff ln series, signalling stationarity. The ARCH-LM test indicates heteroscedasticity in all series.

To test the AH hypothesis, or more generally to compare distributions of heteroscedastic data, the heterocedasticity must be eliminated. The filter proposed in Stockhammar and Öller (2007) enables us to work with mean and variance stationary time series, and thus to make a fair comparison between the frequency distributions of the three series and various probability distributions, notably some asymmetric ones related to the Poisson distribution. The filtering procedure is described in the next section.

3. Data preparation

Heteroscedasticity is removed using the filter proposed in Stockhammar and Öller (2007):

$$\widetilde{z}_t = s_y \left[\frac{\left(z_t^{(i)} \right)^d}{HP^{(\gamma)} \left(\sqrt{\sum_{\tau=t-\nu}^{t+\nu} \left(z_\tau^{(i)} \right)^{2d} / 2\nu} \right)} \right] + \overline{y}, \qquad (3.1)$$

where $t = \max[k-\eta, l-\nu], \max[k-\eta+1, l-\nu+1], ...$ with k and l (both odd) as the window lengths in the numerator and denominator, respectively.

$\eta = (k-1)/2$, $\nu = (l-1)/2$, $\widetilde{z}_t$ is the filtered series and $i = a, b$ from the detrending operations

$$(a) \qquad z_t^{(a)} = \Delta y_t - \sum_{\tau=t-\eta}^{t+\eta} \Delta y_\tau \, /k \, , \qquad t = \eta+1, \eta+2, ..., n-\eta \qquad (3.2a)$$

and with y_τ delayed one period:

$$(b) \qquad z_t^{(b)} = \Delta y_t - \sum_{\tau=t-\eta}^{t+\eta} \Delta y_{\tau-1} \, /k \, , \quad t = \eta+2, \eta+3, ..., n-\eta+1 \qquad (3.2b)$$

where $\Delta y_t = y_t - y_{t-1}$, y_t is the logarithmic series at time t. The transformations in (3.2) are generalized by raising $z_t^{(i)}$ to the power d (which might not be an integer). Candelon and Gil-Alana (2004) concluded that the US and UK GDP series are integrated of order around $I(1.4)$. In this study, however, $d = 1$ has been used. Different degrees of integration can also be achieved using different values on η in (3.2). There are two extremes. For $\eta = (n-1)/2$, the term $\sum_{\tau=t-\eta}^{t+\eta} \Delta y_\tau \, /k$ equals $\overline{\Delta y}$. The other extreme appears when k equals one, that is $\eta = 0$. Operation (3.2b) is used only in the latter case and is equivalent to the second difference operation, $\Delta^2 y_t$. The best choice of η depends on the properties of the series studied. If the series is close to $I(1)$ then you should just choose η close to $(n-1)/2$, and if the series is close to $I(2)$ then choose $\eta = 0$ in (b) or a small value on η in (a). Stockhammar and Öller (2007) proposed using window length $k = l = 15$ (or $\eta = \nu = 7$) and the standard value used for quarterly data, $\gamma = 1\ 600$, see below.

$HP^{(\gamma)}$ in (3.1) is the Hodrick-Prescott (1997) filter designed to decompose a macroeconomic time series into a nonstationary trend component and a stationary cyclical residual. Given a seasonally adjusted time series x_t (in this case the time-varying variance), the decomposition into unobserved components is

$$x_t = g_t + c_t,$$

where g_t denotes the unobserved trend component at time t, and c_t the unobserved cyclical residual at time t. Estimates of the trend and cyclical components are obtained as the solution to the following minimization

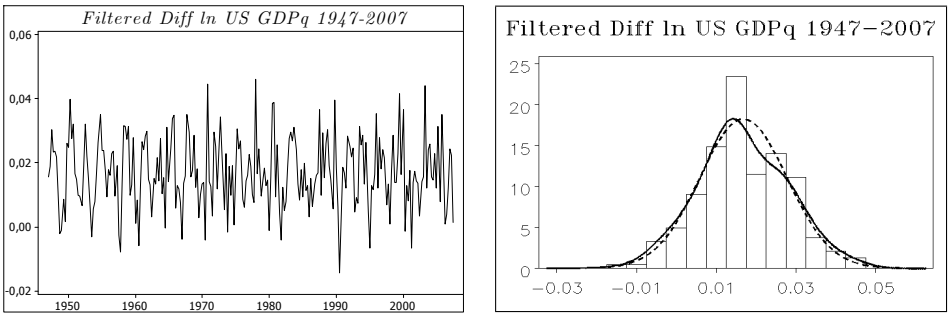
problem

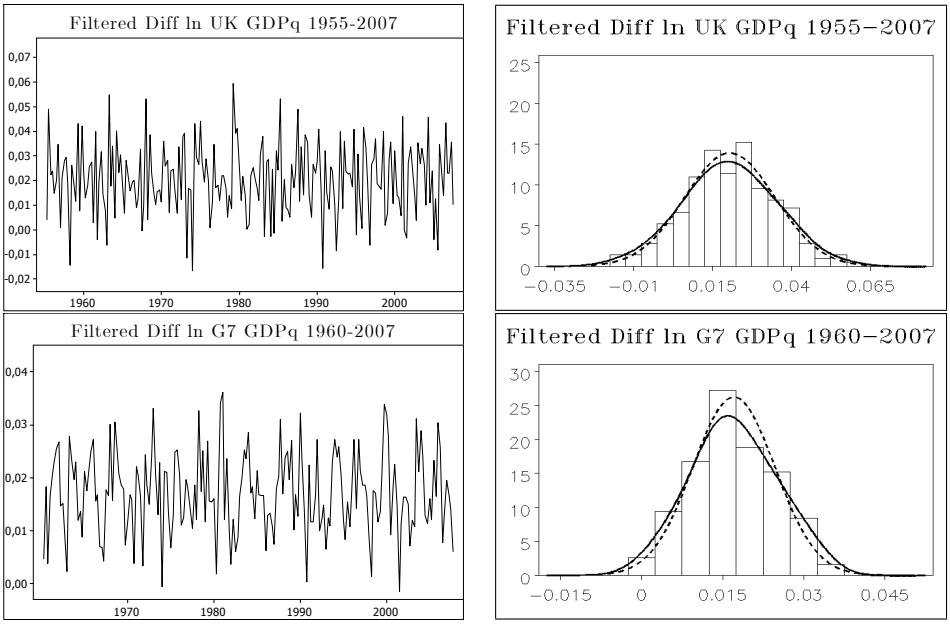
$$\min_{[g_t]_{t=1}^N} \left\{ \sum_{t=1}^{N} c_t^2 + \gamma \sum_{t=3}^{N} (\triangle^2 g_t)^2 \right\}, \qquad (3.3)$$

where $\triangle g_t = g_t - g_{t-1}$ and $g_{\min}$ is the HP-filter. The first sum of (3.3) accounts for the accuracy of the estimation, while the second sum represents the smoothness of the trend. The positive smoothing parameter γ controls the weight between the two components. As γ increases, the HP trend become smoother and vice versa. Note that the second sum, $(\triangle^2 g_t)$, is an approximation to the second derivative of g at time t. For quarterly data (the frequency used in most business-cycle studies) there seems to be a consensus in employing the value $\gamma = 1\ 600$ as originally suggested in Hodrick-Prescott (1997). In this study, the HP filter is applied in order to smooth the moving standard deviations.

Figure 3.1 shows the Diff ln US, UK and G7 GDP series after the heteroscedasticity filtering. The filtering was done using $k = l = 15$, $d = 1$ and $\gamma = 1\ 600$.

Figure 3.1: The heteroscedasticity filtered Diff ln GDP series. The right hand panels show the frequency distributions of the data. The dashed line is the N distribution with the same mean and variance as those of the series. The solid line is the Kernel density





The filter effects on the four moments of the three series can be seen in Table 3.1.

Table 3.1: Filter effects on the moments of the Diff ln US, UK and G7 GDP series. Period 1 represents the quarters 1947q1-1977q2 (US), 1955q1-1980q2 (UK) and 1960q1-1983q4 (G7). Period 2 contains 1977q3-2007q4 (US), 1980q3-2007q4 (UK) and 1984q1-2007q4 (G7)

	$\widehat{\mu}$	$\widehat{\mu}$	$\widehat{\sigma}$	$\widehat{\sigma}$	$\widehat{\tau}$	$\widehat{\tau}$	$\widehat{\kappa}$	$\widehat{\kappa}$
Period	1	2	1	2	1	2	1	2
$y_{t,US}$	0.018	0.016	0.013	0.008	−0.27	1.63	1.00	5.51
$\widetilde{z}_{t,US}$	0.017	0.017	0.011	0.011	−0.05	0.22	−0.30	0.42
$y_{t,UK}$	0.024	0.016	0.018	0.007	0.45	0.53	0.40	0.49
$\widetilde{z}_{t,UK}$	0.020	0.020	0.014	0.014	−0.11	−0.16	0.71	−0.35
$y_{t,G7}$	0.022	0.013	0.008	0.004	−0.06	0.51	0.16	0.40
$\widetilde{z}_{t,G7}$	0.017	0.017	0.008	0.008	−0.01	0.16	−0.31	−0.24

The filter succeeds in stabilizing the means ($\widehat{\mu}$) and the standard deviations ($\widehat{\sigma}$) of the series. The estimates of skewness ($\widehat{\tau}$) and excess kurtosis ($\widehat{\kappa}$) are

also more stable in the filtered series. Significant skewness and leptokurtocity were found in the entire $\widetilde{z}_{t,US}$ and $\widetilde{z}_{t,G7}$. In Stockhammar and Öller (2007) we showed that this filter does not distort white noise, and thus preserves the dynamics of the time series.

The unfiltered series in Figure 2.1 do not appear to be normal. Table 3.2 shows that the filter brings them closer to normality

Table 3.2: Filter effects on the normality of the Diff ln US, UK and G7 GDP series

	A-D	S-W	K-S	J-B
$y_{t,US}$	$***$	$***$	$***$	$***$
$\widetilde{z}_{t,US}$	$*$		$**$	$*$
$y_{t,UK}$	$***$	$***$	$***$	$***$
$\widetilde{z}_{t,UK}$				
$y_{t,G7}$	$***$	$***$	$***$	$***$
$\widetilde{z}_{t,G7}$				

In Table 3.2 $*$, $**$ and $***$ represent significance at the 10%, 5% and 1% levels, respectively, for the null hypothesis of normality. Four commonly used normality tests are reported, where A-D, S-W, K-S and J-B are the Anderson-Darling, Shapiro-Wilk, Kolmogorov-Smirnov and Jarque-Bera test, respectively. These tests are based on very different measures and can therefore lead to different conclusions.

The difficulty to reject normality is surprising if we take Figure 3.1 into consideration, where filtered data seem to have fatter tails than the normal distribution. Also, there is little trace of the significant skewness and leptokurtocity corroborated in the US and G7. According to e.g. Dyer (1974) the power of normality tests is generally low, especially for small samples. Note that the K-S, A-D and J-B statistics for the US reject the null hypotheses of normality. At least for the US series it seems worthwhile to see if there are other distributions that better fit the data. Considering the low power of the tests we will try the same for the UK and the G7 series. The normal distribution remains an alternative hypothesis.

4. Models for the shock distributions

In the AH model, endogenous growth is driven by creative destruction in which the underlying source is innovations, assumed to be the result of the stochastic arrival of new technologies modelled as a Poisson process. The arrival rate itself is affected by the share of the labour force engaged in research as well as by the Poisson probability of an innovation (research productivity). Each owner of a patent is assumed to have a temporary monopoly of the product lasting until it is replaced and destroyed by a better product.

AH specifies an entire simultaneous model. Our main hypothesis for growth is a reduced univariate form of the AH model, where both positive and *negative* shocks hit production exogenously. The drawback with this approach is that the origin of the innovations cannot be identified. The Poisson assumption in the AH model leads to asymmetric shock distributions, already confirmed in both filtered and unfiltered data, cf. Tables 2.1 and 3.1. In this section some asymmetric distributions are described, notably some related to the Poisson distribution.

With long time series there is a nonnegligible risk of distributional changes over time. One can argue that data have passed through a number of different regimes, not completely eliminated by filter (3.1). Every such regime could be N distributed but with different means and variances. The filtered US GDP in Figure 3.1 still shows a small hump in the right tail, which may indicate that the data are characterized by at least two regimes, each one N distributed. Given the relatively few observations, the number of regimes is here restricted to two. Moreover, the homoscedasticity test did not detect non-constancy of variances, so even two regimes with different variances could be hard to detect. The introduction of different means and variances for the regimes render it possible to introduce skewness and excess kurtosis in the NM distribution. The probability distribution function (pdf) of the NM distribution is:

$$f_{NM}(\widetilde{z}_t;\boldsymbol{\theta}) = \frac{w}{\sigma_1\sqrt{2\pi}}\exp\left\{-\frac{(\widetilde{z}_t-\mu_1)^2}{2\sigma_1^2}\right\}+\frac{1-w}{\sigma_2\sqrt{2\pi}}\exp\left\{-\frac{(\widetilde{z}_t-\mu_2)^2}{2\sigma_2^2}\right\}, \quad (4.1)$$

where $\boldsymbol{\theta}$ consists of the parameters $(w,\mu_1,\mu_2,\sigma_1,\sigma_2)$ and where $0 \leq w \leq 1$ is the weight parameter.

The distribution closest to represent the AH model is the exponential distribution. In order to allow for negative or below average shocks this distribution is modified by simply assuming that growth is driven by a process, which is the sum of two (one positive, one negative) exponentially distributed random shocks. If these shocks have the same mean we arrive at the Laplace (L) distribution with pdf:

$$f_L(\widetilde{z}_t;\boldsymbol{\theta}) = \frac{1}{2\phi}\exp\left\{-\frac{|\widetilde{z}_t-\mu|}{\phi}\right\}, \qquad (4.2)$$

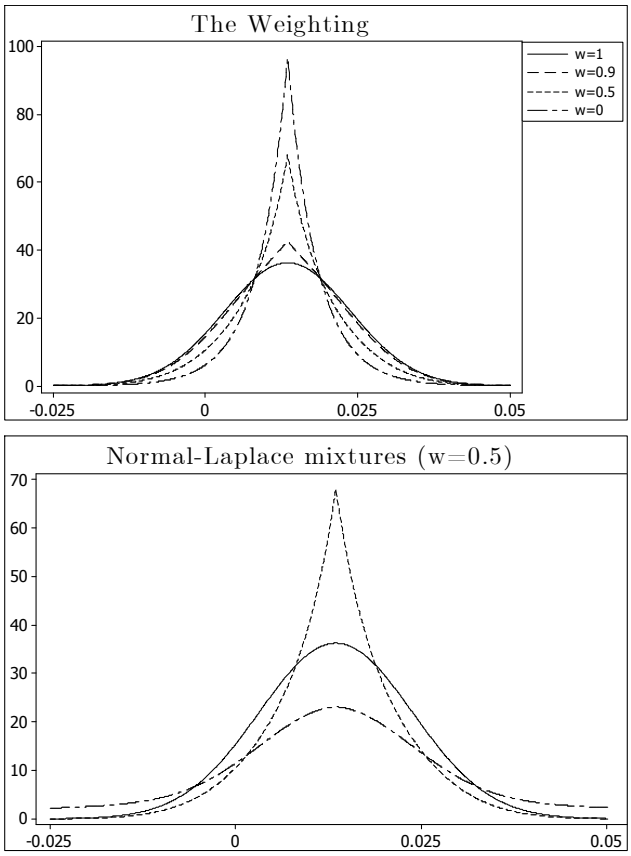
where $\boldsymbol{\theta} = (\mu,\phi)$, $\mu \in \mathbb{R}$ is the location parameter and $\phi > 0$ is the scale parameter. The L distribution (which is sometimes also called the double exponential distribution) has been used in many fields: engineering, finance, electronics etc, see Kotz et al. (2001), and the references therein. The L distribution is symmetric around its mean (μ) with $Var(y) = 2\phi^2$ and excess kurtosis $\kappa = 3$. It has fatter tails than the N distribution, but it lacks an explicit shape parameter, making it rather inflexible. Also, the excess kurtosis is restricted to the constant value (3), no matter what the kurtosis in the data. Table 3.1 shows the kurtosis in Laplace variables is way too large for the filtered growth series in this study ($\widehat{\kappa} = 0.07$ for the US, $\widehat{\kappa} = 0.19$ for the UK and $\widehat{\kappa} = -0.29$ for the G7). Clearly, the L distribution cannot alone explain the data.

The L distribution can, however, be modified by allowing it to have a second stochastic component in the sense that its empirical counterpart is buried in Gaussian noise. We thus combine (4.2) with a N distribution via a weight parameter w. This mixture was introduced by Kanji (1985) to fit wind shear data using the Normal - Laplace (NL) mixture distribution specified by:

$$f_{NL}(\widetilde{z}_t;\boldsymbol{\theta}) = \frac{w}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{(\widetilde{z}_t-\mu)^2}{2\sigma^2}\right\}+\frac{(1-w)}{2\sigma}\exp\left\{-\frac{|\widetilde{z}_t-\mu|}{\sigma}\right\}, \qquad (4.3)$$

for $-\infty < \widetilde{z}_t < \infty$ and for the parameters: $-\infty < \mu < \infty$, $0 \leq w \leq 1$ and $\sigma > 0$. In (4.3) the N and L distributions have the same mean. Jones and McLachlan (1990) generalized (4.3) and showed that this may lead to an even better fit than Kanji's. The characteristics of the NL density are shown in Figure 4.1.

Figure 4.1: NL densities



The upper panel in Figure 4.1 shows different weightings of the two components in the NL distribution (with μ=0.017, σ=0.011, ϕ=0.005). The solid line in the lower panel shows the pure $N(0.017,0.011)$ distribution together with two mixture distributions with w=0.5, ϕ=0.05 (dashed/dotted line) and ϕ=0.005 (dashed line), respectively.

The above L and NL mixture distributions do not account for skewness in the data making them poor AH model representatives. A suitable skewed generalization of the L distribution is presented in McGill (1962) who proposes an asymmetric Laplace (AL) distribution of the form

$$f_{AL}(\widetilde{z}_t;\boldsymbol{\theta}) = \begin{cases} \frac{1}{2\psi}\exp\left\{\frac{\widetilde{z}_t-\mu}{\psi}\right\} & \text{if } \widetilde{z}_t \leq \mu \\ \frac{1}{2\phi}\exp\left\{\frac{\mu-\widetilde{z}_t}{\phi}\right\} & \text{if } \widetilde{z}_t > \mu \end{cases}, \qquad (4.4)$$

where again μ is the location parameter, for which the median is the Maximum Likelihood (ML) estimate, and $\boldsymbol{\theta} = (\mu, \phi, \psi)$. This distribution is negatively skewed if $\psi > \phi$, and vice versa for $\psi < \phi$. If $\psi = \phi$ the AL collapses to the L distribution. In AL, ψ is the parameter of shocks weaker than the trend and ϕ that of stronger shocks than the trend. If $\psi \neq \phi$ then Schumpeterian shocks that lead to weaker than trend growth behave differently from stronger growth shocks. During the last couple of decades, various forms and applications of AL distributions have appeared in the literature, see Kotz et al. (2001) for an exposé. Linden (2001) used an AL distribution to model the returns of 20 stocks, where ψ and ϕ were shown to be highly significant. Another recent paper is Yu and Zhang (2005) who used a three-parameter AL distribution to fit flood data.

An advantage of the AL distribution is that, unlike the L distribution, the kurtosis is not fixed. The AL distribution is even more leptokurtic than the L distribution with an excess kurtosis that varies between three and six (the smallest value for the L distribution, and the largest value for the exponential distribution). Another advantage of the AL distribution is that it is skewed (for $\psi \neq \phi$).

Because of the large leptokurtocity of the AL distribution, we will add Gaussian noise. To the authors´ best knowledge this distribution has not been used before for macroeconomic time series data. We assume that each shock is an independent drawing from either a N or an AL distribution. The probability density distribution of the filtered growth series ($\widetilde{z}_t$) can then be described by a weighted sum of N and AL random shocks, i.e:

$$f_{NAL}(\widetilde{z}_t; \boldsymbol{\theta}) = \frac{w}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(\widetilde{z}_t - \mu)^2}{2\sigma^2}\right\} + (1-w) \begin{cases} \frac{1}{2\psi} \exp\left\{\frac{\widetilde{z}_t - \mu}{\psi}\right\} & \text{if } \widetilde{z}_t \leq \mu \\ \frac{1}{2\phi} \exp\left\{\frac{\mu - \widetilde{z}_t}{\phi}\right\} & \text{if } \widetilde{z}_t > \mu \end{cases}, \quad (4.5)$$

where $\boldsymbol{\theta}$ consists of the five parameters $(w, \mu, \sigma, \phi, \psi)$. Equation (4.5) is referred to as the mixed Normal - Asymmetric Laplace (NAL) distribution and is our main AH model representative. Note that (as in Jones and McLachlan, 1990) equal medians, but unequal variances, are assumed for the components in the proposed distribution. It has a jump discontinuity at μ when $\psi \neq \phi$, see Figure 4.2. Looking at the smoothed empirical distributions in Figure

3.1, the discontinuity seems counterintuitive. However, the histograms in Figure 3.1 lend some support to a jump close to μ. The cumulative distribution function (cdf) of 4.5 is given in the appendix. Figure 4.2 shows NAL densities for three different values of the weight parameter w.

