Accurate Risk and Density Prediction of Asset Prices

Dissertation

der Wirtschaftswissenschaftlichen Fakultät
der Universität Zürich

zur Erlangung der Würde
eines Doktors der Ökonomie

vorgelegt von
Sven Christian Steude
von Heikendorf/Deutschland

genehmigt auf Antrag von
Prof. Dr. Marc S. Paolella und Prof. Michael Wolf, Ph.D.

Die Wirtschaftswissenschaftliche Fakultät der Universität Zürich gestattet hierdurch die Drucklegung der vorliegenden Dissertation, ohne damit zu den darin ausgesprochenen Anschauungen Stellung zu nehmen.

Zürich, 13. Juni 2007

Der Dekan: Prof. Dr. H. P. Wehrli

Table of Contents

List of Figures

List of Tables

Summary

For most investors and banking regulators the down side loss potential of a financial position is of great interest, which can be quantified though the widely used measure Value–at–Risk. Chapter 2 introduces a method in which this down side of the return distribution receives more attention by using special weighting functions. Up to now, in empirical finance, the overwhelming majority of time series models for density and Value–at–Risk forecasting are estimated with traditional maximum likelihood methods. By definition, the likelihood implicitly involves placing equal weight on each of the observations in the sample, but this need not be the case. For example, Mittnik and Paolella (2000) have demonstrated that using a weighted likelihood scheme in which more weight is placed on observations in the recent past results in considerable forecast improvement compared to the default of equal weights. This is presumably because the data generating process is not constant through time. Now, if instead of accurate forecasting of the entire density, interest is restricted to just Value–at–Risk, then it would seem wise to place more weight on the negative observations in the sample which, as shown in this Chapter, yields to a considerable improvement in forecast accuracy.

In Chapter 3 a new and quite general class of models for asset returns is proposed. The structure of the model is such that the two most characteristic stylized facts of financial time series are modelled, which are i) that risk, measured for example through volatility, comes in clusters and is therefore not constant through time and ii) that more severe market movements are observed (so called fat tails) than suggested by the Gaussian assumption which served as a convenient model for many years.

The proposed model combines a dynamic multi–component generalized autoregressive conditional heteroscedastic (GARCH) structure (to tackle volatility cluster) with the stable Paretian distributional assumption (to tackle fat tails). It is well known that the stable Paretian is the only valid distribution that arises as a limiting distribution of sums of independently, identically distributed random variables. Given that unexpected innovations in econometric models are usually interpreted as random variables that represent the sum of the external effects not being captured by the model, the use of the stable Paretian assumption should therefore be highly desirable and lends some theoretical justification for the model and its economic interpretations.

A common property of the dynamic multi–component GARCH structure is constancy of the mixing weights of the component densities, which often allows for a straightforward interpretation of the contributions of the individual components (for example one component for the firm level, one for the sector level and one economy wide component). However, constancy of the distributional proportions may not be a realistic assumption in general, and, as is demonstrated in Chapter 4 leads to less accurate forecasts compared with a more general class of models which allows for time variation in the weights. In particular, by relating current weights to past returns via sigmoid response functions, an empirically reasonable representation of Engle and Ng's (1993) news impact curve with an asymmetric impact of unexpected return shocks on future volatility is obtained.

In Chapter 5 experimental stock markets are used to add more evidence that Black‘s (1976) leverage effect in financial markets does not necessarily stem from the financial leverage of the firm. The original explanation of the leverage effect (a negative correlation between volatility and stock price level) goes back to Modigliani and Miller's (1958) classical work and the firm's capital structure, e.g.

its ratio of debt to equity. We surprisingly observe the leverage effect also in a large number of markets although the underlying assets do not exhibit any financial leverage, and therefore have a capital structure without any debt.

Preface

Thanks a lot to all of you colleagues, friends and family who contributed to this thesis! Especially to my Mom who already asked me after 6 months when I will finally hand in my thesis. Thanks to Prof. Mittnik and Dr. Haas, with whom we worked together very closely, although Munich and Zurich are several miles way from each other. Thanks to Simon Broda for discussing Matlab programming issues even though I sometimes called you late at night. Thanks to Prof. Wolf for very helpful comments while acting as co-referee.

Many thanks to Prof. Hens, especially for providing me with my first salary in Zurich and introducing me to the Finrisk Resarch Network. I really learned a lot about the big picture!

But the biggest thanks goes to my doctoral adviser Prof. Marc Paolella! Marc not only has been the best employer I could possible imagine, you also showed me the necessary tools to do real research. You made it possible for me to come to Zurich and doing research with you was great fun most of the time. I shall never forget the excellent brainstorming sessions in the office and at Starbucks!

Chapter 1
Introduction

The world is getting more and more quantitative as computer science further develops in terms of computing power and data warehouses pop up everywhere. Naturally, this development influences the way empirical finance is answering today's theoretical and practical questions raised by policymakers, analysts and researchers alike. New quantitative techniques and models answering selected questions of empirical finance are the topic of the next four Chapters.

For example, for most investors and banking regulators the down side loss potential of a financial position is of great interest, which can be quantified though the widely used measure Value–at–Risk. Chapter 2, coauthored with Marc S. Paolella, introduces a method in which this down side of the return distribution receives more attention by using special weighting functions. Up to now, in empirical finance, the overwhelming majority of time series models for density and Value–at–Risk forecasting are estimated with traditional maximum likelihood methods. By definition, the likelihood implicitly involves placing equal weight on each of the observations in the sample, but this need not be the case. For example, Mittnik and Paolella (2000) have demonstrated that using a weighted likelihood scheme in which more weight is placed on observations in the recent past results in considerable forecast improvement compared to the default of equal weights. This is presumably because the data generating process is not constant through time. Now, if instead of accurate forecasting of the entire density, interest is restricted to just Value–at–Risk, then it would seem wise to place more weight on the negative observations in the sample which, as shown in this Chapter, yields to a considerable improvement in forecast accuracy.

In Chapter 3, coauthored with Markus Haas, Stefan Mittnik and Marc S. Paolella, a new and quite general class of models for asset returns is proposed. The structure of the model is such that the two most characteristic stylized facts of financial time series are modelled, which are i) that risk, measured for example through volatility, comes in clusters and is therefore not constant through time and ii) that more severe market movements are observed (so called fat tails) than suggested by the Gaussian assumption which served as a convenient model for many years.

The proposed model combines a dynamic multi–component generalized autoregressive conditional heteroscedastic (GARCH) structure (to tackle volatility cluster) with the stable Paretian distributional assumption (to tackle fat tails). It is well known that the stable Paretian is the only valid distribution that arises as a limiting distribution of sums of independently, identically distributed random variables. Given that unexpected innovations in econometric models are usually interpreted as random variables that represent the sum of the external effects not being captured by the model, the use of the stable Paretian assumption should therefore be highly desirable and lends some theoretical justification for the model and its economic interpretations.

A common property of the dynamic multi–component GARCH structure is constancy of the mixing weights of the component densities, which often allows for a straightforward interpretation of the contributions of the individual components (for example one component for the firm level, one for the sector level and one economy wide component). However, constancy of the distributional proportions may not be a realistic assumption in general, and, as is demonstrated in Chapter 4, also coauthored with Marc S. Paolella, Markus Haas and Stefan Mittnik, leads to less accurate forecasts compared with a more general class of models which allows for time variation in the weights. In particular, by relating current weights to past returns via sigmoid response functions, an empirically reasonable representation of Engle and Ng's (1993) news impact curve with an asymmetric impact of unexpected return shocks on future volatility is obtained.

In Chapter 5, coauthored with Thorsten Hens, experimental stock markets are used to add more evidence that Black's (1976) leverage effect in financial markets does not necessarily stem from the financial leverage of the firm. The original explanation of the leverage effect (a negative correlation between volatility and stock price level) goes back to Modigliani and Miller's (1958) classical work and the firm's capital structure, e.g. its ratio of debt to equity. We surprisingly observe the leverage effect also in a large number of markets although the underlying assets do not exhibit any financial leverage, and therefore have a capital structure without any debt.

Chapter 2
Multi Weighted Maximum Likelihood

2.1 Introduction

In many quantitative disciplines, particularly economics, estimation of time series models plays a fundamental role for both theoretical researchers and applied analysts. Among the major estimation methods (including Bayesian, GMM, indirect inference, nonparametric, and least squares), likelihood based inference, with emphasis on the maximum likelihood estimator (MLE), is at least among, if not the, most popular and important. Similar to the other methods of inference, in its traditional form, the likelihood places equal weight on all observations in the sample. Similar to the well-known use of weighted least squares to account for discrepancies in the variance of the dependent variable, weighted likelihood is an established method for estimation and hypothesis testing when the elements of the sample have differing amounts of information. Typically, this is used in the context of robust estimation, to avoid the destructive effect of outliers; see, e.g. Hadi and Luceno (1997) or Cheng (2005). In a related strand of literature, weighted likelihood methodology is used in the area of mixture distributions, see, e.g. Markatou (2000), who uses weighting functions to downweight certain observations.

In the current context, we consider time series models, and generalize the usual likelihood paradigm of implicitly placing equal weight on each of the filtered innovations in, e.g. ARMA or GARCH models. The first generalization of the standard procedure we consider, when applied to time series, is to apply a weighting function such that more recent observations receive a higher relative weight than values far in the past. Observe that, with time weighted ML (or, in short, TiWML), parameter estimation is conducted using all available data, instead of just the notorious choice of "moving window" procedures which completely ignore observations beyond a certain arbitrarily–chosen distance in the past, while equally weighting recent ones. Indeed, this popular scheme is just a (poor) special case of TiWML.

The method of TiWML should be particularly appealing when dealing with the modelling and forecasting of time series of financial returns, with their quite nonstandard properties including rich, nonlinear dynamics which change through time, as well as high kurtosis and time-varying skewness. The most notable stylized fact of asset returns measured at a weekly or daily, or higher, frequency, is volatility clustering. It is well-known that recent volatility shocks are good predictors of volatility in subsequent periods, which would suggest that placing more weight on recent observations could enhance predictive power. Also, for a model whose parameters are changing over time or—more likely in the context of modelling financial returns data—for a much more complicated data generating process which cannot be embodied in either a parametric or nonparametric setup, the method of choosing a tractable parametric structure which reasonably captures the salient features of the data generating process and estimating it with more weight given to recent observations can lead to considerable forecasting improvements. This was demonstrated to be the case by Mittnik and Paolella (2000) in the—for finance, very relevant—context of density forecasting; see, e.g. Diebold et al. (1998), Tay and Wallis (2000), Bao et al. (2006), Amisano and Giacomini (2006), and the references therein.

While TiWML responds to the window size issue through heavier weights for recent observations, it still places equal weight on both negative and positive observations in the sample. With financial risk management applications in mind, standard likelihood inference can be further generalized by placing relatively more weight on negative returns. This will be advantageous if, instead of accurate forecasting of the entire density, interest is restricted to some risk measure that just takes the downside of the distribution into account, such as Value–At–Risk (VaR), or Expected Shortfall (ES).

One source of inaccuracies in the prediction of downside risk in this context is likely to stem from asymmetries in the data which are not adequately captured by the chosen model, even when the model allows for (i) an asymmetric response to shocks in the GARCH equation and (ii) a flexible asymmetric innovations distributional assumption. This is the case, for example, with the asymmetric-power-GARCH (or APARCH) model coupled with an asymmetric generalized Student's t distribution, which has been shown independently by Mittnik and Paolella (2000) and Giot and Laurent (2004) to deliver relatively (compared to other models) very accurate VaR forecasts, but which still have room for improvement. We conjecture, and show in our empirical study below, that with weighting functions designed to place more weight on negative observations, this asymmetry problem can be mitigated. We refer to the proposed method as tail weighted ML (or, in short, TaWML).

It should be mentioned that there are now numerous ways of computing the VaR of a particular financial asset, with comparisons of some of the most promising methods detailed in Bao et al. (2006) and Kuester et al. (2006). As emphasized in the methodology developed in Hartz et al. (2006), the search for ever-more-complicated models might not be as productive as starting with simpler, easily-estimated models and changing instead something other than the parametric form of the model. In Hartz et al. (2006), the authors demonstrate the viability of using the bootstrap and a bias-correction step in conjunction with the rudimentary (and otherwise inadequate, but easily estimated) normal-GARCH model. In this paper, we use parametric GARCH models which are only slightly more sophisticated and still very easy to program and estimate, but change the weights associated with the observations as they enter the likelihood.

A completely different approach to modelling in this context is to use nonparametric methods. Kuester et al. (2006) demonstrate that, even for the successful nonparametric methods (which, interestingly, all have, as part of their methodology, a parametric component, including EVT-GARCH and filtered historical simulation), the choice of the parametric modelling component associated with the GARCH filter and distributional assumption for the innovations is critical for the out–of–sample performance. As such, parametric methods still appear to be of great relevance in this context. We do not consider other types of nonparametric GARCH modelling schemes, such as that developed in Bühlmann and McNeil (2002).

The remainder of this paper is organized as follows. Section 2.2 introduces the two weighted likelihood families, TiWML and TaWML. A short description of the parametric models that are used for illustrative purpose is contained in Section 2.3, while Section 2.4 discusses the empirical results. Because the two weighting families address different shortcomings, we also show how a combination of them results in further gains from synergies. Section 2.5 concludes.

2.2 Weighting Families

The fundamental concept of TiWML is to places more weight on recent observations. To implement the weighting scheme, a vector of standardized weights $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_T)$, is used such that $\sum_{t=1}^{T} \tau_t = 1$, where T is the length of the time series under study. The model parameters are then estimated by maximizing the weighted likelihood, whereby the likelihood component associated with period t, $t = 1, \ldots, T$, is multiplied by τ_t.

In Mittnik and Paolella (2000), the two weighting schemes, geometric and hyperbolic, are proposed and studied, given respectively by $\tau_t \propto \rho^{T-t}$ and $\tau_t \propto (T - t + 1)^{\rho-1}$, where the single parameter ρ dictates the shape of the weighting function. In both cases, values of $\rho < 1$ ($\rho > 1$) cause more recent observations to be given relatively more (less) weight than those values further in the past; $\rho = 1$ corresponds to standard ML estimation.

The method of TaWML, proposed herein, places relatively more weight on negative (or, possibly, only on the *extreme* negative) returns in the sample. As already mentioned, this would seem promising when, for example as in risk management applications, interest centers on the downside risk potential of a financial position and density forecasts are not necessarily needed.

There are several possible strategies of placing more weight on negative observations. As in the time weighting scheme TiWML, each of the T components of the likelihood gets multiplied by the tth component of a standardized weight vector $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_T)$, with $\sum_{t=1}^{T} \omega_t = 1$. Figure 2.1 shows some promising tail weighting candidates. Except for the upper left plot, which shows the default setting of equal weights on positive and negative returns, all weighting schemes place relative more weight on negative observations in the sample, with the special case of the candidate labelled number 4, which is such that both extremes (negative and positive) receive relatively more weight, while medium sized returns receive less. The following is a description of the most promising weighting schemes in risk management applications.

The first extension of the traditional approach of equal weights is straightforward and obvious: split the sample into two groups of negative and positive observations, and assign relatively more weight assigned to the former group, such that all negative observations receive the same weight, and likewise for the positive ones. This can be graphically seen in the upper right panel of Figure 2.1, labelled "Weighting Scheme 1", subsequently referred to as WS-1. Let w_n (w_p) be the weight on the negative (positive) values, of which there are n_n (n_p) of them. For the weights to sum to one, we require $w_n n_n + w_p n_p = 1$. In the standard (unweighted) setup, clearly $w_n = w_p = 1/\left(n_n + n_p\right)$, and to characterize the weighted case, we use parameter γ and take

$$w_n = \frac{\gamma}{n_n + n_p}, \quad w_p = \frac{1 - w_n n_n}{n_p}, \quad \text{where} \quad \gamma \in \left[0, \frac{n_n + n_p}{n_n}\right].$$

Observe that, if $\gamma = 0$, then $w_n = 0$ and $w_p = n_p^{-1}$, i.e., all weight is placed on positive observations, and if $\gamma = \left(n_n + n_p\right)/n_n$, then $w_n = n_n^{-1}$, $w_p = 0$, and all weight is placed on the negative observations. For financial returns data, $n_n \approx n_p$, so the upper bound on γ is approximately 2.

The way WS-1 is constructed, it is not smooth in the sense that there is a discrete jump in the weights where the filtered innovations of the model are split into positives and negatives. Our second weighting scheme, WS-2, avoids this jump by making use of a cumulative density function (CDF) to dictate the weights (see Figure 2.1). The CDF of a Beta(p, q) random variable is used,

$$F_{\mathrm{Beta}}(x; p, q) = \frac{B_x(p, q)}{B(p, q)}, \tag{2.1}$$

where $B_x(p, q) = \int_0^x t^{p-1}(1-t)^{q-1}\mathrm{d}t$ is the incomplete beta function and $B(p, q) = B_1(p, q)$ is the beta function.

WS-2 is constructed as follows. Let $\omega_t, t = 1, 2, \ldots, T$ denote the weight corresponding to $r_t, \ldots, r_T$, respectively. Mapping the returns on the domain of the beta CDF by means of the transformation $\pi_t = r_t/\max_t(-r_t) + 1$, we have the following relation between the weights and the returns r_t,

$$\omega_t = \begin{cases} \gamma(1 - F_{\mathrm{Beta}}(\pi_t; p, q)) + 1 & \text{if } r_t < 0 \\ 1 & \text{otherwise,} \end{cases}$$

resulting in the continuous weighting scheme shown in Figure 2.1. In the above equation weighting factor $\gamma \in \mathbf{R}_{\geq \mathbf{0}}$ is the scale constant determining how much weight is assigned to negative weights. A value of $\gamma > 0$ means negative returns receive more weight, while $\gamma = 0$ is the default case of equal weights. In the final step the weights are standardized so that the sum of all weights (negative and positive) equals one, $\sum_{t=1}^{T} \omega_t = 1$.

When assigning different weights to negative returns we keep the parameters of the beta CDF (p and q) constant in order to preserve the shape shown in Figure 2.1. In order to get the desired (and economically reasonable) shape for the time weights, we used $p = 250$ and $q = 2$. The CDF corresponding to these values gives rise to the plot in Figure 2.1 although with varying maxima governed by the constant γ. We also tested different values for p and q but the results were qualitatively the same.

In the empirical analysis we limit ourselves to WS-1 and WS-2, but there are many more possible variants of weighting schemes. For example, WS-1 is easily extended to the case in which (like for WS-2) more weight is placed on extreme negatives than on moderate negatives, which can be seen in Figure 2.1 in the second row of subplots labelled 'Weighting Scheme 3'. Furthermore, in WS-1 and WS-2, we formed two groups of returns, e.g. positive and negative ones. It is also possible to distinguish extreme negative returns and moderate negative returns on the one hand and positive returns on the other hand, as can be seen in Figure 2.1 for 'Weighting Scheme 3' in the third row.

There are more potentially interesting schemes we have not yet empirically tested (schemes 4, 5, and 6). Candidates can be seen in Figure 2.1. Weighting scheme 4 places more weight on both extremes, positive and negative observations, not necessarily equally. Weighting scheme 5 places more weight on negative returns but not on very extreme negative returns. And finally scheme 6 even further reduces the relative weight of *very* positive returns compared to moderate positive returns.

In the empirical section we show how the results change when we combine the weighting families TiWML and TaWML. The returns are then multiplied by the product of each weighting family and re–standardized so that the weights sum to one.

Note that, for both weighting scheme families (including the time weighting schemes), an "optimal" value of ρ and γ cannot be estimated with the model parameters, but must be obtained with respect to some criterion outside of the likelihood function of the T observations. In Section 2.4 we will make use of out–of–sample VaR predictions for this purpose.

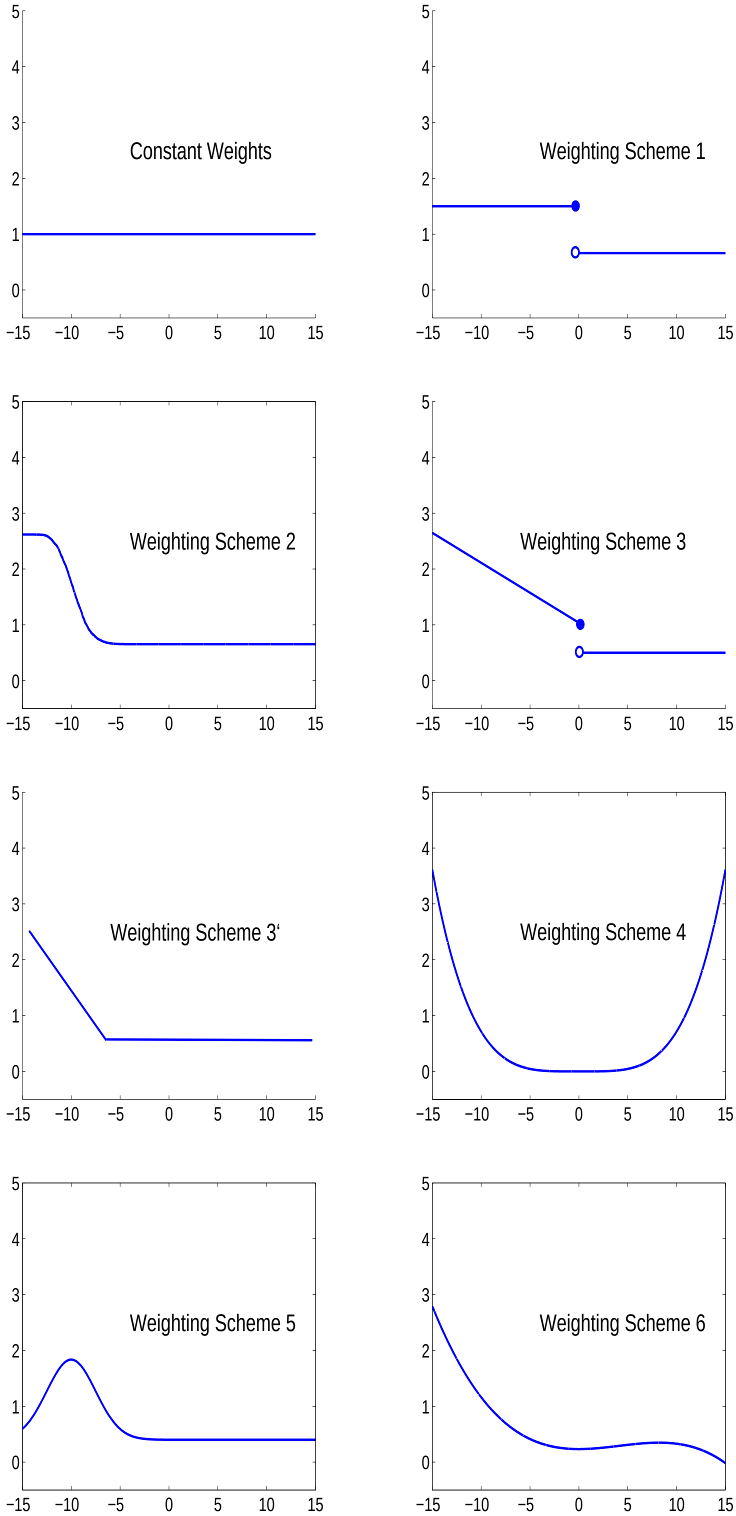

FIGURE 2.1: Different possible weighting schemes through which more weight is placed on negative observations in the sample. On the horizontal axes are the returns, the vertical axes shows the weights associated with them. Weighting scheme 'Constant Weights' corresponds to the default case of equal weights on negative and positive returns.

2.3 VaR Models

In order to show the general usefulness of the aforementioned weighting schemes, we explore how the forecasting ability of different fully parametric models is influenced by the proposed weighting functions.

In particular, we entertain five different models in the empirical section:

1. As a basic reference model we apply the original GARCH model by Bollerslev (1986) with normally distributed innovations (hereafter normal–GARCH).

2. We use the same GARCH structure as in 1. but with the Student's t distribution to accommodate the fat tailed behavior (t–GARCH), and

3. also with an asymmetric generalized Student's t distribution (t_3) to allow for skewness in the distributional assumption (t_3–GARCH).

4. To capture the asymmetric impact of unexpected return shocks on future volatility, Black's (1976) so called leverage effect, we apply the asymmetric power GARCH by Ding et al. (1993) coupled with the t_3 distribution (t_3–APARCH).

5. We further analyze the mixed normal GARCH model (MixN) by Haas et al. (2004a), through which, owing to its great flexibility, the skewness in the data generating process is captured in a completely different way then by the t_3 distribution for example.

We will describe the models in more detail in the remainder of this section. All entertained models belong to a location–scale family of probability distributions of the form

$$r_t = \mu_t + \epsilon_t = a_0 + \sum_{i=1}^{p} a_i r_{t-i} + \sigma_t z_t, \tag{2.2}$$

where location μ_t and scale σ_t are $\mathcal{F}_{t-1}$–measurable parameters and $z_t \overset{\text{iid}}{\sim} f_Z(\cdot)$, where f_Z is a zero–location, unit–scale probability density which could have additional shape parameters (such as the degrees of freedom parameter in the Student's

t distribution). Concerning μ_t we limit ourselves in the empirical analysis to an AR(1) process.

