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

Pär Stockhammar

Stockholm University

Contents

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

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

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 $\left|h\left(e^{jw}\right)\right|^2$ 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

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

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 $E\left(\left|Y_t^2\right|\right) < \infty$. 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.

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

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

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

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

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

which is obviously different from the OLS variance, $\sigma^2(\mathbf{X'X})^{-1}$. 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 $g(v_t) = \theta v_t + \delta \left[|v_t| - E\,|v_t| \right]$ and $v_t = \varepsilon_t \,/ \sqrt{h_t}$. Because of the log-linear 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.