Figure 4.2: NAL densities

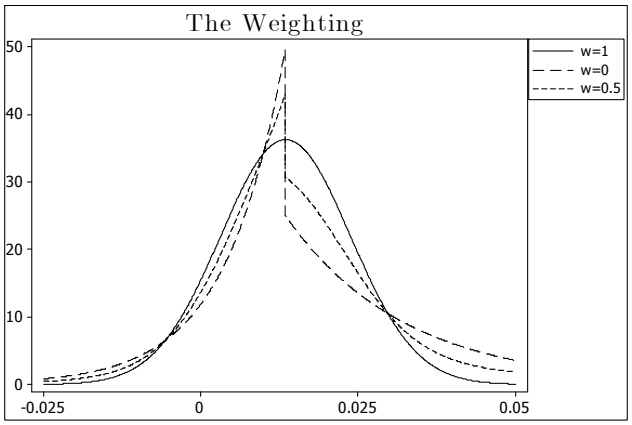


Figure 4.2 shows a pure $N(0.017,0.011)$, an $AL(\psi{=}0.01,\phi{=}0.02$ and $\mu{=}0.017)$ distribution ($w{=}1$ and $w{=}0$ respectively) and a compound of these two components with $w{=}0.5$. Note the discontinuity at μ.

To avoid the discontinuity in μ, we may also assume that each shock is a random mixture of a N and an AL distributed component. We then arrive at the convoluted version suggested by Reed and Jorgensen (2004). Instead of using the AL parameterization in (4.4) they used:

$$f_{AL^*}(\widetilde{z}_t;\boldsymbol{\theta}) = \begin{cases} \frac{\alpha\beta}{\alpha+\beta}\exp\{\beta\widetilde{z}_t\} & \text{if } \widetilde{z}_t \leq 0 \\ \frac{\alpha\beta}{\alpha+\beta}\exp\{-\alpha\widetilde{z}_t\} & \text{if } \widetilde{z}_t > 0 \end{cases} \tag{4.6}$$

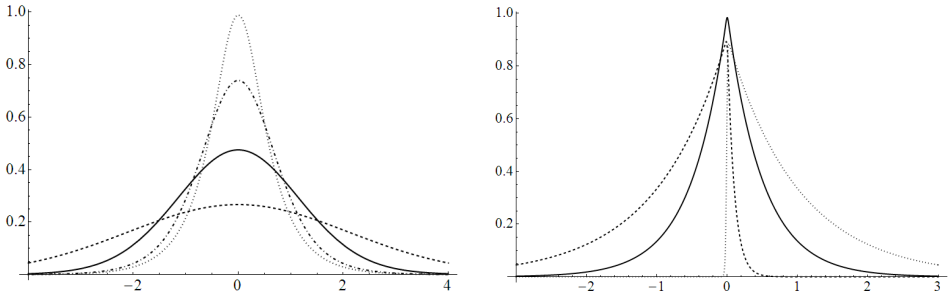
which was convoluted with a N distribution giving the following pdf:

$$f_{c-NAL}(\widetilde{z}_t;\boldsymbol{\theta}) = \frac{\alpha\beta}{\alpha+\beta}\phi\left(\frac{\widetilde{z}_t-\mu}{\sigma}\right)\left[R\left(\alpha\sigma-\frac{\widetilde{z}_t-\mu}{\sigma}\right)+R\left(\beta\sigma+\frac{\widetilde{z}_t-\mu}{\sigma}\right)\right], \tag{4.7}$$

where $\boldsymbol{\theta}=(\alpha,\beta,\mu,\sigma)$ and $R(z) = \frac{\Phi^c(z)}{\phi(z)}$ is the Mill´s ratio. Ibid. called

this distribution the four parameter Normal Laplace distribution, but it is here called the convoluted NAL, c-NAL. As can be seen from Figure 4.3 this distribution lacks the jump in the mode.

Figure 4.3: c-NAL densities.



The left panel in Figure 4.3 shows c-NAL densities using $(\alpha,\beta)=(2,2)$, $\mu=0$ and $\sigma=1/4$ (dotted line), $\sigma=1/2$ (dashed/dotted), $\sigma=1$ (solid) and $\sigma=2$ (dashed). The right panel shows c-NAL densities using $\mu=0$, $\sigma=0.01$, $(\alpha,\beta)=(10,1,$ dashed line), $(1,10,$ dotted) and $(\alpha,\beta)=(2,2,$ solid), respectively.

The c-NAL distribution has the advantage of being more parsimonious than NAL. Whether it is more suitable to describe the probability distribution of economic growth is the issue of the next section.

5. Estimation and distributional accuracy

In this section we will fit all four distributions (N, NM, NAL and c-NAL) in order to find out which one best describes the data. The five parameters in the NM distribution (4.1) will be estimated using the method of moments (MM) for the first four moments of the same. A close distributional fit is important in density forecasting. As noted by e.g. Fryer and Robertson (1972), the method of maximum likelihood might break down for this distribution. Unfortunately the choice of MM excludes model selection criteria such as the AIC and BIC. This is compensated by elaborating distributional comparisons using several statistical techniques. The noncentral moments of (4.1) are given in the appendix. Equating the theoretical and the observed first

four moments using the five parameters yields infinitely many solutions[9]. A way around this dilemma is to estimate μ_1 by i.e. the observed mode, here approximated by the maximum value of the Kernel function estimator of the empirical distribution ($\max f_K(\widetilde{z}_i)$). In the presence of positive skewness we can expect μ_1 to be smaller than μ, and vice versa. Here $\widehat{\mu}_{1,US}$, $\widehat{\mu}_{1,UK}$ and $\widehat{\mu}_{1,G7}$ are substituted for $\max f_K(\widetilde{z}_{t,US}) = 0.0142$, $\max f_K(\widetilde{z}_{t,UK}) = 0.0196$ and $\max f_K(\widetilde{z}_{t,G7}) = 0.0159$. The observed moments and the corresponding MM parameter estimates for the filtered series (using the above values for $\mu_{1,US}$, $\mu_{1,UK}$ and $\mu_{1,G7}$) are given in Table 5.1.

Table 5.1: Sample moments and estimated parameters of the NM assumption

	Sample noncentral moments				Estimated NM parameters			
	$E(\widetilde{z}_t)$	$E(\widetilde{z}_t^2)$	$E(\widetilde{z}_t^3)$	$E(\widetilde{z}_t^4)$	$\widehat{w}$	$\widehat{\mu}_2$	$\widehat{\sigma}_1$	$\widehat{\sigma}_2$
US	0.0168	0.0004	$1.1 \cdot 10^{-5}$	$3.4 \cdot 10^{-7}$	0.833	0.030	0.010	0.005
UK	0.0205	0.0006	$2.1 \cdot 10^{-5}$	$8.1 \cdot 10^{-7}$	0.919	0.030	0.015	0.011
G7	0.0171	0.0003	$8.0 \cdot 10^{-6}$	$2.0 \cdot 10^{-7}$	0.866	0.024	0.008	0.006

The likelihood function, $L(\boldsymbol{\theta})$, of the NAL distribution is:

$$
\begin{aligned}
L(\boldsymbol{\theta}) \;=\; \prod_{t=1}^{n} \Bigg[& w \left(\left(2\pi\sigma^2\right)^{-1/2} \exp\left(\frac{\left(\widetilde{z}_t - \mu\right)^2}{2\sigma^2} \right) \right) \\
& + (1-w) \begin{pmatrix} (2\psi)^{-1} \exp\left(\frac{1}{2\psi}\left(\widetilde{z}_t - \mu\right) \right) I\left(\widetilde{z}_t \leq E(\widetilde{z}_t)\right) \\ (2\phi)^{-1} \exp\left(\frac{1}{2\phi}\left(\mu - \widetilde{z}_t\right) \right) I\left(\widetilde{z}_t > E(\widetilde{z}_t)\right) \end{pmatrix} \Bigg],
\end{aligned}
$$

where I is the indicator function. The ML estimates might be found by numerical optimization of the above likelihood function. In order to enable a fair comparison with the NM and the c-NAL distributions, the parameters are again estimated using MM. The formulae of the noncentral and central moments of (4.5) are given in the appendix. There are five parameters and only four moment conditions, so again equating the theoretical and the observed first four moments will not give a unique solution. We now estimate

[9] We tried to make use of the fifth moment, but in none of the series did it even at a 10% significance level differ from zero. For convenience, also the fifth moment is included in the appendix.

μ by the ML estimate of μ in the AL distribution, that is the observed median, $\widehat{md}$. Here $\widehat{\mu}_{US}$, $\widehat{\mu}_{UK}$ and $\widehat{\mu}_{G7}$ are substituted for $\widehat{md}_{US} = 0.0156$, $\widehat{md}_{UK} = 0.0203$ and $\widehat{md}_{G7} = 0.0167$. The parameter values that satisfy the moment conditions are:

Table 5.2: Estimated NAL parameters

	Estimated parameters			
	$\widehat{w}$	$\widehat{\sigma}$	$\widehat{\psi}$	$\widehat{\phi}$
US	0.711	0.012	0.006	0.014
UK	0.828	0.015	0.018	0.017
G7	0.939	0.008	0.008	0.020

Table 5.2 shows that the Gaussian noise component dominates. In the US and G7 series $\widehat{\psi}$ is much smaller than $\widehat{\phi}$, which indicates that growth shocks that are weaker than trend have a smaller spread than above trend shocks. Together with a mean growth larger than zero this ensures long-term economic growth.

Reed and Jorgensen (2004) provided some guidelines on how to estimate the c-NAL parameters in (4.7) using ML techniques. To be consistent and to make fair comparisons, the parameters are here again estimated using MM. Ibid. also supplied the first four cumulants, and in order to find the MM parameter estimates we provide the first four noncentral moments in the appendix. Note that the c-NAL distribution has four parameters so there is no need to fix one of the parameters to find a unique MM solution. The MM estimates are given in Table 5.3.

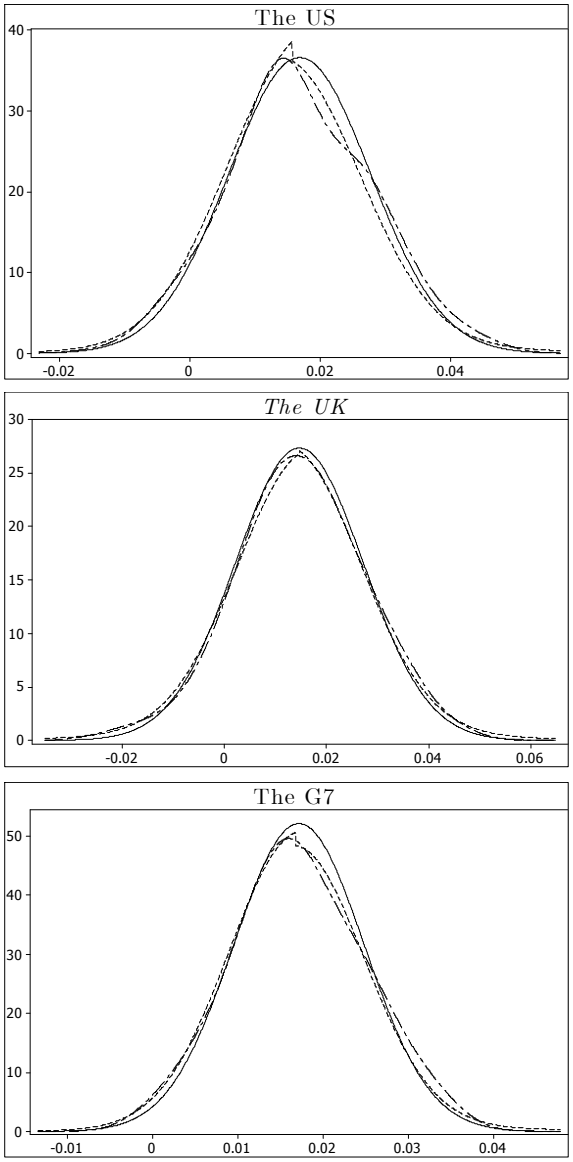
Table 5.3: Estimated c-NAL parameters

	Estimated parameters			
	$\widehat{\mu}$	$\widehat{\sigma}$	$\widehat{\alpha}$	$\widehat{\beta}$
US	0.017	0.009	197.0	198.0
UK	0.021	0.010	140.1	136.0
G7	0.017	0.007	498.2	498.8

Figure 5.1 shows the estimated NAL distributions together with the benchmark N distributions and the smoothed empirical GDP series. The graphs,

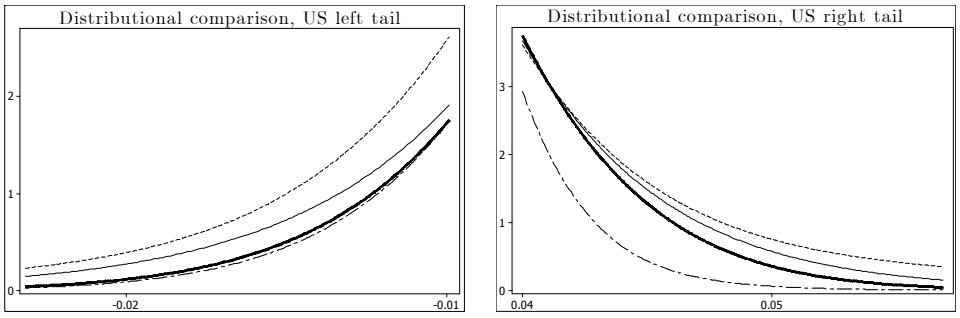
especially of the US distribution, reveal the discontinuity at $\widehat{\mu}$, resulting in a poignant peak, as discussed in connection with Figure 4.2.

Figure 5.1: Distributional comparison of the N (solid line) and NAL (dashed line) distributions to the Kernel distribution (dashed/dotted line)



The NM and c-NAL distributions were omitted in Figure 5.1 in order to improve readability. A left and right tail distributional comparison of all four distributions for the US are shown in Figure 5.2. The tail distributions in the UK and, especially G7, are very similar.

Figure 5.2: Distributional comparison of the US left and right tail of the N (thick solid line), NM (dashed/dotted), NAL (dashed) and c-NAL (thin solid) distributions



The NAL distribution thus has fatter tails than the alternatives.

Table 5.4 shows the distributional fit for the three hypotheses. We have chosen to compare the ordinates of the empirical distributions and the hypothetical N, NM, NAL and c-NAL distribution for 1 000 equidistant points, however dropping points outside the interval $(\widehat{\mu} - 4\widehat{\sigma}, \widehat{\mu} + 4\widehat{\sigma})$. Four accuracy measures are used.

The *Root Mean Square Error, RMSE*, is here defined as:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{1\ 000} \left[f_K\left(\widetilde{z}_i\right) - \widehat{f}\left(\widetilde{z}_i\right) \right]^2}{1\ 000}},$$

where $f_K\left(\widetilde{z}_i\right)$ is the Kernel density estimator, and $\widehat{f}\left(\widetilde{z}_i\right)$ is the hypothetical distribution. The sum is taken over ordinates of equidistant points on the horizontal axis, and hence there are more points where the distributions are almost parallel to the x-axis, providing these points with more weight. The peak to the left of the median considerably affects RMSE for the US data.

Percentage error measures are widely used but they also have their disadvantages. They are undefined at $f_K(\widetilde{z}_i) = 0$, and they have a very skewed distribution for $f_K(\widetilde{z}_i)$ close to zero. The *Median Absolute Percentage Error measure, MdAPE* is defined as:

$$MdAPE = median\left(\frac{100 \times \left|f_K(\widetilde{z}_i) - \widehat{f}(\widetilde{z}_i)\right|}{f_K(\widetilde{z}_i)}\right),$$

it is because of the asymmetry, a better measure than its close relative, the *Mean Absolute Percentage Error, MAPE*[10]. Yet an advantage over MAPE is that positive errors are not counted heavier than negative ones. This is the reason why so-called "symmetric" measures have been suggested (Makridakis, 1993). One is the *Symmetric Median Absolute Percentage Error, sMdAPE*

$$sMdAPE = median\left(\frac{200 \times \left|f_K(\widetilde{z}_i) - \widehat{f}(\widetilde{z}_i)\right|}{f_K(\widetilde{z}_i) + \widehat{f}(\widetilde{z}_i)}\right).$$

Hyndman and Koehler (2006) suggested the *Mean Absolute Scaled Error, MASE,* defined as:

$$MASE = \frac{1}{1\ 000}\left|\frac{f_K(\widetilde{z}_i) - \widehat{f}(\widetilde{z}_i)}{\frac{1}{999}\sum_{i=2}^{1\ 000}\left|f_K(\widetilde{z}_i) - f_K(\widetilde{z}_{i-1})\right|}\right|.$$

Ibid. showed that this measure is widely applicable and less sensitive to outliers and small samples than the other measures.

All the above four measures are reported in Table 5.4.

[10] defined as: $\frac{100}{n} \times \sum_{i=1}^{n}\left|f_K(\widetilde{z}_i) - \widehat{f}(\widetilde{z}_i)\right| / f_K(\widetilde{z}_i))$.

Table 5.4: Distributional accuracy comparison

		RMSE	MdAPE	sMdAPE	MASE
US	$N(0.0168, 0.0111)$	1.76	16.65	16.73	15.38
US	NM	**1.41**	13.98	13.19	15.59
US	NAL	1.73	**9.07**	**9.12**	**12.68**
US	c-NAL	1.59	14.25	14.43	14.11
UK	$N(0.0205, 0.0145)$	0.65	13.00	13.63	9.46
UK	NM	0.59	15.46	14.99	8.66
UK	NAL	**0.45**	**9.47**	**9.50**	**6.77**
UK	c-NAL	0.58	15.02	14.80	7.99
G7	$N(0.0171, 0.0077)$	2.21	18.69	20.13	14.60
G7	NM	1.64	11.07	10.55	11.78
G7	NAL	**1.20**	**8.65**	**8.99**	**9.00**
G7	c-NAL	1.70	12.27	11.61	12.55

The NAL distribution using the parameter values in Table 5.2 is superior to the N, NM and the c-NAL distribution according to every measure, except RMSE for the US where, as expected, the relatively large discontinuity peak has a large impact on the measure. NAL shows on average 27.6%, 29.2% and 48.3% better fit for the US, UK and G7, respectively (comparing with the benchmark N distribution). Comparing to the NM distribution, NAL is an improvement with on average 15.5% , 30.2% and 21.7% for the US, UK and G7. Finally, the NAL is on average a 18.8%, 27.6% and a 27.4% improvement over the c-NAL distribution. According to this numerical comparison, the US, UK and G7 GDP series could be looked upon as samples from a NAL distribution (as representing the AH model) with the parameter estimates in Table 5.2. In other words, the AH hypothesis of economic growth could be correct, if we accept that shocks are either AL (Poisson) *or* N distributed, with N dominating.

Kernel estimation is based on subjective choices both of function and of bandwidth. But so are goodness of fit tests and it is well known that tests based on both approaches have low power. To be on safer ground we have chosen also to test the histograms using three different numbers of bins. We then get unique critical values for these tests which enable calculation of p-values. Table 5.5 reports the p-values of the χ^2 tests using 10, 15 and 20 bins when testing the null hypotheses $H_{0,1} : \widetilde{z}_t \sim$N, $H_{0,2} : \widetilde{z}_t \sim$NM,

$H_{0,3} : \widetilde{z}_t \sim$NAL and $H_{0,4} : \widetilde{z}_t \sim$c-NAL. The number of degrees of freedom are given in parentheses.

Table 5.5: The χ^2 goodness of fit test

	Bins(df.)	US	UK	G7
$H_{0,1}$: $\widetilde{z}_t \sim$ N	10(7)	0.01	0.78	0.09
	15(12)	0.00	0.61	0.15
	20(17)	0.01	0.17	0.10
$H_{0,2}$: $\widetilde{z}_t \sim$ NM	10(4)	0.01	0.41	0.02
	15(9)	0.00	0.09	0.05
	20(14)	0.02	0.27	0.01
$H_{0,3}$: $\widetilde{z}_t \sim$ NAL	10(4)	0.03	0.38	0.10
	15(9)	0.04	0.45	0.17
	20(14)	0.11	0.13	0.10
$H_{0,4}$: $\widetilde{z}_t \sim$ c-NAL	10(5)	0.01	0.48	0.07
	15(10)	0.01	0.59	0.13
	20(15)	0.02	0.70	0.07

While the power of these tests is low, they still indicate that the NAL distribution fits US and G7 growth best whereas the N and c-NAL distributions are most accurate for the UK.

6. Conclusions

The hypothesis that economic growth could be described by a reduced and modified AH model is not strongly contradicted by data. Asymmetries arises from the AH model, where innovations are assumed to arrive according to a Poisson process. The Laplace and asymmetric Laplace distributions are unable to describe the asymmetric and just slightly leptokurtic shape. A mixed Normal - Asymmetric Laplace (NAL) distribution is introduced and is shown to better describe the density distribution of growth than the Normal, Normal Mixture, convoluted NAL and Laplace distributions. This paper thus, from a new angle, supports the hypothesis that innovations arriving according to a Poisson process play an important role in economic growth, as

suggested by e.g. Helpman and Trajtenberg (1994) and Maliar and Maliar (2004). According to the AH hypothesis, λ measures the intensity of only positive shocks (research productivity). Thus, our technique provides a way to estimate related quantities, and perhaps to compare different economies.

The mean, variance, skewness and the fatness of the tails stand in relation to the five parameters in the NAL distribution, and the parameters are estimated using MM on the first four moments. The moment generating function and the first four central and noncentral moments of the NAL distribution are provided. Because of the close distributional fit, the NAL distribution is a good choice for density forecasting of GDP growth series, or for that matter of any series with these features. The NAL distribution could also be used for conditional density forecasts applying priors on the parameters ϕ and ψ, but that merits another study.