In Bollerslev's (1986) GARCH(r,s) model the scale parameter evolves as,

$$\sigma_t^2 = c_0 + \sum_{i=1}^{r} c_i \epsilon_{t-i}^2 + \sum_{j=1}^{s} d_j \sigma_{t-j}^2. \tag{2.3}$$

As mentioned above we utilize three different assumptions for the innovation distribution, f_Z, which are the normal, the Student's t with $\nu \in \mathbb{R}_+$ degrees of freedom (in short, t distribution), and the generalized asymmetric t, (in short, t_3) with density

$$f(z; d, \nu, \theta) = C\left(1 + \frac{(-z\theta)^d}{\nu}\right)^{-\frac{\nu+1}{d}} \mathcal{I}(z < 0) + C\left(1 + \frac{(z/\theta)^d}{\nu}\right)^{-\frac{\nu+1}{d}} \mathcal{I}(z \geq 0), \tag{2.4}$$

where $d, v, \theta \in \mathbb{R}_+$, $\mathcal{I}(\cdot)$ is the indicator function and $C = \left[\left(\theta + \theta^{-1}\right) d^{-1} \nu^{1/d} B\left(d^{-1}, \nu\right)\right]^{-1}$. The r^{th} raw integer moment, $0 \leq r < \nu d$, for the t_3 is

$$\frac{(-1)^r \theta^{-(r+1)} + \theta^{r+1}}{\theta^{-1} + \theta} \frac{B\left(\frac{r+1}{d}, \nu - \frac{r}{d}\right)}{B\left(\frac{1}{d}, \nu\right)} \nu^{r/d},$$

from which, for example, variance, skewness and kurtosis can be computed if they exist. The cumulative distribution function (CDF) of the t_3 (as required for VaR calculation) is given by

$$F(z) = \begin{cases} \dfrac{I_L(\nu, 1/d)}{1+\theta^2}, & \text{if } z \leq 0, \\ \dfrac{I_U(1/d, \nu)}{1+\theta^{-2}} + \left(1+\theta^2\right)^{-1}, & \text{if } z > 0, \end{cases}$$

where $L = \nu \,/\, \left[\nu + (-z\theta)^d\right]$, $U = (z/\theta)^d \,/\, \left[\nu + (z/\theta)^d\right]$, and

$$I_x(a,b) = \frac{B_x(a,b)}{B(a,b)}$$

is the incomplete beta ratio.

With Bollerslev's original GARCH model (and any distributional assumption) it is not possible to let positive and negative shocks to the series have different impacts on future volatility. In order to show that our results hold also for asymmetric models in which negative shocks have greater impact on future volatility than do positive shocks of the same size, we also entertain the asymmetric power ARCH, or A–PARCH (r, s). Then (2.3) is extended to

$$\sigma_t^\delta = c_0 + \sum_{i=1}^{r} c_i \left(|\epsilon_{t-i}| - \gamma_i \epsilon_{t-i}\right)^\delta + \sum_{i=1}^{s} d_i \sigma_{t-i}^\delta, \qquad |\gamma_i| < 1, \tag{2.5}$$

which, as is shown in Ding et al. (1993), nests at least seven GARCH(r, s) models proposed in the literature. The special case $\delta = 1$ and $\gamma = 0$ corresponds to (2.3); parameter δ serves as a symmetric power transformation of σ_t; while $\gamma \neq 0$ allows σ_t to respond asymmetrically to positive and negative shocks.

We now briefly turn to the last model for which we will provide forecasting comparisons in the empirical section. This is the mixed normal GARCH model, denoted MixN, that has recently been proposed by Haas et al. (2004a) and Alexander and Lazar (2006), and generalizes the classic normal GARCH model of Bollerslev (1986) to the normal mixture setting. In general, a random variable is said to follow a k–component normal mixture distribution if its density is given by

$$f_{\text{MixN}}(y; \boldsymbol{\lambda}, \boldsymbol{\eta}, \boldsymbol{\sigma}^{(2)}) = \sum_{j=1}^{k} \lambda_j \phi(y; \eta_j, \sigma_j^2), \tag{2.6}$$

where $\phi(y; \eta_j, \sigma_j^2)$ are normal densities; $\boldsymbol{\lambda} = [\lambda_1, \ldots, \lambda_k]'$ is the vector of strictly positive mixing weights which satisfy $\sum_j \lambda_j = 1$; and the elements of $\boldsymbol{\eta} = [\eta_1, \ldots, \eta_k]'$ and $\boldsymbol{\sigma}^{(2)} = [\sigma_1^2, \ldots, \sigma_k^2]'$ are the component means and variances, respectively. If $Y \sim \text{MixN}(\boldsymbol{\lambda}, \boldsymbol{\eta}, \boldsymbol{\sigma}^{(2)})$, then

$$\mathrm{E}[Y] = \sum_{j=1}^{k} \lambda_j \eta_j, \quad \text{and} \quad \mathrm{Var}(Y) = \sum_{j=1}^{k} \lambda_j (\sigma_j^2 + \eta_j^2) - \left(\sum_{j=1}^{k} \lambda_j \eta_j \right)^2. \tag{2.7}$$

In the MixN model for asset returns, it is assumed that the *conditional* distribution of the return at time t, r_t, is MixN, that is,

$$r_t | \mathcal{F}_{t-1} \sim \text{MixN}(\boldsymbol{\lambda}_t, \boldsymbol{\eta}_t, \boldsymbol{\sigma}_t^{(2)}), \tag{2.8}$$

where $\mathcal{F}_t$ is the common information set at time t. The vector of component variances, $\boldsymbol{\sigma}_t^2$, evolves according to the recursion

$$\boldsymbol{\sigma}_t^{(2)} = \alpha_0 + \sum_{i=1}^{r} \boldsymbol{\alpha}_i \epsilon_{t-i}^2 + \sum_{j=1}^{s} \boldsymbol{\beta}_j \boldsymbol{\sigma}_{t-j}^{(2)}, \tag{2.9}$$

where $\boldsymbol{\alpha}_i = (\alpha_{i1}, \alpha_{i2}, \ldots, \alpha_{ik})'$, $i = 0, \ldots, r$, are $k \times 1$ vectors; $\boldsymbol{\beta}_j$, $j = 1, \ldots, s$, are diagonal $k \times k$ matrices[1] with the component–specific persistence parameters on the main diagonal; and from (2.2), ϵ_t, is

$$\epsilon_t = r_t - \mathrm{E}(r_t|\mathcal{F}_{t-1}) = r_t - \sum_{j=1}^{k} \lambda_{j,t}\eta_{j,t}. \tag{2.10}$$

As in Haas et al. (2004a) we will not pursue the case of component–specific mean dynamics, but assume that we have the same AR(p) structure in each component. That is, the conditional mean in component j can be written as

$$\mu_{j,t} = a_{0,j} + \sum_{i=1}^{p} a_i r_{t-i}, \quad j = 1, \ldots, k, \tag{2.11}$$

where only the constant $a_{0,j}$ may differ across components in order to allow for skewness of the conditional distribution. As an example in the empirical section, we will use the mixed normal GARCH model with three components, denoted MixN, because it might capture different aspects of the skewness than the t_3, and also because several analyses support their superiority in terms of forecasting ability. In the empirical analysis that follows we use a GARCH$(1,1)$ as a basis for all models.

2.4 Empirical Results of various Stock Indices and Exchange Rates

For the empirical analysis we use 9 sets of daily returns, of which 6 are stock indices (NASDAQ Composite, S&P 500, DJIA, Nikkei 225, DAX 30 and FTSE 500) and 3 are foreign exchange rates (US$–Swiss Franc, US$–JEN and US$–German Mark and US$–Euro after 2001). The stock indices range from January 1970 to January 2005 except for NASDAQ and FTSE whose inception dates are February 1971 and January 1978, respectively. The exchange rates range from January 1983 to January 2005 except the US$–Swiss Frank which starts in January 1986.[2]

[1] The full–matrix specification is considered in Haas et al. (2004a), though, as discussed there and confirmed with other data sets, the diagonal restriction is usually favored in empirical applications.

[2] Datastream provides no US$–Swiss Frank data beyond 1986.

For the analysis we calculate the continuously compounded percentage returns, $r_t = 100\,(\log P_t - \log P_{t-1})$, where P_t denotes the index or exchange rate level at time t.

2.4.1 Forecasting Performance

Because a major concern of risk management professionals is the downside loss potential of a financial position, we study the out–of–sample forecast performance with respect to Value–at–Risk (VaR). Our primary aim is to assess the potential improvement in forecasting ability when using the proposed weighting functions, and so we limit ourselves in this study to one–step–ahead forecasts. The VaR with shortfall probability ξ is calculated as $\hat{F}^{\mathcal{M}}_{t|t-1}(\mathrm{VaR}_t(\xi)) = \xi$, where $\hat{F}^{\mathcal{M}}_{t|t-1}$ is the predicted return distribution function at time t based on the information set up to time $t-1$ and use of model $\mathcal{M}$. While numerous tests for the efficiency of VaR forecasts are available (see, e.g. Christoffersen and Pelletier (2004); Kuester et al. (2006); and the references therein), we consider only the empirical coverage probabilities associated with the VaR forecasts. Our forecasting exercise for the 9 returns series uses a moving window of length $1,000$ days with parameter re–estimation for each window. Given the length of the series we archive roughly 8000 forecasts per model that can be compared.

Most forecasting exercises show the percentage of violations associated with the VaR forecasts for the 1%, 2.5%, 5% and 10% ξ-level. Clearly, the model which achieves the closest coverage probability performs best for the respective data set and VaR–level. In order to illustrate the forecasting performance of different models, Kuester et al. (2006) illustrate their results for a spectrum of VaR levels up to $\xi = 10\%$ with a convenient graphical depiction. They plot the forecast CDF against the deviation from a uniform CDF. The VaR levels can therefore be read off the horizontal axis, while the vertical axis depicts, for each VaR-level, the excess of percentage violations over the VaR-level. Thus, the relative deviation from the correct coverage can be compared across VaR levels and competing models. Theoretically, an ideal model would exhibit a flat line at zero. Using this concept as

	$\gamma = 0$	$\gamma = .25$	$\gamma = 5$	$\gamma = 10$
a_0	.0469 (.0462)	.0489 (.0463)	.0668 (.0465)	.0660 (.0465)
a_1	.2638 (.0139)	.2576 (.0139)	.2100 (.0135)	.1747 (.0132)
c_0	.0207 (.0001)	.0207 (.0001)	.0206 (.0001)	.0206 (.0001)
c_1	.0531 (.0003)	.0532 (.0003)	.0541 (.0003)	.0541 (.0003)
d_1	.9112 (.0059)	.9119 (.0059)	.9217 (.0057)	.9287 (.0055)

TABLE 2.1: Normal–GARCH parameter estimates with WS-2 (standard errors in parentheses) as they change when more weight is placed on negative observations. Parameter notation is from (2.2). $\gamma = 0$ means negative and positive observations receive the same weight. As γ increases, more weight is given to the negative observations.

a basis, we also compute the mean absolute deviation (MAD) of the actual violation frequencies from the corresponding theoretical VaR–level over the intervals $[0 - 0.01]$, $[0 - 0.025]$ and $[0 - 0.05]$ and compare these deviations for different weights and weighting functions.

For WS-1 this can be seen in Figure 2.2 for each model and different values of the weight parameter γ. Given a particular VaR level, the best model for this data set is the one which has smallest MAD at $\gamma = 1$, emphasized though the dotted horizontal line, which shows the point where negative and positive observations receive the same weight. The closer the MAD is to zero, the better the model is for the respective value of γ. For example, in Figure 2.2 we see in the upper left subplot that, for increasing γ, the MAD decreases. That is, the forecasting performance of the models improves as more weight is given to negative observations. When we look at the third row we observe, for example in the third column, that for γ bigger than 1.3 the MAD starts rising again, indicating that too much weight on negative observations leads to an increase of the MAD eventually.

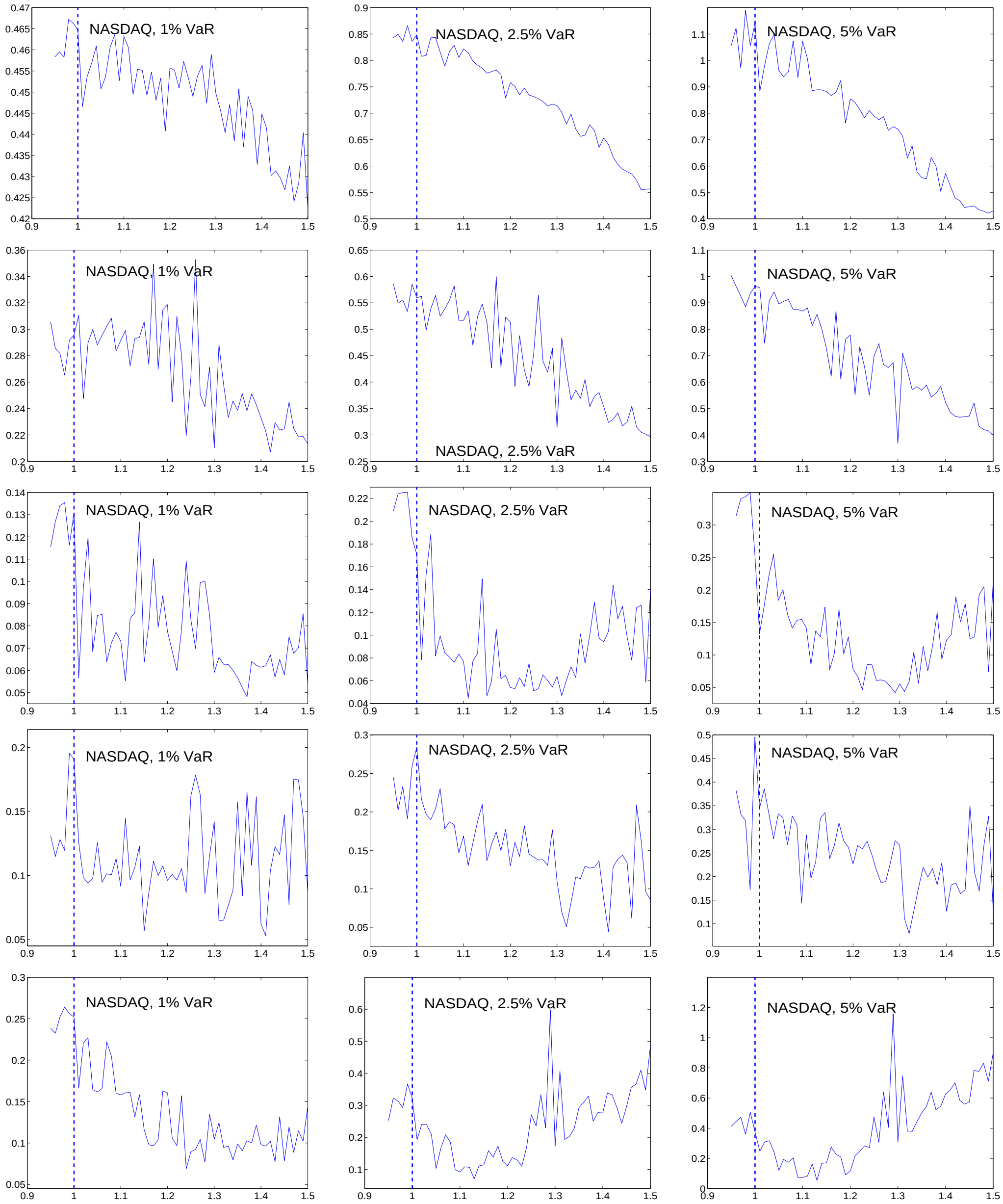

FIGURE 2.2: Mean absolute deviation (MAD) of the empirical from the theoretical tail probability for different weights, γ, ranging from .95 to 1.5 (x–axis) for WS-1. A weight equal to 1 indicates positive and negative observations get the same weight (see vertical dashed line) while a weight bigger than 1 (smaller than 1) results in more (less) weight for negative observations than for positive observations. The upper row depicts the results for GARCH with Gaussian innovations, then follows GARCH with Student's t innovations (second row) and GARCH with skewed t innovations (third row). The fourth row contains the results for the APARCH model with skewed t innovations, while the last row shows how the forecasting performance of the MixN model changes when more weight is placed on negative weights. The left, middle and right columns show the MAD up to 1%, 2.5% and 5% VaR–level, respectively. Data base is the NASDAQ Composite from January 1970 to January 2005.

2.4.2 Empirical Results

We first describe the results for the first weighting scheme, WS-1, for the stock indices and exchange rates and then turn to the second weighting scheme, WS-2, both introduced in Section 2.2. Concerning WS-1, Figure 2.2 contains all entertained models and their MAD for VaR levels of 1%, 2.5% and 5% for the NASDAQ index. For the two weighting schemes, WS-1 and WS-2, different intervals for the weighting factor γ are used. After some experimenting we chose a range for γ for WS-1 of $[.95; 1.5]$ and for γ of WS-2 of $[0; 10]$.

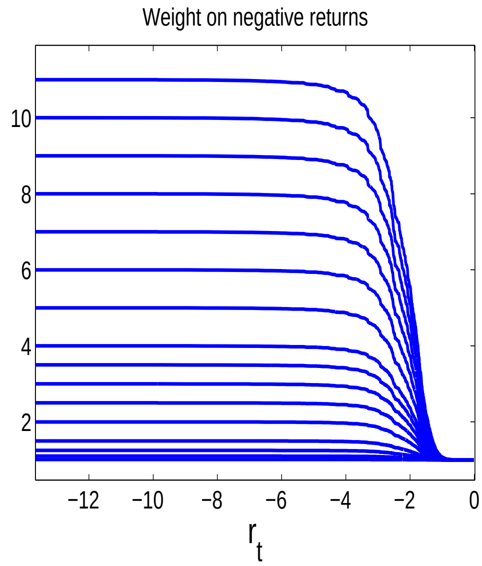

FIGURE 2.3: Weighting functions for the negative observations or returns for the NASDAQ data set and the used grid for WS-2, $\gamma = [.01, .1, .25, .5, 1, 1.5, 2, 2.5, 3, 4, 5, 6, 7, 8, 9, 10]$. As γ increases more weight is placed on negative observations, i.e. the bigger the area under the negative side of the weighting function.

For WS-1 we used steps of .01 resulting in 56 runs with different weights. Values for $\gamma < 1$ mean more weight is placed on positive observations. $\gamma = 1$ is the non–weighted case, while a $\gamma > 1$ means more weight is placed on negative observations.

For WS-2, $\gamma = 0$ refers to the non weighted case and $\gamma > 0$ means more weight is placed on negative than on positive observations. The following grid for γ was used: $[.01, .1, .25, .5, 1, 1.5, 2, 2.5, 3, 4, 5, 6, 7, 8, 9, 10]$, which results in the different weighting functions as shown in Figure 2.3.

To illustrate how the parameters actually change when we place more and more weight on negative observations Table 2.1 provides an overview of their realizations for WS-2 for the normal–GARCH model from (2.2) and (2.3) for the NASDAQ data set. As expected we observe that the AR and the GARCH parameters change when more weight is placed on negative observations. For the AR parameters there seems to be no tendency for a_0, but a_1 decreases monotonically as γ increases. For the GARCH parameter we observe that c_0 decreases by a very little amount. However, at the same time both parameters that govern the volatility clusters, c_1 and d_1, increase quite substantially, thus increasing the GARCH effects in the model and making extreme values more likely to occur.

There are numerous papers comparing the forecasting performance of competing VaR models and so we do not go into detail comparing the models for the equal weight case, for which, unsurprisingly, our results are in line with the literature. Going though all data sets and models we find that the plain vanilla GARCH is outperformed by all fatter tailed models.

Turning to WS-1 and Figure 2.2 the first striking result for the NASDAQ data set is that for all entertained models the forecasting performance can be improved by allocating more weight on negative observations. Even the powerful models that incorporate fat tails and skewness, the t_3–GARCH, the t_3–APARCH and the MixN, can be improved by placing more weight on negative observations. The behavior of the MAD, however, is different among the models. First and not surprisingly, we see different absolute and percentage changes of the MAD for different models. Second, within the range of γ for some models their MAD starts getting worse again after some optimal γ while for other models the turning point seems to lie outside the applied range. Third, the MAD changes are erratic for some models, but for other models it changes quite smoothly.

Concerning the turning point of the MAD we find that, for example, for the normal-GARCH (upper row) and the 2.5%-VaR level in the middle row of Figure 2.2 the MAD starts at a value of about .85 at $\gamma = 1$ and we receive a lowest MAD of about .55 for $\gamma = 1.5$. For the MixN the lowest MAD is at about $\gamma = 1.15$ with a MAD value of less than .1. So, in order to archive the same reduction in MAD we need more weight on negative observations for the normal–GARCH than for the MixN, for example.

Table 2.2 gives an overview of the MAD of the other data sets. For every data set and VaR–level the table contains the MAD value for $\gamma = 1$ and the lowest MAD that was obtained in the γ–interval $[0.95, 1.5]$. For example, for the 1%–VaR level for the normal–GARCH in the first row we have a MAD of .33 for the Dow Jones index in the equal weight case and can achieve a value of at least .17 when more weight is placed on negative observations, meaning we more than half the respective MAD. Because of the erratic behavior of the MAD that can be seen in Figure 2.2 the table actually just provides an indication of the MAD improvement that can be achieved with WS-1. To save space pictures have been omitted for the other data sets because they look similar to the NASDAQ case. The general picture of these data sets is, however, qualitatively the same as we see for the NASDAQ index.

For the Jen–US\$ exchange rate the potential improvement of the MAD is principally comparable with the NASDAQ index as can be seen from Figure 2.4. For example, for the normal–GARCH and the 1% and 2.5% VaR level we do not even achieve the minimum MAD within the chosen interval for γ. Also, we observe the usual turning point for the more complex models, for example, the t_3–GARCH in the third row, from which on the MAD starts increasing again. Results for the US\$-German Mark and US\$-SFR exchange rates are also contained in Table 2.2.

We now turn to WS-2. Up to this point we have seen results for the most simple WS-1. This weighting scheme comprises a characteristic jump at $\epsilon = 0$ as can be seen in Figure 2.1. During the estimation process this jump can lead to convergency problems due to the fact that we actually weight filtered innovations. AR–parameters that change during the estimation process influence the group composition of negative or positive observations and thus effecting the estimation.

To use a weighting scheme with no such jump in the weighting function at $\epsilon = 0$ we also show the results for WS-2. As an example, Figure 2.5 depicts the results for the known VaR levels of 1%, 2.5% and 5% for the NASDAQ data set and the t_3–GARCH model. What is striking is that the picture looks very similar to the plots for WS-1. We observe that for growing γ the models provide better forecasts and we also observe the typical turning point at a value of γ between 2 and 6.

		DOW		Dax		FTSE		S&P		Nikkei		DM-US$		US$-SFR	
Model	VaR–level	$\gamma = 1$	min	$\gamma = 1$	min	$\gamma = 1$	min	$\gamma = 1$	min	$\gamma = 1$	min	$\gamma = 1$	min	$\gamma = 1$	min
normal-GARCH	1%	.33	.17	.27	.11	.35	.21	.32	.23	.37	.30	.34	.21	.25	.10
	2.5%	.32	.14	.30	.15	.46	.17	.38	.17	.57	.26	.37	.17	.24	.13
	5%	.28	.24	.23	.17	.30	.18	.29	.26	.65	.26	.27	.25	.28	.26
$t-$GARCH	1%	.09	.03	.13	.04	.24	.02	.08	.05	.10	.02	.06	.03	.14	.12
	2.5%	.12	.05	.15	.08	.34	.08	.10	.05	.26	.03	.08	.05	.20	.14
	5%	.14	.05	.19	.06	.40	.06	.18	.04	.51	.05	.10	.06	.21	.13
t_3-GARCH	1%	.04	.03	.05	.03	.05	.08	.06	.05	.10	.10	.12	.10	.08	.04
	2.5%	.05	.04	.05	.05	.07	.03	.07	.04	.12	.06	.13	.10	.08	.07
	5%	.05	.04	.08	.05	.15	.06	.08	.04	.14	.08	.13	.08	.12	.09
t_3-APARCH	1%	.05	.02	.06	.03	.09	.03	.06	.03	.07	.02	.05	.05	.08	.04
	2.5%	.05	.05	.09	.03	.09	.04	.05	.03	.12	.03	.09	.09	.07	.07
	5%	.13	.04	.13	.05	.30	.07	.06	.05	.12	.07	.13	.08	.13	.06
MixN	1%	.09	.04	.10	.03	.11	.04	.13	.02	.04	.02	.12	.07	.04	.03
	2.5%	.12	.06	.10	.05	.17	.08	.08	.03	.09	.03	.27	.13	.06	.06
	5%	.13	.12	.17	.06	.22	.10	.08	.05	.21	.05	.37	.14	.05	.05

TABLE 2.2: Mean absolute deviation (MAD) of the empirical from the theoretical tail probability for WS-1. The entry '$\gamma = 1$' contains the MAD for the equal weighted case (as a reference) and the entry named 'min' stands for the minimal achieved MAD when more weight is given to the negative observations. This table is basically a summary of Figure 2.2 for other data sets.

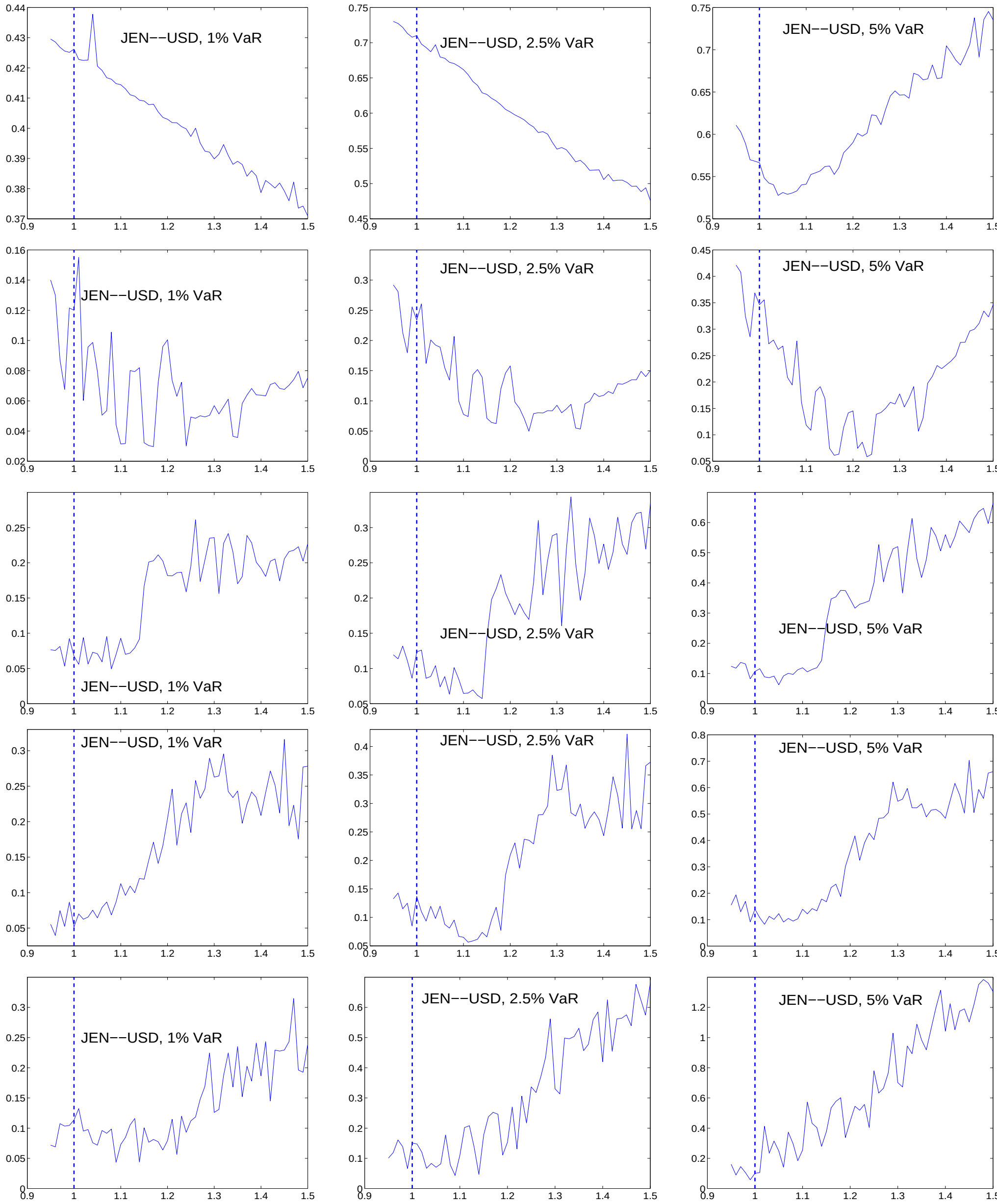

FIGURE 2.4: Same as Figure 2.2 but for JEN-US exchange rate data.