Acknowledgments

This research was supported by the Department of Statistics at Stockholm University, Royal Swedish Academy of Sciences, the International Institute of Forecasters and by the Societas Scientiarum Fennica. We gratefully acknowledge helpful comments from Daniel Thorburn of Stockholm University, Mattias Villani of the Swedish Riksbank and from Roy Batchelor of Cass Business School, London. Parts of this paper have been presented at the International Symposium on Forecasting in 2007 and 2008 and also at various other venues. We are grateful for the many suggestions from seminar participants.

Appendix, Theoretical moments

The noncentral moments of the Normal mixture distribution (4.1) are given by

$$\begin{aligned} E(Y^n) &= \frac{w}{\sigma_1\sqrt{2\pi}}\sum_{k=0}^{n}\binom{n}{k}\left\{1+(-1)^k\right\}\mu_1^{n-k}2^{(k-1)/2}\sigma_1^{k+1}\Gamma\left(\frac{k+1}{2}\right)+ \\ &\quad \frac{1-w}{\sigma_2\sqrt{2\pi}}\sum_{k=0}^{n}\binom{n}{k}\left\{1+(-1)^k\right\}\mu_2^{n-k}2^{(k-1)/2}\sigma_2^{k+1}\Gamma\left(\frac{k+1}{2}\right) \end{aligned}$$

and specifically the first five moments are

$$\begin{aligned} E(Y) &= w\mu_1+(1-w)\,\mu_2 \\ E(Y^2) &= w\left(\mu_1^2+\sigma_1^2\right)+(1-w)\left[\mu_2^2+\sigma_2^2\right] \\ E(Y^3) &= w\left(\mu_1^3+3\mu_1\sigma_1^2\right)+(1-w)\left[\mu_2^3+3\mu_2\sigma_2^2\right] \\ E(Y^4) &= w\left(\mu_1^4+6\mu_1^2\sigma_1^2+3\sigma_1^4\right)+(1-w)\left[\mu_2^4+6\mu_2^2\sigma_2^2+3\sigma_2^4\right] \\ E(Y^5) &= w\left(\mu_1^5+10\mu_1^3\sigma_1^2+15\mu_1\sigma_1^4\right)+(1-w)\left[\mu_2^5+10\mu_2^3\sigma_2^2+15\mu_2\sigma_2^4\right]. \end{aligned}$$

The cdf of the NAL distribution (4.5) is given by:

$$F\left(\widetilde{z}_t;\boldsymbol{\theta}\right)=\begin{cases} w\Phi\left(\frac{\widetilde{z}_t-\mu}{\sigma}\right)+\frac{1-w}{2}\exp\left\{\frac{\widetilde{z}_t-\mu}{\psi}\right\} & \text{if } \widetilde{z}_t\leq\mu \\ w\Phi\left(\frac{\widetilde{z}_t-\mu}{\sigma}\right)+(1-w)\left(1-\frac{1}{2}\exp\left\{\frac{\widetilde{z}_t-\mu}{\phi}\right\}\right) & \text{if } \widetilde{z}_t>\mu \end{cases},$$

where $\Phi\left(\cdot\right)$ denotes the cdf of the standard normal distribution.

The noncentral moments of (4.5) are given by

$$\begin{aligned} E(Y^n) &= \frac{w}{\sigma\sqrt{2\pi}}\sum_{k=0}^{n}\binom{n}{k}\left\{1+(-1)^k\right\}\mu^{n-k}2^{(k-1)/2}\sigma^{k+1}\Gamma\left(\frac{k+1}{2}\right)+ \\ &\quad (1-w)\left[\frac{1}{2\psi}\sum_{k=0}^{n}\binom{n}{k}(-1)^k\mu^{n-k}\psi^{k+1}k!+\frac{1}{2\phi}\sum_{k=0}^{n}\binom{n}{k}\mu^{n-k}\phi^{k+1}k!\right]. \end{aligned}$$

Specifically,

$$
\begin{aligned}
E(Y) &= w\mu + (1-w)\left[\mu + \frac{\phi-\psi}{2}\right] \\
E(Y^2) &= w\left(\mu^2+\sigma^2\right) + (1-w)\left[\mu^2 + \psi(\psi-\mu) + \phi(\phi+\mu)\right] \\
E(Y^3) &= w\left(\mu^3+3\mu\sigma^2\right) + (1-w)\left[\mu^3 + \frac{3\psi}{2}\left(2\mu\psi - 2\psi^2 - \mu^2\right) + \right. \\
&\quad \left. \frac{3\phi}{2}\left(\mu^2 + 2\mu\phi + 2\phi^2\right)\right] \\
E(Y^4) &= w\left(\mu^4 + 6\mu^2\sigma^2 + 3\sigma^4\right) + (1-w)\left[2\psi\left(3\mu^2\psi - 6\mu\psi^2 + 6\psi^3 - \mu^3\right) + \right. \\
&\quad \left. 2\phi\left(\mu^3 + 3\mu^2\phi + 6\mu\phi^2 + 6\phi^3\right) + \mu^4\right] \\
E(Y^5) &= w\left(\mu^5 + 10\mu^3\sigma^2 + 15\mu\sigma^4\right) + (1-w)\left[\frac{5\psi}{2}\left(4\mu^3\psi - 12\mu^2\psi^2 + 24\mu\psi^3 - 24\psi^4 - \mu^4\right) + \right. \\
&\quad \left. \frac{5\phi}{2}\left(\mu^4 + 4\mu^3\phi + 12\mu^2\phi^2 + 24\mu\phi^3 + 24\phi^4\right) + \mu^5\right]
\end{aligned}
$$

Note that the mean of the NAL distribution can be written as

$$
E(Y) = \mu + (1-w)\left[\frac{\phi-\psi}{2}\right],
$$

which clearly shows the obvious fact that positively skewed distributions ($\phi > \psi$) will have a mean larger than μ and vice versa. The central moments are

$$
\begin{aligned}
E(Y) &= 0 \\
Var(Y) &= w\sigma^2 + (1-w)\left[\frac{\phi\psi(1-w)}{2} + \frac{\left(\phi^2+\psi^2\right)(3+w)}{4}\right] \\
\tau(Y) &= \frac{1}{4}\left[7\phi^3 + 3\phi^2\psi - 3\phi\psi^2 - 7\psi^3\right] - \frac{3w(\phi-\psi)}{4}\left[(\phi+\psi)^2 + 2\sigma^2\right] \\
&\quad -\frac{w^2(\phi-\psi)}{4}\left[(\phi+\psi)^2 - 2\sigma^2\right] - \frac{w^3(\phi-\psi)^3}{4} \\
\kappa(Y) &= \frac{3}{16}\left[39\phi^4 + 20\phi^3\psi + 10\phi^2\psi^2 + 20\phi\psi^3 + 39\psi^4\right] \\
&\quad -\frac{3w}{4}\left[5\phi^4 + 8\phi^3\psi + 5\psi^4 - 2\phi^2\psi^2 - 4\sigma^4 + \phi^2\left(6\psi^2 - 2\sigma^2\right)\right. \\
&\quad \left. +4\phi\psi\left(2\phi^2 + \sigma^2\right)\right] - \frac{3w^2(\phi-\psi)^2}{8}\left[7\phi^2 + 10\phi\psi + 7\psi^2 + 8\sigma^2\right] \\
&\quad -\frac{3w^3(\phi-\psi)^2}{4}\left[(\phi+\psi)^2 - 2\sigma^2\right] - \frac{3w^4(\phi-\psi)^4}{16}.
\end{aligned}
$$

The first four noncentral moments of the c-NAL (4.7) are:

$$
\begin{aligned}
E(Y) &= \mu + \frac{1}{\alpha} - \frac{1}{\beta} \\
E(Y^2) &= \mu^2 + \sigma^2 - \frac{2\mu}{\beta} + \frac{2\left(\mu\beta - 1\right)}{\alpha\beta} + \frac{2}{\alpha^2} + \frac{2}{\beta^2} \\
E(Y^3) &= \frac{1}{\alpha^3\beta^3}\left[6\beta^3 + 6\alpha\beta^2\left(\beta\mu - 1\right) + 3\alpha^2\beta\left(2 - 2\beta\mu + \beta^2\left(\mu^2 + \sigma^2\right)\right)\right. \\
&\quad \left. + \alpha^3\left(6\beta\mu - 6 - 3\beta^2\left(\mu^2 + \sigma^2\right) + \beta^3\left(\mu^3 + 3\mu\sigma^2\right)\right)\right] \\
E(Y^4) &= \frac{1}{\alpha^4\beta^4}\left[24\beta^4 + 24\alpha\beta^3\left(\beta\mu - 1\right) + 12\alpha^2\beta^2\left(2 - 2\beta\mu + \beta^2\left(\mu^2 + \sigma^2\right)\right)\right. \\
&\quad + 4\alpha^3\beta\left(6\beta\mu - 6 - 3\beta^2\left(\mu^2 + \sigma^2\right) + \beta^3\left(\mu^3 + 3\mu\sigma^2\right)\right) \\
&\quad + \alpha^4\left(24 - 24\beta\mu + 12\beta^2\left(\mu^2 + \sigma^2\right) - 4\beta^3\left(\mu^3 + 3\mu\sigma^2\right)\right) \\
&\quad \left. + \beta^4\left(\mu^4 + 3\sigma^4 + 6\mu^2\sigma^2\right)\right]
\end{aligned}
$$

The first four central moments are

$$
\begin{aligned}
E(Y) &= 0 \\
Var(Y) &= \sigma^2 + \frac{1}{\alpha^2} + \frac{1}{\beta^2} \\
\tau(Y) &= \frac{2}{\alpha^3} - \frac{2}{\beta^3} \\
\kappa(Y) &= 3t^2\left(t^2 + \frac{2}{\alpha^2} + \frac{2}{\beta^2}\right) + 6\left(\frac{1}{\alpha^2\beta^2} + \frac{3}{2\alpha^4} + \frac{3}{2\beta^4}\right).
\end{aligned}
$$

References

Aghion, P. and Howitt, P. (1992) A model of growth through creative destruction. *Econometrica,* 60, 323-351.

Aghion, P. and Howitt, P. (1998) *Endogenous growth theory.* The MIT press, Cambridge.

Candelon, B. and Gil-Alana, L. A. (2004) Fractional integration and business cycle features. *Empirical Economics,* 60, 343-359.

Dyer, A. R. (1974). Comparisons of tests for normality with a cautionary note. *Biometrika*, 61, 185-189.

Fryer, J. G. and Robertson, C. A. (1972) A comparison of some methods for estimating mixed normal distributions, *Biometrika*, 59, 639-648.

Grossman, G. and Helpman, E. (1991) Quality ladders in the theory of growth. *Review of Economic Studies*, 63, 43-61.

Helpman, E. and Trajtenberg, M. (1994) A Time to Sow and a Time to Reap: Growth Based on the General Purpose Technologies. Centre for Economic Research Policy. *Working paper no. 1080.*

Hodrick, R. J. and Prescott, E. C. (1997) Postwar U.S. business cycles: An empirical investigation. *Journal of Money, Credit and Banking*, 29, 1-16.

Hyndman, R. J. and Koehler, A. B. (2006) Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22, 679-688.

Jones, P. N. and McLachlan, G. J. (1990) Laplace-normal mixtures fitted to wind shear data. *Journal of Applied Statistics,* 17, 271-276.

Kanji, G. K. (1985) A mixture model for wind shear data. *Journal of Applied Statistics*, 12, 49-58.

Kotz, S. Kozubowski, T. J. and Podgorski, K. (2001) *The Laplace distribution and generalizations: a revisit with applications to communications, economics, engineering, and finance.* Birkhäuser, Boston.

Kozubowski, T. J. and Podgorski, K. (1999) A class of asymmetric distributions. *Actuarial Research Clearing House*, 1, 113-134.

Kozubowski, T. J. and Podgorski, K. (2000) Asymmetric Laplace distributions. *The Mathematical Scientist*, 25, 37-46.

Linden, M. (2001) A model for stock return distribution. *International Journal of Finance and Economics*, 6, 159-169.

Lucas, R. E. (1988) On the Mechanics of Economic Development. *Journal of Monetary Economy*, 22, 3-42.

Makridakis, S. (1993) Accuracy measures: theoretical and practical concerns. *International Journal of Forecasting*, 9, 527-529.

Maliar, L. and Maliar, S. (2004) Endogenous growth and endogenous business cycles. *Macroeconomic Dynamics*, 8, 559-581.

McGill, W. J. (1962) Random fluctuations of response rate. *Psychometrika*, 27, 3-17.

Pearson, K. (1895) Contributions to the mathematical theory of evolution, 2: skew variation in homogeneous material. *Philosophical Transactions of the Royal Society of London (A)*, 186, 343-414.

Reed, W. J. and Jorgensen, M. A. (2004) The double Pareto-lognormal distribution - A new parametric model for size distributions. *Communications in Statistics - Theory and Methods,* 33, 1733-1753.

Romer, P. M. (1986) Increasing returns and long-run growth. *Journal of Political Economy*, 94, 1002-1037.

Schumpeter, J. A. (1942) *Capitalism, Socialism and Democracy.* Harper, New York.

Segerstrom, P. Anant, T. Dinopoulos, E. (1990) A Schumpeterian model of the product life cycle. *American Economic Review*, 80, 1077-1091.

Silverman, B. W. (1986) *Density estimation for statistics and data analysis.* Chapman and Hall, London.

Stockhammar, P. and Öller, L.-E. (2007) A simple heteroscedasticity removing filter. *Research Report 2007:1*, Department of Statistics, Stockholm University.

Yu, K. and Zhang, J. (2005) A three-parameter asymmetric Laplace distribution and its extension. *Communications in Statistics - Theory and Methods*, 34, 1867-1879.

Density Forecasting of the Dow Jones Stock Index

Pär Stockhammar and Lars-Erik Öller

Department of Statistics, Stockholm University
S-106 91 Stockholm, Sweden
E-mail: par.stockhammar@stat.su.se

Abstract

The distribution of differences in logarithms of the Dow Jones stock index is compared to the Normal (N), Normal Mixture (NM) and a weighted sum of a normal and an asymmetric Laplace distribution (NAL). It is found that the NAL fits best. We came to this result by studying samples with high, medium and low volatility, thus circumventing strong heteroscedasticity in the entire series. The NAL distribution also fitted economic growth, thus revealing a new analogy between financial data and real growth.

Keywords: Density forecasting, heteroscedasticity, mixed Normal - Asymmetric Laplace distribution, Method of Moments estimation, connection with economic growth.

1. Introduction

In some fields, including economics and finance, series exhibit heteroscedasticity, asymmetry and leptokurtocity. Ways to account for these features have been suggested in the literature and also used in some applications. The Bank of England uses the two-piece normal distribution (see John, 1982 and Britton et al., 1998) when calculating interval and density forecasts of

macroeconomic variables in the UK. The close relative, the Normal Mixture (NM) distribution has been used in e.g. Wallis (2005) and recently in Mitchell and Wallis (2010). Another increasingly popular distribution to describe data with fatter than Normal (N) tails is the Laplace (L) distribution. In the finance literature it has been applied to model interest rate data (Kozubowski and Podgórsky, 1999), currency exchange data (Kozubowski and Podgórsky, 2000), stock market returns (Madan and Senata, 1990) and option pricing (Madan et al., 1998), to name a few applications. Stockhammar and Öller (2008) showed that the L distribution may be too leptokurtic for economic growth data. Instead, allowing for asymmetry, a mixed Normal - Asymmetric Laplace (NAL) distribution was proposed and in ibid. it was shown that this distribution more accurately describes GDP growth data of the US, the UK and the G7 countries than N, NM and L distributions. The convoluted version of the NAL, suggested by Reed and Jorgensen (2004) was also examined there, but proved inferior to the weighted sum of probabilities of the NAL.

In the present study, the density of the Dow Jones Industrial Average (DJIA) is investigated. This series is significantly skewed, leptokurtic and heteroscedastic. Diebold et al. (1998) showed that a MA(1) - t-GARCH$(1,1)$ model is suitable to forecast the density of the heteroscedastic S&P 500 return series. Here another approach is employed. Instead of modeling the conditional variance, the data are divided into parts according to local volatility (each part being roughly homoscedastic). For every part we estimate and compare the density forecasting ability of the N, NM and the NAL distributions. If the NAL distribution would fit both stock index data and GDP growth, this would hint at a new analogy between the financial sphere and the real economy.

This paper is organized as follows. Section 2 provides some theoretical underpinnings. The data are presented in Section 3 and a distributional discussion in Section 4. Section 5 contains the estimation set-up and a density forecasting accuracy comparison. Section 6 contains an illustrative example and Section 7 concludes.

2. Density forecast evaluation

The key tool in the recent literature on density forecast evaluation is the probability integral transform (PIT). It goes back at least to Rosenblatt (1952), with contributions by e.g. Shepard (1994) and Diebold et al. (1998). The PIT is defined as

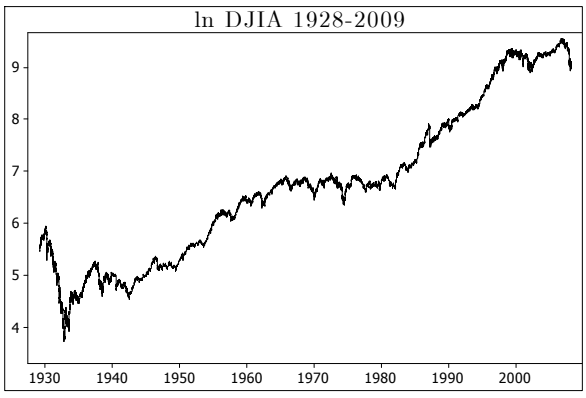
$$z_t = \int_{-\infty}^{y_t} p_t(u)du, \qquad (2.1)$$

where y_t is the realization of the process and $p_t(u)$ is the assumed forecast density. If $p_t(u)$ equals the true density, $f_t(u)$, then $z_t \sim$ i.i.d. $U(0,1)$. This suggests that we can evaluate density forecasts by assessing whether z_t are i.i.d. $U(0,1)$. This enables joint testing of both uniformity and independence in Section 4.

3. The data

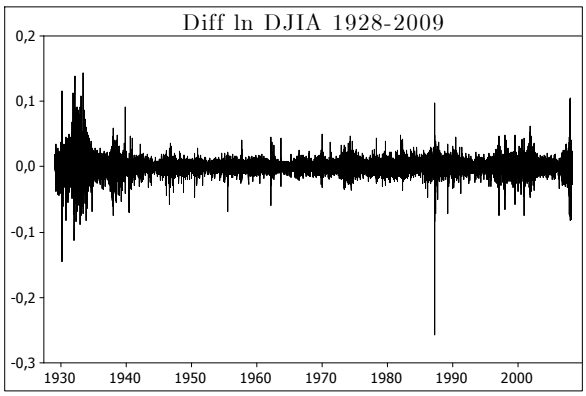
In this paper the Dow Jones Industrial average index (daily closing prices) Oct. 1, 1928 to Jan. 31, 2009 (20 172 observations) is studied as appearing on the website www.finance.yahoo.com. The natural logarithm of the series is shown in Figure 3.1.

Figure 3.1: The ln Dow Jones Industrial Average Oct. 1, 1928 to Jan. 31, 2009



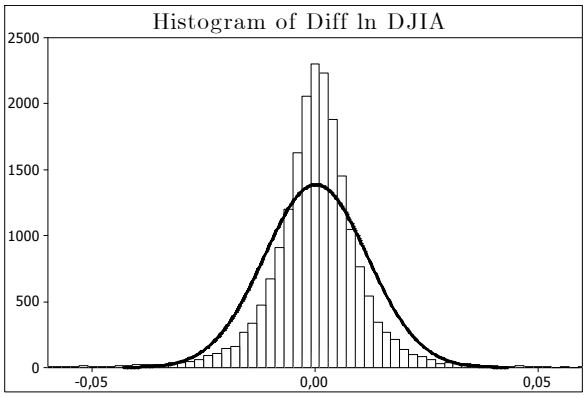
Taking the first difference of the logarithmic data (Diff ln) gives Figure 3.2, which reveals the heteroscedasticity.

Figure 3.2: Diff ln Dow Jones Industrial Average Oct. 1, 1928 to Jan. 31, 2009



As seen in Figure 3.3, the Diff ln series seems to be leptokurtic. Significant both leptokurtocity and skewness were found in tests.

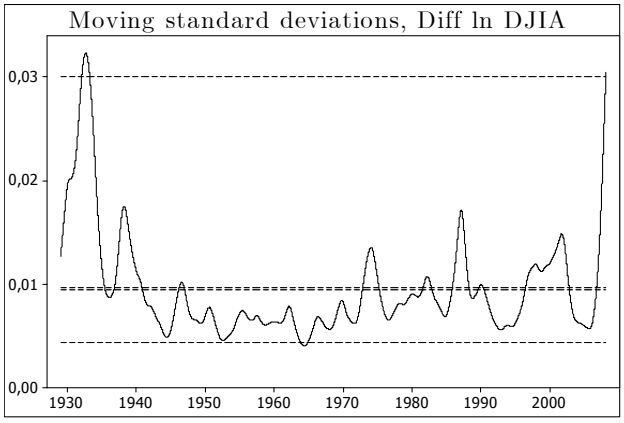
Figure 3.3: Histogram of Diff ln DJIA Oct. 1, 1928 to Jan. 31, 2009. The solid line is the Normal distribution using the same mean and variance as in the series



The heteroscedasticity is also evident in Figure 3.4, which shows moving

standard deviations using a window of length $k = 45$ smoothed with the Hodrick-Prescott (HP) (1997) filter (using smoothing parameter $\lambda = 1.6 \times 10^7$).

Figure 3.4: Smoothed moving standard deviations using window k=45 and a HP filter with $\lambda = 1.6 \times 10^7$



The data have been divided into three small groups of volatility, cf. Figure 3.4. The periods denoted as high (H), medium (M) and low (L) volatility ($y_{t,H}$, $y_{t,M}$ and $y_{t,L}$) are defined as times when the smoothed moving standard deviations, $\widehat{\sigma}_t$, (see Figure 3.4) are larger than 0.03, between 0.0095 and 0.0097, and smaller than 0.0044, respectively. These limits were chosen so as to get approximately equally-sized samples, for which in-sample variance is fairly constant. Also, choosing only the very extreme parts of volatility facilitates calibration of the parameters of the distributions described in Section 4. The three periods consist of 308, 267 and 277 observations, respectively. The variables, $y_{t,H}$ and $y_{t,L}$ have been sampled from undivided periods, 1931-11-05 to 1933-01-27 and 1964-03-10 to 1965-04-13, respectively. According to the ARCH-LM, the augmented Dickey-Fuller (ADF) and various normality tests, $y_{t,H}$ and $y_{t,L}$ are homoscedastic, stationary and non-normal. The skewness is significantly nonzero in $y_{t,L}$ and significant leptokurtocity appears in both $y_{t,H}$ and $y_{t,L}$. On the contrary, the medium volatility part, $y_{t,M}$, contains observations from 16 disjoint periods. Standard homoscedasticity, unit-root and normality tests are not available for non-equidistant data.

The proposed procedure of circumventing strong heteroscedasticity in the entire series is aimed at finding the most accurate density forecast distribution

for each part of local volatility. The result is then used to provide guidelines for the intervening situations of local volatility. Using the simplified NAL distribution described in Section 4, a strict judgmental estimation of the parameters is facilitated using the estimated distributions for the high, medium and low volatility parts as guidelines. It is also possible to constantly reestimate the parameters using the techniques described in Section 5.

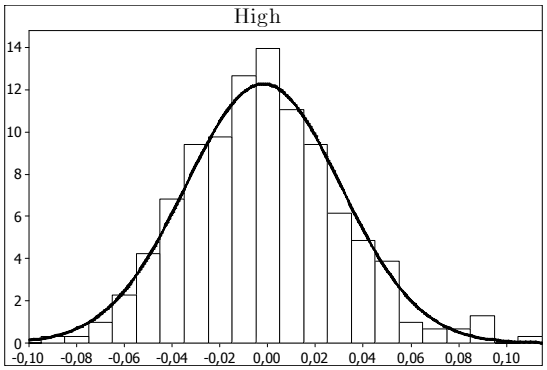
Table 3.1 shows the first four sample central and noncentral moments of the high, medium and low volatility observations.

Table 3.1: The sample central and noncentral moments of $y_{t,H}$, $y_{t,M}$ and $y_{t,L}$

	$y_{t,H}$	$y_{t,M}$	$y_{t,L}$		$y_{t,H}$	$y_{t,M}$	$y_{t,L}$
$\widehat{\mu}$	-0.0018	0.0012	0.0004	$E(y_t)$	-0.00184	0.00121	0.00043
$\widehat{\sigma}$	0.0325	0.0088	0.0039	$E(y_t^2)$	0.001057	0.000079	0.000015
$\widehat{\tau}$	0.33	0.15	-0.47	$E(y_t^3)$	0.000006	0.000000	0.000000
$\widehat{\kappa}$	0.35	0.98	0.54	$E(y_t^4)$	0.000004	0.000000	0.000000

In Table 3.1, $\widehat{\tau}$ and $\widehat{\kappa}$ are the sample skewness and excess kurtosis, respectively. As expected the variance is very different in the three samples. Note that the mean of $y_{t,H}$ is negative, the volatility thus tends to increase when DJIA declines. Figure 3.5 shows the distributions of $y_{t,H}$, $y_{t,M}$ and $y_{t,L}$.

Figure 3.5: The distributions of $y_{t,H}$, $y_{t,M}$ and $y_{t,L}$



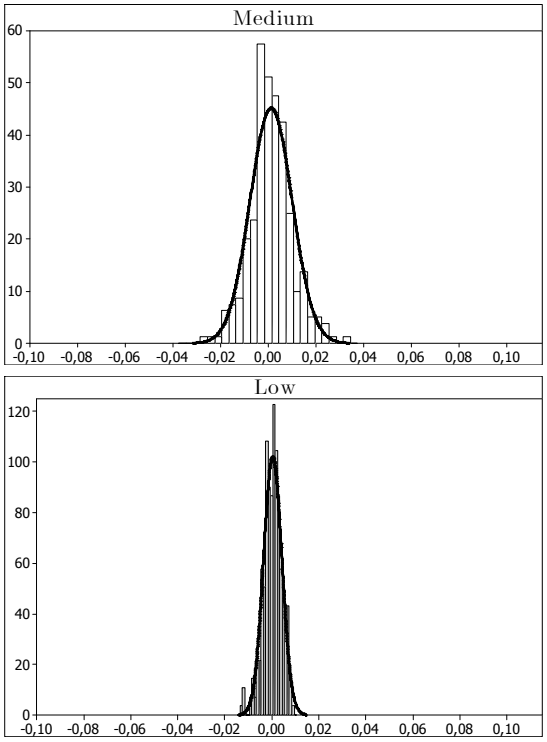


Figure 3.5 indicates that the distribution of $y_{t,M}$ could also be non-normal (significant non-normality were found in $y_{t,H}$ and $y_{t,L}$). But in order to vindicate the conclusions, we keep the Gaussian distribution as a benchmark. This will be compared with the NM and the NAL distributions. That is the topic of the next section.

4. Distributional discussion[11]

The use of different means and variances for the regimes enables introducing skewness and excess kurtosis in the NM distribution. The probability distribution function (pdf) of the NM distribution is:

$$f_{NM}(y_t;\boldsymbol{\theta}_1) = \frac{w}{\sigma_1\sqrt{2\pi}}\exp\left\{-\frac{(y_t-\mu_1)^2}{2\sigma_1^2}\right\}+\frac{1-w}{\sigma_2\sqrt{2\pi}}\exp\left\{-\frac{(y_t-\mu_2)^2}{2\sigma_2^2}\right\}, \quad (4.1)$$

[11] See Stockhammar and Öller (2008) for a more detailed description of the distributions.

where $\boldsymbol{\theta}_1$ consists of the parameters $(w, \mu_1, \mu_2, \sigma_1, \sigma_2)$ and where $0 \leq w \leq 1$ is the weight parameter. Another distribution often used to describe fatter than normal tails is the double (two-sided) exponential, or the Laplace (L) distribution. It arises as the difference between two exponential random variables with the same parameter value. The pdf of the L distribution is:

$$f_L(y_t; \boldsymbol{\theta}_2) = \frac{1}{2\phi} \exp\left\{-\frac{|y_t - \mu|}{\phi}\right\}, \qquad (4.2)$$

where $\boldsymbol{\theta}_2 = (\mu, \phi)$, $\mu \in \mathbb{R}$ is the location parameter and $\phi > 0$ is the scale parameter. Again studying Figure 3.3 the L distribution seems promising, but it cannot describe the significant skewness in the data. Instead we choose the asymmetric Laplace (AL) distribution with pdf:

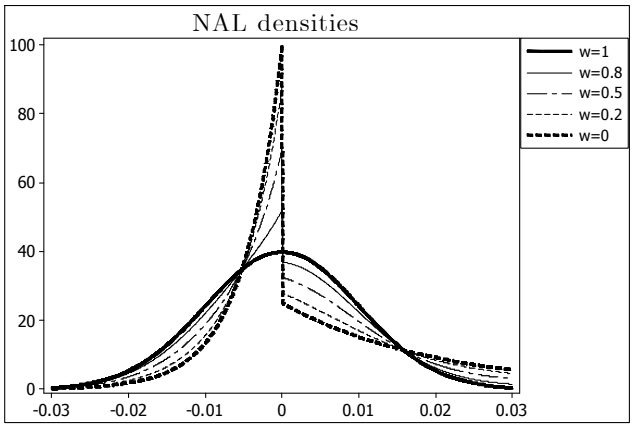
$$f_{AL}(y_t; \boldsymbol{\theta}_3) = \begin{cases} \frac{1}{2\psi} \exp\left\{\frac{y_t - \mu}{\psi}\right\} & \text{if } y_t \leq \mu \\ \frac{1}{2\phi} \exp\left\{\frac{\mu - y_t}{\phi}\right\} & \text{if } y_t > \mu \end{cases}, \qquad (4.3)$$

where $\boldsymbol{\theta}_3$ consists of the three parameters (μ, ϕ, ψ). The main advantage of the AL distribution is that it is skewed (when $\psi \neq \phi$), conforming with the empirical evidence in Table 3.1. Another property of this distribution is that, unlike the pure L distribution, the kurtosis is not fixed. The AL distribution also has a discontinuity at μ. To further improve flexibility, Gaussian noise is added. To the author's best knowledge this distribution has not been used before to model financial time series data. We assume that the probability density distribution of the Diff ln Dow Jones series (y_t) can be described as a weighted sum of Normal and AL random densities, i.e.:

$$f_{NAL}(y_t; \boldsymbol{\theta}_4) = \frac{w}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(y_t - \mu)^2}{2\sigma^2}\right\} + (1 - w) \begin{cases} \frac{1}{2\psi} \exp\left\{\frac{y_t - \mu}{\psi}\right\} & \text{if } y_t \leq \mu \\ \frac{1}{2\phi} \exp\left\{\frac{\mu - y_t}{\phi}\right\} & \text{if } y_t > \mu \end{cases}, \qquad (4.4)$$

where $\boldsymbol{\theta}_4 = (w, \mu, \sigma, \phi, \psi)$. Distribution (4.4) is referred to as the mixed Normal-Asymmetric Laplace (NAL) distribution. Note that equal medians but unequal variances are assumed for the components. Figure 4.1 shows NAL densities for five different values of the weight parameter w.

Figure 4.1: NAL densities using a $N(0,0.01)$ and an $AL(\psi{=}0.005, \phi{=}0.02)$ and weightings of them using $w{=}1$, 0.8, 0.5, 0.2 and 0.

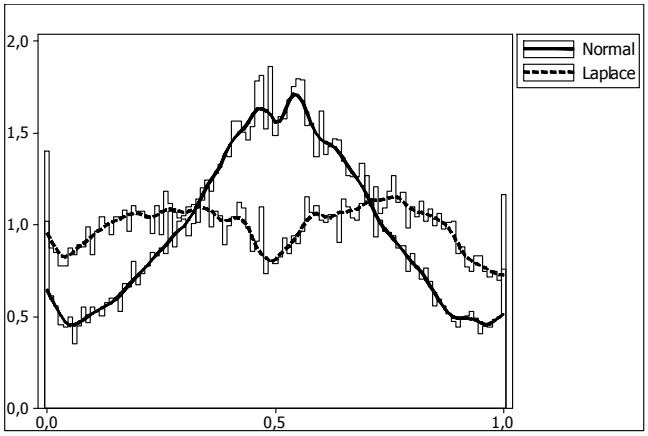


A graphical examination of the PIT histograms (see Section 2) might serve as a first guide when determining the density forecasting accuracy of the above distributions. One intuitive way to assess uniformity is to test whether the empirical cumulative distribution function (cdf) of $\{z_t\}$ is significantly different from the 45° line (the theoretical cdf). This is done using e.g. the Kolmogorov-Smirnov (K-S) statistic or χ^2-tests.

Assessing whether z_t is i.i.d. can be made visually by examining the correlogram of $\{z_t - \overline{z}\}^i$ (with $i = 1, 2, 3, 4$) and the corresponding Bartlett confidence intervals. Thus, we examine not only the correlogram of $\{z_t - \overline{z}\}$ but also check for autocorrelations in higher moments. Using $i = 1, 2, 3$ and 4 will reveal dependence in the (conditional) mean, variance, skewness and kurtosis. This way to evaluate density forecasts was advocated by Diebold et al. (1998).

In order to illustrate why the NAL distribution (4.4) is a plausible choice we once more study the entire series. Figure 4.2 shows the contours of calculated PIT histograms together with Kernel estimates (using the Gaussian Kernel function and Silverman´s bandwidth) for the L and the cumulative benchmark N distribution.

Figure 4.2 Density estimates[12] *of* z_t

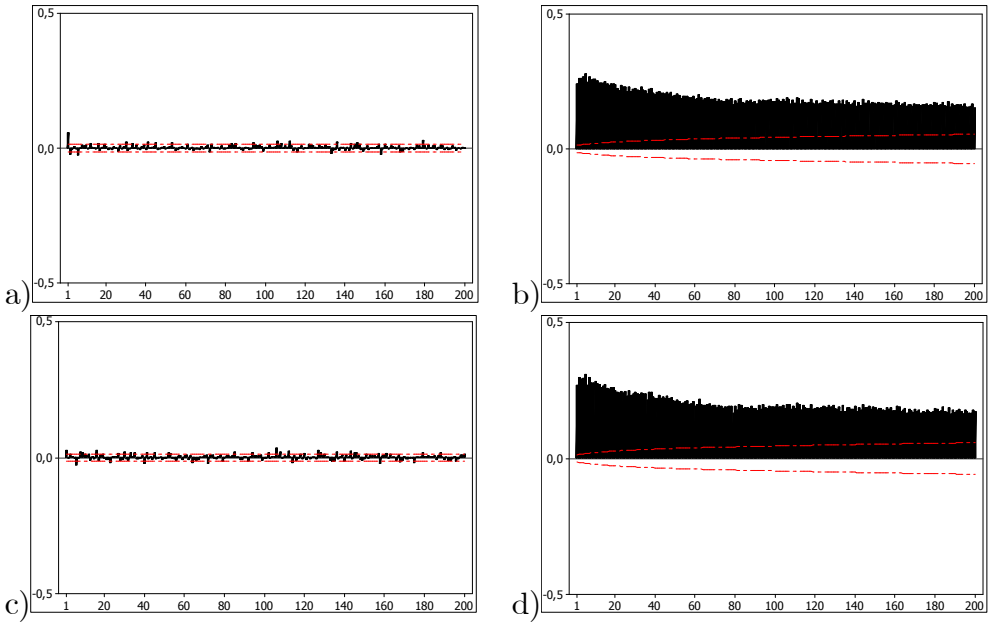


The N histogram has a distinct non-uniform "moustache" shape – a hump in the middle and upturns on both sides. This indicates that too many of the realizations fell in the middle and in the tails, relative to what we would expect if the data were normally distributed, see Figure 3.3. The "seagull" shape of the L histogram is flatter than that of N, but is nevertheless non-uniform. The L histogram is the complete opposite of the N histogram with too few observations in the middle and in the tails.

Neither of the two distributions is appropriate to use as forecast density function. It may be possible to find a suitable weighted average of them (the Normal-Laplace (NL) distribution) or, accounting for the asymmetry, the NAL as defined in (4.4). However, assessing whether $z_t \sim$ i.i.d. $U(0,1)$ shows the disadvantages with the above distributions. Neither of them is particularly suitable to describe heteroscedastic data (such as the entire Diff ln series), see Figures 4.3 a-d) of the autocorrelation functions (ACF) of $\{z_t - \overline{z}\}^i$ using the N distribution as forecast density.

[12] 100 bins were used. If the forecast density were true we would expect one percent of the observations in each of the 100 classes, with a standard error of 0.0295 percent.

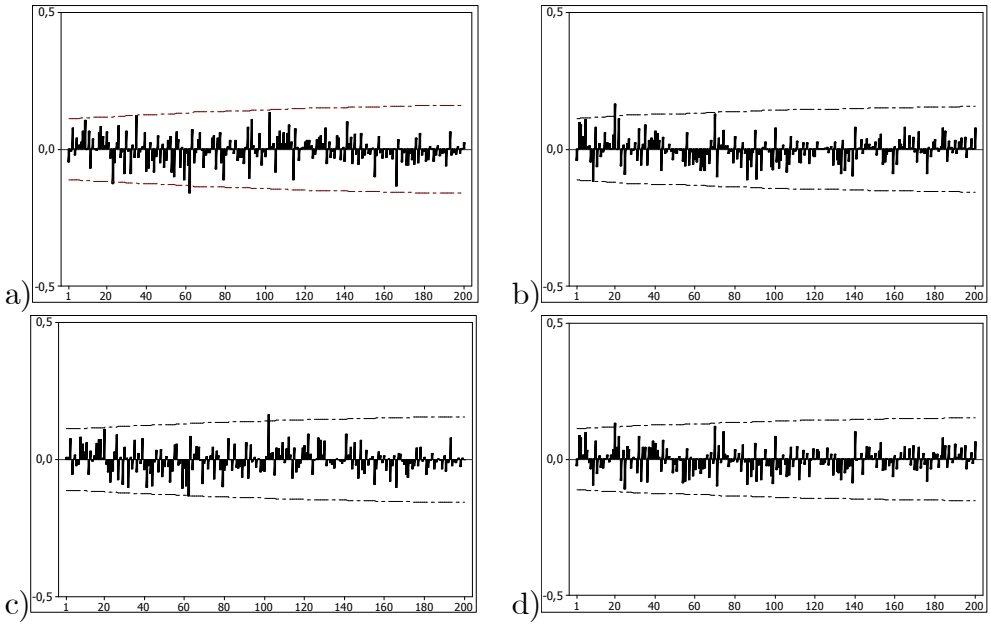
Figure 4.3: Estimates of the ACF of $\{z_t - \overline{z}\}^i$ *, i=1,2,3 and 4, for* y_t *assuming normality*



The strong serial correlation in $\{z_t - \overline{z}\}^2$ and $\{z_t - \overline{z}\}^4$ (panels b and d) shows another key deficiency of using the N density - it fails to capture the volatility dynamics in the process. Also, the L correlograms indicate neglected volatility dynamics. This was expected. Neither single (N or L), nor mixed distributions (NM, NAL) are able to capture the volatility dynamics in the entire series. One could model the conditional variance using e.g. GARCH type models (as in Diebold et al., 1998), or State Space exponential smoothing methods, see Hyndman et al. (2008). Here we are more interested in finding an appropriate distribution to describe the data. Instead of modeling the conditional variance, as said in Section 3 the data are divided into three parts according to their local volatility (each of which is homoscedastic).

Figure 4.4 further supports the homoscedasticity assumption in the high volatility data ($y_{t,H}$), assuming normality.

Figure 4.4: Estimates of the ACF of $\{z_t - \overline{z}\}^i$, *i=1,2,3 and 4, for* $y_{t,H}$ *assuming normality*



The series of medium and low volatility assuming the N, L, NM and NAL distributions give similar ACF:s. Standard tests do not signal autocorrelation in these series assuming any of the distributions. This means that our demand for independence is satisfied, and finding the most suitable distribution for density forecasts is a matter of finding the distribution with the most uniform PIT histogram. This is done using the K-S and χ^2 tests for $y_{t,H}$, $y_{t,M}$ and $y_{t,L}$ separately, when the parameters have first been estimated. These are issues of the next section.

5. Estimation

The parameters are estimated for the three periods of high, medium and low volatility, respectively. For each part separately, the five parameters in the NM and NAL distributions (4.1 and 4.4) will be estimated using the method of moments (MM) for the first four moments. The noncentral and central moments and the cumulative distribution function (cdf) of (4.1) and (4.4) were derived in Stockhammar and Öller (2008). Equating the theoretical

and the observed first four moments in Table 3.1 using the five parameters yields infinitely many solutions[13]. A way around this dilemma is to fix μ_1 in the NM to be equal to the observed mode, which is here approximated by the maximum value of the Kernel function of the empirical distribution, $\max f_K(y_i)$ where $i = H, M, L$. Here, $\widehat{\mu}_{1,H}$, $\widehat{\mu}_{1,M}$ and $\widehat{\mu}_{1,L}$ are substituted for $\max f_K(y_{t,H}) = -0.0025$, $\max f_K(y_{t,M}) = -0.0001$ and $\max f_K(y_{t,L}) = 0.0011$. In the NAL distribution, μ is fixed to be equal to the MLE with respect to μ in the AL distribution, that is the observed median, $\widehat{md}$. Here, $\widehat{\mu}_H = \widehat{md}_H = -0.00359$, $\widehat{\mu}_M = \widehat{md}_M = 0.00081$ and $\widehat{\mu}_L = \widehat{md}_L = 0.00070$. Fixing one of the parameter in each distribution makes it easier to provide guidelines to forecasters concerning which parameter values to use, and when. With the above parameters fixed, the NM and NAL parameter values that satisfy the moment conditions are:

Table 5.1: Parameter estimates

	NM_H	NM_M	NM_L		NAL_H	NAL_M	NAL_L
$\widehat{w}$	0.8312	0.7803	0.7898	$\widehat{w}$	0.8447	0.7651	0.7994
$\widehat{\mu}_2$	0.0141	0.0059	−0.0021	$\widehat{\sigma}$	0.0292	0.0091	0.0041
$\widehat{\sigma}_1$	0.0229	0.0081	0.0041	$\widehat{\psi}$	0.0365	0.0036	0.0042
$\widehat{\sigma}_2$	0.0604	0.0098	0.0011	$\widehat{\phi}$	0.0563	0.0070	0.0015

Note that the estimated weights in all cases are close to 0.8. To further improve user-friendliness, it is tempting to also fix the weights to that value. If this can be done without losing too much in accuracy it is worth further consideration. With $w = 0.8$ (and the μ´s fixed as above), the remaining three MM estimates are:

Table 5.2: Parameter estimates

	NM_H	NM_M	NM_L		NAL_H	NAL_M	NAL_L
$\widehat{\mu}_2$	0.0008	0.0065	−0.0023	$\widehat{\sigma}$	0.0321	0.0088	0.0041
$\widehat{\sigma}_1$	0.0217	0.0081	0.0040	$\widehat{\psi}$	0.0137	0.0040	0.0042
$\widehat{\sigma}_2$	0.0582	0.0097	0.0018	$\widehat{\phi}$	0.0312	0.0079	0.0015

[13]We tried to make use of the fifth moment, but in none of the three parts did it significantly differ from zero.