2.4.3 Combining the Weighting Families

As mentioned earlier the two weighting families: tail and time weighting (TaWML and TiWML) address different shortcomings of the traditional approach. To combine both weighting schemes we multiply the two respective weights belonging to the same likelihood component and, as usual, divide them by their joint sum. For the analysis of the combined approach we use a two dimensional grid. The first dimension contains the know grid for WS-2 that we used in the analysis above for TaWML. The second dimension contains a grid for the 'damping factor' ρ to capture the time dimension. We use a similar grid for ρ as in Mittnik and Paolella (2000) for their geometric decay function, $[.6, .7, .8, .9, 1, 1.1, 1.2]$. For values of $\rho < 1$ ($\rho > 1$), recent observations are given relatively more (less) weight than those values far in the past, while $\rho = 1$ corresponds to standard ML estimation. So we have 16 realizations of γ for TaWML and 6 realizations of ρ for TiWML leading to 86 different forecasting exercises. To safe computing time we just run the t_3–GARCH model for the NASDAQ index. With Matlab Version 7 and a standard Pentium 4 PC this alone takes about a week of number crunching.

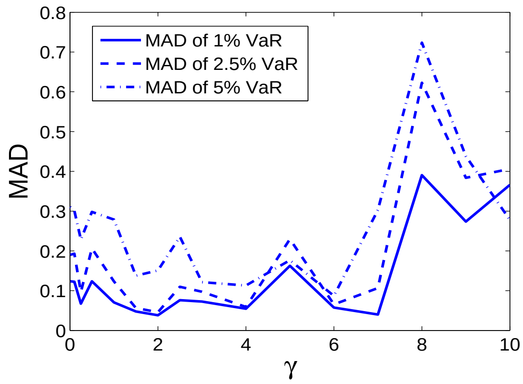

FIGURE 2.5: Same as in Figure 2.2 but for WS-2 and just for the t_3–GARCH model.

Figure 2.6 shows the results. It basically shows 8 subplots with different values of γ for every subplot. On each subplot we have ρ on the X–axis and the corresponding MAD on the Y–axis. We observe that for lower ρ (as more weight is placed on recent observations) MAD tends to be lower. Furthermore, it seems

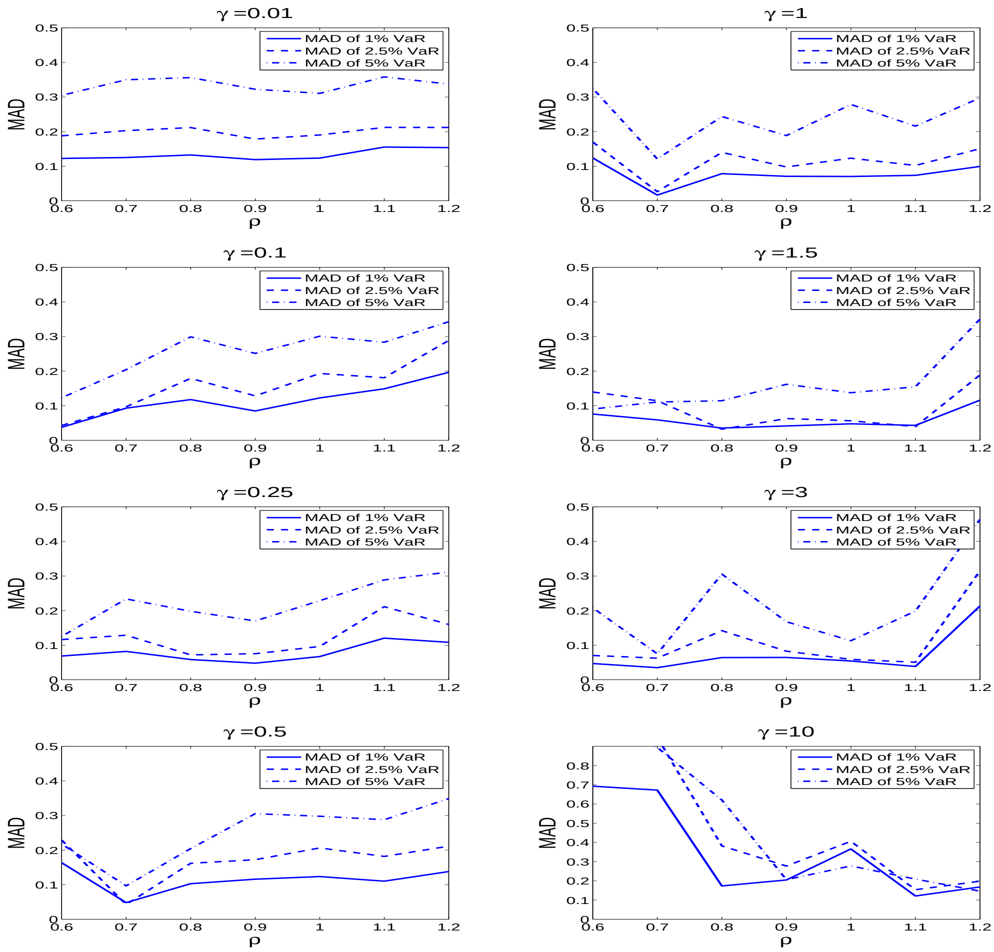

FIGURE 2.6: MAD for three VaR levels (1% (solid), 2.5% (dashed) and 5% (dashdot)) and time varying weighting parameter ρ, which determines how much weight is allocated between recent and past observations. $\rho = 1$ refers to the case in which all observations are equally weighted. The smaller ρ the more weight is placed on recent values.

clear that for increasing γ the graphs move down, however, this is not true for very high values of γ, for example $\gamma = 10$. In that case there is even no improvement by varying ρ anymore.

Figure 2.7 can be seen as an aggregated version of Figure 2.6 in which the synergies of both weightings are more apparent. Different combinations of γ and ρ for the 1% VaR level are shown and we find an optimal MAD for $\gamma = 1.5$ and $\rho = .8$. So we indeed see that the MADs improve for smaller ρ (more weight on recent observations) and as γ increases (more weight on negative observations). Both weighting families are capable of decreasing the model MADs, and their combination leads to the best forecasting performance independent of the model choice.

2.5 Conclusion

Almost all time series models used in empirical finance for density and Value–at–Risk (VaR) forecasting are estimated via maximum likelihood methods. As known to the literature using a weighted likelihood scheme in which more weight is placed on observations in the recent past results in considerable out–of–sample one step–ahead forecast improvement compared to the default of equal weights. Extending the concept of weighted ML we introduce a weighting scheme in which negative observations are valued heavier than positive ones. We propose several potential weighting schemes in which the 'left side' of the return distribution is treated more important and demonstrate that for a variety of GARCH models using such weighting schemes leads to an even further improvement in terms of VaR forecasts.

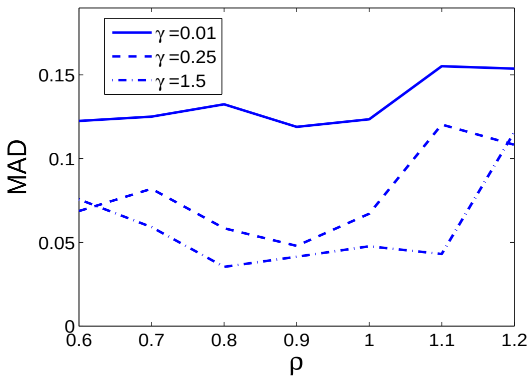

FIGURE 2.7: Same as Figure 2.6 but MAD just for the 1% VaR level and different values of γ.

Chapter 3

Stable Mixture GARCH Models

3.1 Introduction

Starting with the pioneering works of Mandelbrot (1963) and Fama (1965), a variety of studies have investigated the use of the stable Paretian distribution for modelling the unconditional—and, later, the conditional—distribution of asset returns. Given the skewed and very fat-tailed nature of weekly, daily, and higher-frequency financial returns data, it is not surprising that the asymmetric stable distribution has been very successful in this regard. However, an array of other fat-tailed, skewed distributions exist, some of which also fit asset returns remarkably well (see, e.g., McDonald (1996); Mittnik and Paolella (2000); Knight and Satchell (2001); and the references therein), but none of these are closed under summation. It is well known that the stable Paretian is the only valid distribution that arises as a limiting distribution of sums of independently, identically distributed (iid) random variables. Given that error terms in econometric models are usually interpreted as random variables that represent the sum of the external effects not being captured by the model, the use of the stable Paretian assumption should be highly desirable. For an overview of the use of the stable Paretian distribution in financial and other econometric applications, see Mittnik and Rachev (1993a,b), McCulloch (1996), Rachev et al. (1997), Rachev et al. (1999), Rachev and Mittnik (2000), Bams et al. (2005), and the contributions in Adler et al. (1998) and Rachev (2003).

Another popular and successful approach to the unconditional modelling of asset returns and VaR prediction involves the use of discrete mixtures of normal distributions. Owing to its great flexibility, a normal mixture, even with just two components, is well-suited for capturing the usual stylized facts typical in a financial context. This model has been investigated by numerous authors, including

Fama (1965), Kon (1984), Akgiray and Booth (1987), Tucker and Pond (1988) and Kim and Kon (1994). For example, Kon (1984) suggests that returns may be influenced by a series of different information flows including macroeconomic, institutional specific and seasonal information announcements, thus justifying the discrete mixture.

A different economic motivation for the presence of a mixture of distributions is provided by Vigfusson (1997). He builds on theoretical work which attempts to explain the stylized facts of financial time series by the interaction of heterogeneous groups of agents, with the groups processing market information differently; see, e.g., Samanidou et al. (2002) for an overview of such models. In doing so, he identifies a "chartist" and a "fundamentalist" component. The fact that, for each component, a central limit theorem argument can be used to justify the use of the normal distribution is appealing, and lends some theoretical justification for the model and its economic interpretations.

In recent years, attention has moved somewhat away from the unconditional modelling of the distribution of asset returns, and more towards accurate assessment of predicting the Value-at-Risk, or VaR, of a financial portfolio. For a given, small target probability λ, typically chosen between 1% and 5%, the VaR delivers an upper bound on losses such that it will be exceeded with probability λ. In particular, conditional on the information given up to time t, the VaR for period $t+h$ of one unit of investment is the negative λ-quantile of the conditional return distribution, i.e.,

$$\mathrm{VaR}_{t+h} := -Q_\lambda(r_{t+h} \mid \mathcal{F}_t) = -\inf_x\{x \in \mathbb{R} : \Pr(r_{t+h} \leq x \mid \mathcal{F}_t) \geq \lambda\}, \quad 0 < \lambda < 1, \tag{3.1}$$

where $Q_\lambda(\cdot)$ denotes the quantile function; r_t is the return on an asset or portfolio in period t; and $\mathcal{F}_t$ represents the information available at date t. While it has been argued that VaR is not an adequate risk measure because it does not satisfy the essential property of subadditivity,[1] its importance and practical relevance is

[1]See, e.g., Artzner et al. (1999), Acerbi and Tasche (2002a,b), Tasche (2002) and Dowd (2002, Sec. 2.2.3).

undoubted, due to its central role in banking regulation and its use for internal risk management, which has stimulated a great deal of academic interest in calculating reliable VaR measures. The fact that VaR involves the quantile of a predictive distribution means that all the unconditional and conditional parametric distributional models of (portfolio) asset returns using the stable Paretian and/or the normal mixture assumption are relevant.

This paper proposes a model which nests the stable Paretian and the normal mixture assumptions, combines them with a rich conditional heteroscedastic structure, and demonstrates its effectiveness for VaR prediction. The remainder of this paper is as follows. Section 3.2 summarizes the existing conditional heteroscedastic models and their relation to the stable Paretian and mixed normal models discussed above. Section 3.3 introduces the new model, Section 3.4 illustrates its use with the DAX stock index, and Section 3.5 provides some concluding remarks.

3.2 Conditional Heteroscedastic Models

Of all the conditional models in use for VaR prediction (see Kuester et al. (2005), and the references therein for surveys of the various methods), the class of GARCH models has proven itself to be highly effective and is by far the most popular; see Bollerslev, Engle and Nelson (1994), Palm (1997) and Gourieroux (1997) for surveys. The standard GARCH model of Engle (1982) and Bollerslev (1986), is given by

$$r_t = \mu + \sigma_t \varepsilon_t, \quad \sigma_t^2 = \theta_0 + \sum_{i=1}^{r} \theta_i |y_{t-i} - \mu|^2 + \sum_{j=1}^{s} \phi_j \sigma_{t-j}^2, \tag{3.2}$$

where $\varepsilon_t \overset{\text{iid}}{\sim} \mathrm{N}(0,1)$. Despite their success in capturing a very high percentage of the volatility movement, countless applications have confirmed that the residuals, or filtered innovations, from (3.2), when applied to weekly, daily, or higher frequency asset return data, still deviate considerably from normality. This has given rise to a large number of alternative models which replace the normality distributional assumption in the GARCH model with a fat-tailed, possibly skewed, distribution, the most popular of which being the t-GARCH, which assumes a Student's t distribution whose degrees of freedom parameter, $v \in \mathbb{R}_+$, is jointly estimated with the other model parameters.

With regard to the aforementioned appropriateness of the stable distribution instead of ad-hoc assumptions such as the Student's t, it would seem natural to consider its use in conjunction with a GARCH-type model. This was indeed done, first by McCulloch (1985) in a simplified ARCH-type structure, and later in the more general GARCH setting by Liu and Brorsen (1995), Panorska et al. (1995), Mittnik et al. (1998a,b), and Mittnik and Paolella (2003), all of whom have demonstrated its effectiveness in a variety of applications. A more general power-stable-GARCH model is proposed in Mittnik et al. (2002), where sufficient, easily computable stationarity conditions are derived. Their model is given as follows. The sequence y_t is said to be a stable Paretian power GARCH process or, in short, an $\mathrm{S}^{\delta}_{\alpha,\beta}\mathrm{GARCH}(r,s)$ process, if

$$y_t = \mu + c_t \varepsilon_t, \quad c_t^{\delta} = \theta_0 + \sum_{i=1}^{r} \theta_i |y_{t-i} - \mu|^{\delta} + \sum_{j=1}^{s} \phi_j c_{t-j}^{\delta}, \tag{3.3}$$

where $\varepsilon_t \overset{\text{iid}}{\sim} S_{\alpha,\beta}(0,1)$, and $S_{\alpha,\beta}(0,1)$ denotes the standard asymmetric stable Paretian distribution with stable index α, skewness parameter $\beta \in [-1,1]$, zero location parameter, and unit scale parameter. The power parameter δ is introduced for the same reasons as in the power GARCH model of Ding et al. (1993), and, in this context, is constrained such that $0 < \delta < \alpha$. This differs from the model of Liu and Brorsen (1995), who constrain $\delta = \alpha$, and Panorska et al. (1995), who set $\delta = 1$, for both theoretical and practical reasons; see Mittnik et al. (2002) for further details. As in Samorodnitsky and Taqqu (1994) and Rachev and Mittnik (2000),

$$\int_{-\infty}^{\infty} e^{itx} f_S(x;\alpha,\beta,0,1)\,dx = \begin{cases} \exp\{-|t|^{\alpha}[1 - i\beta \operatorname{sign}(t) \tan \frac{\pi\alpha}{2}]\}, & \text{if } \alpha \neq 1, \\ \exp\{-|t|[1 + i\beta \frac{2}{\pi} \operatorname{sign}(t) \ln|t|]\}, & \text{if } \alpha = 1, \end{cases} \tag{3.4}$$

is the characteristic function of the standard (location zero, scale one) asymmetric stable Paretian distribution and f_S denotes its density function. For $\alpha = 2$, the standard stable Paretian distribution coincides with normal distribution $\mathrm{N}(0,2)$, while for $\alpha < 2$, ε_t does not possess moments of order α or higher.

If the true data generating process of the observed data is actually given by (3.3) and the model is estimated with a consistent estimator such as maximum likelihood, then the residuals of the model should be (approximately) iid standard asymmetric stable, and thus obey the summability property. A test for the summability property in the context of stable GARCH models with correct size and reasonable power against alternatives (such as Student's t and mixed normal innovations) has been proposed and studied in Paolella (2001), and further applied in Mittnik et al. (2000). The results from those studies indicate that, for several stock indices and exchange rate series, the null hypothesis of stable innovations is tenable, though for many series, the null hypothesis can be rejected.

Building on the success of the mixed normal for capturing the skewness and excess kurtosis of asset returns, several ways have been proposed which combine mixture distributions with GARCH-type structures. These are reviewed in Haas et al. (2004a), hereafter HMP, in which a general model structure is proposed and its stationarity conditions are derived. The model nests previously proposed models which worked with various parameter-restricted, two-component structures, such as those in Vlaar and Palm (1993), Palm and Vlaar (1997) and Bauwens et al. (1999), and also reduces to Bollerslev's (1986) original GARCH model in the single-component case. The model allows for conditional heteroscedasticity in each of the components as well as dynamic feedback between the components. This mixed–normal GARCH model also inherits the economic appeal attributed to the normal mixture models, and, irrespective of economic interpretation, has been shown in HMP and Kuester et al. (2005) to produce very competitive out-of-sample VaR predictions.

As in HMP, we say that time series $\{\epsilon_t\}$ is generated by an n–component mixed normal GARCH(r, s) process, denoted MixN-GARCH, if the conditional distribution of ϵ_t is an n–component mixed normal with zero mean, i.e.,

$$\epsilon_t | \mathcal{F}_{t-1} \sim \mathrm{MN}\left(\boldsymbol{\omega}, \boldsymbol{\mu}, \boldsymbol{\sigma}_t^2\right), \tag{3.5}$$

where $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_n)'$, $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)'$ and $\boldsymbol{\sigma}_t^2 = (\sigma_{1t}^2, \ldots, \sigma_{nt}^2)'$ are column vectors, and the mixed normal density is given by

$$f_{\mathrm{MN}}\left(y; \boldsymbol{\omega}, \boldsymbol{\mu}, \boldsymbol{\sigma}_t^2\right) = \sum_{j=1}^{n} \omega_j \phi\left(y; \mu_j, \sigma_{jt}^2\right),$$

ϕ is the normal density, $\omega_j \in (0,1)$ with $\sum_{j=1}^{n} \omega_j = 1$ and, to ensure zero mean, $\mu_n = -\sum_{j=1}^{n-1} (\omega_j/\omega_n)\, \mu_j$. The component variances, $\boldsymbol{\sigma}_t^{(2)}$, follow the GARCH–like structure

$$\boldsymbol{\sigma}_t^{(2)} = \gamma_0 + \sum_{i=1}^{r} \boldsymbol{\gamma}_i \epsilon_{t-i}^2 + \sum_{j=1}^{s} \boldsymbol{\Psi}_j \boldsymbol{\sigma}_{t-j}^{(2)}, \tag{3.6}$$

where $\boldsymbol{\gamma}_i = (\gamma_{i1}, \gamma_{i2}, \ldots, \gamma_{in})'$, $i = 0, \ldots, r$, are $n \times 1$ vectors, and $\boldsymbol{\Psi}_j$, $j = 1, \ldots, s$, are $n \times n$ matrices. In virtually all applications, $r = s = 1$ suffices, which agrees with the vast number of papers which implement the regular normal GARCH model. In the remainder of this paper, we take $r = s = 1$. As discussed and demonstrated in HMP, it is also reasonable to restrict $\boldsymbol{\Psi}_j$ to be diagonal because it results in a much more parsimonious model which is superior to the fully parameterized one according to the usual model selection criteria AIC and BIC. It is noteworthy that, for $n > 1$, the MixN-GARCH model (with or without the diagonal restriction on the $\boldsymbol{\Psi}_j$ matrices) gives rise to time varying skewness, a stylized fact which has been noted by several authors (see e.g., Harvey and Siddique (1999); Rockinger and Jondeau (2002)).

It is plausible that the component of the mixture assigned to the most volatile observations does not require a GARCH structure, i.e., occasionally occurring jumps in the level of volatility may be captured by a component with a relatively large, but constant, variance. We denote by MixN(n, g) the model given by Equations (3.5) and (3.6), with $r = s = 1$ and diagonal $\boldsymbol{\Psi}_1$, and with n component densities, but such that only g, $g \leq n$, follow a GARCH(1,1) process (and $n - g$ components restricted to be constant). For capturing the weak correlation in the returns, the AR(1) model $r_t = a_0 + a_1 r_{t-1} + \epsilon_t$ is used jointly with the GARCH structure (3.5) and (3.6).

Kuester et al. (2005) demonstrate that one or more of the component densities of the MixN model can still exhibit tails which are fatter than the normal. Similar to replacing the normal assumption in a standard GARCH model, one ad-hoc remedy is to replace the mixed normality assumption by a more flexible (symmetric) distribution. For example, HMP used a Student's t for each component while Kuester et al. (2005) applied the generalized exponential distribution, or GED, which was found to be superior to the use of the Student's t. The location–zero, scale–one GED density with exponent p is given by

$$f\left(x;p\right)=\frac{p}{2\Gamma\left(p^{-1}\right)}\exp\left\{-\left|x\right|^{p}\right\},\quad p\in\mathbb{R}_{+}. \tag{3.7}$$

After rescaling, the normal and Laplace distributions arise as special cases for $p = 2$ and $p = 1$, respectively. As $p \to \infty$, the GED approaches a uniform distribution. The cumulative distribution function (CDF) is required for VaR calculations; straightforward calculation shows that the CDF for $x \leq 0$ is given by

$$F(x;p)=\frac{1}{2}\left(1-\bar{\Gamma}_{(-x)^{p}}\left(p^{-1}\right)\right),\quad x\leq 0, \tag{3.8}$$

where $\bar{\Gamma}$ is the incomplete gamma ratio. The symmetry of the density implies that $F(x) = 1 - F(-x)$, from which $F(x)$ for $x > 0$ can be computed using Equation (3.8). We denote by MixGED(n, g) the model similar to how we define MixN(n, g), but such that the n distributional components are GED instead of normally distributed, each with its own shape parameter p_i, $i = 1, \ldots, n$. These additional n parameters are (as usual) not pre-specified, but jointly estimated along with the remaining ones.

While the MixGED leads to improved model fit and forecasts, the GED is still just an ad-hoc distribution not in accordance with the central limit theorem. In light of these facts, this paper extends the MixN model by replacing the normality assumption with the symmetric stable Paretian distribution—which also nests the normal, but is still in line with the (generalized) CLT. We show in Section 3.4 that, based on almost 8000 out-of-sample VaR forecasts for the daily returns of the DAX stock index, our new model yields a highly significant improvement in VaR prediction compared to its special cases, and also improves upon the MixGED.

3.3 The MixStab Model

Analogous to the model given in Equations (3.5) and (3.6), time series $\{\epsilon_t\}$ follows an n–component mixed stable GARCH(r, s) process, denoted MixStab-GARCH, if the distribution of $\epsilon_t \mid \mathcal{F}_{t-1}$ is a weighted mixture of symmetric stable distributions, i.e., its density at some real value x is given by

$$f_{\epsilon_t|\mathcal{F}_{t-1}}(x; \boldsymbol{\alpha}, \boldsymbol{\omega}, \boldsymbol{\mu}, \boldsymbol{\sigma}_t^{(\delta)}) = \sum_{j=1}^{n} \omega_j f_S(x; \alpha_j, 0, \mu_j, \sigma_{jt}^{\delta}), \tag{3.9}$$

where $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)'$ is the set of tail indices corresponding to the n symmetric stable distributional components, and, as before, $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_n)'$ is the set of nonnegative weights which sum to one, $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)$ is the set of component means, and now $\boldsymbol{\sigma}_t^{(\delta)} = (\sigma_{1t}^{\delta}, \ldots, \sigma_{nt}^{\delta})'$ is the set of strictly positive scale parameters, and $f_S(x; \alpha, 0, \mu, c)$ is the location-μ, scale-c, symmetric stable Paretian density function with tail index α. We assume that $1 < \alpha \leq 2$, so that the mean exists,[2] and restrict $0 < \delta < \min_i \alpha_i$, which is a natural extension of the power restriction in the (single component) stable GARCH model (3.5) and (3.6). As with the MixN model, to ensure zero mean, $\mu_n = -\sum_{j=1}^{n-1} (\omega_j/\omega_n)\, \mu_j$. The component scale terms, analogous to the variance term in the MixN model, evolve according to

$$\boldsymbol{\sigma}_t^{(\delta)} = \gamma_0 + \sum_{i=1}^{r} \boldsymbol{\gamma}_i \epsilon_{t-i}^{\delta} + \sum_{j=1}^{s} \boldsymbol{\Psi}_j \boldsymbol{\sigma}_{t-j}^{(\delta)}. \tag{3.10}$$

Similar to the MixN model, MixStab(n, g) denotes the n–component mixed stable GARCH$(1, 1)$ process with diagonal $\boldsymbol{\Psi}_1$ matrix, the above constraints on the α_i and δ, and such that only g of the n components have a GARCH structure.