Table 5.2 shows that not much changes if we fix w. The exception is for the NAL estimates of high volatility data, where both the magnitude and the ratio of $\widehat{\psi}$ to $\widehat{\phi}$ changes dramatically. Giving less weight to the N distribution is compensated for by a larger $\widehat{\sigma}$ and decreasing $\widehat{\psi}$ and $\widehat{\phi}$ and vice versa. Because of the positive skewness in $y_{t,H}$, $y_{t,M}$, $\widehat{\psi}_H < \widehat{\phi}_H$ and $\widehat{\psi}_M < \widehat{\phi}_M$. That $\widehat{\psi}_L > \widehat{\phi}_L$ accords well with the results in Table 3.1. Note that $y_{t,H}$ and $y_{t,L}$ have completely opposite properties in Table 3.1, $y_{t,H}$ having a mean below zero and positive skewness and the other way around for $y_{t,L}$. The relative difference between $\widehat{\psi}$ and $\widehat{\phi}$ is approximately the same in Table 5.2. $y_{t,M}$ shows yet another pattern with above zero mean and positive skewness ($\widehat{\psi}$ about half the value of $\widehat{\phi}$).

In order to compare the distributional accuracy of the above empirical distributions we make use of the K-S test. Because of the low power of this test, as with all goodness of fit tests, this is supplemented with χ^2 tests. The K-S test statistic (D) is defined as

$$D = \sup \left| F_E\left(x\right) - F_H(x) \right|,$$

where $F_E(x)$ and $F_H(x)$ are the empirical and hypothetical or theoretical distribution functions, respectively. Note that $F_E(x)$ is a step function that takes a step of height $\frac{1}{n}$ at each observation. The D statistic can be computed as

$$D = \max_i \left(\frac{i}{n} - F(x_i),\ F(x_i) - \frac{i-1}{n} \right),$$

where we have made use of the PIT (2.1) and ordered the values in increasing order to get $F(x_i)$. If $F_E(x)$ is the true distribution function, the random variable $F(x_i)$ is $U(0,1)$.

Table 5.3 reports the value of the D statistics (in parentheses), and also the p-values of the χ^2 test using 10 and 20 bins when testing $H_{0,1} : y_{t,k} \sim \text{N}$, $H_{0,2} : y_{t,k} \sim \text{NM}^{(1)}$, $H_{0,3} : y_{t,k} \sim \text{NM}^{(2)}$, $H_{0,4} : y_{t,k} \sim \text{NAL}^{(1)}$ and $H_{0,5} : y_{t,k} \sim \text{NAL}^{(2)}$ (k =high, medium or low). The number of degrees of freedom when calculating the p-values are in parentheses. Note that the fixed parameters in the NM and NAL distributions are treated as estimated, resulting in the reported number of degrees of freedom. The $\text{NM}^{(1)}$ and $\text{NAL}^{(1)}$ distributions are based on the parameter estimates in Table 5.1 while $\text{NM}^{(2)}$ and $\text{NAL}^{(2)}$

are based on the estimates in Table 5.2.

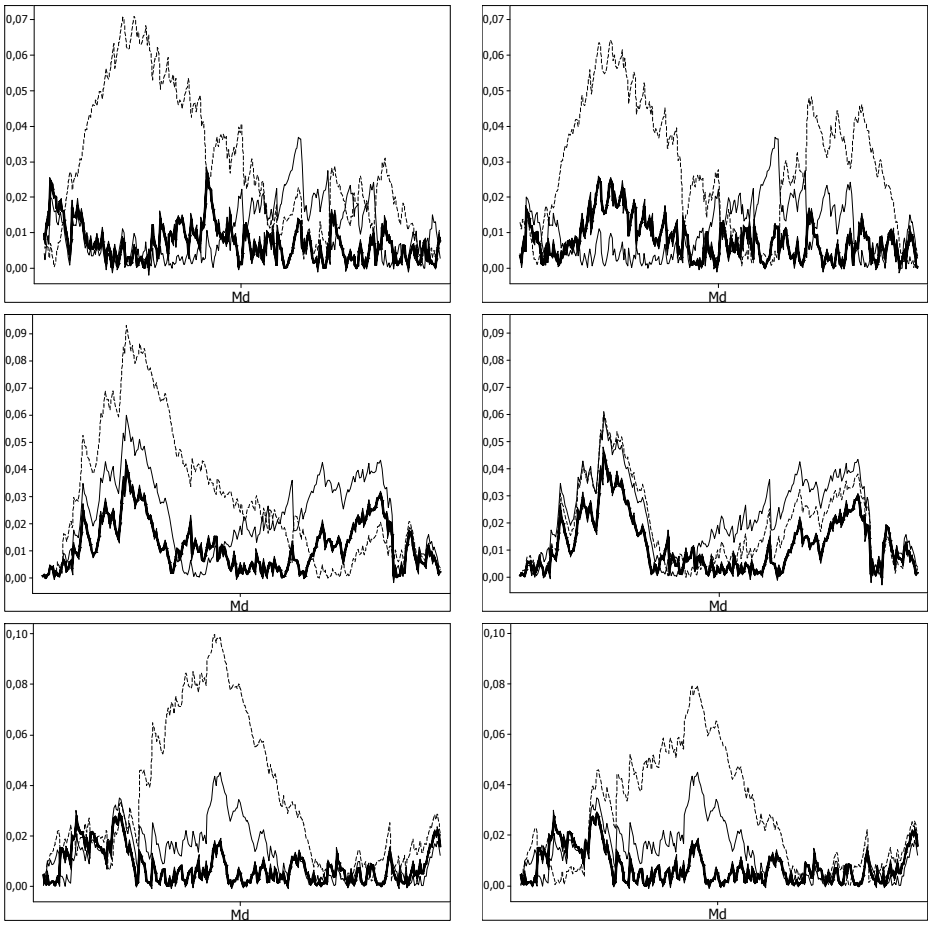
Table 5.3: Goodness of fit tests

		High	Medium	Low
$H_{0,1} : y_{t,k} \sim$ N	K-S	(0.038)	(0.060)	(0.045)
	$\chi^2(7)$	0.56	0.04	0.78
	$\chi^2(17)$	0.73	0.21	0.88
$H_{0,2} : y_{t,k} \sim$ NM$^{(1)}$	K-S	(0.071)	(0.092)	(0.099)
	$\chi^2(4)$	0.01	0.00	0.00
	$\chi^2(14)$	0.07	0.03	0.01
$H_{0,3} : y_{t,k} \sim$ NM$^{(2)}$	K-S	(0.064)	(0.062)	(0.079)
	$\chi^2(4)$	0.00	0.03	0.01
	$\chi^2(14)$	0.06	0.24	0.10
$H_{0,4} : y_{t,k} \sim$NAL$^{(1)}$	K-S	(0.027)	(0.041)	(0.028)
	$\chi^2(4)$	0.65	0.25	0.48
	$\chi^2(14)$	0.93	0.19	0.67
$H_{0,5} : y_{t,k} \sim$NAL$^{(2)}$	K-S	(0.025)	(0.046)	(0.028)
	$\chi^2(4)$	0.81	0.34	0.53
	$\chi^2(14)$	0.94	0.36	0.75

Table 5.3 shows that the NAL distributions are superior to the N and NM on average. Also, there is no great loss of information by fixing the weight parameter. In fact the NM fit was improved after fixing w, but the fit was nevertheless inferior to both the NAL and (surprisingly) the N distribution. Also the NAL$^{(2)}$ fit is slightly superior to the NAL$^{(1)}$. The NM distributions thus have a relatively poor fit to the extreme volatility parts of Diff ln DJIA. In general the N fit is, contrary to earlier results, quite good, particularly for the high and low volatility observations but, because of the skewness in the data, the NAL fits even better. Interestingly, In Stockhammar and Öller (2008) the NAL was also found to accurately describe GDP growth data.

Figure 5.1 shows the absolute deviations of the empirical distribution functions of the probability integral transforms from the theoretical 45° lines (the measure the K-S test is based on, cf. the K-S values in Table 5.3).

Figure 5.1: Absolute deviations of the N, $NM^{(1)}$,$NAL^{(1)}$ and N, $NM^{(2)}$, $NAL^{(2)}$ from the theoretical distributions

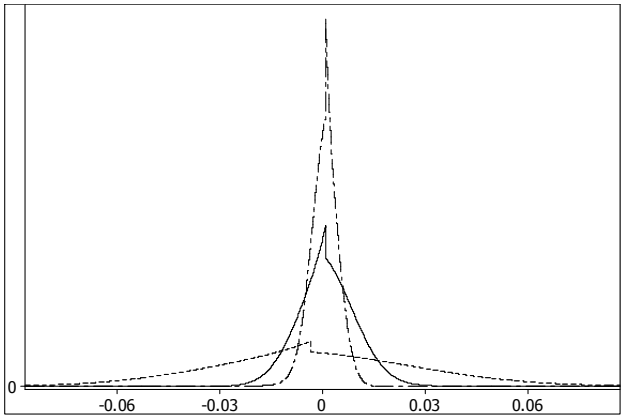


The N, NM and NAL distributions are marked with thin solid, dashed and thick solid lines, respectively, and the upper, centre and lower panels are the high, medium and low parts of the series. The panels to the left and right hand side are the distributions in Table 5.1 and 5.2, respectively.

Figure 5.1 adds further information about the fit. The left tail fit is inferior to the right tail fit. This is particularly prominent for the NM. It conforms well with Bao and Lee (2006) who came to the same conclusion using various nonlinear models for the S&P daily closing returns. Except for the low volatility part the fit close to the median is generally acceptable. Because of the similarity in distributional accuracy between the $NAL^{(1)}$ and $NAL^{(2)}$

the latter distribution is the obvious choice. With both μ and w fixed it is easier to interpret the remaining parameters. Figure 5.2 shows the forecast densities of the NAL$^{(2)}$ distributions for the three parts of extreme volatility.

Figure 5.2: Forecasting densities of the NAL$^{(2)}$ distributions for $y_{t,H}$ (dashed), $y_{t,M}$ (solid) and $y_{t,L}$ (dashed/dotted)



Here a jump at the median of each distribution is evident[14]. But this is of little importance when it comes to density forecasting where the tail behaviour is more interesting. The negative median in $y_{t,H}$ means that for high volatility data we expect a negative trend. But due to the skewness, large positive shocks will be more frequent than large negative cf. Table 3.1. Positive skewness is apparent also in the medium volatility data but for low volatility data, large negative shocks are more frequent than large positive.

In a situation of a very large local variance, here defined as $\widehat{\sigma}_t > 0.03$ for the last 45 days, we propose the use of the high volatility NAL distribution and the corresponding estimates in Table 5.2. Similarly we suggest the $\text{NAL}_{\text{M}}^{(2)}$ and $\text{NAL}_{\text{L}}^{(2)}$ estimates in Table 5.2 if the local variance falls between 0.0095 and 0.0097, or fall below 0.0044. For the intervening values a subjective choice is encouraged using the estimates in Table 5.2 and their corresponding distributions in Figure 5.2 as guidelines. Note that this approach is facilitated

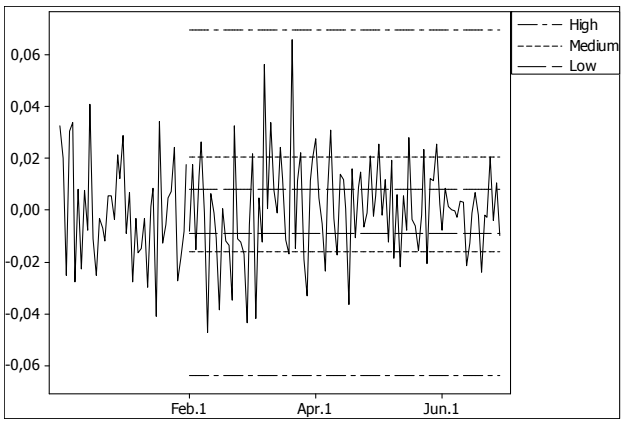
[14]The discontinuity at the median can be avoided using e.g. the convoluted NAL version of Reed and Jorgensen (2004). Since this approach did not prove promising in Stockhammar and Öller (2008), we do not pursue it here.

using the simplified NAL distribution. Another option is to regularly reestimate the parameters using the MM method and the latest set of moment estimates. During the worldwide financial crises of 2008 and 2009 we would most often use the $\mathrm{NAL_H}$ estimates (or values close to them). On the contrary we suggest the use of the $\mathrm{NAL_L}$ estimates during calm, or "business as usual" periods. This is exemplified in the following section.

6. Application

The proposed density forecast method is here applied to the Diff ln DJIA series Feb. 1, 2009 to Jun. 30, 2009, thus showing a realistic forecasting scenario. According to Figure 3.4 the local volatility at the end of January 2009 is very large ($\widehat{\sigma}_t \approx 0.03$). Following the earlier discussion we should in this situation choose the $\mathrm{NAL_H^{(2)}}$ distribution when calculating density forecasts, but to serve as comparisons we will also include the density forecasts made using the $\mathrm{NAL_M^{(2)}}$ and $\mathrm{NAL_L^{(2)}}$ distributions. We have used the (neutral) median in each distribution as point forecasts. Other point forecasts could, and probably should, be used in real life practice. Figure 6.1 shows the original Diff ln series Dec. 1, 2008 to Jun. 30, 2009 together with the 95 percent confidence intervals for the point forecasts using the $\mathrm{NAL_H^{(2)}}$, $\mathrm{NAL_M^{(2)}}$ and $\mathrm{NAL_L^{(2)}}$ distributions, calculated from Feb. 1, 2009.

Figure 6.1: Interval forecast comparison, Dec. 1, 2008 - Jun. 30, 2009



The forecasting horizon (5 months) in the above example is too long to be

classified as a high volatility period. The corresponding distribution works best only for the first half of the period. For the later half it is probably better to use parameter values closer to the $\mathrm{NAL}_{\mathrm{M}}^{(2)}$ distribution. In practice, frequent updates of the forecasts are recommended.

7. Conclusions

In this paper we have looked at a way to deal with the asymmetric and heteroscedastic features of the DJIA. The heteroscedasticity problem is solved by dividing the data into volatility groups. A mixed Normal - Asymmetric Laplace (NAL) distribution is proposed to describe the data in each group. Comparing with the Normal and the Normal Mixture distributions the NAL distributional fit is superior, making it a good choice for density forecasting Dow Jones stock index data. On top of good fit the simplicity of this distribution is particularly desirable since it enables easy-to-use guidelines for the forecaster. Subjective choices of the parameter values is encouraged, using the given parameter values for scaling.

The fact that the same distribution fits both stock index data and GDP growth indicates a analogy between financial and growth data not known before. The NAL distribution was derived as a representative of a Schumpeterian model of growth, the driving mechanism for which was Poisson (Aghion and Howitt, 1992) distributed innovations, see Stockhammar and Öller (2008). Interestingly, the same mechanism seems to work with stock index data.

Acknowledgments

This research was supported by the Department of Statistics at Stockholm University, Royal Swedish Academy of Sciences, the International Institute of Forecasters and by the Societas Scientiarum Fennica. We gratefully acknowledge helpful comments from Daniel Thorburn of Stockholm University and from Mattias Villani of the Swedish Riksbank. This paper has been presented in parts or in full at the International Symposium on Forecasting in 2008 and at Helsinki, Turku, Örebro, Uppsala and Stockholm Universities. We are grateful for the many suggestions from seminar participants.

References

Aghion, P. and Howitt, P. (1992) A model of growth through creative destruction. *Econometrica,* 60, 323-351.

Bao, Y. and Lee, T-H. (2006) Asymmetric predictive abilities of nonlinear models for stock returns: evidence from density forecast comparison. *Econometric Analysis of Financial and Economic Time Series / Part B, Advances in Econometrics*, 20, 41-62.

Britton, E., Fisher, P. G. and Whitley, J. D. (1998) The inflation report projections: understanding the fan chart. *Bank of England Quarterly Bulletin*, 38, 30-37.

Diebold, F. X., Gunther, T. A. and Tay. A. S. (1998) Evaluating density forecasts with applications to financial risk management. *International Economic Review*, 39, 863-883.

Hodrick, R. J. and Prescott, E. C. (1997) Postwar U.S. business cycles: An empirical investigation. *Journal of Money, Credit and Banking*, 29, 1-16.

Hyndman, R. J., Koehler, A. B., Ord, J. K. and Snyder, R. D. (2008) *Forecasting with exponential smoothing.* Springer Verlag, Berlin.

John, S. (1982) The three parameter two-piece normal family and its fitting. *Communications in Statistics - Theory and Methods*, 11, 879-885.

Kozubowski, T. J. and Podgorski, K. (1999) A class of asymmetric distributions. *Actuarial Research Clearing House*, 1, 113-134.

Kozubowski, T. J. and Podgorski, K. (2000) Asymmetric Laplace distributions. *The Mathematical Scientist*, 25, 37-46.

Madan, D. B. and Senata, E. (1990) The Variance Gamma (V.G.) model for share market returns. *Journal of Business*, 63, 511-524.

Madan, D. B., Carr, P. and Chang, E. C. (1998) The variance gamma process and option pricing. *European Finance Review*, 2, 74-105.

Mitchell, J. and Wallis, K. F. (2010) Evaluating density forecasts: Forecast combinations, model mixtures, calibration and sharpness. *International Journal of Forecasting*, forthcoming.

Reed, W. J. and Jorgensen, M. A. (2004) The double Pareto-lognormal distribution - A new parametric model for size distributions. *Communications in Statistics - Theory and Methods,* 33, 1733-1753.

Rosenblatt, M. (1952) Remarks on a multivariate transformation. *Annals of Mathematical Statistics*, 23, 470-472.

Shepard, N. (1994) Partial non-gaussian state space. *Biometrika*, 81, 115-131.

Stockhammar, P. and Öller, L.-E. (2008) On the probability distribution of economic growth. *Research Report 2008:5*, Department of Statistics, Stockholm University.

Wallis, K. F. (2005) Combining density and interval forecasts: a modest proposal. *Oxford Bulletin of Economics and Statistics*, 67, 983-994.

Comovements of the Dow Jones Stock Index and US GDP

Pär Stockhammar

Department of Statistics, Stockholm University
S-106 91 Stockholm, Sweden
E-mail: par.stockhammar@stat.su.se

Abstract

This paper explores the connection between Dow Jones industrial average (DJIA) stock prices and the US GDP growth. Both series are heteroscedastic, making standard detrending procedures, such as Hodrick-Prescott or Baxter-King, inadequate. The results from these procedures are compared to the results from heteroscedasticity corrected data, thus the effect of the neglected heteroscedasticity is measured. The analysis is mainly done in the frequency domain but relevant time domain results are also reported.

Keywords: Spectral analysis, detrending filters, heteroscedasticity, the connection between stock prices and economic growth.

1. Introduction

Numerous time domain studies have described the relationship between economic variables. Some other studies have investigated the relationship between, say economic growth, and non-economic variables such as the current age distribution in a country or energy consumption, see e.g. Lee (2005). In the frequency domain similar studies are also quite common. Öller (1990) used a frequency domain approach to investigate the fit and comovements

of business survey data and industrial production data in Finland. The exchange rate comovements of 12 countries were studied by Orlov (2009).

National product series, such as GDP, typically contain a unit root (Granger, 1966). Trends and unit roots show up as low or infinite frequency variations in the spectral density. Standard analysis requires stationarity and hence economic time series are detrended prior to further analysis. Done properly, detrending eliminates an infinite peak at zero frequency. Given a finite time series, it is impossible to design an ideal filter, and one has to make a good approximation. Filters may distort the frequency content of the cyclical part. Simple first-differencing, for instance, amplifies the higher frequencies at the expense of lower frequencies. Moreover, in the case of short series, abrupt variations in the frequency response give rise to Gibbs´ phenomenon, see e.g. Priestley (1981, pp. 561).

The most widely used detrending filters are the ones suggested by Hodrick and Prescott (HP) (1997), Beveridge and Nelson (BN) (1981) and Baxter and King (BK) (1999). Ma and Park (2004) used the HP filter in a comovement study of the interest rates in US, Japan and Korea. The HP filter was also applied in Uebele and Ritschl (2009), prior to a comovement study of stock markets and business cycles in Germany before World War I. The BK filter has been used by i.a. Stock and Watson (1999). Most studies focus solely on business cycle frequencies suggested by Burns and Mitchell (1946) of between 6 and 32 quarters. But much information may be extracted also outside this frequency band. This is further discussed in Section 4.

To the author´s best knowledge no detrending filter exists, which takes the highly possible event of heteroscedasticity into consideration. This is surprising because in spectral analysis contributions to the variance at specific frequencies are of prime interest. Neglecting heteroscedasticity will distort frequency domain results, see the discussion in e.g. Engle (1974). Because of this, the heteroscedasticity removing filter of Stockhammar and Öller (2007) will be considered here. The univariate and comovement frequency domain results from the filter proposed in ibid. will be compared with the results from the ones that do not take heteroscedasticity into account. The detrending filters will be further discussed in Section 3.

In Stockhammar and Öller (2008) it was shown that a Normal - Asymmetric Laplace (NAL) mixture distribution accurately describes the frequency distributions of US, UK and Australian GDP quarterly series. Interestingly, in

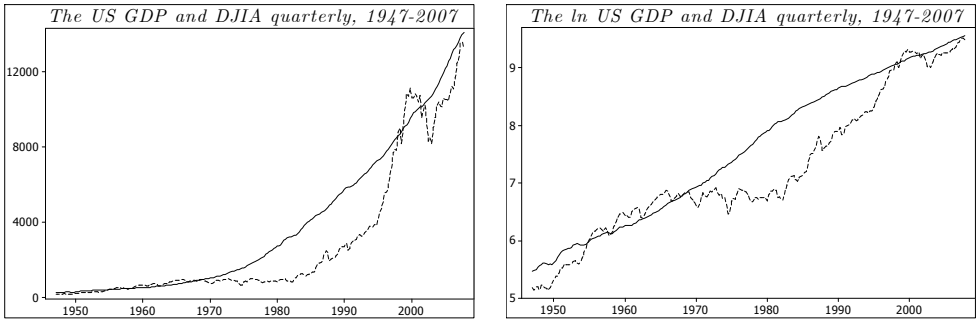
Stockhammar and Öller (2010) the same distribution was found also to work well for Dow Jones industrial average (DJIA) daily closing prices. This fact encouraged a closer look at the movements of stock indexes and GDP series. This will be pursued here using the above filters to detrend the series prior to spectral analysis of their relations. Frequency domain techniques allow a study of the correlations differentiated by frequency (coherency), and thus to concentrate on cycles. Because if the two series are related, the strongest coherency is expected to occur on business cycle frequencies. Comovements of two series may also be studied in the time domain using cross correlation coefficients. As a check of the results, relevant time domain estimates are also reported.

This paper is organized as follows. Section 2 presents the data. In Section 3 the filters are described then used to detrend the series prior to the comovement investigation of Dow Jones stock index data and US GDP in Section 4. Section 5 concludes.

2. The data

Here, quarterly figures 1947-2007 (244 observations) of the seasonally adjusted DJIA and the US GDP are studied as appearing on the website www.finance.yahoo.com and the website of the *Bureau of Economic Analysis*, www.bea.gov, respectively. The DJIA series was converted to quarterly figures from daily closing prices and calendar effects have been accounted for. These series together with their logarithms are presented in Figure 2.1.

Figure 2.1: DJIA (dashed line) and US GDP (solid line), the original series (left panel), logarithmic series (right panel)



The issue of detrending the above series to enable spectral analysis of their relationships is discussed in the next section.

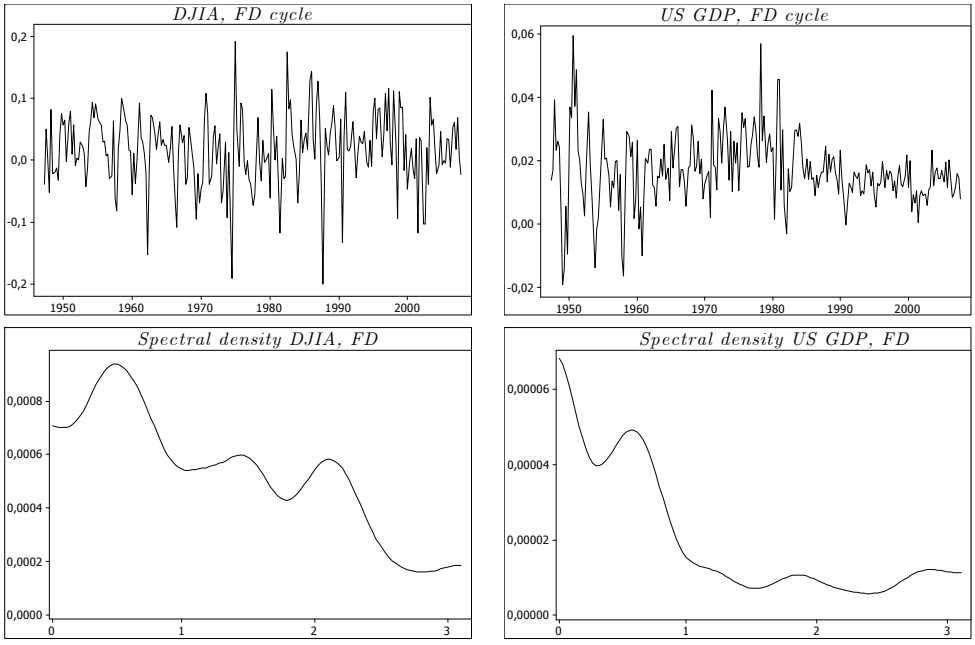
3. Detrending filters

As emphasized by Granger (1966), business cycle peaks in spectral densities are often buried in the massive share of low frequency (trend) variations. Several filters have been suggested to reduce the trend domination and to isolate the business cycle. The components of a time series can be defined in at least two ways (Cogley, 2001). One is the filter-design approach and the other is the model-based approach. For completeness, members of both approaches are compared in this study. In the filter-design approaches, the trend and business cycle are defined as components passing through an ideal[15], low, high or band-pass filter, whose bands are predetermined according to the assumed variation at specific frequencies. The approaches are typically ad hoc by nature, in the sense that the statistical properties of the business cycle are not specified. Here, in the presence of finite-length time series, it is impossible to design an ideal filter and a good approximation will have to suffice. The HP and BK filters described below are examples of this approach. To overcome some of the criticism mentioned below of the filter-design approach (see e.g. Harvey and Jaeger, 1993), the model-based approach has been suggested. The BN filter described in Section 3.3 is an example of the model-based approach.

[15] An ideal filter completely eliminates the frequencies outside the predetermined ones, while passing the remaining ones unchanged.

The simplest way to detrend a time series y_t is to calculate the *first differences*[16] *(FD)*, $y_t^{FD} = \Delta y_t$ where y_t^{FD} is the detrended series from the FD filter and $\Delta = (1 - B)$, where B is the backshift operator such that $B^j y_t = y_{t-j}$. The FD of the logarithmic (Diff ln) DJIA and US GDP are shown in Figure 3.1 (upper panel) together with their spectral densities[17] (lower panel).

Figure 3.1: The Diff ln series (upper panel) and the corresponding spectral densities (lower panel)



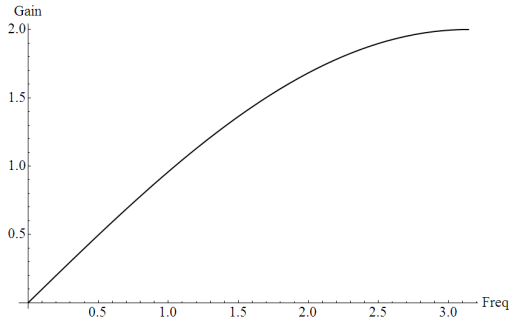
Heteroscedasticity is evident in the upper panels in Figure 3.1, especially in the filtered US GDP. There are some drawbacks using first differences. First, it is not a symmetric filter. This is however of no consequence when applying cross-spectral functions on two FD detrended series. As indicated by the lower panels in Figure 3.1, the densities of the FD series are still dominated by low frequency variations. This is because the true integrating

[16] Another way to get rid of unit roots in bivariate studies is to model in error correction form, given that the root is present in both series.

[17] See e.g. Jenkins and Watts (1968) for a thorough treatment of the spectral and cross-spectral functions used in this paper.

order of the series is somewhere between one and one and a half, see the discussion in Stockhammar and Öller (2007). Candelon and Gil-Alana (2004) concluded that the US GDP series is integrated of order $I(1.4)$. That is, an additional fractional difference of order 0.4 will eliminate the spectral density domination at low frequencies. Also, the FD filter reweights the densities towards higher frequencies as indicated by the gain function in Figure 3.2.

Figure 3.2: The gain function of the FD filter[18]



Despite the drawbacks with FDs they have been used in comovement studies like this one, e.g. by Wilson and Okunev (1999), and recently by Orlov (2009). Also, Knif et al. (1995) used the FD filter prior to cross-spectral analysis of the Finnish and Swedish stock markets.

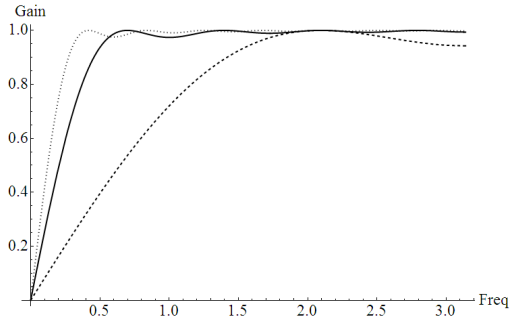
Calculating deviations from *centered moving averages (MA)* generated as

$$y_t^{c,MA} = y_t - \frac{1}{2p+1}\sum_{i=-p}^{p} y_{t+i},$$

where $k = 2p + 1$ is the window length and $y_t^{c,MA}$ is the cyclical component calculated from the MA filter, is another method of detrending time series. Figure 3.3 shows the gain of various values for k.

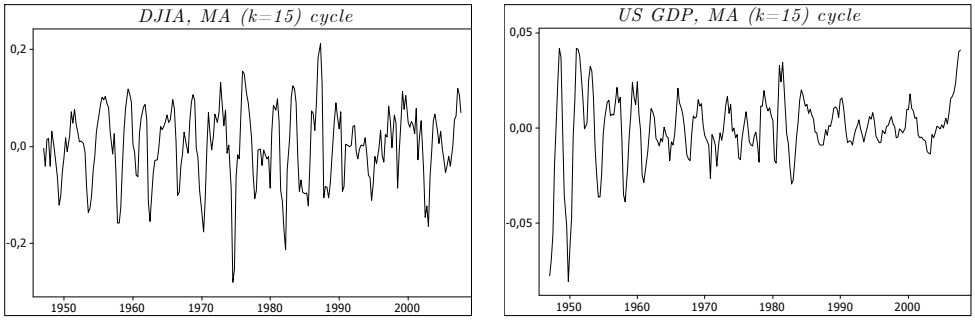
[18] The gain function of the first difference filter is $G(w) = \sqrt{2(1-\cos w)}$, where w is the frequency.

Figure 3.3: The gain functions of $y_t^{c,MA}$ using k=3 (dashed line), k=9 (solid line) and k=15 (dotted line)[19]

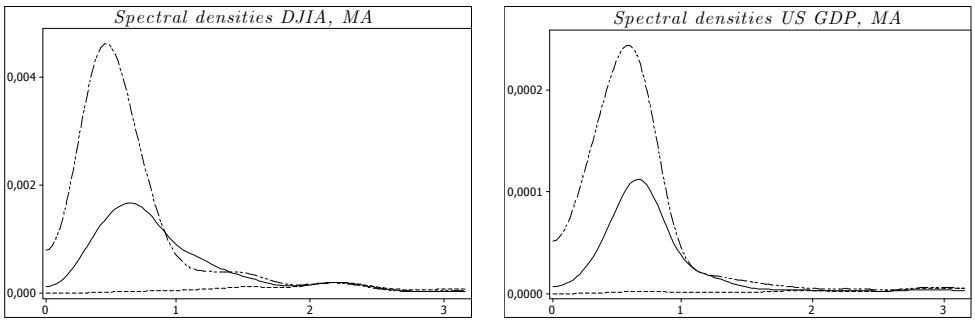


The MA is a symmetric filter but p observations are lost in both ends. Applying the MA($k = 15$) filter on the DJIA and US GDP series yields the filtered series in Figure 3.4 (upper panel). The corresponding spectral density and the spectral densities of the MA($k = 3$) and the MA($k = 9$) filter are also included (lower panel).

Figure 3.4: The MA(k=15) detrended DJIA and US GDP (upper panel). The lower panel shows the spectral densities of the cyclical components calculated from the MA(k=3) (dashed line), MA(k=9) (solid line) and MA(k=15) (dashed/dotted line)



[19] The gain function of the moving average filter that estimates the cycle is $G(w,k) = \sqrt{1 - \frac{1-\cos kw}{k^2(1-\cos w)}}$.



Comparing with the FD filter, the MA filter removes more of the low frequencies in the series, see Figures 3.1 and 3.4. As the window length gets wider, the spectral peaks are shifted to the left. Both the FD and MA filters produce series found to be stationary using the Phillips-Perron (PP) or the augmented Dickey-Fuller (ADF) tests. But they are still significantly heteroscedastic according to the ARCH-LM test.

The most widely used detrending filters are described in subsections 3.1-3.3[20]. The detrending and heteroscedasticity removing filter of Stockhammar and Öller (2007) is described in 3.4.

3.1 The Hodrick-Prescott filter

Perhaps the most commonly used filter to detrend economic time series is the one suggested by Hodrick and Prescott (1997). The HP-filter was designed to decompose a macroeconomic time series into a nonstationary trend component and a stationary cyclical component. Given a non-seasonal time series, y_t, the decomposition into unobserved components is

$$y_t = g_t + c_t,$$

where g_t denotes the unobserved trend component at time t, and c_t the unobserved cyclical component at time t. Estimates of the trend and cyclical

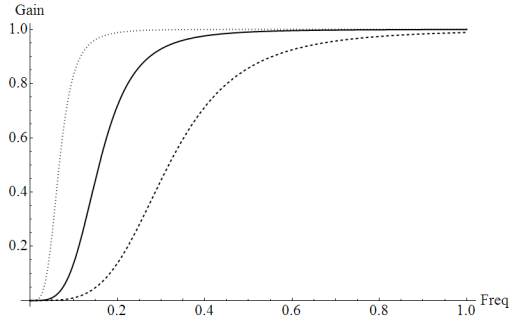
[20]The filter proposed by Christiano and Fitzgerald (2003) is here omitted. The asymmetric and time-varying features of this filter generate phase shifts, and nothing can be said about the stationarity of the output (even if the input is stationary).

components are obtained as the solution to the following minimization problem

$$\min_{[g_t]_{t=1}^N} \left\{ \sum_{t=1}^{N} c_t^2 + \gamma \sum_{t=3}^{N} (\triangle^2 g_t)^2 \right\}, \qquad (3.1)$$

where $\triangle g_t = g_t - g_{t-1}$ and $g_{\min}$ is the HP-filter and the cyclical component is calculated as: $y_t^{c,HP} = y_t - g_t$. The first sum of (3.1) accounts for the accuracy of the estimation, while the second sum represents the smoothness of the trend. The positive smoothing parameter γ controls the weight between the two components. As γ increases, the HP trend becomes smoother and vice versa. Note that the second sum, $(\triangle^2 g_t)$, is an approximation to the second derivative of g at time t. The HP-filter is symmetric and can eliminate up to four unit roots in the data, see e.g. Cogley and Nason (1995). For quarterly data (the frequency used in most business-cycle studies) there seems to be a consensus in employing the value $\gamma = 1\ 600$. The gain of deviations from the HP trend using various values on γ is presented in Figure 3.5.

Figure 3.5: The gain of the HP filter using γ=100 (dashed line), γ=1 600 (solid line) and γ=50 000 (dotted line)[21]



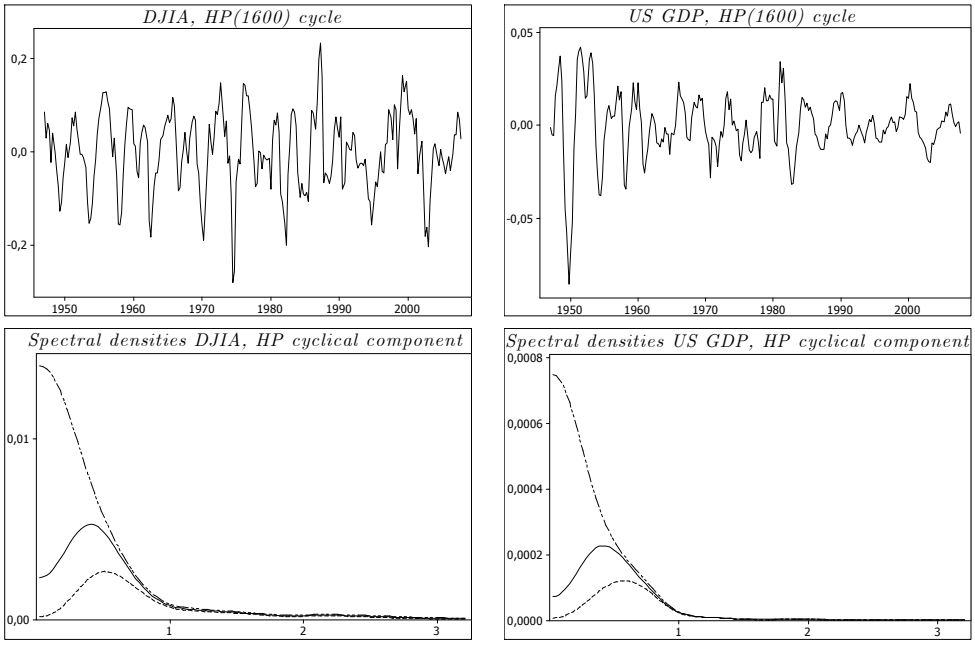
As with first differences and deviations from moving averages, the HP-filter dampens frequencies close to zero. King and Rebelo (1993) criticized the HP-filter and provided examples of how it alters measures of persistence, variability, and comovement when it is applied to observed time series. In addition, Harvey and Jaeger (1993) and Cogley and Nason (1995) showed

[21] The gain function of the HP filter that estimates the cycle is $G(w,\gamma) = 1 - \frac{1}{1+4\gamma(1-\cos w)^2}$.

that the HP filter induces spurious cycles when applied to the level of a random walk process. Criticism is also found in Maravall (1995) and Canova (1998). Because of this, Kaiser and Maravall (1999) provided a computationally convenient modification of the HP filter by including two model-based features.

The upper panel in Figure 3.6 shows the HP(1 600) detrended series and the corresponding spectral densities (lower panel). The spectral densities from the HP(100) and HP(50 000) filters are also included.

Figure 3.6: The HP(1 600) detrended DJIA and US GDP (upper panel). The lower panel shows the spectral densities of the cyclical components calculated from the HP filter using γ=100 (dashed line), γ=1 600 (solid line) and γ=50 000 (dashed/dotted line)



As shown in Figure 3.6, detrending using the HP($\gamma = 50\ 000$) failed. That is, much of the trend remains in the HP($\gamma = 50\ 000$) cyclical component. Using $\gamma = 100$, the cyclical components have almost no density at zero frequency and peaks at 12 quarters both for DJIA and US GDP. Using the standard value for quarterly data $\gamma = 1\ 600$, the spectral densities are larger at zero

frequency with the effect that the peaks are shifted towards lower frequencies and peaks at 16.5 quarters both for DJIA and US GDP (equivalent of frequency $w = 0.38$ in cross-spectral formulas, see again Jenkins and Watts (1968)). All series were found to be stationary and heteroscedastic.

3.2 The Baxter-King filter

The above filters are all approximations of an ideal high-pass filter, which would remove *only* the lowest frequencies from the data. The ideal band-pass filter removes *both* very low and very high frequencies, passing the intervening frequencies. Baxter and King (1999) proposed a moving average type approximation of the business cycle band defined by Burns and Mitchell (1946). That is, the BK filter is designed to pass through time series components with frequencies between 6 and 32 quarters, while dampening higher and lower frequencies. This is done using a symmetric finite odd-order $k = 2p+1$ moving average. The cyclical component using the BK-filter takes the form

$$y_t^{c,BK} = \sum_{j=-p}^{p} h_j B^j y_t, \qquad (3.2)$$

where $y_t^{c,BK}$ is the BK filtered series and h_j are the filter weights obtained by solving the optimization problem

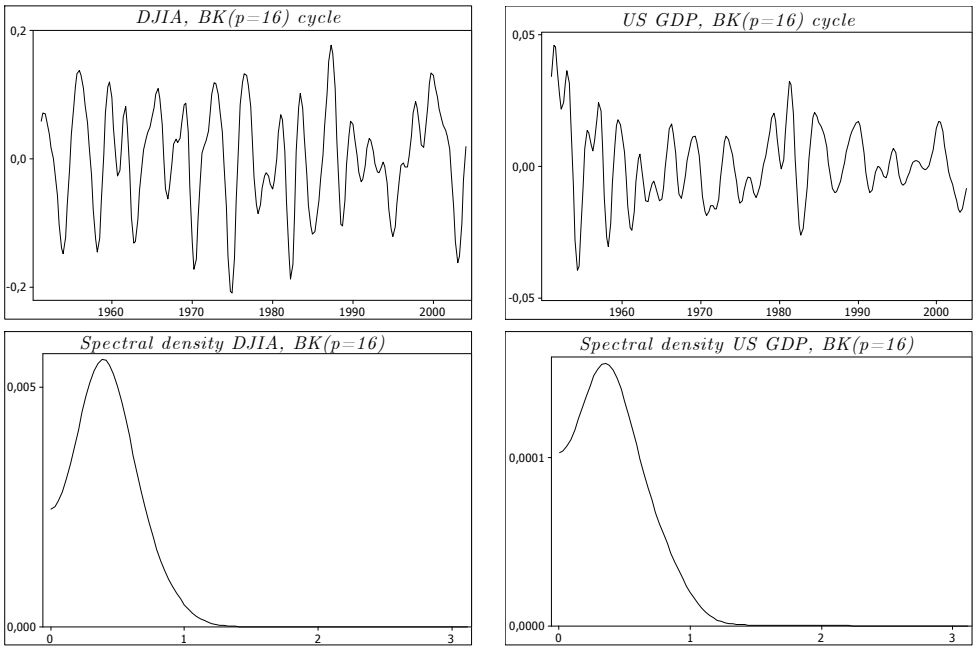
$$\min_{h_j} Q = \int_{-\pi}^{\pi} |\delta(w)|^2 \, dw, \qquad (3.3)$$

where $\delta(w) = \beta(w) - \alpha(w)$ is the error arising from approximating the Fourier transform of the ideal filter, $\beta(w)$, by an approximation of the same, $\alpha(w)$. For the solutions of (3.3), see Baxter and King (1999). Since the BK-filter is symmetric it does not induce phase shifts. Also, the filter is designed to produce stationary output. The BK filter has the ability to remove up to two unit roots. Ibid. suggested a value of $p = 12$ for the frequency band 6 to 32 quarters, and argued that the filtering is basically equivalent for larger values of p. Iacobucci and Noullez (2005) showed that the size of p beyond $p = 12$ matters, and suggested that one should select a value $p > 12$ irrespective of the sample size and the band to be extracted. For this reason, the properties

of the filtered series using $p = 12, 16, 24$ and 36 are studied here.