The resulting model nests, among others, the following models:

[2] This has never been a binding constraint; in all applications, for the single-component stable-GARCH model, the estimated value of α has always been well above one, and this will certainly be the case for n α_i–values in the MixStab-GARCH, because the richer GARCH structure accounts for even more of the fat-tailed behavior of the series.

1. The unconditional stable Paretian model, as proposed for stock returns by Mandelbrot (1963) and Fama (1965), by taking $n = 1$ component and no GARCH structure.

2. The unconditional mixed normal model from Fama (1965), Kon (1984) and others, by taking α restricted to be 2 in each of the n components and no GARCH structure.

3. The plain normal–GARCH model (3.2), by taking $n = 1$ component, and α and δ restricted to be 2.

4. The stable-GARCH model (3.3) of Mittnik et al. (2002), by taking $n = 1$ component.

5. The MixN(n, g) GARCH model (3.5) and (3.6) of HMP, by taking δ and each α_i, $i = 1, \ldots, n$, to be 2.

We have found, for a variety of data sets including the one under study in this paper, that restraining δ to be one results in very little loss of goodness of fit.[3] Given the already large number of parameters in the MixStab model, such a model simplification is welcome.[4] As such, in all MixStab models discussed below, $\delta = 1$.

From Equations (3.9) and (3.10), it is readily apparent that the likelihood of the MixStab model is straightforward to calculate, provided that a computable expression for the density of the stable Paretian distribution is available. This has been addressed in different ways by McCulloch (1998), Doganoglu and Mittnik (1998), and Mittnik, Doganoglu and Chenyao (1999). We use the fast Fourier transform–based method of the latter paper, which, by the nature of the FFT algorithm, is particularly well–suited for computing the density at a large number of (data) points. This was used in conjunction with the (in Matlab Version 7

[3]This agrees with the findings in Panorska et al. (1995) and, for a variety of other distributional assumptions, GARCH-structures, and data sets, Paolella (1997).

[4]Other constraints were tried, such as restricting all α_i to be equal, but this was clearly rejected using likelihood ratio tests and both penalty criteria AIC and BIC.

noticeably improved and very reliable) quasi-Newton multivariate minimization routines available in Matlab for computing and maximizing the likelihood of the model. As computing the likelihood entails evaluating the n stable densities at all T sample observations (in our empirical study, we use a moving window of $T = 1000$ observations), significant computational time is unavoidable, but, despite the large parameterization, numeric estimation problems did not arise, even with the use of relatively naive starting values.

As the stable distribution does not possess a finite second moment, the MixStab–GARCH process will not be covariance stationary. It may have finite unconditional moments of lower order, however. As mentioned above, in the applications, we will always assume $\delta = 1$, i.e., the absolute power GARCH specification. For this model, a sufficient condition for the existence of the unconditional first moment can be obtained by arguments similar to Bollerslev (1986) and HMP. We will only consider the GARCH(1,1) case, which certainly possesses the greatest practical relevance.

We can write the MixStab–GARCH process as

$$\epsilon_t = z_{j,t}\sigma_{j,t} + \mu_j \quad \text{with probability } \omega_j, \quad j = 1, \ldots, k, \tag{3.11}$$

where $\{z_{j,t}\}$, $j = 1, \ldots, k$, is an iid sequence of stable random variables with index α_j. It is well known (see, e.g., Samorodnitsky and Taqqu (1994); Mittnik et al. (2002)), that $\mathrm{E}(|z_{j,t}|) = \kappa_{\alpha_j}$, where

$$\kappa_{\alpha_j} = \frac{2}{\pi}\Gamma\left(1 - \frac{1}{\alpha_j}\right).$$

As $|z_{j,t}\sigma_{j,t} + \mu_j| \leq |z_{j,t}|\sigma_{j,t} + |\mu_j|$, this implies

$$\begin{aligned} \mathrm{E}(\boldsymbol{\sigma}_t|\mathcal{F}_{t-2}) &\leq \boldsymbol{\gamma}_0 + \boldsymbol{\gamma_1}\sum_{j=1}^{k}\omega_j(\kappa_{\alpha_j}\sigma_{j,t-1} + |\mu_j|) + \boldsymbol{\Psi}_1\boldsymbol{\sigma}_{t-1} && (3.12)\\ &= \mathbf{b} + \mathbf{C}\boldsymbol{\sigma}_{t-1}, && (3.13) \end{aligned}$$

where

$$\mathbf{b} := \boldsymbol{\gamma}_0 + \boldsymbol{\gamma}_1\boldsymbol{\omega}'|\boldsymbol{\mu}|, \quad |\boldsymbol{\mu}| := [|\mu_1|, \ldots, |\mu_k|]', \quad \mathbf{C} := \boldsymbol{\gamma}_1(\boldsymbol{\omega} \odot \boldsymbol{\kappa})' + \boldsymbol{\Psi}_1,$$

and $\boldsymbol{\kappa} := (\kappa_{\alpha_1}, \ldots, \kappa_{\alpha_k})'$. From (3.12), we can further deduce that

$$\mathrm{E}(\boldsymbol{\sigma}_t|\mathcal{F}_{t-3}) = \mathrm{E}[\mathrm{E}(\boldsymbol{\sigma}_t|\mathcal{F}_{t-2})|\mathcal{F}_{t-3}] \leq \mathbf{b} + \mathbf{C}\mathrm{E}(\boldsymbol{\sigma}_{t-1}|\mathcal{F}_{t-3}) \leq (\mathbf{I}_k + \mathbf{C})\mathbf{b} + \mathbf{C}^2\boldsymbol{\sigma}_{t-2}.$$

Proceeding in this fashion, we get $\mathrm{E}(\boldsymbol{\sigma}_t|\mathcal{F}_{t-\tau}) \leq \sum_{i=0}^{\tau-2} \mathbf{C}^i\mathbf{b} + \mathbf{C}^{\tau-1}\boldsymbol{\sigma}_{t-\tau+1}$, which tends to $(\mathbf{I}_k - \mathbf{C})^{-1}\mathbf{b} < \infty$ with $\tau \to \infty$, given that the largest eigenvalue of $\mathbf{C}$ is below unity. Assume that the process started indefinitely far in the past with finite first moment, then it has a finite unconditional first moment, which is bounded by $(\boldsymbol{\omega} \odot \boldsymbol{\kappa})'(\mathbf{I}_k - \mathbf{C})^{-1}\mathbf{b} + \boldsymbol{\omega}'|\boldsymbol{\mu}|$. If all component means are zero, implying a (symmetric) scale mixture of stable distributions, the condition on the eigenvalues of $\mathbf{C}$ is also necessary and sufficient for $\mathrm{E}(\epsilon_t)$ to remain finite, and the unconditional expectation can be computed exactly.

Given that the component means are constant, i.e., they do not contribute to the dynamics of the process, we suspect that the condition on $\mathbf{C}$ is also necessary in the more general case, but this is yet to be established. We may also note that the autocorrelation function of $|\epsilon_t|$ will never exist, because this requires $\mathrm{E}(\epsilon_t^2) < \infty$.

3.4 Empirical Results

The VaR forecasting analysis is carried out using the DAX 30 index, which covers the 30 largest German public companies, from Jan. 26th, 1970, to Jan. 25th, 2005. For daily closing index price p_t at time t, we work with the percentage log–returns, $r_t := 100(\ln p_t - \ln p_{t-1})$ of the index. The daily closing index price and returns are shown in Figure 3.1. Using a rolling window of length 1,000 days to account for the fact that the model parameters most likely change through time, our analysis comprises 7,758 one-step out–of–sample forecasts. For all models considered, we re–estimate the model parameters every 20 trading days. This resulted in a massive computing time reduction, in particular for the stable mixtures model (this being an admitted drawback of the model).

As discussed in the Introduction, we restrict attention to the out–of–sample forecast performance with respect to VaR, and also to one-step ahead forecasts, which is common in the literature (though not always what is desired in practice).

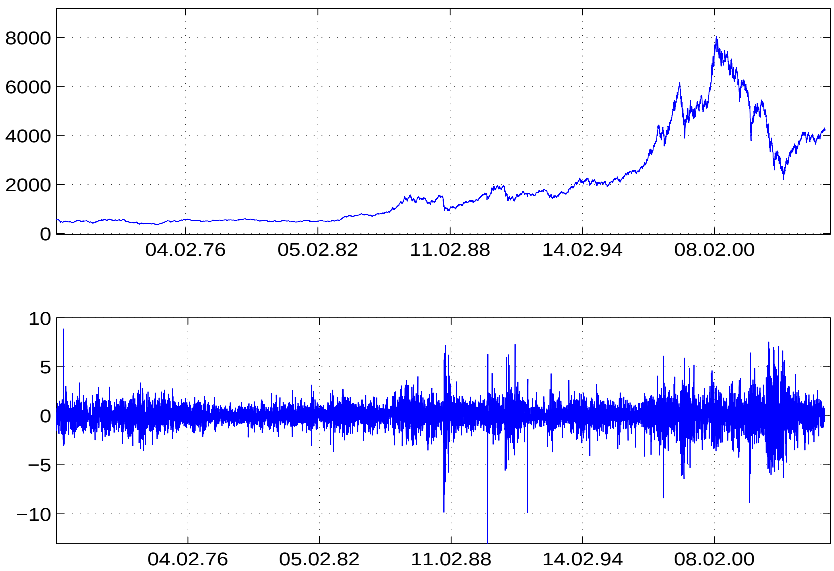

FIGURE 3.1: Percentage returns (lower panel) and levels (upper panel) of the DAX 30 Index; daily closing quotes from January 26th, 1970 to January 25th, 2005.

Multi-step ahead VaR forecasts require simulation, and are straightforward to conduct with all the models considered herein, but we limit ourselves in this study to one-step, which is adequate for demonstrating the benefit of our newly proposed model class. Also, while numerous tests for the efficiency of VaR forecasts are available (see, e.g., Christoffersen (1998); Kuester et al. (2005); and the references therein), we study only the empirical coverage probabilities associated with the VaR forecasts, the accuracy of which is by far the most important criteria.

3.4.1 Model Comparison

For the empirical out–of–sample forecasting comparison, we restrict ourselves to the conditional models because, as numerous studies including Brooks et al. (2005), Kuester et al. (2005), and Bao et al. (2004) have shown, the *un*conditional model specifications suffer from poor VaR performance across all λ–levels. These include the models 1 and 2 listed in Section 3.3.

Besides the mixture models MixN(n, g), MixGED(n, g), and MixStab(n, g), the conditional models under investigation include the normal–GARCH model (number 3 in our list) and the stable–GARCH model (number 4). Regarding the former, we take the power parameter in the GARCH equation of (3.2) to be one instead of two. This is done firstly because it enables a more direct comparison with the stable models, and secondly, because it generally leads to a better fit and better volatility forecasts than the usual choice of two; see Taylor (1986), Schwert (1989) and Nelson and Foster (1994). (The same idea can also be applied to the MixN recursion (3.6), though in our empirical studies, the difference was very minor and not worth reporting.) Regarding the stable–GARCH model, we use two forms; one restricting the stable innovations to be symmetric (so that it is precisely the one-component, MixStab$(1, 1)$ model), and unrestricted, i.e., with asymmetric stable innovations. This allows us to determine the benefit of asymmetric innovations in the one-component model. For all GARCH-type models used in our study, we always take $r = s = 1$.

In sample fit: As is common in the GARCH literature, we use the standard model selection criteria AIC and BIC (see, e.g., Burnham and Anderson, 2003, for an excellent presentation and derivation of such criteria) to rank the models with different numbers of parameters. For a K–parameter model, based on T observations and with log likelihood L at the MLE, AIC $= -2L + 2K$ and BIC $= -2L + K \log T$.

Consider first the results for the one component models of normal–GARCH, symmetric stable–GARCH, and asymmetric stable–GARCH. The top segment of Table 3.1 shows the empirical coverage probabilities (as percentages) for the 1%, 2.5%, 5% and 10% VaR–levels. The results across the three models are similar, and not particularly good. The symmetric and asymmetric stable GARCH are, as expected, better than the normal GARCH model for the more extreme VaR values, including 1% and 2.5%. The normal–GARCH performs best for the 10% VaR level.

For a more convenient overview of the performance for the three models, Figure 3.2 plots the forecast CDF value (i.e., $\hat{F}_{t+1|t}(r_{t+1})$, where $\hat{F}_{t+1|t}$ is the CDF of the predictive distribution) against the deviation from a uniform CDF for all VaR levels up to $\lambda = 0.1$. The VaR levels can be read off the horizontal axis, while the vertical axis depicts, for each VaR-level, the excess of percentage violations over the VaR-level. Thus, the relative deviation from the correct coverage can be compared across VaR levels and competing models. It is clear from Figure 3.2 that asymmetric stable–GARCH perform best for the more extreme (and indeed, most commonly used) VaR–levels, up to roughly 6%, while normal–GARCH outperforms both stable–GARCH models for the less interesting VaR–levels from 6% to 10%.[5]

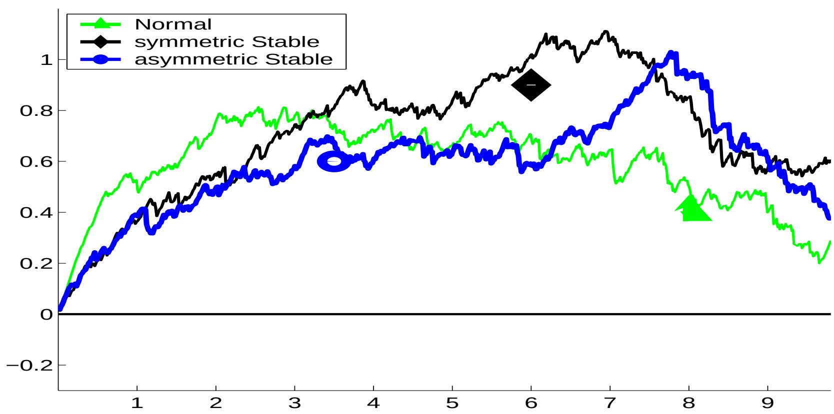

FIGURE 3.2: Deviation probability plot for the one component GARCH models with different distributional assumptions. Plotted values are "Deviation" := $100(F_U - \hat{F})$ (vertical axis) versus $100\hat{F}$ (horizontal axis), where F_U is the CDF of a uniform random variable; $\hat{F}$ refers to the empirical CDF formed from evaluating the 7,758 one–step, out–of–sample distribution forecasts at the true, observed return.

From Figure 3.2, it is readily apparent for which VaR levels, and how much, improvement is gained in going from the (one-component) symmetric stable GARCH model to the asymmetric: for virtually the whole range of VaR levels considered, and notably from 2.5% to 7.5%, there is considerable improvement. However, as we now demonstrate, the additional accuracy achieved by the mixture models is far greater than this. This underscores our finding that, for this data set and class of models, changing the innovations distribution in a (single component) GARCH model is not enough to reap large improvements in VaR quality; what is required is

[5]That the normal–GARCH model performs reasonably well for larger values of λ agrees with the findings of other empirical studies such as McNeil and Frey (2000) and Danielsson and Morimoto (2000).

a richer GARCH structure which can account for the heteroscedasticity dynamics which the usual GARCH structure (irrespective of r and s) cannot capture.

We now turn to the mixture models, which, as just mentioned, are superior to their simple, one-component counterparts. An important question inherent in these models regards the optimal number of mixture components to be used. For this data set and the chosen window length of 1,000 days, we found that, for all three distributional assumptions, the two–component models ($n = 2$) are significantly outperformed by use of $n = 3$. This agrees precisely with the findings of Kuester et al. (2005) for the NASDAQ returns data set. We thus restrict attention to the three–component models and the two choices $g = \{2, 3\}$. The VaR violation frequencies for the MixN(n, g), MixGED(n, g) and MixStab(n, g) models can be found in Table 3.1 for the 1%, 2.5%, 5% and 10% VaR levels. It is evident that the MixGED(n, g) outperforms the MixN(n, g) models, indicating that substantial nonnormality in the filtered mixture components still exists. While this was to be expected, it is less obvious that the MixStab(n, g) should emerge as the overall best performer.

As with the results for the one-component models, a broader overview of the quality for all VaR levels up to 10% is provided by the deviation plot in Figure 3.3. To serve as a reference, the normal–GARCH result from Figure 3.2 is overlaid. The most obvious (and least surprising) observation is that all mixture models perform far better than the normal–GARCH. Next, a clear ranking between the mixture models emerges from the plot: The forecasting performance of the MixGED$(3, g)$ and MixStab$(3, g)$ models is better than that of the MixN$(3, g)$, while MixStab$(3, g)$ is seen to be the best. The MixN$(3, g)$ and MixGED$(3, g)$ prediction performance, for both $g = 2$ and $g = 3$, is such that the deviation plots are above the horizontal line, i.e., these models, on average, underestimate the true frequency of extreme returns. The one–component GARCH models, regardless of the distributional assumption, underestimate even more (see Figure 3.2). On the other hand, the

MixStab($3, g$) models tend to slightly overestimate VaR, which is not surprising given their inherently very heavy–tailed nature.

We also report a measure of fit that summarizes the VaR performance of each model. As in Kuester et al. (2005), we report the mean absolute deviation (MAD) and mean squared deviation (MSD) of the actual violation frequencies from the corresponding theoretical VaR-level. Table 3.2 provides this summary information by averaging over all deviations in the (practically more interesting) interval for λ $(0, 0.05]$ and also in $(0, 0.1]$. Both summary measures indicate the same results revealed in the deviation plots, namely that the MixStab($3, g$) models provide the lowest deviation over the entire tail. For MAD(5%) and MSD(5%), MixStab(3,2) is the best performer, very closely followed by MixStab(3,3) and the MixGED models. Note that, for reasons of simplicity, both measures treat under– and overestimation of risk equally, which might not be reasonable for risk–management purposes. The performance of the stable models would profit even further from an adjusted measure which favors overestimation.

3.4.2 Further Analysis of the Mixture Models

As the MixStab and MixN models coincide by taking all the stable indices α_j to be two, it is interesting to examine how far from two the values of the estimated α_j deviate. For convenience, we just write $g = 2$ and $g = 3$ to refer to the MixStab($3, 2$) and MixStab($3, 3$) models, respectively. Figure 3.4 and 3.5 plot the estimated tail index values, though the moving window, for $g = 2$ and $g = 3$, respectively. The clear upward pattern in Figure 3.5 for $g = 3$ would seem to indicate that the returns on the DAX have become less heavy–tailed through time. Notice, however, that because of the scaling in the graph, only the first (least fat–tailed) component increases significantly.[6] The $\hat{\alpha}_j$ for $g = 2$ also indicate that the

[6] From a practitioners point of view, this would suggest that, when working with recent DAX data, the constraint $\alpha_1 = 2$ could be entertained, which would save estimation time and, depending on its validity, potentially lead to more accurate point estimates of the remaining model parameters. We leave this exercise to future, more applied work.

tails of the DAX have become less heavy. However, in stark contrast to the $g = 3$ case, all three $\hat{\alpha}_j$ sequences decrease at the beginning of the second half of the sample (which roughly coincides with the years 1987 to 1991). From Figure 3.1, we see that this window comprises three very negative returns, each of more than -9% (Oct. 19th, 1987; Oct. 16th, 1989; and Aug. 19th, 1991). A straightforward and plausible reason why the $\hat{\alpha}_j$ decrease in this period for $g = 2$ and not for $g = 3$ is that the added GARCH structure in the latter is compensating for the increased occurrence of tail events (which is exactly what the original GARCH model was designed to do).

We now briefly look at the estimated component weights for some of the models. Figure 3.6 shows the evolution of the three component weights $\hat{\omega}_j$, $j = 1, 2, 3$, for the MixStab$(3, 3)$ model, estimated throughout the moving window of 1,000 trading days, and having been updated (re-estimated) 20 days. We see immediately that the first two component weights are, in terms of relative changes, virtually constant through time, and the third component lies below 4%, and is never zero. This relative constancy of the weights contributes some evidence that the DAX market index is actually comprised of mixtures of different components.[7] Figures 3.7 and 3.8 are similar to Figure 3.6, but show the weights for the MixN(3,3) and MixGED(3,3) models, respectively. In both cases, the weights are far more volatile than with the stable mixture model. While this is not evidence, per se, that the stable model is preferable to the normal and GED, it certainly lends favor to it, in the sense that weights which are highly erratic complicate the interpretation of the model and weaken the case supporting the usage of mixtures.

[7]As the weights do move somewhat, their relative constancy is not an artefact of faulty numerical likelihood maximization (whereby the parameters "stick" to their initial values). This was also confirmed by using different starting values for some data segments.

Model	VaR(1%)	VaR(2.5%)	VaR(5%)	VaR(10%)
normal–GARCH	1.68	3.28	5.71	10.23
sym-stable–GARCH	1.43	3.27	6.01	10.63
asym-stable–GARCH	1.41	3.16	5.68	10.39
MixN$(3,2)$	1.25	2.99	5.45	10.30
MixN$(3,3)$	1.12	2.75	5.31	10.32
MixGED$(3,2)$	1.15	2.73	5.40	10.30
MixGED$(3,3)$	1.22	2.69	5.17	9.99
MixStab$(3,2)$	1.13	2.55	5.03	9.72
MixStab$(3,3)$	0.88	2.40	4.86	9.89

TABLE 3.1: Empirical VaR coverage percentages of the mixture models, MixN(n,g), MixGED(n,g) and MixStab(n,g), as well as the (one component) GARCH model with normal, symmetric stable and asymmetric stable distribution. For the mixture models, parameters (n,g) indicate a model with n components, g of which follow a GARCH process and $n-g$ components are restricted to having constant variances.

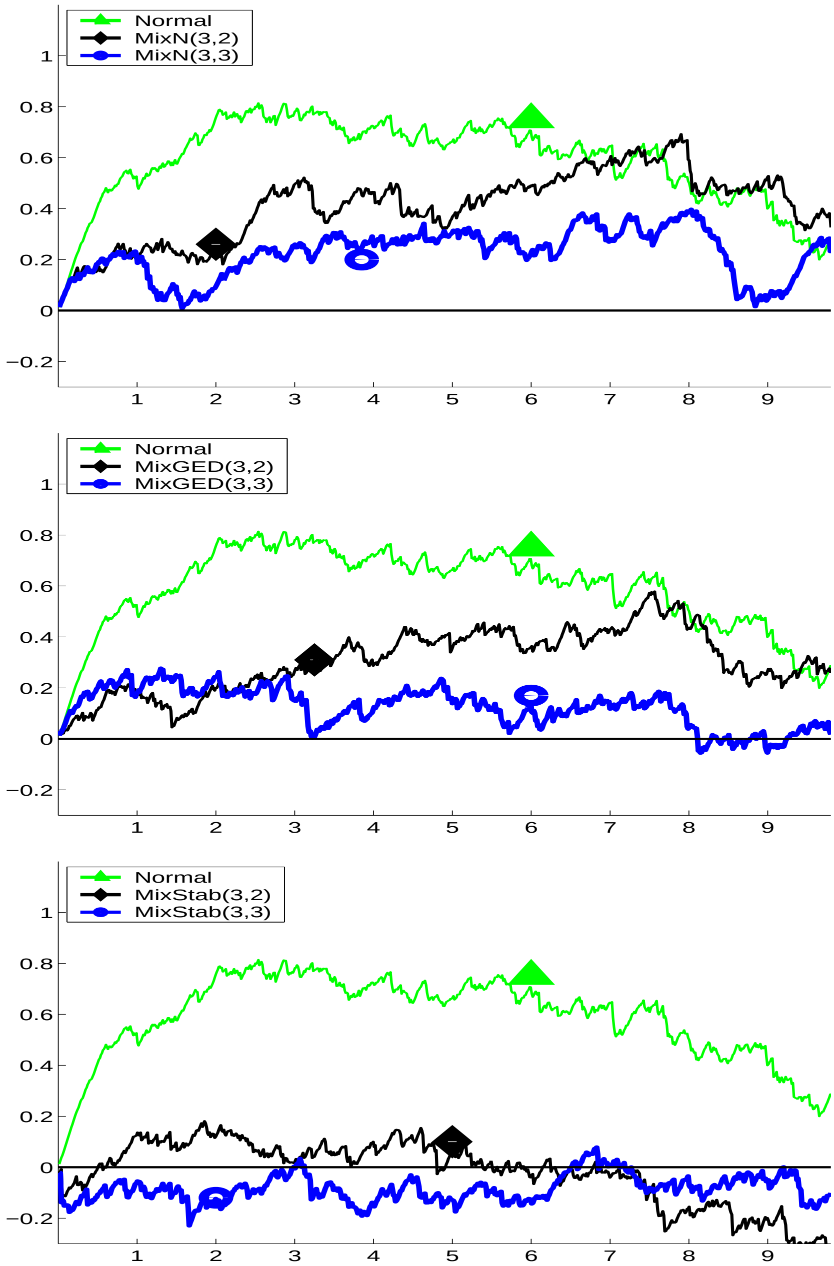

FIGURE 3.3: Same deviation probability plot as in Figure 3.2 for the GARCH models with a mixture of normal distributions (upper plot), a mixture of GED distributions (middle plot), and a mixture of stable distributions (lower plot). Mix*(n, g) is short for a mixture of distributions for a GARCH(1,1) process with n components, g of which follow a GARCH process and $n - g$ are restricted to having constant variance.