As with all moving average smoothers, p observations will be lost at the beginning and at the end of the filtered series. Figure 3.7 shows the cycles and spectral densities of the BK ($p = 16$) filtered series (using band-pass 6 to 32 quarters).

Figure 3.7: The cycle and spectral densities of the cyclical components calculated from the BK filter using p=16



The shape of the spectral densities using $p = 12$, 24 and 36 are very similar. As the window length becomes larger the peak is shifted slightly to the right. The spectral densities of the cyclical components from the BK filter are very similar in shape to ones from the HP filter.

The standard frequency band of 6 to 32 quarters used to extract business cycles seems to work well when applied to individual series. Section 5 will reveal that this choice of band greatly influences the shape of the coherency and phase functions, especially at frequencies shorter than 6 quarters. This might remove important high frequency comovements encouraging an extension of

the BK frequency band to between 2 to 32 quarters. This does not change the frequency domain properties of the individual filtered series much, see Table 3.1. The two BK filter alternatives are hereafter denoted $\text{BK}^{(6,32)}(p)$ and $\text{BK}^{(2,32)}(p)$. The spectral densities of the BK filter using frequency band 1 to 32, showed peaks at zero frequency for $p = 24$ and $p = 36$, and will therefore not be included in this study.

3.3 The Beveridge-Nelson filter

The model-based decomposition suggested by Beveridge and Nelson (1981) is based on Wold's representation theorem and separates a time series into a permanent (P) and transitory (T) component. A shock at time t results in a permanent change to the series if it affects the permanent component, while the effect of the shock will dampen down over time if it affects the transitory component. The series, y_t is thus decomposed as follows:

$$y_t = P_t + T_t.$$

It is further assumed that y_t is an $ARIMA(p, 1, q)$ process (and thus $\Delta y_t = \Delta P_t + \Delta T_t$). The first difference, Δy_t, of an $ARIMA(p, 1, q)$ process can be expressed as an infinite order moving average process

$$\begin{aligned} \Delta y_t &= c(B)a_t \\ &= c_0 a_t + \psi(B)(1 - B)a_t, \end{aligned}$$

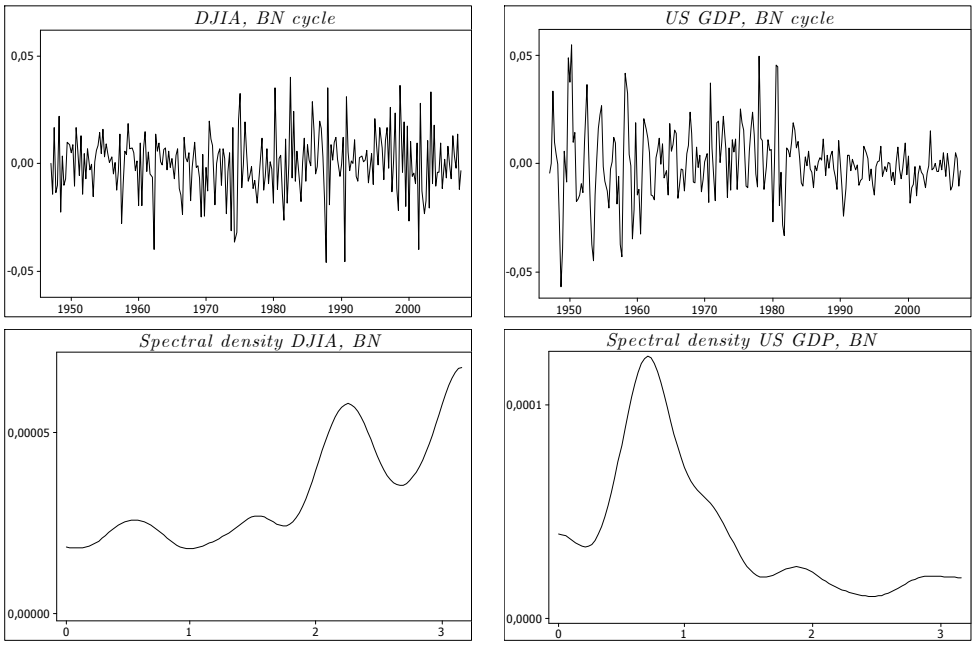
where $\psi(B) = \psi_0 + \psi_1(B) + ...$, is an infinite order polynomial. The permanent and transitory components are identified as

$$\begin{aligned} \Delta P_t &= c_0 a_t \\ \Delta T_t &= \psi(B)(1 - B)a_t, \end{aligned}$$

Note that $T_t = \psi(B)a_t$, thus P_t is a process of integrating order one, $I(1)$, whereas T_t is $I(0)$. That is, one difference is required to make P_t stationary whereas T_t is stationary by definition. There are basically two ways to estimate the BN components (Morley, 2007). Here the approach suggested by

Beveridge and Nelson (1981) is used. Figure 3.8 shows the cyclical series and the corresponding spectral densities.

Figure 3.8: Cyclical components calculated from the BN filter (upper panel) and the corresponding spectral densities (lower panel)[22]



Note the high-pass properties of the BN filter for the DJIA, for which the filter eliminates most of the low frequency variation, including the business cycles. The first hump represents the business cycle, almost overshadowed by the high frequency variations. The spectral densities of the US GDP series are quite similar in shape to the ones in Figures 3.6 and 3.7. As before, heteroscedasticity is revealed in the upper panel of Figure 3.8, as noted before, especially in the US GDP series.

[22] The models with the smallest AIC among the adequate ones were an ARIMA(1,1,1) and ARIMA(2,1,2), for DJIA and US GDP, respectively.

3.4 A trend and heteroscedasticity removing filter

None of the above filters accounts for the highly possible event of heteroscedasticity in economic and financial series. Despite first order stationarity in the detrended series, the null hypothesis of homoscedasticity is rejected for every one of them. This is a major drawback when applying a frequency domain approach (as in this study), see e.g. Engle (1974). Because of this, Stockhammar and Öller (2007) proposed the following detrending and heteroscedasticity removing filter

$$\widetilde{z}_t = s_y \left[\frac{\left(z_t^{(i)}\right)^d}{HP^{(\gamma)} \left(\sqrt{\sum_{\tau=t-\nu}^{t+\nu} \left(z_\tau^{(i)}\right)^{2d} / 2\nu} \right)} \right] + \overline{y}, \qquad (3.4)$$

where $t = \max[k-\eta, l-\nu], \max[k-\eta+1, l-\nu+1], ...$ with k and l (both odd) as the window lengths in the numerator and denominator, respectively. $\eta = (k-1)/2$, $\nu = (l-1)/2$ and $i = a, b$ from the detrending operations

$$(a) \qquad z_t^{(a)} = \Delta y_t - \sum_{\tau=t-\eta}^{t+\eta} \Delta y_\tau \, / k \, , \qquad t = \eta+1, \eta+2, ..., n-\eta \qquad (3.5a)$$

and with y_τ delayed one period:

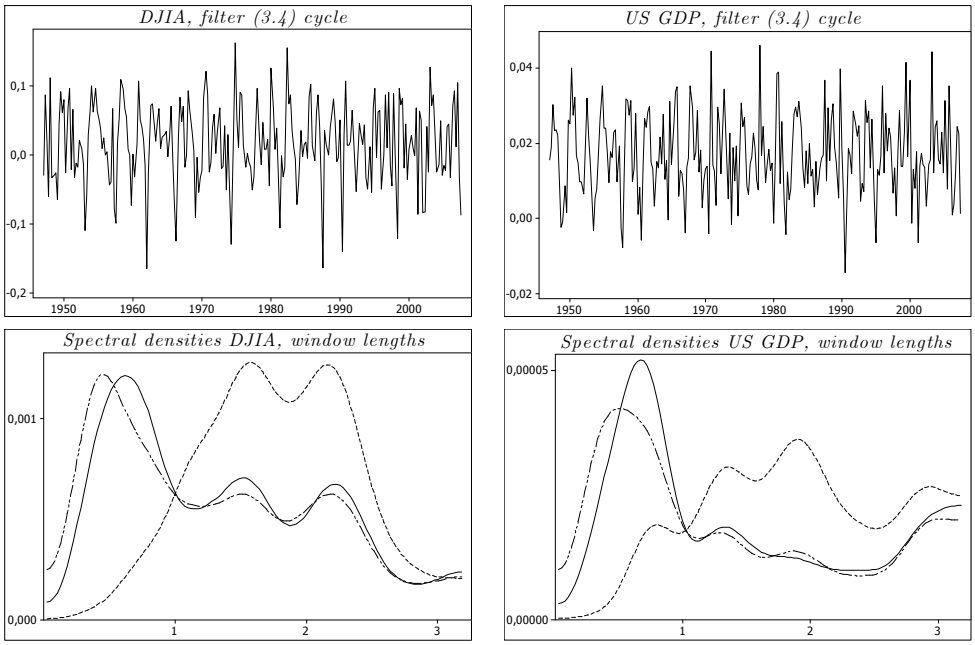
$$(b) \qquad z_t^{(b)} = \Delta y_t - \sum_{\tau=t-\eta}^{t+\eta} \Delta y_{\tau-1} \, / k \, , \qquad t = \eta+2, \eta+3, ..., n-\eta+1 \qquad (3.5b)$$

where y_t is the logarithmic series at time t. This transformation is generalized by raising $z_t^{(i)}$ to the possibly non-integer power d, $\left(z_t^{(i)}\right)^d$. In this study, however, $d = 1$ has been used. Using different values on η in (3.5), different degrees of integration are achieved. There are two extremes. For $\eta = (n-1)/2$, the term $\sum_{\tau=t-\eta}^{t+\eta} \Delta y_\tau \, / k$ equals $\overline{\Delta y}$ assuming that the original series is $I(1)$ centered at zero. The other extreme appears when k equals one, that is $\eta = 0$. Operation (3.5b) is used only in the latter case and is equivalent to the second difference operation, $\Delta^2 y_t$. The choice of η depends

on the series studied. If it is close to $I(1)$ then you should just choose η close to $(n-1)/2$, and if the series is close to $I(2)$ then choose $\eta = 0$ in (b) or a small value on η in (a). The operations in (3.5) contain first differencing implying a small phase shift, but everything else is symmetrical.

Filter (3.4) was designed to remove heteroscedasticity in time series data in a simple, yet efficient way. As proposed by Hodrick and Prescott (1997), $\gamma = 1\ 600$ is suitable for quarterly data, Stockhammar and Öller (2007) further suggest the use of $k = l = 15$ and $d = 1$ in (3.4). The window length, l, is hereafter assumed to be equal to 15. The upper panel in Figure 3.9 shows the filtered series using the proposed filter. The spectral density functions of the two time series filtered by (3.4) (using $\gamma = 1\ 600$ and different window lengths, k) are presented in the lower panel.

Figure 3.9: The detrended and heteroscedasticity corrected filtered series using γ=1 600, k=l=15 and d=1 (upper panel). The lower panel shows the corresponding spectral densities using k=5 (dashed line), k=15 (solid line) and k=25 (dashed/dotted line)



Both filtered series are found to be stationary according to ADF and PP

tests, and (contrary to all other filters discussed in this section) homoscedastic according to the ARCH-LM test. The window length proposed in Stockhammar and Öller (2007), $k = 15$, results in a spectral peak at 10.4 quarters both for DJIA and US GDP. Increasing the window length to $k = 25$ shifts the peaks close to 14 quarters.

Table 3.1 summarizes the main spectral properties of the above filters. The peak frequencies are measured in quarters.

Table 3.1: Spectral density peaks

	DJIA		US GDP	
	Peak freq. (quarters)	value	Peak freq. (quarters)	value
FD	12.8	0.0009	10.6	0.00005
MA($k = 3$)	7.9*	0.0002	9.0*	0.00001
MA($k = 9$)	9.8	0.0017	9.8	0.00011
MA($k = 15$)	13.5	0.0046	11.1	0.00024
HP($\gamma = 100$)	12.0	0.0025	12.0	0.00013
HP($\gamma = 1\ 600$)	16.5	0.0052	16.5	0.00023
HP($\gamma = 50\ 000$)	–	–	–	–
$BK^{(6,32)}(12)$, $BK^{(2,32)}(12)$	16.9 15.7	0.0054 0.0054	18.3 18.3	0.0002 0.0002
$BK^{(6,32)}(16)$, $BK^{(2,32)}(16)$	16.5 16.3	0.0056 0.0055	17.8 17.7	0.0002 0.0002
$BK^{(6,32)}(24)$, $BK^{(2,32)}(24)$	14.9 15.1	0.0048 0.0049	16.2 15.1	0.0001 0.0001
$BK^{(6,32)}(36)$, $BK^{(2,32)}(36)$	14.5 14.3	0.0045 0.0044	15.8 15.6	0.0001 0.0001
BN	11.6*	0.0000	9.0	0.00012
(3.4),($k = 5$,$\gamma = 1\ 600$)	4.2	0.0012	7.6*	0.00002
(3.4),($k = 15$,$\gamma = 1\ 600$)	10.4	0.0011	10.4	0.00005
(3.4),($k = 25$,$\gamma = 1\ 600$)	14.1	0.0011	14.1	0.00004
Mean	**13.2**		**13.6**	

*The high frequency peaks are assumed to be spurious, cf. Figures 3.4, 3.8 and 3.9.

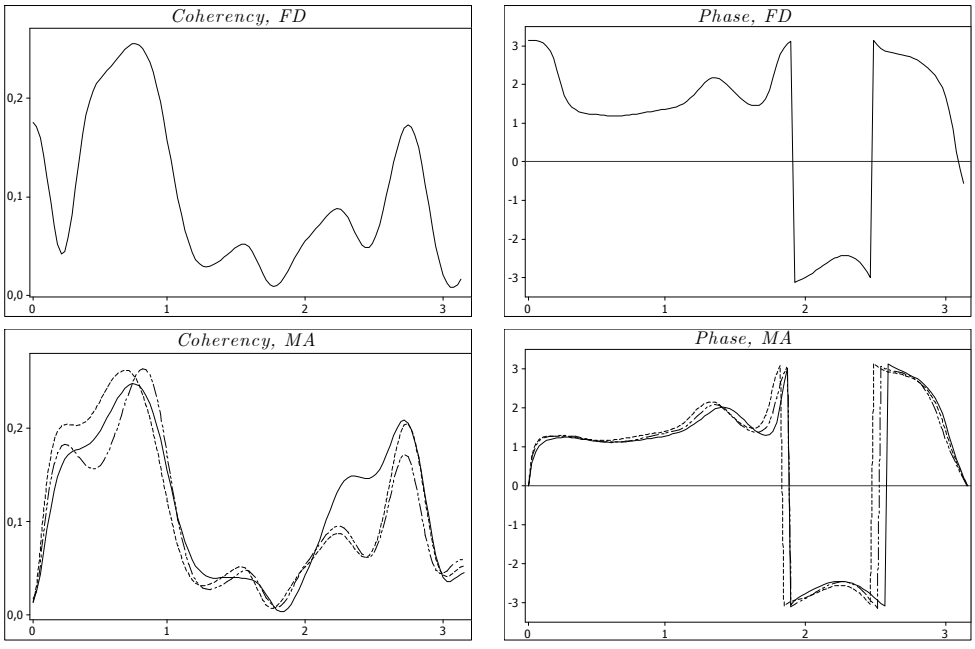
As summarized in Table 3.1, the length of the business cycle depends on the choice of detrending filter. The spectral peaks for MA($k = 9$), HP($\gamma = 100$), HP($\gamma = 1\ 600$) and filter (3.4) (using $\gamma = 1\ 600$, and $k = 15, 25$), are located at the same frequency. That is a favourable feature improving the estimates of cross-spectral densities. The individual spectral densities of the $BK^{(6,32)}(p)$ and $BK^{(2,32)}(p)$ filtered series are similar, but both seem rather nonrobust to different window lengths (with peak frequencies located between 14 and 18 quarters). This accords well with the HP($\gamma = 1\ 600$) filter. Accounting for

heteroscedasticity using filter (3.4) and the suggested window length $k = 15$, slightly shortens the cycle.

4. Comovements between the two series

The choice of detrending filter not only affects the shape of the individual spectral densities, but also the cross-spectral functions. The cross-spectral differences of the detrending procedures are the issue of this section. Figure 4.1 shows the coherency and phase spectra between the cyclical components of DJIA and US GDP using the FD filter and the MA filter. In creating all subsequent phase spectra, the DJIA series have been put before US GDP.

Figure 4.1: Cross-spectral densities of the cyclical components of DJIA and US GDP using the FD filter (upper panel) and the MA(k =3) (dashed line), MA(k=9) (solid line) and MA(k=15) (dashed/dotted line) (lower panel)

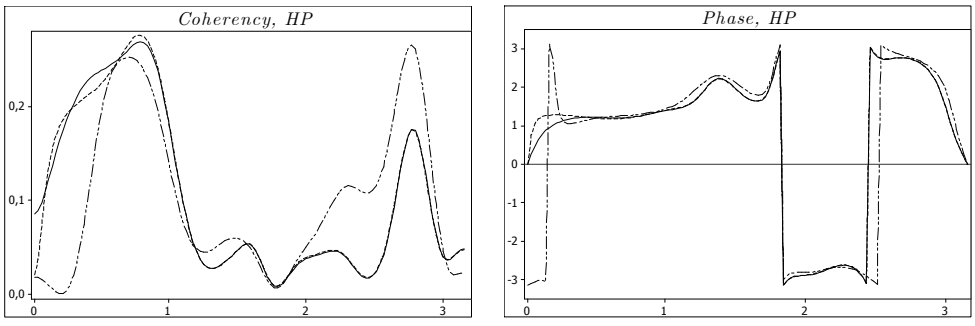


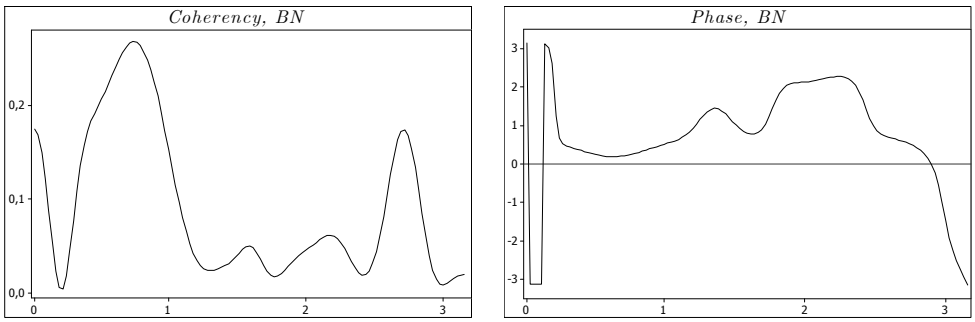
The coherency function using the FD filter in Figure 4.1 (upper panel) peaks ($K^2_{1,2}(w = 0.72) = 0.26$) at 8.7 quarters, which means that the relationship

between the two series is closest at a frequency of just over two years. The spectra of the two individual series, $f_1(w)$ and $f_2(w)$, peak at $10-12$ quarters, essentially the frequency to focus on in the coherency plots. At these frequencies, the approximate coherency is $K^2_{1,2}(w) = 0.23$. Extending the window length in the MA filter shifts the coherency peaks to the right.

The frequency to focus on in the phase spectrum is the peak coherency frequency. In most cases in this study this corresponds to a relatively linear (positive) part of the phase spectrum, see e.g. Figure 4.1. The slope is estimated using linear regression on the frequency of interest and four observations on each side. Most filtered series also indicate rather high coherency at high frequencies (around $2-2.5$ quarters). At this frequency, the trend in the phase spectra is typically negative and is again estimated using a linear regression on nine observations surrounding the (high frequency) peak coherency frequency. Put in practice, the phase spectrum for the FD filtered series indicates that DJIA leads US GDP on the business cycle frequencies by on the average of 0.45 quarters. It also shows signs of feedback at high frequencies, where US GDP leads DJIA by on average 1.34 quarters. This could be due to the fact that share holders tend to follow early indicators and react on unexpected values. It might also be an alias feedback, or perhaps a combination of them both. Figure 4.2 shows the same measures using the HP and BN filters.

Figure 4.2: Cross-spectral densities of the cyclical components of DJIA and US GDP using the HP filter with γ=100 (dashed line), γ=1 600 (solid line) and γ=50 000 (dashed/dotted line) (upper panel) and using the BN filter (lower panel)

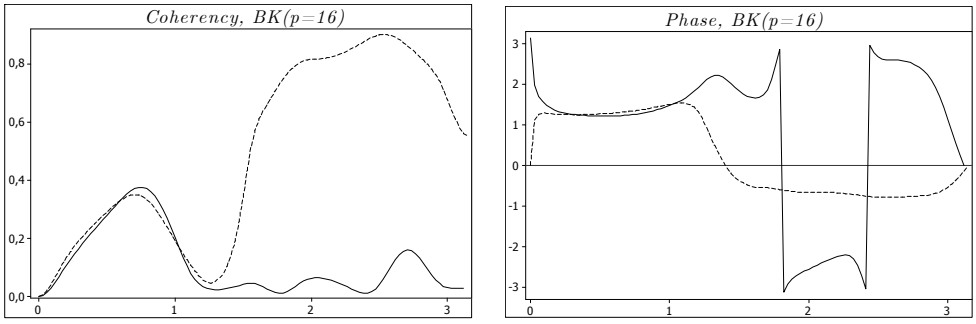




In Figure 4.2 (upper panel), the coherency function using the HP(100) and HP(1 600) peaks at 8.1 quarters. Increasing γ to 50 000 slightly shifts the peak to the left. Using HP(1 600), the spectra of the two individual series, $f_1(w)$ and $f_2(w)$, peak at around 16.5 quarters, for which the coherency is $K^2_{1,2}(w = 0.38) = 0.23$. The coherency function using the BN filter is similar, but note that its phase spectrum does not have the typical discontinuities around $w = 1.8$ and $w = 2.5$. Also, both the HP(50 000) and the BN filtered series have low frequency discontinuities in the phase spectra. The phase function is defined as the arctan of the ratio between the quadrature and the co-spectrum resulting as discontinuities at frequency multiples of $\frac{\pi}{2}$, see e.g. Jenkins and Watts (1968) for details.

In Figure 4.3 the coherency and phase spectra for the $\text{BK}^{(2,32)}(p = 16)$ and $\text{BK}^{(6,32)}(p = 16)$ filtered series are presented. The densities using other values on p show very similar patterns.

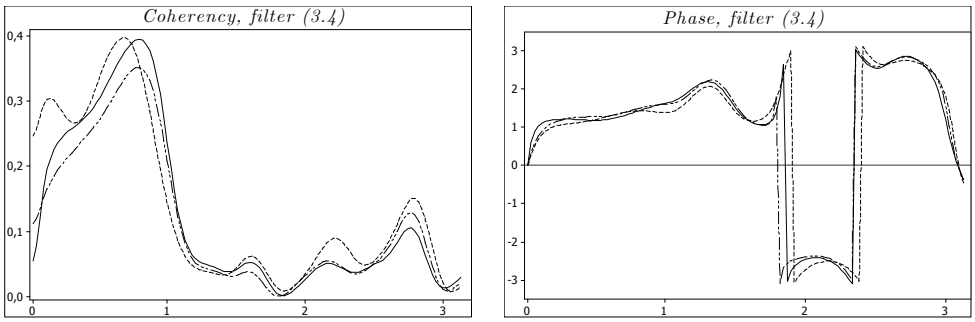
Figure 4.3: Cross-spectral densities of DJIA and US GDP using the $BK^{(2,32)}(p=16)$ (solid line) and $BK^{(6,32)}(p=16)$ (dashed line) filter



The choice of frequency band to extract in the BK filter has a large effect on the coherency and phase. The BK standard frequency band, $6-32$ quarters, by definition completely misses the information that can be extracted at high frequencies. This results in a peculiar shape of the coherency function which peaks far from the business cycle. As with the BN filter, the phase spectrum does not have the typical discontinuity at high frequencies. On the contrary, the cross-spectral densities of the $\mathrm{BK}^{(2,32)}(p = 16)$ filtered series are similar to the filtered series discussed above and below. The smallest coherency peak frequencies in the study were found for the $\mathrm{BK}^{(2,32)}(p = 36)$ and $\mathrm{BK}^{(6,32)}(p = 36)$ filtered series.

The cross-spectral densities between the two series detrended by filter (3.4) with $\gamma = 1\ 600$ and $k = 5$, 15 and 25 are shown in Figure 4.4.

Figure 4.4: Cross-spectral densities of DJIA and US GDP using filter (3.4) with γ=1 600 and k=5 (dashed line), k=15 (solid line) and k=25 (dashed/ dotted line)



The coherency function for the (3.4) filtered series (using $\gamma = 1\ 600$ and $k = 15$) peaks on the average at 8.2 quarters at which $K^2_{1,2}(0.76) = 0.40$. At this frequency, $\phi_{1,2}(0.76) = 1.20$ quarters, with feedback $\phi_{1,2}(2.77) = -1.48$. At the peak spectral frequencies of 10.4 quarters (see Table 3.1), $K^2_{1,2}(0.60) = 0.35$. The results are summarized in Table 4.1.

Table 4.1: Cross-spectral density peaks

	Coherency			Phase	
	Peak freq.(q)	Value	Value at peak spectral freq.	At peak coh. freq.(q)	At high freq.(q)
FD	8.7	0.26	0.23	0.45	−1.34
MA($k = 3$)	9.4	0.26	0.25	0.49	−1.59
MA($k = 9$)	8.7	0.25	0.24	0.33	−1.84
MA($k = 15$)	7.8	0.26	0.17	0.73	−1.51
HP($\gamma = 100$)	8.1	0.28	0.23	0.65	−1.45
HP($\gamma = 1\ 600$)	8.1	0.27	0.23	0.51	−1.40
HP($\gamma = 50\ 000$)	9.0*	0.25	–	0.26	−1.39
$BK^{(6,32)}(12)$	8.5*	0.29	0.21	0.35	−0.23
$BK^{(6,32)}(16)$	8.9*	0.35	0.23	0.41	−0.10
$BK^{(6,32)}(24)$	8.1*	0.39	0.24	0.24	0.14
$BK^{(6,32)}(36)$	7.6*	0.35	0.17	0.09	0.13
$BK^{(2,32)}(12)$	8.5	0.28	0.22	0.16	2.54
$BK^{(2,32)}(16)$	8.2	0.38	0.21	0.51	−1.02
$BK^{(2,32)}(24)$	8.5	0.36	0.26	−0.18	−0.44
$BK^{(2,32)}(36)$	7.8	0.35	0.17	−0.23	−1.45
BN	8.7	0.27	0.24	0.67	−1.49
(3.4),($k = 5$,$\gamma = 1\ 600$)	9.3	0.40	0.10	0.73	−1.35
(3.4),($k = 15$,$\gamma = 1\ 600$)	8.2	0.40	0.35	1.20	−1.48
(3.4),($k = 25$,$\gamma = 1\ 600$)	8.1	0.35	0.26	0.99	−1.02
Mean	**8.4**	**0.32**	**0.22**	**0.44**	**−0.86**

*The high frequency peaks are assumed to be spurious, cf. Figure 4.2 and 4.3.

The column to the far right shows the phase shift at the rightmost coherency peak frequency, see Figures 4.1-4.4. The coherency seems quite robust to different filters. Its peak frequencies vary from 7.6 to 9.4 quarters, with an average of 8.4 quarters. The choice of $BK^{(6,32)}(p)$ or $BK^{(2,32)}(p)$ does not seem to matter much, but this is true only for the Burns and Mitchell (6 to 32 quarters) business cycle frequencies. Due to the extended high frequency band, the $BK^{(2,32)}(p)$ has the ability to also describe variations at shorter frequencies, see Figure 4.3. It is therefore in comovement studies advisable to use this filter (if the series are homoscedastic). The phase at peak coherency frequency is less robust with values varying from −0.23 to 1.20 quarters (average 0.44). All filters (with the exceptions of the $BK^{(2,32)}(24)$ and $BK^{(2,32)}(36)$ filters) reports that DJIA leads US GDP at peak coherency frequency. Both BK filters show scattered phase. Accounting for heteroscedasticity using filter (3.4) shows coherency peaks approximately on average frequency, with

larger than average coherency values. Also, the homoscedastic series induce the longest lead shifts at peak coherency frequency.

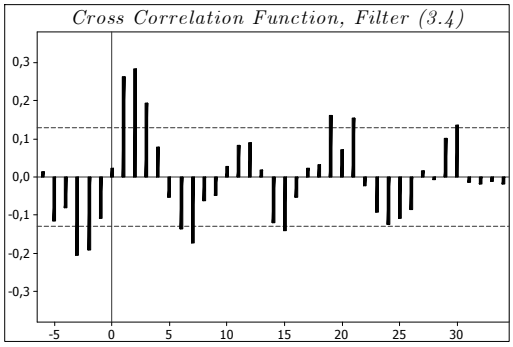
It is possible from cross-spectral analysis to detect lead or lag of the series under a common cyclical period. The phase densities show both the lead times and reveal the direction of the comovements. As a double check of the results, the time domain Granger-causality test was performed on the proposed (3.4) filtered and stationary series (using $\gamma = 1600$, $k = l = 15$ and $d = 1$). The results are presented in table 4.2.

Table 4.2: Granger-causality tests, p-values

\Lag	1	2	3	4	5	6	7	8
DJIA doesn´t Granger-cause US GDP	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00
USGDP doesn´t Granger-cause DJIA	0.08	0.02	0.00	0.12	0.02	0.14	0.36	0.45

Table 4.2 shows that the null hypothesis of Granger-noncausality is rejected for all lags from DJIA to US GDP, using the 0.05 significance level. At lags $2, 3$ and 5, there seem to exist a feedback - US GDP leads DJIA. This further confirms the phase densities in Figures 4.1-4.4 and is also supported by the cross correlation function between the proposed (3.4) filtered series:

Figure 4.5: The cross correlations between DJIA and US GDP



where the dashed lines denotes ± 2 standard errors for the estimates. Table 4.3 presents the cross correlations (at lag 1) between DJIA and US GDP for the entire period, and for three subperiods (standard errors in parentheses

and significant correlations in bold figures). The subperiods were chosen in accordance with US GDP volatility, where 1947-1960, 1961-1983 and 1984-2007 are periods denoted as high, medium and low volatility, respectively. This is also in compliance with Stock and Watson (2003) who reported that the US GDP variance declined over 50 percent from 1960-1983 to 1984-2002 when averaged over four quarters. The decline in volatility was even larger in e.g. Italy and Japan, and also widespread across sectors within the US. As indicated by Stockhammar and Öller (2007 and 2008), the volatility in the US GDP was even larger before 1960. Note that the decreasing volatility does not apply to DJIA, cf. Figures 3.1, 3.4, 3.6, 3.7 and 3.8.

Table 4.3: First order cross correlations between DJIA and US GDP

	Full sample	1947-1960	1961-1983	1984-2007
FD	**0.173** (0.064)	**0.534** (0.135)	0.167 (0.104)	0.125 (0.102)
MA($k=3$)	**0.151** (0.064)	**0.448** (0.134)	0.061 (0.104)	0.060 (0.102)
MA($k=9$)	**0.335** (0.064)	**0.559** (0.134)	**0.339** (0.104)	0.139 (0.102)
MA($k=15$)	**0.336** (0.064)	**0.574** (0.134)	**0.294** (0.104)	0.152 (0.102)
HP($\gamma=100$)	**0.338** (0.064)	**0.642** (0.134)	**0.294** (0.104)	0.074 (0.102)
HP($\gamma=1\ 600$)	**0.329** (0.064)	**0.573** (0.134)	**0.250** (0.104)	0.196 (0.102)
HP($\gamma=50\ 000$)	0.092 (0.064)	**0.309** (0.134)	−0.190 (0.104)	0.112 (0.102)
$BK^{(6,32)}(12)$, $BK^{(2,32)}(12)$	**0.285** (0.067) **0.275** (0.067)	**0.514** (0.151) **0.537** (0.151)	**0.225** (0.104) **0.242** (0.104)	0.083 (0.109) 0.048 (0.109)
$BK^{(6,32)}(16)$, $BK^{(2,32)}(16)$	**0.309** (0.069) **0.288** (0.069)	**0.621** (0.158) **0.597** (0.158)	**0.231** (0.104) **0.208** (0.104)	0.039 (0.112) 0.066 (0.112)
$BK^{(6,32)}(24)$, $BK^{(2,32)}(24)$	**0.223** (0.071) **0.218** (0.071)	**0.606** (0.177) **0.578** (0.177)	0.193 (0.104) 0.150 (0.104)	**−0.321** (0.118) **−0.265** (0.118)
$BK^{(6,32)}(36)$, $BK^{(2,32)}(36)$	0.131 (0.076) 0.133 (0.076)	0.335 (0.224) **0.750** (0.224)	**0.368** (0.104) 0.018 (0.104)	**0.484** (0.129) **−0.427** (0.129)
BN	**0.170** (0.064)	0.192 (0.134)	**0.318** (0.104)	0.080 (0.104)
(3.4),($k=5$,$\gamma=1\ 600$)	**0.168** (0.064)	0.139 (0.134)	0.044 (0.104)	0.066 (0.102)
(3.4),($k=15$,$\gamma=1\ 600$)	**0.263** (0.064)	**0.275** (0.134)	**0.222** (0.104)	0.090 (0.102)
(3.4),($k=25$,$\gamma=1\ 600$)	**0.236** (0.064)	0.210 (0.134)	0.166 (0.104)	0.077 (0.102)

Table 4.3 further corroborates that the detrending method used has effect on estimated cross correlations. This is expected due to the one-to-one relationship between cross covariances and spectral densities. In addition, Table 4.3 shows that the cross correlation generally decreases with US GDP volatility, but less so using filter (3.4). The choice of $BK^{(6,32)}(p)$ or $BK^{(2,32)}(p)$ has

little effect on correlations except when $p = 36$. Using $p = 24$, both filtered series show negative cross correlations in the low volatility part. Odd looking negative correlation is also present in the medium volatility part using the HP(50 000) filter. The BN filter shows for the same period the maximum cross correlation. This could be due to the cross correlation between opposite cyclical phases in the two series (see the feedback discussion above).

5. Conclusions

This paper reveals frequency domain relationships between the Dow Jones industrial average stock prices and US GDP growth. Both series are heteroscedastic, making standard detrending procedures, such as Hodrick-Prescott or Baxter-King, inadequate. Neglecting the heteroscedasticity distorts frequency domain results and induces inefficient estimation of the spectral densities. Surprisingly, many frequency domain studies do not take notice of this and mechanically use standard detrending filters. Prior to the comovement study, the univariate and comovement frequency domain results from these filters are compared to the results from the heteroscedasticity removing filter suggested by Stockhammar and Öller (2007). Thus, the effect of the often neglected heteroscedasticity is measured.

Accounting for the heteroscedasticity somewhat shortens the business cycles. No matter which filter is used, significant comovements exist between the DJIA and US GDP series. The coherency seems quite robust to the different filters. Accounting for heteroscedasticity slightly shifts the coherency peak to the left and with larger than average coherency values. The phase shift is less robust, especially for the BK filtered series. Most filters report that DJIA leads US GDP at peak coherency frequency ($7.6 - 9.4$ quarters), but also reveal a feedback from US GDP to DJIA at around $2 - 2.5$ quarters. The filtered series using the suggested heteroscedasticity removing filter induce the longest lead shifts (1.2 quarters) at peak coherency frequency, and also above average feedback lag (1.48 quarters). Using the BK filter with frequency band 6 to 32 quarters (as first suggested by Burns and Mitchell, 1946) by definition completely misses this information. The same applies to the BN filter. It is therefore advisable to extend the frequency bands to the interval 2 to 32 quarters in comovement studies like this (under the condition that the series are homoscedastic). The frequency domain results were corroborated in the time domain using cross correlations and Granger-causality

tests. When applied on subperiods in accordance with US GDP volatility, most filtered series showed scattered first order cross correlations, but less so in the homoscedastic series.

Thus, the choice of detrending filter affects both univariate and bivariate frequency domain (and time domain) results. More importantly, heteroscedasticity matters and must be eliminated prior to comovement studies like this one.

Acknowledgments

This research was supported by the Department of Statistics at Stockholm University. I gratefully acknowledge helpful comments from Professor Lars-Erik Öller and Daniel Thorburn of Stockholm University.

References

Baxter, M. and King, R. G. (1999) Measuring business-cycles: Approximate band-pass filters for economic time series. *The Review of Economics and Statistics*, 81, 575-593.

Beveridge, S. and Nelson, C. R. (1981) A new approach to the decomposition of economic time series into permanent and transitory components with particular attention to measurement of the business cycle. *Journal of Monetary Economics,* 7, 151-174.

Burns, A. F. and Mitchell, W. C. (1946) Measuring business cycles. *National Bureau of Economic Research*. New York.

Candelon, B. and Gil-Alana, L. A. (2004) Fractional integration and business cycle features. *Empirical Economics,* 60, 343-359.

Canova, F. (1998) Detrending and business cycle facts. *Journal of Monetary Economics*, 41, 475-512.

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.

Cogley, T. (2001) Alternative definitions of the business cycle and their implications for the business cycle models: a reply to Torben Mark Pederson. *Journal of Economic Dynamics and Control*, 25, 1103-1107.

Christiano, L. and Fitzgerald, T.J. (2003) The band-pass filter. *International Economic Review*, 44, 435-465.

Engle, R. F. (1974) Band spectrum regression. *International Economic Review*, 15, 1-11.

Granger, C. W. J. (1966) The typical spectral shape of an economic variable. *Econometrica*, 34, 150-161.

Harvey, A. C. and Jaeger, A. (1993) Detrending, stylized facts and the business cycle. *Journal of Applied Econometrics*, 8, 231-247.

Hodrick, R. J. and Prescott, E. C. (1997) Postwar U.S. business cycles: An empirical investigation. *Journal of Money, Credit and Banking*, 29, 1-16.

Jenkins, G. M. and Watts, D. G. (1968) *Spectral analysis and its applications.* Holden Day, San Francisco.

Kaiser, R. and Maravall, A. (1999) Estimation of the business cycle: A modified Hodrick-Prescott filter. *Spanish Economic Review*, 1, 175-206.

King, R. G. and Rebelo, S. (1993) Low frequency filtering and real business-cycles. *Journal of Economic Dynamics and Control,* 17, 207-31.

Knif, J., Pynnönen, S. and Luoma, M. (1995) An analysis of lead-lag structures using a frequency domain approach: Empirical evidence from the Finnish and Swedish stock markets. *European Journal of Operational Research*, 81, 259-270.

Lee, C.-C. (2005) Energy consumption and GDP in developing countries: A cointegrated panel analysis. *Energy Economics*, 27, 415-427.

Iacobucci, A. and Noullez, A. (2005) A frequency selective filter for short-length time series. *Computational Economics*, 25, 75-102.

Ma, S-R. and Park, S-B. (2004) An analysis of co-movements and causality of international interest rates: The case of Korea, Japan and the U.S. *International Journal of Applied Economics*, 1, 98-114.

Maravall, A. (1995) Unobserved components in economic time series. In: Pesaran, M. H. and Wickens, M. R. (eds.). *The Handbook of Applied Econometrics.* vol. 1. Basil Blackwell, Oxford.

Morley, J. C. (2007) The two interpretations of the Beveridge-Nelson decomposition. *Working paper*, University of Washington.

Öller, L.-E. (1990) Forecasting the business cycle using survey data. *International Journal of Forecasting*, 6, 453-461.

Orlov, A. G. (2009) A cospectral analysis of exchange rate comovements during Asian financial crisis. *Journal of International Financial Markets, Institutions and Money*, 19, 742-758.

Percival, D. B. and Walden, A. T. (1993) *Spectral analysis for physical applications. Multitaper and conventional univariate techniques*, Cambridge University Press, Cambridge.

Priestley, M. B. (1981) *Spectral analysis and time series, vol. 1: univariate series.* Academic Press, London.

Stock, J. H. and Watson, M. W. (1999) *Business cycle fluctuations in US macroeconomic time series*. Handbook of Macroeconomics, John Taylor and Michael Woodford, eds. Elsevier Science, Amsterdam.

Stock, J. H. and Watson, M. W. (2003) *Has the business cycle changed? Evidence and explanation*s. In: Monetary Policy and Uncertainty: Adapting to Changing Economy Symposium Proceedings, Federal Reserve Bank of Kansas City, pp. 9-56.

Stockhammar, P. and Öller, L.-E. (2007) A simple heteroscedasticity removing filter. *Research Report 2007:1*, Department of Statistics, Stockholm University.

Stockhammar, P. and Öller, L.-E. (2008) On the probability distribution of economic growth. *Research Report 2008:5*, Department of Statistics, Stockholm University.

Stockhammar, P. and Öller, L.-E. (2010) Density forecasting of the Dow Jones stock index. *Research Report 2010:1*, Department of Statistics, Stockholm University.

Uebele, M. and Ritschl, A. (2009) Stock markets and business cycle comovement in Germany before World War I: Evidence from spectral analysis. *Journal of Macroeconomics*, 31, 35-57.

Wilson, P. and Okunev, J. (1999) Spectral analysis of real estate and financial assets markets. *Journal of Property Investment and Finance*, 17, 61-74.