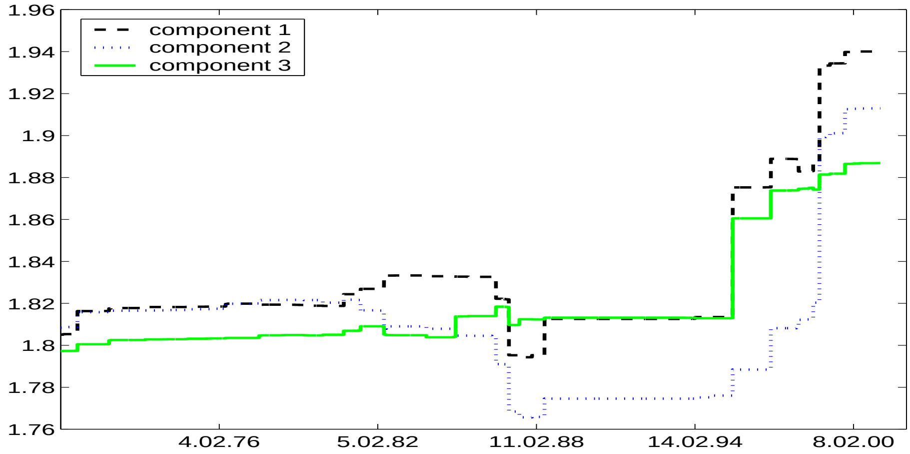

FIGURE 3.4: Evolution of the three stable indices ($\hat{\alpha}_j$, $j = 1, 2, 3$) estimated throughout the moving window of 1,000 trading days (updated only every 20 days) for the MixStab$(3, 2)$ case.

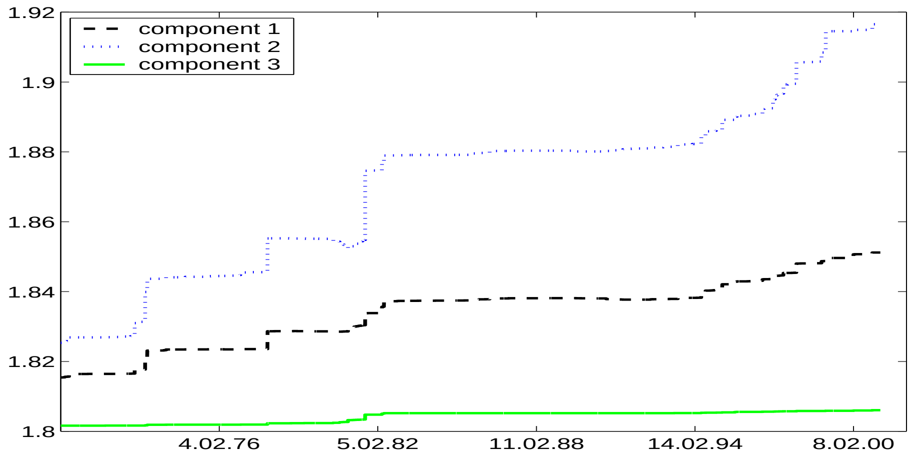

FIGURE 3.5: Similar to Figure 3.4, but with three stable indices for the MixStab$(3, 3)$ case.

3.5 Conclusions and Outlook

We have proposed a new model class which nests a variety of seemingly disparate models used for modelling asset returns and VaR prediction, and demonstrated its superiority to the most important special cases and related models in an extensive out-of-sample forecasting comparison. Our results thus lend further evidence that the stable Paretian hypothesis for the conditional modelling of asset returns is viable, but primarily when the conditional heteroscedasticity is correctly accounted for via a rich dynamic structure such as (3.10).

The fact that the test results in Paolella (2001) and Mittnik et al. (2000) based on use of the single component stable GARCH model (3.3) often rejected the stable hypothesis could potentially be due to having used too simple of a dynamic structure for capturing the conditional heteroscedasticity. The results of this paper and those in Kuester et al. (2005) clearly indicate the inadequacy of single-component models compared to the mixed structure used herein. As such, future considerations include developing testing procedures for the stable assumption which are applicable in the mixture context considered here. Also, methods using the MixStab model to deliver not just a point estimate of VaR, but also a region (which correctly takes into account the parameter uncertainly), along the lines of Bams et al. (2005), would be of interest.

Various model extensions suggest themselves for future consideration. A natural variation on the model used here would use asymmetric stable Paretian components instead of symmetric. Although this source of asymmetry is different than that achieved via the μ_i mean components, it will most likely be required to restrict the μ_i to be zero. Another viable extension is generalizing the law of motion for the scale terms in (3.10) along the lines of the GARCH models in Ding, Granger and Engle (1993) or Sentana (1995). Finally, incorporation of a Markov switching structure, as done in Haas et al. (2004b) based on the normal mixture GARCH model, or extension to time-varying component weights, as recently investigated by Haas et al. (2005) could also be entertained.

Model	MAD(5%)	MSD(5%)	MAD(10%)	MSD(10%)
normal–GARCH	0.63	0.42	0.58	0.36
sym-stable–GARCH	0.58	0.39	0.70	0.56
asym-stable–GARCH	0.49	0.26	0.58	0.38
MixN$(3,2)$	0.32	0.12	0.41	0.18
MixN$(3,3)$	0.20	0.05	0.23	0.06
MixGED$(3,2)$	0.24	0.07	0.31	0.11
MixGED$(3,3)$	0.17	0.03	0.13	0.02
MixStab$(3,2)$	0.08	0.01	0.10	0.01
MixStab$(3,3)$	0.10	0.02	0.09	0.01

TABLE 3.2: The mean of the absolute deviation (MAD) and mean squared deviation (MSD) of the empirical from the theoretical tail probability ("Deviation" in Figure 3.2 and Figure 3.3). This is computed over the first 393 and 786 (up to 5% and 10% VaR, respectively) of the sorted out–of–sample CDF values.

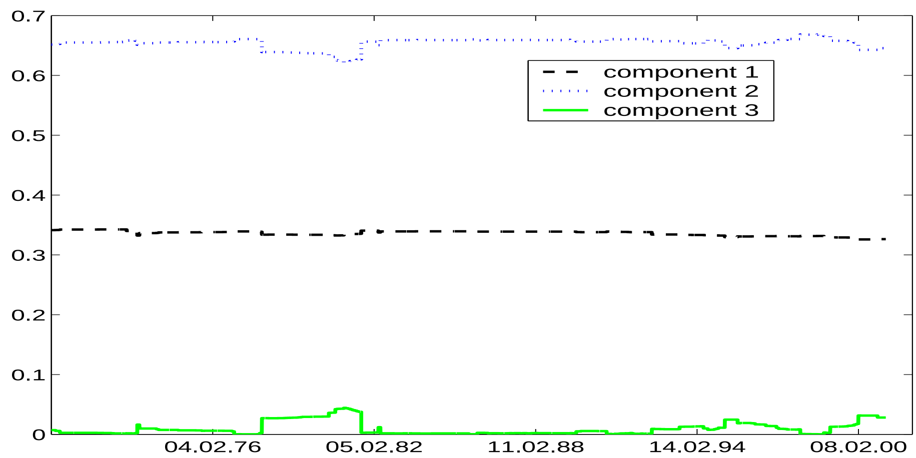

FIGURE 3.6: The evolution of the three component weights ($\hat{\omega}_j$, $j = 1, 2, 3$), estimated throughout the moving window of 1,000 trading days (updated only every 20 days) for the MixStab$(3,3)$ model.

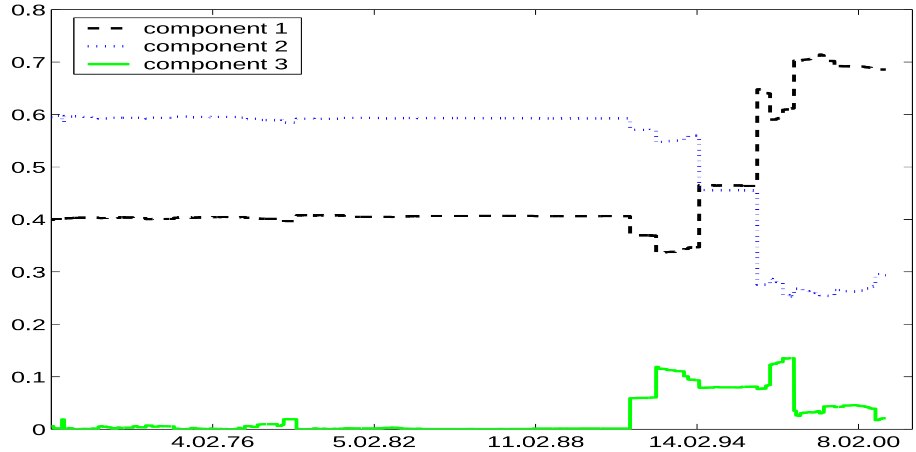

FIGURE 3.7: Same as in Figure 3.6 but for MixN(3, 3).

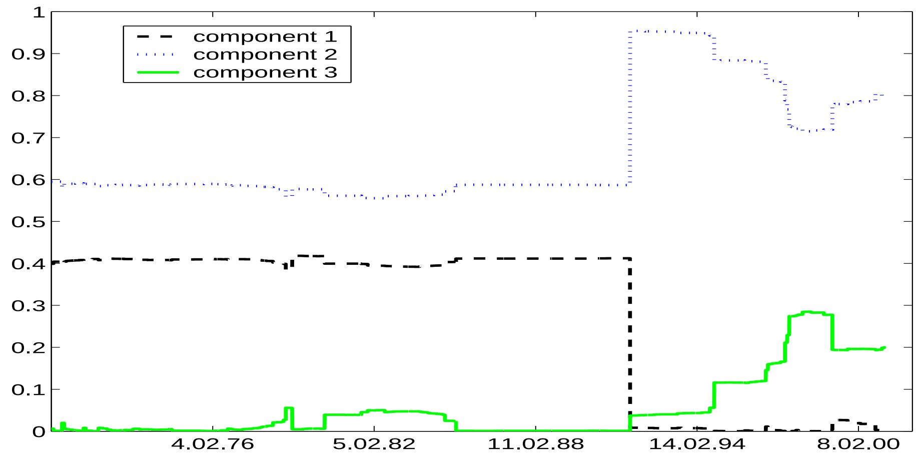

FIGURE 3.8: Same as in Figure 3.6 but for MixGED(3, 3).

Chapter 4

Analyzing and Exploiting Asymmetries in the News Impact Curve

4.1 Introduction

The use of a mixed normal distribution for modelling the unconditional distribution of asset returns is very effective, and has been considered by numerous authors, including Fama (1965), Kon (1984), Tucker and Pond (1988), and Aparicio and Estrada (2001). More recently, Kim and White (2004, p. 72) provide further evidence of the appropriateness of normal mixtures for financial data, stating "[We propose that] it may be more productive to think of the S&P500 index returns studied here as being better described as a mixture containing a predominant component that is nearly symmetric with mild kurtosis and a relatively rare component that generates highly extreme behavior." Along similar lines, Neftci (2000) argues that the extreme movements in asset prices are caused by mechanisms which are "structurally different" from the "routine functioning of markets".

The problem with any unconditional model for asset returns is that they cannot capture the blatant volatility clustering inherent in virtually all return series observed at weekly or higher frequencies, and will suffer appropriately in terms of short–term Value–at–Risk (VaR) forecasting ability. The effectiveness and easy implementation of GARCH models for this purpose is undisputed, and numerous variations and extensions of Bollerslev's (1986) original construct have been proposed and shown to deliver superior forecasts; see, for example, Palm (1996), Alexander (2001, Ch. 4), and Kuester et al. (2006) for surveys.

The mixed normal GARCH, or MixN-GARCH, is a very recent GARCH–type model class which combines the features of normal mixture distributions and a GARCH model, and has been independently proposed and investigated by Alexander and Lazar (2006) and Haas et al. (2004a,b). By judiciously coupling a k–mixture of normal distributions with a GARCH–type dynamic structure which links the k density components, several previously advocated models can be nested, and a variety of stylized facts of asset returns can be successfully modelled, such as the usual fat tails and volatility clustering, but also time-varying skewness and kurtosis. The model has been shown in the aforementioned papers to offer a plausible decomposition of the contributions to market volatility, and also to deliver highly competitive out-of-sample forecasts.

A common property of the MixN–GARCH models discussed in the aforementioned papers is constancy of the mixing weights of the component densities, which often allows for a straightforward interpretation of the contributions of the individual components. However, constancy of the distributional proportions may not be a realistic assumption in general, and, as we demonstrate below, leads to less accurate forecasts compared with a more general class of models which allows for time variation in the weights.

The choices of functional forms, or "laws of motion", of the mixing weights are discussed below; they are shown to give rise to a smooth sigmoid–type response function between the previous time period's innovation and the weights of the mixing components, where sigmoid functions, in their most general from, are smooth monotonic functions bounded between zero and one (as applied, for example, in neural networks).

While in empirical applications of the MixN–GARCH model with constant weights negative component means and higher component volatilities coincide, there is no *dynamic* asymmetry in the sense that negative shocks tend to increase *future* volatility more than positive shocks. This type of dynamic asymmetry is known as Black's (1976) leverage effect, and it is a robust characteristic of stock returns. As will be shown below, an appealing and operationally straightforward way of incorporating an asymmetric response between lagged innovations and future volatility is by relating current mixing weights to past innovations.

In the context of MixN–GARCH models, Alexander and Lazar (2005) discuss several asymmetric extensions. Their approach, however, is to employ existing asymmetric GARCH structures to extend each component's volatility process, and, thus, fundamentally differs from our approach. A discussion of their models will be given in Section 4.2.2, while an empirical comparison with our approach is provided in Section 4.6.

The remainder of this paper is as follows. Section 4.2 briefly introduces the MixN–GARCH model. Section 4.3 discusses its extension to allow for time-varying mixing weights and some choices for the functional form describing the evolution of the weights. Section 4.4 discusses the implications of the model on the news impact curve. Empirical results and out–of–sample forecasting exercise are detailed for the NASDAQ index in Section 4.5, while Section 4.6 briefly summarizes our findings for other financial return series. Section 4.7 provides concluding remarks and some ideas for future research.

4.2 Mixed Normal GARCH Models

In this section, we briefly review the mixed normal GARCH model and the asymmetric extensions considered by Alexander and Lazar (2005).

4.2.1 Mixed Normal GARCH

The mixed normal GARCH model, denoted MixN-GARCH, has recently been proposed by Haas et al. (2004a) and Alexander and Lazar (2006), and generalizes the classic normal GARCH model of Bollerslev (1986) to the normal mixture setting. In general, a random variable is said to follow a k–component normal mixture distribution if its density is given by

$$f_{\mathrm{MixN}}(y; \boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\sigma}^{(2)}) = \sum_{j=1}^{k} \lambda_j \phi(y; \mu_j, \sigma_j^2), \tag{4.1}$$

where $\phi(y; \mu_j, \sigma_j^2)$ are normal densities; $\boldsymbol{\lambda} = [\lambda_1, \ldots, \lambda_k]'$ is the vector of strictly positive mixing weights which satisfy $\sum_j \lambda_j = 1$; and the elements of $\boldsymbol{\mu} = [\mu_1, \ldots, \mu_k]'$ and $\boldsymbol{\sigma}^{(2)} = [\sigma_1^2, \ldots, \sigma_k^2]'$ are the component means and variances, respectively. If

$Y \sim \text{MixN}(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\sigma}^{(2)})$, then

$$\text{E}[Y] = \sum_{j=1}^{k} \lambda_j \mu_j, \quad \text{and} \quad \text{Var}(Y) = \sum_{j=1}^{k} \lambda_j (\sigma_j^2 + \mu_j^2) - \left(\sum_{j=1}^{k} \lambda_j \mu_j \right)^2, \tag{4.2}$$

with the latter being relevant to the discussion in Section 4.3.1 below.

In the MixN-GARCH model for asset returns, it is assumed that the *conditional* distribution of the return at time t, r_t, is MixN, that is,

$$r_t | \mathcal{F}_{t-1} \sim \text{MixN}(\boldsymbol{\lambda}_t, \boldsymbol{\mu}_t, \boldsymbol{\sigma}_t^{(2)}), \tag{4.3}$$

where $\mathcal{F}_t$ is the information set at time t. The vector of component variances, $\boldsymbol{\sigma}_t^2$, evolves according to the recursion

$$\boldsymbol{\sigma}_t^{(2)} = \alpha_0 + \sum_{i=1}^{r} \boldsymbol{\alpha}_i \epsilon_{t-i}^2 + \sum_{j=1}^{s} \boldsymbol{\beta}_j \boldsymbol{\sigma}_{t-j}^{(2)}, \tag{4.4}$$

where $\boldsymbol{\alpha}_i = (\alpha_{i1}, \alpha_{i2}, \ldots, \alpha_{ik})'$, $i = 0, \ldots, r$, are $k \times 1$ vectors; $\boldsymbol{\beta}_j$, $j = 1, \ldots, s$, are diagonal $k \times k$ matrices[1] with the component–specific persistence parameters on the main diagonal; and the innovation, ϵ_t, is

$$\epsilon_t = r_t - \text{E}(r_t | \mathcal{F}_{t-1}) = r_t - \sum_{j=1}^{k} \lambda_{j,t} \mu_{j,t}. \tag{4.5}$$

Haas et al. (2004a) considered the case where the mixing weights, $\lambda_{j,t}$, and the component means, $\mu_{j,t}$, $j = 1, \ldots, k$, are constant over time, but the generalization considered in equations (4.3)–(4.5), with these quantities being time–varying, is straightforward conceptually. In particular, in this paper, we consider MixN-GARCH specifications with time–varying mixing weights, which we discuss in the next section.

While we allow for time–varying mixing laws, we keep the mean equation simple, although it is worth mentioning that mixture models with component–specific dynamics in the conditional mean have gained some popularity in the econometrics literature. For example, mixture autoregressive (AR) models with a different

[1] The full–matrix specification is considered in Haas et al. (2004a), though, as discussed there and confirmed with other data sets, the diagonal restriction is usually favored in empirical applications.

AR structure in each mixture component include the popular Markov–switching autoregressions introduced by Hamilton (1989), as well as the models of Wong and Li (2000, 2001) and Lanne and Saikkonen (2003). However, in view of our focus on stock market returns, and given the limited success of nonlinear models in forecasting such variables (see, e.g., Boero and Marrocu, 2002), we will not pursue the case of component–specific mean dynamics, but assume that we have the same AR(p) structure in each component. That is, the conditional mean in component j can be written as

$$\mu_{j,t} = a_{0,j} + \sum_{i=1}^{p} a_i r_{t-i}, \quad j = 1, \ldots, k, \tag{4.6}$$

where only the constant $a_{0,j}$ may differ across components in order to allow for skewness of the conditional distribution.

An interesting variant of the MixN–GARCH process (4.4) arises when only a subset of the component variances gathered in the vector $\boldsymbol{\sigma}_t^{(2)}$ is subject to GARCH dynamics. For example, occasionally occurring jumps in the level of volatility may be captured by a component with a relatively large, but constant, variance. To discriminate between these model variants, we shall introduce special notation, as follows. We denote by MixN(k, g) the model given by (4.1), (4.3) and (4.4), with k component densities, but such that only g, $g \leq k$, follow a GARCH(1,1) process (and $k - g$ components are restricted to have constant variance). In the empirical work in Section 4.5 below, we take $r = s = 1$, and, regarding k and g, consider the four cases MixN$(2, 2)$, MixN$(3, 2)$, MixN$(3, 3)$ and MixN$(4, 4)$, and compare them with different competing structures that exhibit time–varying component weights.

4.2.2 Asymmetric Mixed Normal GARCH

In order to capture the leverage effect, Alexander and Lazar (2005) propose two asymmetric extensions of the MixN–GARCH model defined by (4.3) and (4.4). As we will consider these in our empirical applications below, we introduce them here.

The first of these extensions, AMixN(1)(k), uses the asymmetric GARCH specification of Engle (1990), i.e., the GARCH(1,1) process driving the variance of mixture component j is

$$\sigma_{jt}^2 = \alpha_{0j} + \alpha_{1j}(\epsilon_{t-1} - \theta_j)^2 + \beta_j \sigma_{j,t-1}^2, \quad j = 1, \ldots, k, \tag{4.7}$$

where the θ_j's are the parameters monitoring the component–specific leverage effect. In particular, if $\theta_j > 0$, then a negative shock will increase the next period's σ_{jt}^2 more than a positive shock.

The second variant, AMixN(2)(k), proposed by Alexander and Lazar (2005), employs the model of Glosten, Jagannathan, and Runkle (1993), widely known as GJR–GARCH, and specifies the variance process of component j as

$$\sigma_{jt}^2 = \alpha_{0j} + \alpha_{1j}\epsilon_{t-1}^2 + \theta_j d_{t-1}^- \epsilon_{t-1}^2 + \beta_j \sigma_{j,t-1}^2, \quad j = 1, \ldots, k, \tag{4.8}$$

where $d_{t-1}^- = 1$ if $\epsilon_{t-1} < 0$ and $d_{t-1}^- = 0$ otherwise. As in (4.7), a positive θ_j implies that $\sigma_{j,t}^2$ reacts more intensely to negative shocks than to positive shocks.

4.3 Time–Varying Weights

The idea of modelling economic variables using mixtures with time–varying mixing weights (or regime probabilities) is not new. Most notably perhaps, the Markov–switching model of Hamilton (1989), which has found many applications in macroeconomics and finance, can be interpreted in this framework. In addition, in a number of applications, mixture models with mixing weights depending on lagged process values as well as exogenous variables have been employed quite successfully. An example is the modelling of exchange rate behavior in target zones, where a jump component reflects the probability of realignments, and the probability of a jump depends on interest differentials and, possibly, further explanatory variables incorporating market expectations (see, e.g., Vlaar and Palm (1993); Bekaert and Gray (1998); Neely (1999); Klaster and Knot (2002); and Haas et al. (2006)). The conditional densities of such mixture models exhibit an enormous flexibility. For example, as illustrated by Haas et al. (2006) in an application to the EMS crisis of 1992, the predictive density may become bimodal when the probability of a realignment as well as the expected jump size are sufficiently large.

A mixture GARCH model with time–varying mixing weights which is closer to our approach, and which is, in fact, nested in our general specification, has recently been proposed by Bauwens et al. (2006). A short discussion of this model is provided in the next section.

In this paper, as mentioned in the introduction, we allow for flexible mixing weights mainly in order to capture the leverage effect, which is a robust feature of many stock return series. As this describes a negative relation between past returns and future volatilities, our specification allows the mixing weights to depend on past shocks, i.e., on the lagged ϵ_t's defined in (4.5).

4.3.1 Sigmoid Structures

A general approach is to relate the weights of the components to past innovations via various sigmoid response functions, also known as dose functions. In its most general form, a sigmoid function $S(u)$ is a smooth monotonic function such that $S(-\infty) = 0$ and $S(\infty) = 1$, e.g., $S(u) = (1 + \exp^{-u})^{-1}$, $u \in \mathbb{R}$, is a simple sigmoid, as is any cumulative distribution function (CDF) corresponding to a strictly continuous random variable.

Our general suggested model structure takes the form

$$\lambda_{jt} = \frac{W_j}{1 + \sum_{i=1}^{k-1} W_i}, \quad j = 1, \ldots, k-1, \quad \lambda_{kt} = 1 - \sum_{j=1}^{k-1} \lambda_{jt}, \tag{4.9}$$

where, mimicking the structure of an asymmetric GARCH-type model,

$$W_j = \exp\left(\gamma_{0j} + \sum_{m=1}^{u} \gamma_{mj}\epsilon_{t-m} + \sum_{m=1}^{v} \kappa_{mj}\lambda_{j,t-m} + \sum_{m=1}^{w} \delta_{mj} \left|\epsilon_{t-m}\right|^d\right), \tag{4.10}$$

and which we denote as TV(u, v, w)MixN(k, g), where u, v and w are the orders of the lagged ϵ_t, lagged λ_j, and lagged $|\epsilon_t|^d$, respectively. In this paper, we will concentrate primarily on the case with $v = w = 0$, in which case we just write the model as TV$(u)(k, g)$ or, shorter, as TV(u), when k and g are not relevant to the discussion.

The first time–varying MixN–GARCH model we consider in detail is TV(1), which relates current mixing weights to lagged innovations, ϵ_{t-1}, via sigmoid functions with two parameters: one for the location, γ_{0j}, and one for the scale, γ_{1j}, $j = 1, \ldots, k-1$. Expressions (4.9) and (4.10) reduce to

$$\lambda_{jt} = \frac{\exp\left(\gamma_{0j} + \gamma_{1j}\epsilon_{t-1}\right)}{1 + \sum_{i=1}^{k-1} \exp\left(\gamma_{0i} + \gamma_{1i}\epsilon_{t-1}\right)}, \quad j = 1, \ldots, k-1. \tag{4.11}$$

This parametrization uses $k-1$ additional parameters, compared to the MixN–GARCH model with constant mixing weights.

To illustrate the appearance of the leverage effect in this framework, consider a two–component mixture, i.e., $k = 2$ in (4.11), and let the second component have the larger variance. We have $\lambda_{1t} = e^{\gamma_0 + \gamma_1 \epsilon_{t-1}}/(1 + e^{\gamma_0 + \gamma_1 \epsilon_{t-1}})$, and $\mathrm{d}\lambda_{1t}/\mathrm{d}\epsilon_{t-1} = \gamma_1 \lambda_{1t}(1 - \lambda_{1t})$, so that, from (4.2), the overall variance at time t decreases with increasing ϵ_{t-1} if $\gamma_1 > 0$.

The second time–varying weight setting we consider in detail is TV(2), which is similar to TV(1), but such that current mixing weights depend on both ϵ_{t-1} and ϵ_{t-2}. Although Engle and Ng (1993) mentioned that the older the "news", the smaller the impact on current and future volatility, we will see below that this model seems quite promising for a variety of data sets. For TV(2), $2(k-1)$ additional parameters are required, compared to the case of constant mixing weights.

An alternative model specification, denoted TV(2^*), is a restricted form of TV(2), such that the weight of component j given by

$$\lambda_{jt} = \frac{\exp\left\{\gamma_{0j} + \gamma_{1j}\left(\epsilon_{t-1} + \epsilon_{t-2}\right)\right\}}{1 + \sum_{j=1}^{k-1} \exp\left\{\gamma_{0j} + \gamma_{1j}\left(\epsilon_{t-1} + \epsilon_{t-2}\right)\right\}}, \quad j = 1, \ldots, k-1, \tag{4.12}$$

and $\lambda_{kt} = 1 - \sum_{j=1}^{k-1} \lambda_{jt}$. The number of parameters in TV(2^*) is the same as in TV(1), but we also capture the influence of ϵ_{t-2}, albeit in a restricted fashion which constrains the impact of ϵ_{t-1} and ϵ_{t-2} to be identical. Anticipating our empirical results below, this specification is preferred for the NASDAQ data over competing models, based on the Bayesian information criterion (BIC).

Quite recently, Bauwens et al. (2006) proposed a mixture GARCH model with time–varying mixing weights which can be nested in (4.9) and (4.10). They specify a two–component model where, in (4.10), $u = v = 0$, $w = 1$, and $d = 2$, so that

$$\lambda_{1t}(\epsilon_{t-1}) = \frac{\exp\{\gamma_{01} + \delta_{11}\epsilon_{t-1}^2\}}{1 + \exp\{\gamma_{01} + \delta_{11}\epsilon_{t-1}^2\}}, \tag{4.13}$$

where $\delta_{11} > 0$, and the first component associated with weight λ_{1t} is assumed to be the low–volatility component. Thus λ_{1t} is a symmetric function of ϵ_{t-1}, and $\lambda_{1t}(-\infty) = \lambda_{1t}(\infty) = 1$. The motivation of this specification is that "large shocks have the effect of relieving pressure by reducing the probability of a large shock in the next period". However, in an application to the NASDAQ index, the authors find that, when using (4.13), the evidence for a time–varying mixing weight is weak. This is in contrast to our results for the NASDAQ when using (4.11) and (4.12), indicating that the leverage effect is stronger for this series than the pressure relieving effect, so that the specifications (4.11) and (4.12) may be preferable.

4.3.2 Non–parametric Forms

The use of the sigmoid structures allows for a fully parametric model fit. Despite their advantage in terms of estimation and inference, the natural drawback of their use is that the data is forced to fit the assumed shape. In order to mitigate this, we consider a more flexible non–parametric form which also serves as a "check" for the adequacy of the sigmoid assumption. In particular, we estimate a piecewise linear function for the weighting scheme consisting of m lines, each with zero slope and estimated intercept. Ideally then, we would have a "staircase" ascending from left to right.[2]

More precisely, consider the case with two mixture components, i.e., $k = 2$. Then for some $m \in \mathbb{N}$ and set of boundary points $\theta_1, \ldots, \theta_{m-1}$, the first weight is

[2]We also used a piecewise linear function such that the lines are forced to be connected, but the slopes get estimated. The results were qualitatively similar.

given by

$$\lambda_{1t}(\epsilon_{t-1}) = \begin{cases} b_1, & \text{if } \epsilon_{t-1} < \theta_1, \\ b_2, & \text{if } \theta_1 \le \epsilon_{t-1} < \theta_2, \\ b_3, & \text{if } \theta_2 \le \epsilon_{t-1} < \theta_3, \\ & \vdots \\ b_{m-1}, & \text{if } \theta_{m-2} \le \epsilon_{t-1} < \theta_{m-1}, \\ b_m, & \text{if } \epsilon_{t-1} \ge \theta_{m-1}, \end{cases} \tag{4.14}$$

where, as usual, $\lambda_{2t} = 1 - \lambda_{1t}$ and $0 \le b_i \le 1$, $i = 1, \ldots, m$ (but the constraint $b_i < b_j$, $i < j$, is *not* imposed). To simplify matters, the boundaries are constructed such that the difference between consecutive θ_i are equal. It is worth emphasizing that the θ_i need to be determined before the estimation is carried out, but the θ_i depend on the domain of $\lambda_{1t}(\epsilon_{t-1})$—which is unknown before the estimation because the ϵ_t result from the filtering of the estimated model. To circumvent this problem and determine an approximate range of the ϵ_t, we first estimate the model with constant weights and then, based on the range of the filtered innovations, specify appropriate values for θ_1 and θ_m. The choice of m involves the usual bias–variance tradeoff and should be chosen as a function of the sample size used for estimation.

Section 4.5.1 below discusses the estimation results for this model.

4.4 News Impact Curve

While volatility in the standard GARCH model responds equally to positive and negative return shocks, asymmetric GARCH models allow positive and negative news surprises to have different impacts on future volatility. The asymmetric response of good and bad news to future volatility, or the leverage effect (Black 1976), is such that, in theory, bad news should increase future volatility while good news should decrease future volatility. The economic rationale behind this links the stock's volatility to the firm's capital structure and goes back to Modigliani and Miller's (1958) classic work. Briefly, for a firm issuing stocks and bonds, its debt to equity ratio changes when, *ceteris paribus*, the stock price moves and, thus, the firm's leverage changes. It is often argued, however, that the degree of asymmetry in volatility is too large to be explained by leverage alone. Thus, the term "leverage effect" is used just to refer to an empirical regularity rather than to imply a theoretical explanation.

The leverage effect can be detected by calculating the correlation of lagged returns and a measure of future volatility; see, for example, Cont (2001), who defines the leverage effect of lag τ as $L(\tau) = \mathrm{Corr}(r_{t-\tau}, r_t^2)$, where $\mathrm{Corr}(a, b)$ is the correlation between a and b, and r_t^2 is used as a measure of volatility at time t.

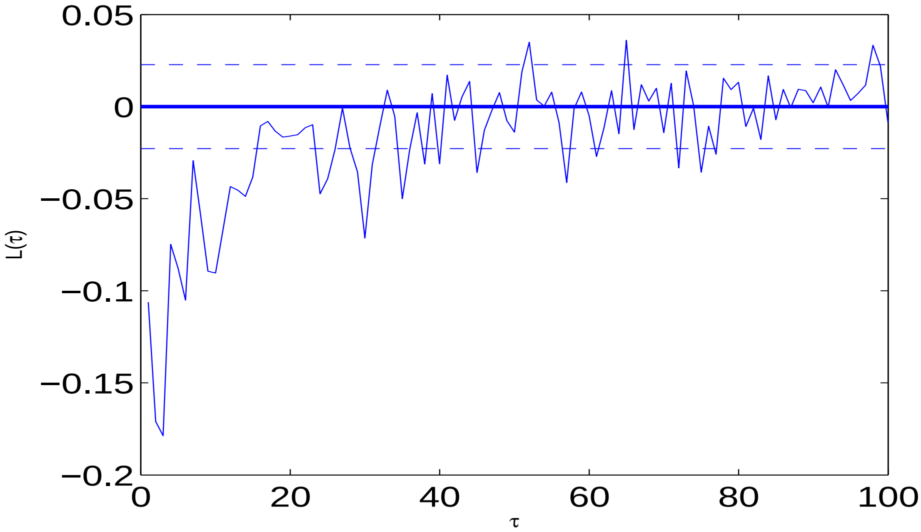

FIGURE 4.1: Estimated leverage effect, $L(\tau) = \mathrm{Corr}(r_{t-\tau}, r_t^2)$, of daily NASDAQ returns from February 1971 to June 2001. The upper and lower 95% confidence bounds are shown as dashed lines.

As is common in the financial econometrics literature, our out-of-sample forecasting exercises below will use moving windows of returns, instead of a growing sample size, to (rudimentarily) account for the fact that the data generating process is unlikely to be constant over time. To illustrate that the leverage effect is not constant throughout the entire data set, Figure 4.2 shows the leverage effect for the same data set of NASDAQ returns as in Figure 4.1, but splitting the data into eight sub–samples, each consisting of roughly four years of data. Most of the graphs exhibit the same characteristic shape of the leverage effect as observed for the whole data set, as depicted in Figure 4.1. However, for some of the sub–

samples, there is no leverage effect at all. For example, in the upper right plot in Figure 4.2, which covers the years 1975 to 1979, the correlations hover around zero for all depicted lags.

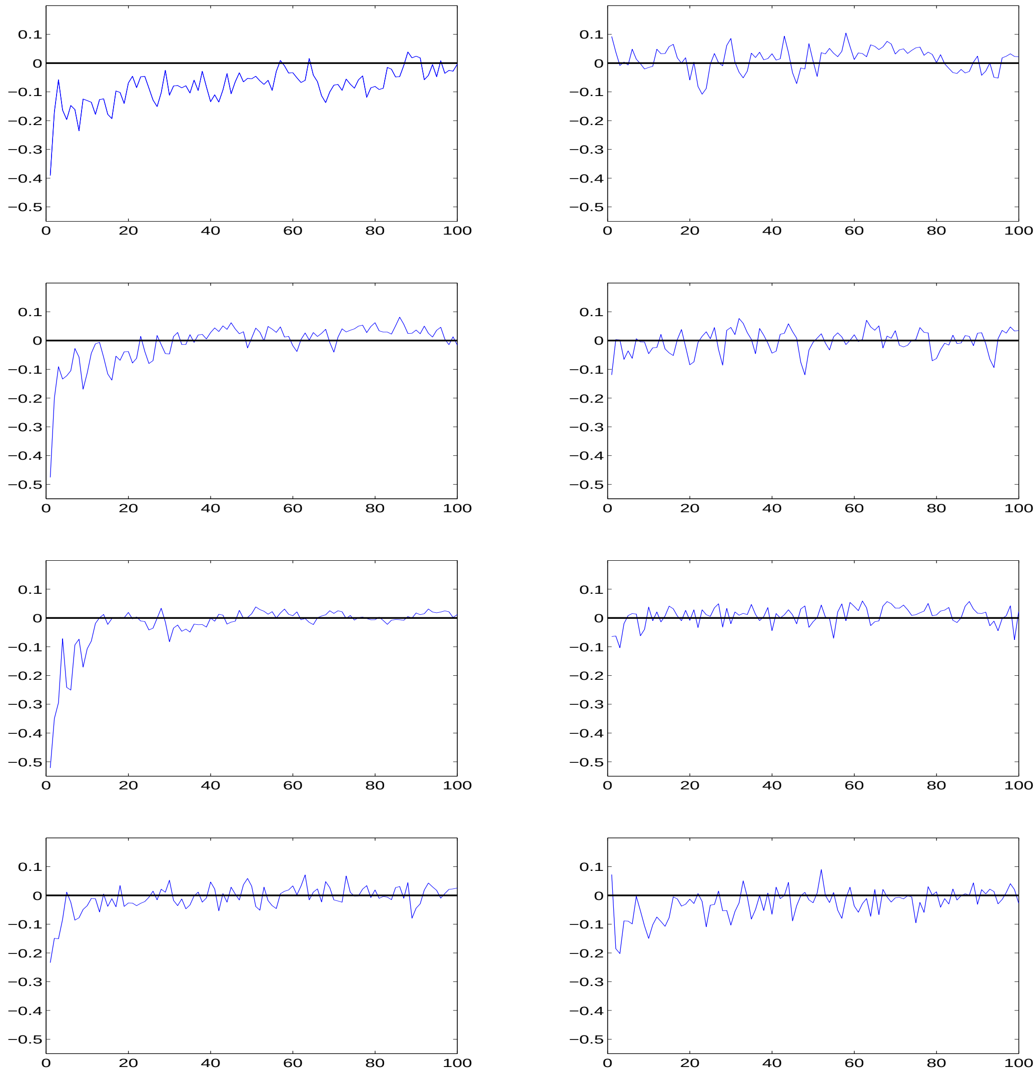

FIGURE 4.2: Same as in Figure 4.1 but divided into eight sub–samples of roughly four years of data, starting from left to right and top to bottom, e.g., the upper left graph is based on the first 1000 data points.

To further illustrate how it changes through time, Figure 4.3 shows the leverage effect for the first, second and fourth lag for a moving window of 1,000 days. Interestingly, all three correlations move somewhat together through time, e.g., in periods with large asymmetry in the market, the leverage effect increases for all lags. A further observation is that, although the leverage changes with the moving window, it is evident from Figure 4.3 that it is less pronounced and less volatile for increasing lags.

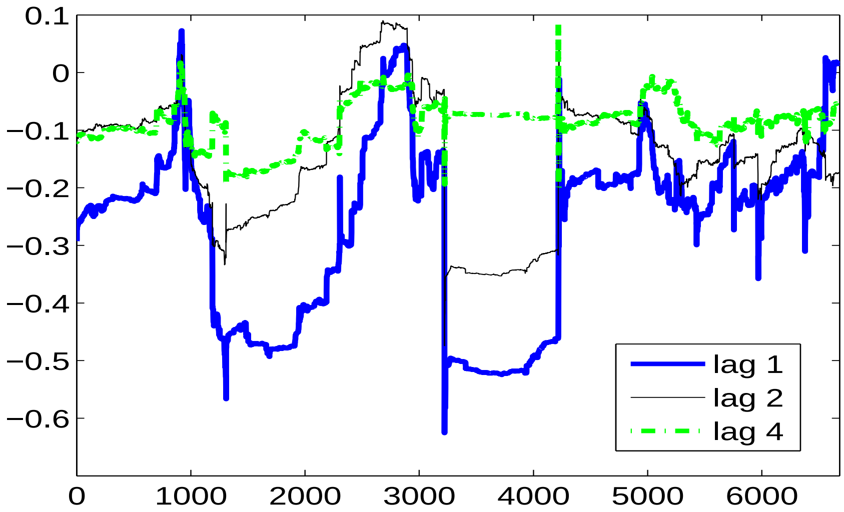

FIGURE 4.3: Leverage effect of daily NASDAQ returns February 1971 to June 2001 for a moving window of length 1000 for lag $\tau = 1, 2$ and 4. The leverage effect, here calculated as the correlation function between r_t and $r^2_{t-\tau}$ is re–calculated every day using the most recent 1000 observations.

4.5 Empirical Results: Detailed Study of the NASDAQ Returns

In this section, the empirical analysis is carried out using a set of daily NASDAQ returns from its inception in February 1971, to June 2001, using continuously compounded percentage returns, $r_t = 100\,(\log P_t - \log P_{t-1})$, where P_t denotes the index level at time t.

4.5.1 In-Sample Fit

As is common in the GARCH literature, we use the standard model selection criteria AIC and BIC (see, e.g., Burnham and Anderson, 2003, for an excellent presentation and derivation of such criteria) to rank the models with different numbers of parameters. For a K–parameter model, based on T observations and with log likelihood L at the MLE, $\text{AIC} = -2L + 2K$ and $\text{BIC} = -2L + K \log T$.

All models entertained and compared below share an AR(1)–MixN–GARCH$(1,1)$ structure with $\mathbf{\Psi}_1$ in (4.4) diagonal. Parameter estimates are obtained by numerically maximizing the conditional likelihood, where we condition on the first p returns and take the initial values of the conditional variance to be a data-driven estimate of the unconditional variance, and the initial values of ϵ_t are set to their unconditional expected values.

Table 4.1 compares the likelihood value and the two likelihood–based information criteria AIC and BIC for the different time–varying structures discussed in Section 4.3.1, as well as for the constant weight MixN–GARCH and the standard one–component GARCH model. Also, the number of parameters, K, and the rank of the fitted models with respect to each criteria are shown.

As can be seen from Table 4.1, the time–varying models perform better relative to their constant counterparts MixN(k,k) and GARCH$(1,1)$. In fact, with respect to each of the three criteria (and most notably the conservative BIC), all constant weight models are ranked last. More striking is the huge improvement of AIC and BIC values associated with the time–varying models. For example, the difference with respect to the BIC between the constant model MixN$(2,2)$ and the basic sigmoid model TV(1)$(2,2)$ is larger than 40, and the difference between MixN(k,k) and TV$(2^*)(k,k)$ is generally above 100 for $k = 2,3,4$. According to the BIC, the best performing model is TV$(2^*)(3,3)$, while for the AIC, it is TV(2)$(4,4)$. Both models also take into account the relation between current mixing weights and past innovations at time $t-1$ and $t-2$. Not surprisingly, the BIC favors the more parsimonious structure TV(2^*), for which just one parameter governs the impact of ϵ_{t-1} and ϵ_{t-2}, while the AIC favors TV(2), which allows for different coefficients of the two lagged innovations.

To illustrate the sigmoid model, Figure 4.4 shows the relation between ϵ_{t-1} and the first mixing weight, λ_{1t}, for the daily returns on the NASDAQ index from 1971 to 2001, for a time–varying weight AR(1)–MixN–GARCH$(1,1)$ model with two component densities. In this example, having a second component with a *higher*

Distributional		L		AIC		BIC	
Model	K	Value	Rank	Value	Rank	Value	Rank
GARCH(1,1)	7	–9142.8	13	18299.5	13	18348.1	13
MixN$(2, 2)$	12	–8874.5	12	17773.0	12	17856.4	11
MixN$(3, 3)$	17	–8847.6	10	17729.2	11	17847.3	10
MixN$(4, 4)$	22	–8833.6	9	17711.2	9	17864.0	12
TV(1)$(2, 2)$	13	–8849.2	11	17724.4	10	17814.7	9
TV(1)$(3, 3)$	19	–8820.5	8	17679.0	8	17811.0	8
TV(1)$(4, 4)$	25	–8792.4	5	17634.8	5	17808.5	7
TV(2)$(2, 2)$	14	–8818.4	7	17664.8	7	17762.0	4
TV(2)$(3, 3)$	22	–8786.3	3	17616.8	4	17769.6	6
TV(2)$(4, 4)$	28	**–8758.5**	**1**	**17573.0**	**1**	17767.5	5
TV$(2^*)(2, 2)$	13	–8818.4	6	17662.8	6	17753.1	3
TV$(2^*)(3, 3)$	19	–8786.9	4	17611.8	3	**17743.8**	**1**
TV$(2^*)(4, 4)$	25	–8763.0	2	17576.0	2	17749.6	2

TABLE 4.1: Likelihood–based model selection criteria and model rankings for the different time-varying model classes, as well as the MixN–GARCH model with constant weights, called MixN(k, k), and the normal-GARCH(1,1) model. All fitted models share a common AR(1)–MixN–GARCH$(1, 1)$ structure. The best model with respect to the particular criteria is highlighted boldface.

variance than the first component, this is exactly what is expected: A negative shock increases the weight of the higher variance component in the next period and a positive shock reduces the weight of the higher variance component. Thus, negative and positive shocks have an asymmetric impact on future volatility in the sense that negative news surprises increase volatility more than positive news surprises.

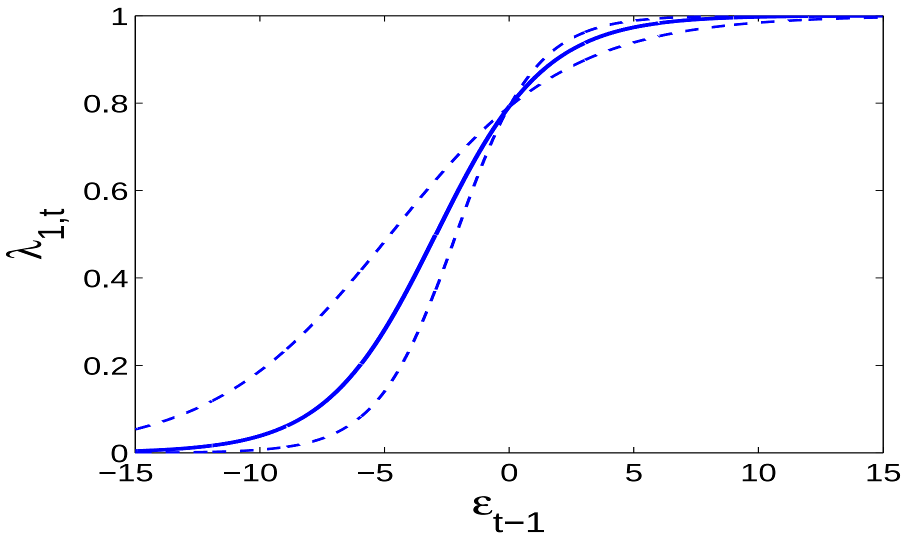

FIGURE 4.4: Estimated relation (from joint ML estimation of the full model) between lagged innovation at time $t-1$, ϵ_{t-1}, and the first mixing weight at time t, λ_{1t}, for continuously compounded daily NASDAQ returns from 1971 to 2001 for the TV(1) model (4.11) with $k = 2$ mixture components (and with an AR(1) term for the mean and $r = s = 1$); the sigmoid function has location parameter $\hat{\gamma}_{01} = 1.33$ and scale parameter $\hat{\gamma}_{11} = 0.45$. Dashed lines show the 95% confidence interval for $\hat{\gamma}_{11}$.

Figure 4.6 shows the corresponding News Impact Curve (NIC) by Engle and Ng (1993) which displays the functional relation between an unexpected return shock at time $t-1$ and the conditional variance at time t.

The dashed line refers to the NIC of the time–varying setting TV(1), and the solid line to the NIC of the constant weight MixN–GARCH model. When calculating the NIC, the variance at time $t-1$ is set to its unconditional value. Because the unconditional variance in the time–varying setting is not known, we have taken the sample mean in Figure 4.6 instead. The NIC of TV(1) clearly reveals an asymmetric behavior, while for the original MixN–GARCH model, the NIC is symmetric

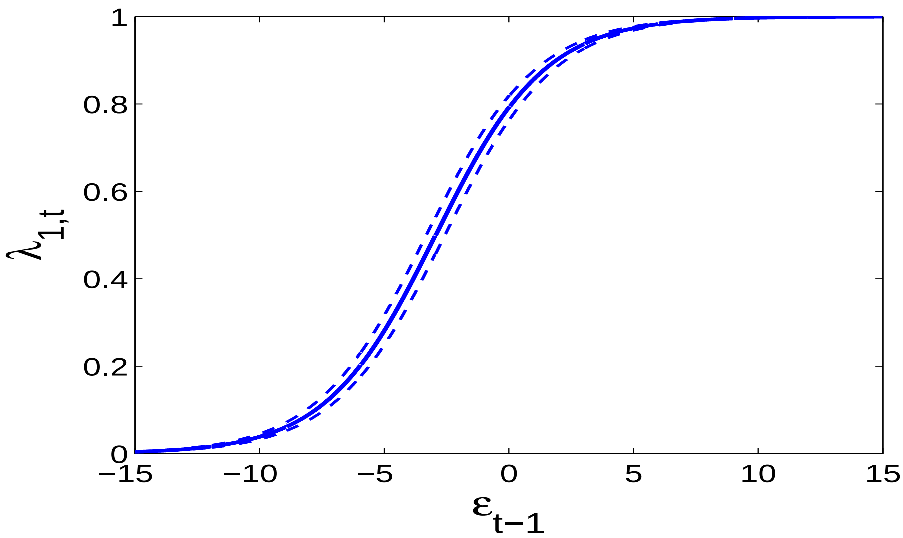

FIGURE 4.5: Same as Figure 4.4 with confidence interval for $\hat{\gamma}_{01}$.

and centered at $\epsilon_{t-1} = 0$. Another source of asymmetry in the NIC (which as can be shown our model is capable of capturing) is to shift its minimum away from $\epsilon_{t-1} = 0$, as is done, for example, in the non–linear NGARCH model of Engle and Ng (1993).

4.5.2 Non–parametric

We now turn to the results when using the non–parametric setting introduced in Section 4.3.2. After some experimenting, a reasonable choice of m in (4.14) is between 5 and 10 for the NASDAQ data set. The resulting fit using $m = 6$ is shown in Figure 4.7 for the $k = 2$ mixture model. Encouragingly, we indeed obtain the "staircase formation" as the theory predicts, further justifying our use of the sigmoid structures. For the original, constant-weight MixN–GARCH model, the graph would be a straight line at 0.822. The intercept values b_1 and b_m, which correspond to the ends of the innovation range, cannot be accurately measured because of the lack of observations in this range, as is seen by the number of overlaid innovations in the figures. This is also confirmed from the estimated standard errors of the b_i (not reported), which increase as we move into the tail of

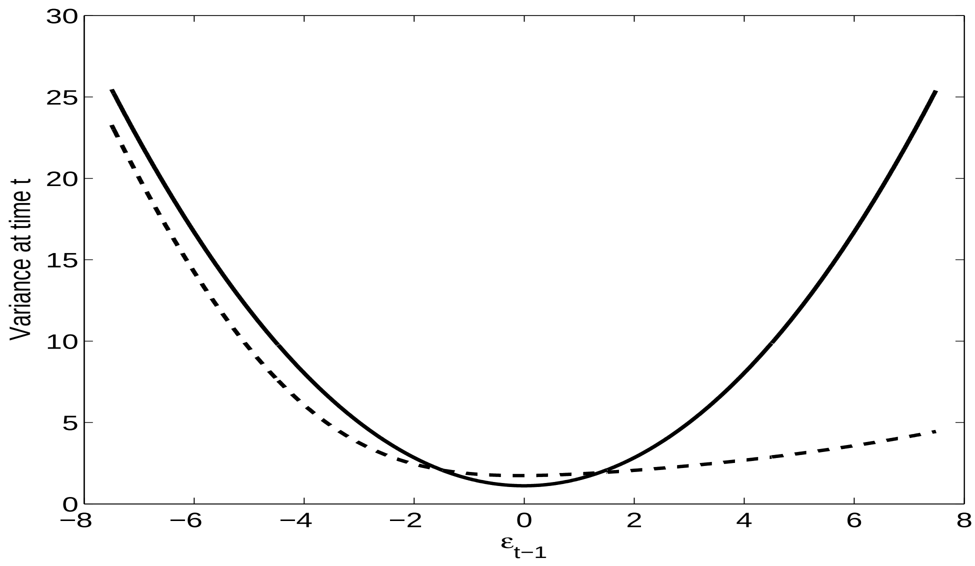

FIGURE 4.6: The asymmetric news impact curve of the fitted MixN–GARCH model with time–varying mixing weights (dashed line) and the symmetric news impact curve for the constant weight MixN–GARCH model (solid line) using the same data set as in Figure 4.1. Both models share two components and an AR(1)–MixN–GARCH$(1,1)$ process.

the innovation distribution. For the data set under study, the range of ± 8 is dense enough to yield reasonably accurate measures of the b_i.

To ensure fair evaluation of the method, the starting values of all the intercepts used in the numeric optimization of the likelihood were set to the value that results from the constant weight assumption (0.822). Use of other starting values were tried, and also resulted in the final values depicted in Figure 4.7, though in general, we have noticed that the choice of starting values can lead to local likelihood maxima.[3] Figure 4.8 is similar, but uses $m = 8$.

To help understand where the shape of the piecewise linear function is most important, Figures 4.7 and 4.8 also overlay a scaled histogram of the innovations. In the range of ± 5, the weighting function clearly exhibits an asymmetric shape, allocating more weight to the first (mild) component for positive shocks than for negative shocks. Thus, a negative shock increases the weight of the higher variance component.

[3]This is, of course, also a consequence of the algorithms chosen for optimization (in our case, the quasi-Newton methods implemented in Matlab), and more robust and sophisticated optimization techniques might obviate the need for trying several sets of starting values, though we have not pursued this.

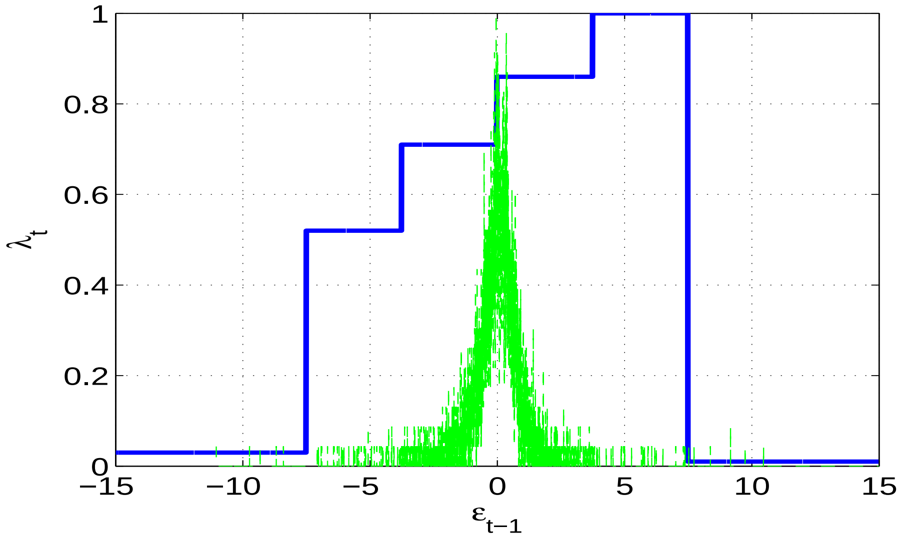

FIGURE 4.7: The non–parametric fit of the weighting function for the NASDAQ data using $m = 6$ for the $k = 2$ mixture model. Superimposed is a scaled histogram of the fitted innovations.

A potentially new stylized fact which might deserve future investigation emerges from Figures 4.7 and 4.8, regarding the weight of the "mild component" when big positive shocks hit the market. With respect to the leverage effect, positive news surprises should result in a higher weight of the low volatility component than negative news surprises, but, for big positive shocks, Figures 4.7 and 4.8 depict a very *low* weight allocated to the low volatility component. Interestingly, the corresponding NIC is flat in the area where innovations are positive (as in Figure 4.6 for the basic sigmoid structure) but sharply increases for *very large* positive innovations, which is in line with the findings of Linton and Mammen (2005), who report a very similar shape for the NIC of S&P 500 data. A potential explanation of this phenomena could be that the GARCH dynamics cannot account for the volatility clusters in every respect, and so the time-varying model is allocating more weight to the high-volatility component in both cases of big negative and big positive shocks. This might also be a result of the fact that big positive shocks also occur in volatile times, typically after a big negative shock (market rebound).

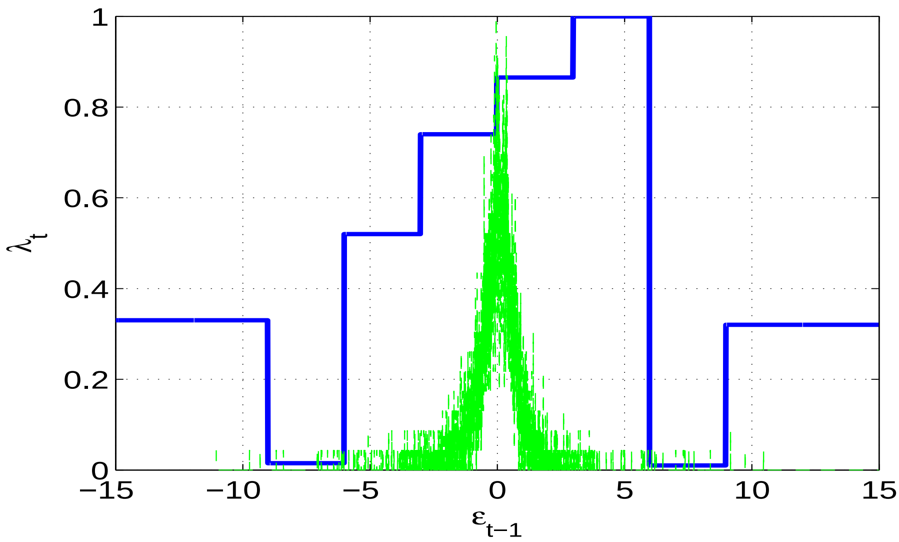

FIGURE 4.8: Same as Figure 4.7 but 8 lines instead of 6.

4.5.3 Forecasting Performance

Given that a major concern of risk management professionals is the downside loss potential of a financial position, we also study the out–of–sample forecast performance with respect to Value–at–Risk (VaR). Our primary aim is to assess the potential improvement in forecasting ability compared to constant-weight models, and so we limit ourselves in this study to one-step-ahead forecasts. The VaR with shortfall probability ξ is calculated as $\hat{F}^{\mathcal{M}}_{t|t-1}(\mathrm{VaR}_t(\xi)) = \xi$, where $\hat{F}^{\mathcal{M}}_{t|t-1}$ is the predicted return distribution function at time t based on the information set up to time $t-1$ and use of model $\mathcal{M}$. While numerous tests for the efficiency of VaR forecasts are available (see, e.g., Christoffersen and Pelletier (2004); Kuester et al. (2006); and the references therein), we consider only the empirical coverage probabilities associated with the VaR forecasts. Our forecasting exercise for the NASDAQ returns uses a rolling window of length $1,000$ days with parameter re-estimation for each window, and based on the MixN(k, g) and TV$(1)(k, g)$ models for three sets of k, g values.

Concerning the VaR forecast, the percentage of violations can be seen from Table 4.2 for the 1%, 2.5%, 5% and 10% ξ-level. Clearly, the time–varying models perform best, agreeing with the in–sample results. In fact, the time–varying models outperform their constant weight counterparts for all reported VaR levels, except for the case of the (practically less important) 10% VaR–level and $k = 3$ mixture components, in which case both structures lead to the same percentage of violations. In particular, comparing the three component models, MixN$(3, 3)$ has 1.18%, 2.93%, 5.55% and 10.19% violations for $\xi = 0.01$, 0.025, 0.05 and 0.1, respectively, and TV(1)$(3, 3)$ has 1.05%, 2.44%, 5.16% and 10.19%. Observe that, in addition to being closer to the desired nominal level, the time–varying models tend to be more conservative with respect to VaR forecasts than their constant–weight counterparts. This is a welcome fact, given that the vast majority of VaR forecasting models tend to underestimate risk (see, e.g., Kuester et al. (2005), and the references therein), which might be more costly to financial institutions than a slight overestimation of risk.

Model	VaR(1%)	VaR(2.5%)	VaR(5%)	VaR(10%)
MixN$(2, 2)$	0.91	2.86	5.78	11.14
MixN$(3, 2)$	1.29	2.86	5.66	10.43
MixN$(3, 3)$	1.18	2.93	5.55	10.19
TV(1)$(2, 2)$	0.94	2.25	5.46	10.45
TV(1)$(3, 2)$	1.27	2.47	5.37	9.96
TV(1)$(3, 3)$	1.05	2.44	5.16	10.19

TABLE 4.2: Actual percentage VaR coverage of the time–varying models, or TV(1)(k, k), as well as the constant weight models, MixN(k, k). Parameters (k, g) indicate a model with k components, g of which follow a GARCH process and $k - g$ components being restricted to having constant variances.

To further illustrate this point, the results for a spectrum of VaR levels up to $\xi = 10\%$ can be seen from Figure 4.9, which is a convenient graphical depiction of the coverage results. It plots the forecast CDF against the deviation from a uniform CDF. The VaR levels can be read off the horizontal axis, while the vertical axis depicts, for each VaR-level, the excess of percentage violations over the VaR-level. Thus, the relative deviation from the correct coverage can be compared across VaR levels and competing models. Theoretically, an ideal model would exhibit a flat line at zero. One can immediately spot that the deviations tend to be smaller for the time–varying structures. In fact, it is apparent that the time-varying structures have roughly half the deviations compared to their constant weight counterparts for all VaR levels, while TV(1)(3, 3) performs best overall.

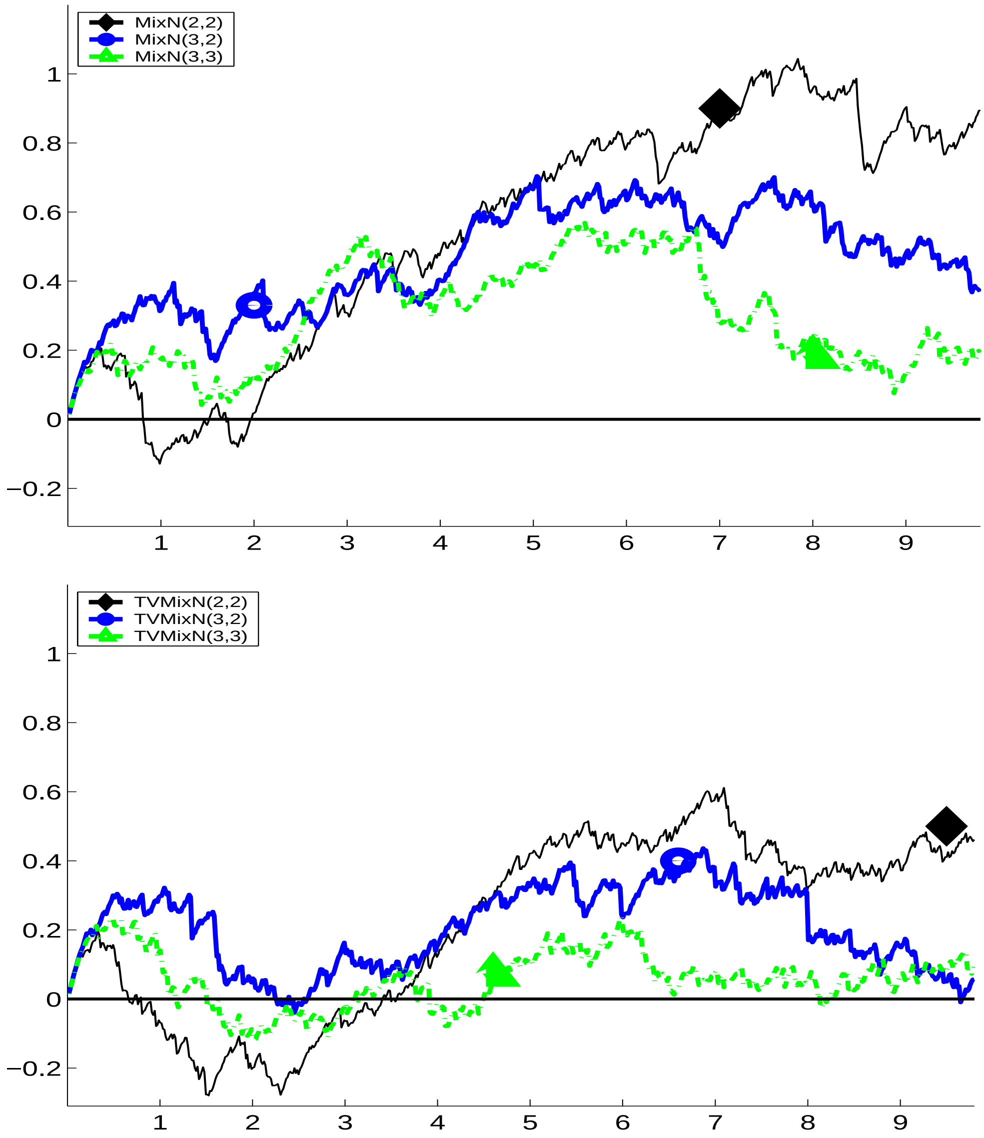

FIGURE 4.9: Deviation probability plot for the NASDAQ forecasted VaR results. The values are "Deviation" := $100(F_\xi - \hat{F})$ (vertical axis) versus $100\hat{F}$ (horizontal axis), where F_ξ is the CDF of a uniform random variable; $\hat{F}$ refers to the empirical CDF formed from evaluating the 6,681 one–step, out–of–sample distribution forecasts at the true, observed return.

Table 4.3 provides summary information for each model by averaging over all deviations in $(0, 0.05)$ and in $(0, 0.1)$. To construct a measure of fit, we computed the mean absolute deviation (MAD) and mean squared deviation (MSD) of the actual violation frequencies from the corresponding theoretical VaR-level. Both summary measures indicate the same result and underpin our previous findings: The time–varying settings provide lowest deviation over the entire tail.

Model	MAD(5%)	MSD(5%)	MAD(10%)	MSD(10%)
MixN$(2,2)$	0.28	0.12	0.56	0.42
MixN$(3,2)$	0.36	0.15	0.46	0.24
MixN$(3,3)$	0.27	0.09	0.30	0.11
TV(1)$(2,2)$	0.15	0.03	0.30	0.12
TV(1)$(3,2)$	0.16	0.04	0.20	0.05
TV(1)$(3,3)$	0.08	0.01	0.08	0.01

TABLE 4.3: The mean of the absolute deviation (MAD) and mean squared deviation (MSD) of the empirical from the theoretical tail probability ("Deviation" in Figure 4.9). This is computed over the first 334 and 668 (up to 5% and 10% VaR, respectively) of the sorted out–of–sample CDF values.

4.6 Empirical Results: Analysis of Other Data Sets

To compliment our previous findings, which were limited to just a single index (NASDAQ), we use daily data from four additional major stock indices, S&P 500, DJIA, Nikkei 225 and DAX 30, ranging from January 1970 to January 2005. While using the same model setup as for the NASDAQ, we concentrate on the out–of–sample forecasting performance and provide their MAD and MSD for varying VaR levels. Because of the enormous amount of computation required, we restrict the model class to just the simplest time–varying models TV(1)(k, g) and compare them with MixN(k, g) for three sets of k, g values. In practice, the other time–varying models would be entertained and would most likely lead to further improvement.

Table 4.4 provides an overview of the results. The best model for each data set and deviation measure is depicted in boldface. A first result is that, for all data sets, TV(1)(k, g) almost always outperforms MixN(k, g). There are some exceptions from this general scheme: For the S&P 500, MixN$(3,3)$ exhibits the lowest MAD(2.5%), but it is closely followed by that of TV(1)$(3,2)$. For the Nikkei index, the deviation measures at the 1% and 10% reported levels favor the MixN$(3,3)$ and MixN$(3,2)$ model, respectively, though as in the S&P 500 case, the differences are relatively very small. For all other deviation measures and all other data sets, the time–varying structures perform best, often with a very sizable improvement.

It is interesting to note that, overall, the TV(1)$(3,3)$ exhibits excellent performance for all data sets. This is in agreement with our results for the NASDAQ data. In particular, for the DAX index, TV(1)$(3,3)$ is the best model with respect to all reported criteria. Another noteworthy finding visible from the table is that the results of TV(1)$(3,3)$ and TV(1)$(3,2)$ are quite close, though the former is almost always (slightly) better.

Turning to the two asymmetric extensions of the mixed normal model by Alexander and Lazar (2005) the results concerning their forecasting performance can be seen in Table 4.5 for the two component structures AMixN(1)(2) and AMixN(2)(2) together with MixN$(2,2)$ and TV(1)$(2,2)$. The overall picture is that, except for higher VaR-levels and the DAX index, AMixN(1)(2) and AMixN(2)(2) are always inferior to the MixN$(2,2)$ and TV(1)$(2,2)$ in terms of forecasting various VaR-levels.

4.7 Conclusions and Further Extensions

We have relaxed the constant weights assumption in the class of mixed normal GARCH processes to allow for a more flexible time–varying weight setting. We concentrated on relating current mixing weights to past innovations via sigmoid response functions via (4.9) and (4.10) with $v = w = 0$, and limited most of the empirical work to $u = 1$. We have shown that this gives rise to a more realistic and highly asymmetric News Impact Curve, and also that in–sample tests and out–of–sample Value–at–Risk forecasting performance favor their use.

Data	Model	MAD(1%)	MSD(1%)	MAD(2.5%)	MSD(2.5%)	MAD(5%)	MSD(5%)	MAD(10%)	MSD(10%)
DAX	MixN(2, 2)	0.14	0.02	0.14	0.02	0.19	0.04	0.20	0.05
	MixN(3, 2)	0.16	0.03	0.16	0.03	0.17	0.03	0.23	0.06
	MixN(3, 3)	0.23	0.06	0.20	0.04	0.18	0.04	0.21	0.06
	TV(1)(2, 2)	0.29	0.10	0.38	0.16	0.36	0.14	0.27	0.09
	TV(1)(3, 2)	0.16	0.03	0.17	0.04	0.15	0.03	0.13	0.02
	TV(1)(3, 3)	**0.05**	**0.00**	**0.10**	**0.01**	**0.08**	**0.01**	**0.11**	**0.02**
Dow	MixN(2, 2)	0.17	0.03	0.08	0.01	0.10	0.02	0.13	0.03
	MixN(3, 2)	0.15	0.02	0.11	0.02	0.20	0.06	0.23	0.08
	MixN(3, 3)	0.21	0.05	0.16	0.03	0.13	0.02	0.14	0.02
	TV(1)(2, 2)	0.06	0.00	0.12	0.02	0.15	0.03	0.19	0.05
	TV(1)(3, 2)	0.06	0.00	0.07	0.01	**0.10**	0.01	**0.09**	**0.01**
	TV(1)(3, 3)	**0.04**	**0.00**	**0.05**	**0.00**	0.11	**0.02**	0.11	0.02
Nikkei	MixN(2, 2)	0.06	0.00	0.11	0.02	0.23	0.08	0.41	0.21
	MixN(3, 2)	0.13	0.02	0.11	0.02	0.10	0.01	**0.09**	**0.01**
	MixN(3, 3)	**0.06**	**0.00**	0.12	0.02	0.23	0.07	0.41	0.21
	TV(1)(2, 2)	0.21	0.05	0.21	0.05	0.36	0.16	0.44	0.23
	TV(1)(3, 2)	0.07	0.01	0.07	0.01	0.14	0.03	0.20	0.06
	TV(1)(3, 3)	0.08	0.01	**0.06**	**0.01**	**0.07**	**0.01**	0.14	0.03
S&P	MixN(2, 2)	0.28	0.09	0.27	0.08	0.18	0.05	0.12	0.03
	MixN(3, 2)	0.15	0.02	0.25	0.07	0.30	0.10	0.32	0.11
	MixN(3, 3)	0.09	0.01	**0.05**	0.01	0.10	0.02	0.14	0.03
	TV(1)(2, 2)	0.10	0.01	0.23	0.07	0.32	0.12	0.25	0.08
	TV(1)(3, 2)	0.10	0.01	0.07	0.01	0.06	0.01	0.08	0.01
	TV(1)(3, 3)	**0.05**	**0.00**	0.08	**0.01**	**0.05**	**0.00**	**0.06**	**0.00**

TABLE 4.4: Same as in Table 4.3 for different data sets. MAD and MSD are also shown for VaR–levels of 1% and 2.5%. Entries in boldface type indicate the best model for each criteria and data set.

Data	Model	MAD(1%)	MSD(1%)	MAD(2.5%)	MSD(2.5%)	MAD(5%)	MSD(5%)	MAD(10%)	MSD(10%)
DAX	MixN(2, 2)	**0.14**	**0.02**	**0.14**	**0.02**	0.19	0.04	0.20	0.05
	TV(1)(2, 2)	0.29	0.10	0.38	0.16	0.36	0.14	0.27	0.09
	AMixN(1)(2)	0.18	0.04	0.19	0.04	**0.13**	**0.03**	0.20	0.06
	AMixN(2)(2)	0.24	0.07	0.29	0.09	0.26	0.08	**0.19**	**0.05**
Dow	MixN(2, 2)	0.17	0.03	**0.08**	**0.01**	**0.10**	**0.02**	**0.13**	**0.03**
	TV(1)(2, 2)	**0.06**	**0.00**	0.12	0.02	0.15	0.03	0.19	0.05
	AMixN(1)(2)	0.28	0.10	0.43	0.21	0.44	0.21	0.28	0.12
	AMixN(2)(2)	0.29	0.10	0.48	0.26	0.44	0.22	0.30	0.13
Nikkei	MixN(2, 2)	**0.06**	**0.00**	**0.11**	**0.02**	**0.23**	**0.08**	**0.41**	**0.21**
	TV(1)(2, 2)	0.21	0.05	0.21	0.05	0.36	0.16	0.44	0.23
	AMixN(1)(2)	0.28	0.10	0.56	0.38	0.73	0.60	0.63	0.47
	AMixN(2)(2)	0.27	0.09	0.57	0.40	0.74	0.62	0.64	0.48
S&P	MixN(2, 2)	0.28	0.09	0.27	0.08	**0.18**	**0.05**	**0.12**	**0.03**
	TV(1)(2, 2)	**0.10**	**0.01**	**0.23**	**0.07**	0.32	0.12	0.25	0.08
	AMixN(1)(2)	0.30	0.11	0.42	0.20	0.41	0.18	0.29	0.11
	AMixN(2)(2)	0.29	0.10	0.43	0.21	0.50	0.27	0.40	0.20

TABLE 4.5: Same as in Table 4.3 for different data sets. MAD and MSD are also shown for VaR–levels of 1% and 2.5%. Entries in boldface type indicate the best model for each criteria and data set.

The model class is quite rich, and future applications should entertain assessing other choices of the sigmoid orders u, v and w. For example, we have shown that, for the NASDAQ data set, relating current mixing weights to both innovations at time $t-1$ and $t-2$ (i.e., taking $u=2$) further improves the in–sample–fit, and the dynamics could obviously be extended to include more lagged innovations beyond ϵ_{t-2} in order to account for a more flexible decay of the leverage effect. As mentioned above, Engle and Ng (1993) show that the older the news, the smaller the impact on current and future volatility. Also for the leverage effect, it is well known from Bouchaud et al. (2001) that its decay time differs across assets, with stocks requiring about 10 days, and indices about 50 days. These authors also show that the correlation function $L(\tau)$ describing the leverage effect can be fit with a (single) exponential. This fitted exponential can be directly transferred to the dynamics between current mixing weights and past $\epsilon_{t-\tau}$ (e.g., $\tau = 1, \ldots, 10$ or 50 days). The impact of $\epsilon_{t-\tau}$ at time $t-\tau$ on current mixing weights then decays exponentially with increasing lag τ. An advantage of this approach is that just two parameters are needed to model the exponential decay, instead of using one parameter for each lag.

In addition, more flexible and more asymmetric sigmoid functions might be useful in order to further account for the "down–market effect" or "panic effect", i.e., a "one–sided" leverage effect related to falling stock prices. In fact, according to Figlewski and Wang (2000), a rise in the stock price does not affect volatility at all. They find the leverage effect is just a "down-market effect" and not existent for positive news surprises which could be easily incorporated in our model by extending the constant weight assumption just for negative innovations and/or using non–parametric response functions.

Finally, in addition to, or instead of, relating current mixing weights to the past innovations and past mixing weights as suggested in (4.10), it might be advantageous to consider use of the conditional variance, skewness or kurtosis.

Chapter 5

The Leverage Effect without Leverage: An Experimental Study

5.1 Introduction

Financial time series exhibit several so called 'stylized facts' that are present in most markets. Uncorrelated returns, non-constant volatility through time and fat tailed return distributions are probably the most prominent (see for example, Cont (2001) and Granger (2005)). In addition, there are several asymmetry properties such as the skewness of the return distribution and the asymmetric impact of good and bad news to future volatility. With regards to the latter property - the so called leverage effect - Black (1976) and Christie (1982) found that volatility appears to rise when stock prices go down and to decrease when stock prices go up.

One economic rationale behind this stylized fact links the stock's volatility to the firm's capital structure. This reasoning goes back to Modigliani and Miller's (1958) classical work. Briefly, for a firm issuing stocks and bonds, its debt to equity ratio changes when, *ceteris paribus*, the stock price moves. Its leverage changes because the claims of the debt holders are limited so that (almost) all the variation in total firm value is transmitted to equity. Thus, when the stock price increases, the value of equity increases more than the value of debt so that its debt to equity ratio decreases and the firm is less risky. This results in an drop in volatility. Following the same line of reasoning, falling stock prices should lead to an increase in future volatility.

However, there is also evidence that the leverage effect observed in financial time series is not fully explained by the firm's leverage. Figlewski and Wang (2000) find a strong leverage effect for falling stock prices but for positive returns, they find a very weak or even nonexistent leverage effect. Hence, they claim that it would be more appropriate to call the leverage effect a 'down market effect'. They argue further that the firm's leverage is a level effect rather than a change effect. When the stock price changes and so its financial leverage, volatility should change permanently, too. However, volatility changes stemming from stock returns are not permanent but die out quickly. In almost the same line, Aydemir et al. (2005) quantify the leverage effect by using an equilibrium asset pricing model and find that financial leverage is not economically significant at market level while at the firm level it only partially explains variations in volatility.

For assets with no financial leverage, such as commodities, other effects are at work. Knittel and Roberts (2001) find an inverse leverage effect in electricity markets. For electricity prices volatility tends to *rise* more with *positive* shocks than with negative shocks (the opposite of Black's leverage effect). An inverse leverage effect is also found for intraday stock market data by Platen et al. (2004). Richter and Soerensen (2002) find an inverse leverage effect for soybeans. In particular, they estimate the correlation to be above 0.4, implying that when spot prices are high, volatility is also high on average.

From the discussion about the down market effect and the inverse leverage effect, we learn that financial leverage is therefore not the only driver for the relation between the level and volatility of asset prices.

In this paper we will provide clear evidence that indeed the leverage effect could also stem from different reasons than the capital structure of the firm. We do so by using experimental stock markets with no financial leverage in the underlying asset. While in real stock markets changes in volatility could result from many unobservable characteristics (for example, changes in expectation and changes in market liquidity), in experimental stock markets one has more control on the exogenous determinants. Experiments are by now a well established method in finance, as evidenced by the award of the Nobel Prize in Economics to Vernon Smith, for example.

In particular, in an experiment one can control the degree of leverage while in real settings the leverage is subject to mistakes in accounting, for example. As is common in laboratory stock markets our experiment is such that eight to 24 traders interact with each other using an electronic trading system. Trading takes place in form of a double auction and the traded asset is a single stock that pays a dividend at the end of each period. The dividend process is stochastic but its true data generating process is unknown to the traders. At the beginning of the experiment, traders start with an initial endowment of some experimental currency and some stocks and are rewarded with real money at the end of the experiment depending on their actual trading performance. To prevent falling market prices towards the end of the experiment, we use the following design. Instead of being informed about the exact amount of periods to be run, after each period investors were only informed about the likelihood of continuing the experiment.

With this basic set up, a stock price change does not affect the capital structure of the firm. So if the stock price decreases, there is no increase in the firm's debt/equity ratio nor does the firm become more highly leveraged. Because the underlying level of uncertainty remains unchanged, we expect its uncertainty stays exactly the same so we expect no leverage effect based on the firm's capital structure. However, we find a strong and significant leverage effect in all our markets.

The remainder of the paper is organized as follows. Section 2 describes the basic design of the experimental markets and the experimental setup. In Section 3 we show the results concerning the leverage effect for the generated time series and Section 4 concludes. The Appendix contains an English translation of the instructions given to the traders.

5.2 The Design of the Experiment

At the beginning of the experiment each participant received 10,000 Gulden (an imaginary currency) and five shares of a single risky asset as an initial endowment. The subjects were instructed to act as participants in the real stock markets and to use their initial endowment as well as their potential gains to participate in the one-asset stock market. The subjects were free to buy or sell shares and could transfer gains to a special account which served as the final disbursement at the end of the experiment. The trading orders, which were entered into a computer terminal, were carried out in real time and credited and debited directly to the individuals' trading accounts.

The experiment was designed such that the traded shares yielded a return paid out entirely as a dividend at the end of each period. The dividends were generated by an iid process with a higher probability of a rise in dividends. Both the iid characteristic of the dividend process and the probabilities of an increase or decrease were *unknown* to the subjects. However, all possible realizations of the randomly chosen dividends were common knowledge. In addition to the dividend information, the subjects had a chart on the trading screen which showed the development of the dividends and average trading prices in each round.

The subjects were students of the University of Zurich and ETH Zurich and were recruited from the database of interested students maintained by the Institute for Empirical Research in Economics (IEW).[1] Participants were invited for a two to three hour interactive decision-making experiment in which they could earn real money. To allow reserves for no-show ups, 26 subjects were invited for each session although just 24 subjects were needed per session. In the event more than 24 participants showed up volunteers were invited to leave the experiment and were offered a show-up fee of SFR 10. At the beginning of each session the subjects were reminded on the importance of absolute silence and individual decision-making during the entire experiment. We ensured that subjects had enough time to study the instructions at their own pace and to answer the questions at the end of the instructions. These questions served the sole purpose of making sure that everybody understood the procedure and the rules of the experiment. After all the subjects had read the instructions and answered the questions, the experimenter read a summary of the experiment. Several possible random dividend paths were

[1] The database is created and increased annually by recruiting students in 'freshman classrooms' as well as via e-mails.

simulated for five minutes before the experiment started in order to give the participants an idea of the random characteristics.

Each experiment had the following two stages. First investors had to decide how much money they wanted to save for 'consumption' and second they had to decide how many assets they wanted to buy or sell. Regarding this consumption decision, the investor was asked to decide how much money he or she wanted to transfer to the disbursement account, and how much to keep available for trading shares. It is important to mention that the final disbursement was such that the money transferred to it could not be used for trading in the future. In other words the final disbursement account was a 'one way account': the money remained in this account until the end of the session and constituted the sole payoff to the investor. All stocks were worthless in the event that the experiment ends.

Trading took place in the form of a double auction that lasted 120 seconds. Each trader could make offers and close agreements according to their own wishes, although each new bid must have been higher than the previous bid, and each new ask must have been lower than the previous ask. Throughout the whole experiment short selling was not allowed. Individual account statements were shown at the end of each period showing the amount of Gulden in the trading account, the amount of Gulden in the final disbursement account, and the amount of shares belonging to the investor.

The experiment was run in two sessions each with 24 subjects. In one of the sessions the 24 participants were divided into three groups of eight participants and in the other session all 24 participants acted in one market such that we obtained 4 independent time series. When constructing the experiment, we had to choose values for the termination probability of the sessions and the volatility of the dividend process. As already mentioned the experiment ended at random. With the termination probability of .03 for each period we obtained time series with 85 periods for the first session and with 61 periods for the second session. The dividend process we applied started with a dividend $d_{t=0}$ of 100 Gulden. It was constructed such that in the periods $t = 1, 2, \ldots, T$, the dividend d_t changed by +20%, +5%, -5% or -20% with probabilities 10%, 50%, 35% and 5%, respectively, i.e., d_t followed a stochastic process with positive drift resulting from the higher probability of rising dividends.

5.3 Data and Empirical Results

In order to give a general overview of the data of the experiment, Figures 5.1 to 5.4 illustrate the time series of the dividend process and the traded stock prices. For illustrative purposes, the figures also contain the plots of the log returns, the liquidity for all markets as well as the trading volume. Turning to the log–returns, they are shown here in order to spot some high volatility periods in the series as volatility clusters can be better seen when plotting the log-returns. Market liquidity is shown as additional information. We observe that liquidity is quite stable after some 'adjustment' periods (approximately, 5-10 periods) in the beginning.

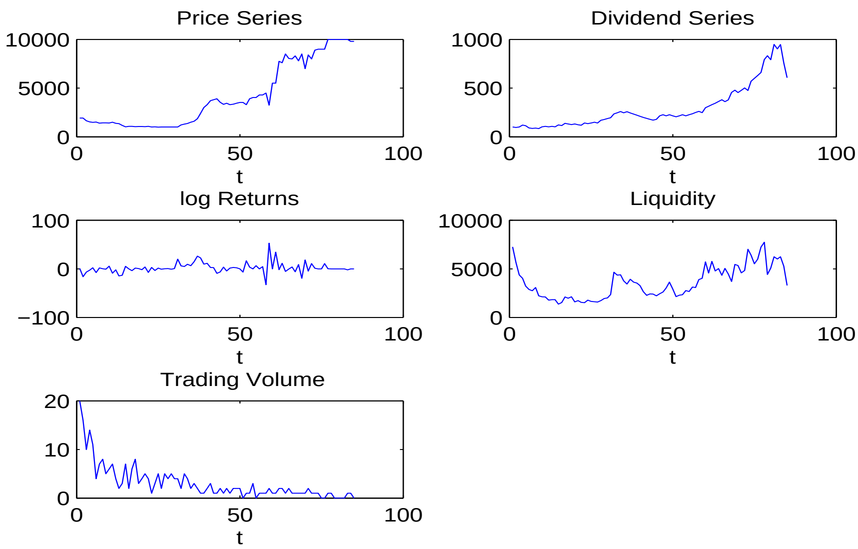

FIGURE 5.1: Price series together with the dividend, return and liquidity series of market 1.

Naturally, one can not see the leverage effect in these raw series just by eyeballing. We have to calculate it for each market. To do so, we calculate the correlation of lagged returns and a measure of future volatility. For instance, Cont

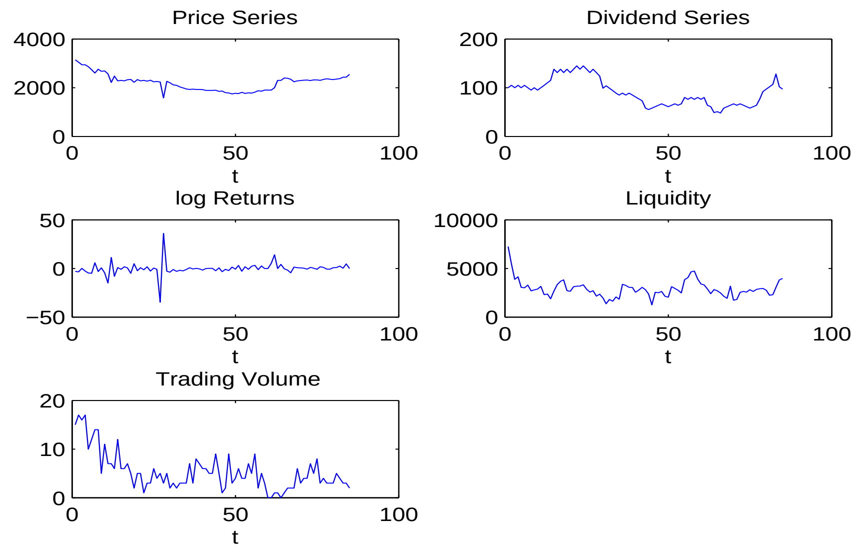

FIGURE 5.2: Same as Figure 5.1 but for market 2.

(2001), defines the leverage effect of lag τ as $L(\tau) = \text{Corr}(r_{t-\tau}, r_t^2)$, where $\text{Corr}(a, b)$ is the (linear) correlation between a and b, and r_t^2 is used as a measure of volatility at time t. The leverage is present when L is significantly negative, so that past returns and future volatility are negatively correlated.

Using the same measure as in Cont (2001) we observe a leverage effect at a significance level of 5% in the four markets, for which, the correlations of lag one, $L(1)$, are: -0.26 (0.0158), -0.44 (0.0000), -0.52 (0.0000), and -0.39 (0.0026). The associated p-values are given in parentheses. It can be seen that correlations for all four markets are negative and also statistically significant at the 5% level. Three out of four markets exhibit correlations that are even significant at the 1% level. For real stock markets, the leverage effect is also present over more than one lag but dies out after some periods, averaging 10 days in daily stock data. For two of the series we also see a leverage effect for lag 2 and 3, but the correlations are not significant.

Even when we split the time series in halve and calculate the leverage effect for both series we obtain a significant leverage for five of the eight series. As mentioned earlier the underlying asset does not exhibit a financial leverage. It is

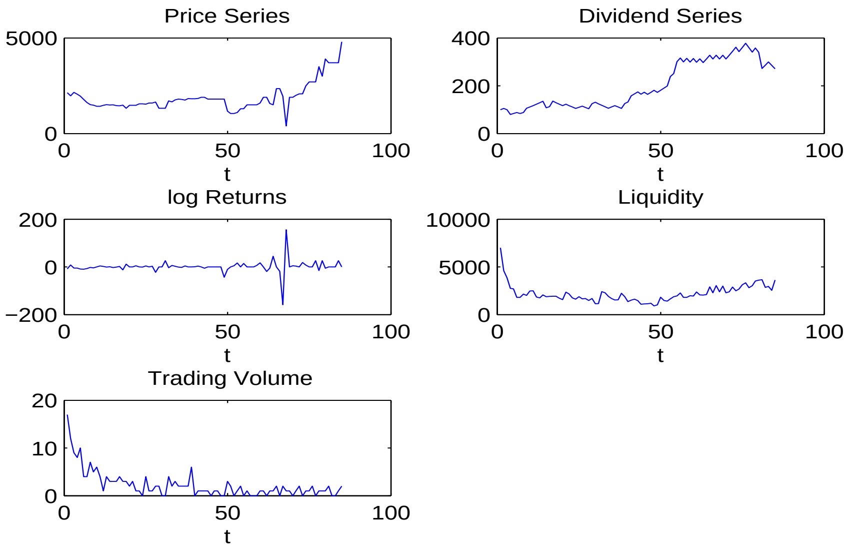

FIGURE 5.3: Same as Figure 5.1 but for market 3.

just a financial security with a dividend process given by a random walk with drift. Thus, there is no increase in leverage nor risk when the stock goes down nor is there a decrease in leverage or risk when the stock price goes up: It is therefore rather unexpected that we observe such a significant leverage effect in all four markets.

The observation of the leverage effect in these markets with an asset that is free of any financial leverage is not even restricted to a particular time series pattern. Irrespective of whether the price series go up or down in the experiment, we observe the leverage effect.

The termination probability has (theoretically) no effect on the leverage effect even if we consider the fact that subjects loose the entire amount they hold in stocks when the experiment ends. We therefore consider the end of the experiment as a state of bankruptcy. This is because although there is no predetermined length of the experiment, the termination probability is constant throughout. In the case of the leverage effect observed for truly levered stocks, this is different. The termination probability of the company, e.g., the probability of the state of bankruptcy, will increase when the stock price drops, *ceteris paribus*. As mentioned earlier, the leverage effect occurs because the state of bankruptcy becomes

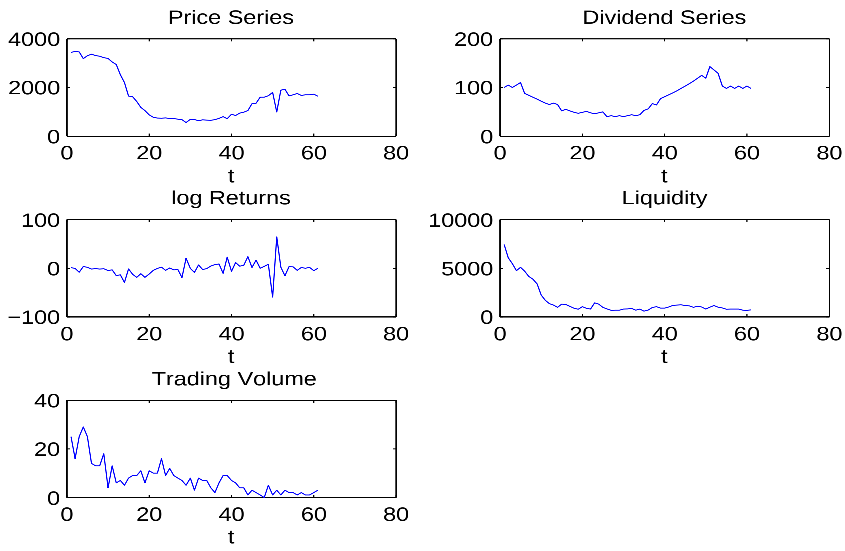

FIGURE 5.4: Same as Figure 5.1 but for market 4.

more likely when stock prices fall. As a result, the company is more risky and so volatility rises. This effect is not present in our experiment because a rise or fall in the price of the stock does not effect the termination probability of the session. Volatility should therefore stay constant, too. One can easily construct an experiment wherein the termination probability is increased (decreased) as the stock prices goes down (up). We leave this for future research.

5.4 Conclusion

We use experimental stock markets to provide evidence that the leverage effect in financial markets does not necessarily stem from the financial leverage of the firm. In all four experimental markets we explore, we find a significant leverage effect although the underlying asset does not exhibit any financial leverage. Put differently, although the capital structure of the underlying firm never changes, we observe a leverage effect in traded asset prices. It would be interesting to see if the magnitude of the leverage effect changes when we introduce an asset which exhibits different degrees of financial leverage.

5.5 Appendix: Experimental Instructions

Experiment Instructions

Overview

You are now participating in an economic experiment. Please read the following instructions carefully. You will find every information you need to take part in this experiment. If you do have further questions feel free to show hands so that we can answer your questions at your place.

Contingent on your decisions in this experiment, you can earn real money by collecting Gulden. The experiment takes place in several periods and in every period you can earn Gulden. The amount of Gulden you will earn depends on your decisions as well as on the decisions of the other participants. It is thus important that you read theses instructions carefully. We will exchange your Gulden to Franken[2] at the end of the experiment at a rate of:

100 Gulden to 15 Rappen.

Please refrain form talking for the duration of the experiment. Furthermore, you are only allowed to use the functions of the computer that are designed for the experiment. If you do not observe these rules, we will have to exclude you from the experiment and all payments, and ask you to leave. If you have questions please feel free to ask us.

At the end of these instructions you will find some control questions. Please answer them all and let us know that you are finished.

[2]One dollar was about 1.8 Franken at the time the experiment took place. And one Franken is 100 Rappen

Basic Structure of the experiment

The experiment deals with consumption decisions and trading in financial markets in which you can invest your money in a single stock or transfer it to a final disbursement account. The stock pays a dividend, the value of which is determined by chance. Be aware that the experiment ends at random. At the beginning of the experiment, i.e., at the beginning of the first period, you receive an endowment in money and stocks of:

5 stocks and 100000 Gulden.

All following periods are structured in the same way. You first decide how much money you want to transfer to your final disbursement account and how much money you want to put into stocks. Please note the following two important points. First: The money transferred to the final disbursement account can not be transferred back to the trading account during the remainder of the experiment, so it can not be used for trading stocks again. Second: Only the money on the final disbursement account will be disbursed, that means, if the experiment ends the money in the trading account and the stocks are worthless.

After you chose how much money to transfer to your final disbursement account you can start trading the stock. We will describe how the actual trading takes place further below. Keep in mind that for trading you just have the money on your trading account. For all stocks you buy you receive a dividend. This dividend payment and the money that is left on your trading account can be used to trade in the next period or can be transferred to your final disbursement account.

At the end of each period the computer decides randomly if the experiment ends or continues. The probability of ending is 3%.

Some details

Trading

Trading takes place in form of double auction lasting 120 Seconds. During this time you can make offers and close agreements according to your own wishes. The rules for trading are easy, each new bid must be higher than the previous bid, and each new ask must be lower than the previous ask. The contracts are closed via pressing the button. Gains or losses will be immediately credited or debited to your account.

Shares

All shares are based on one company and **every share pays a dividend** which is paid out at the end of each period. **The dividend payment is 100 Gulden in the first period and after that the dividend payment changes at random in every period.** The dividend can rise or fall **and the changes are such that the dividend from the previous period is multiplied by 1.20, 1.05, 0.95 or 0.80.**

However, none of the participants know the exact underlying dividend process. That means, **you do not know with what probability the dividend will rise or fall in the next period.**

Before we start the experiment, several possible **random examples** will be shown to you in a test run. **Attention!** None of the shown dividend paths will be the 'right' one in the actual experiment, they are just examples. During the experiment the computer will determine totally new random dividend paths.

In summary:

- The shares pay a dividend at the end of each period.
- The dividend is 100 Gulden in the first period but changes randomly for all other periods. The dividend can change +20%, +5%, -5% or -20%.
- The dividend payment will be announced at the end of each period.
- The experiment ends at the end of the period with a probability of 3%.

Accounts

Trading account: With the money on this account you can buy shares in each trading period (see below). Dividend payments will be transferred to this account, too.

Final disbursement account: At the beginning of every period you can transfer money to this account but please notice, you can not transfer it back to the trading account. However, just the money on this account will be paid to you as Swiss Franks when the experiment ends.

Experiment duration

The experiment lasts several periods but ends randomly. At the end of every period it will be determined whether the experiment ends or if there is another period. To give you a better feeling about the experiment duration, we will show you several **examples** with possible **durations and dividend evolutions** before the experiment starts. **Attention!** None of the durations and dividend paths will be the actual ones in the experiment. They are really just examples.

Gains

Your gain will be the actual amount on your final disbursement account at the end of the experiment.

Sequence of a period

In every period you have to decide how much money you want to save for 'consumption' and how many shares you want to buy or sell.

Now we look at each of the steps in more detail:

- Allocation of money

  You decide how much money you want to transfer from your trading account to your final disbursement account.

- Trading

  In every period you have 2 Minutes to trade. You can see the remaining time and the period on the upper part of the screen. In the middle part you can see the actual number of stocks and the amount of money in you trading account. You can also see the dividend payments for the previous and current period.

In the lower part you can trade:

a) In the window on the left hand side you can make your offer for sale to the other participants. You enter the price for which you want to sell one single share and then press 'Offer for sale'. This price will then be shown to the other participants. You can only enter whole numbers that have to be positive. Your selling order has to be lower than the current lowest one.

b) In the next window you can view the selling orders of all other participants. You can buy **one** share for the depicted price. The best offer is highlighted. If you press 'Buy' you will automatically buy from the person with the best offer. The price will be subtracted from your trading account.

c) All actually traded prices of the current period are listed in the middle window.

d) In the fourth window you can see the bids of all other participants. You can sell one of your shares for that price. The best offer is highlighted here, too. If you press 'Sell' you will automatically get the best price and the amount will be credited to your trading account.

e) In the window on the right hand side you can make you bids. Enter the price at which you want to buy one more share and press 'Bid'. This price will be shown to all other participants in the filed 'Bids'. As for selling, you can just enter whole numbers that are positive and your bid must be higher than the current highest bid.

Some trading rules for stocks

You are not allowed to sell shares that you do not own. You are not allowed to sell shares to yourself. You are not allowed to buy shares with borrowed money, e.g., you are not allowed to bid for a share for more money you actually have on your trading account.

The computer will constrain your trading behavior and will automatically enforce these rules. In case you are surprised from a refusal of execution of one of your orders, please check if one of the above rules are violated.

Do you have any questions? If not please turn the page and answer the following control questions.

Control Questions

Please answer all of the following questions. We will not start before we have not checked your answers. False answers do not have any effect on your potential gains (nor do right ones). The questionnaire serves just to make sure that everybody has understood the experiment correctly. The prices and money on the accounts are chosen so that you do not have a hard time to do the maths. In case there are any obscurities, please show hands and we will try to solve your problem.

1. Imagine a situation for participant A at the beginning of the trading. She owns 5000 Gulden on her trading account and 8 shares. Which of the following transaction plans are valid given the following chronology?

   (a) Buying one share at a price of 2500 Gulden and after that buying one more for 3000 Gulden. Then selling one share for 3050 Gulden.

   (b) Selling one share for 2400 Gulden. Later cancellation through buying for her own.

   (c) On the screen are already depicted the following selling orders: 3700, 3600, 3500, 3400 Gulden. You enter your selling order of 3400 Gulden.

2. Another situation for participant A. Before the trading starts A transfers 500 Gulden to her final disbursement account. Until now, she has 4500 Gulden on this account. At the beginning of the trading she has 6500 Gulden on her trading account and 5 shares. She buys one share from participant B for a price of 3100 Gulden and one from participant C for 3300 Gulden. She then sells one for 3500 Gulden. In the end of the period the dividend increases

20% from a level of 110. How do the individual accounts of participant A, B and C look like?

3. Given the dividend is fixed at 200 Gulden for the whole experiment. How much is the sum of all dividends that you expect when you hold one share for the whole experiment and when you know the termination probability is 3%?

   a) 10000 Gulden b) 6666 Gulden c) 1000 Gulden

   d) Own estimate:

4. What is the average price of the share you expect in the first trading period? (This question can not be answered right or wrong, meaning there is no right answer from the instructions given so far. We are just keen on hearing your opinion.)

Ready? Please show hands. We will come to you!

Bibliography

Acerbi, C. and Tasche, D. (2002a). Expected Shortfall: a natural coherent alternative to Value at Risk. *Economic Notes*, 31(2):379–388.

Acerbi, C. and Tasche, D. (2002b). On the coherence of Expected Shortfall. *Journal of Banking and Finance*, 26(7):1487–1503.

Adler, R. J., Feldman, R. E., and Taqqu, M. S., editors (1998). *A Practical Guide to Heavy Tails*. Birkhäuser, Boston.

Akgiray, V. and Booth, G. G. (1987). Compound Distribution Models of Stock Returns: An Empirical Comparison. *The Journal of Financial Research*, 10(2):269–280.

Alexander, C. and Lazar, E. (2005). Asymmetries and Volatility Regimes in the European Equity Market. ICMA Centre Discussion Papers in Finance 2005-14, The Business School for Financial Markets at the University of Reading.

Alexander, C. and Lazar, E. (2006). Normal Mixture GARCH(1,1): Applications to Exchange Rate Modelling. *Journal of Applied Econometrics*, 21:307–336.

Alexander, C. (2001). *Market Models. A Guide to Financial Data Analysis*. John Wiley & Sons, Chichester.

Amisano, G. and Giacomini, R. (2006). Comparing Density Forecasts via Weighted Likelihood Ratio Tests. *Forthcoming: Journal of Business and Economic Statistics*.

Aparicio, F. M. and Estrada, J. (2001). Empirical Distributions of Stock Returns: European Securities Markets, 1990-95. *European Journal of Finance*, 7:1–21.

Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D. (1999). Coherent Measures of Risk. *Mathematical Finance*, 9(3):203–228.

Aydemir, A. C., Gallmeyer, M., and Hollified, B. (2005). Financial Leverage Does Not Cause the Leverage Effect. *Working Paper, Available at: www.mfrc.mcgill.ca/documents/GallmeyerMichael.pdf.*

Bams, D., Lehnert, T., and Wolff, C. C. P. (2005). An Evaluation for Alternative VaR-Models. *Journal of International Money and Finance*, 24(6):944–958.

Bao, Y., Lee, T.-H., and Saltoglu, B. (2004). A Test for Density Forecast Comparison with Applications to Risk Management. Technical report, University of California, Riverside.

Bao, Y., Lee, T.-H., and Saltoglu, B. (2006). Evaluating Predictive Performance of Value-at-Risk Models in Emerging Markets: A Reality Check. *Journal of Forecasting*, 25(2):101–128.

Bauwens, L., Bos, C. S., and van Dijk, H. K. (1999). Adaptive Polar Sampling with an Application to Bayes Measure of Value-at-Risk. Tinbergen Institute Discussion Paper TI 99-082/4, Erasmus University Rotterdam.

Bauwens, L., Preminger, A., and Rombouts, J. (2006). Regime Switching GARCH Models. CORE Discussion Paper 2006/11, Université Catholique de Louvain.

Bekaert, G. and Gray, S. F. (1998). Target Zones and Exchange Rates: An Empirical Investigation. *Journal of International Economics*, 45:1–35.

Black, F. (1976). Studies in Stock Price Volatility Changes. In *American Statistical Association, Proceedings of the Business and Economic Statistics Section*, pages 177–181.

Boero, G. and Marrocu, E. (2002). The Performance of Non–Linear Exchange Rate Models: A Forecasting Comparison. *Journal of Forecasting*, 21:513–542.

Bollerslev, T., Engle, R. F., and Nelson, D. B. (1994). ARCH Models. In Engle, R. and McFadden, D., editors, *Handbook of Econometrics*, volume 4, chapter 49. Elsevier Science B.V., Amsterdam, The Netherlands.

Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, 31:307–327.

Bouchaud (2001). "Leverage Effect in Financial Markets: the Retarded Volatility Model. *Physical Review Letters*, 87:1–4.

Brooks, C., Clare, A. D., Dalle Molle, J. W., and Persand, G. (2005). A Comparison of Extreme Value Theory Approaches for Determining Value at Risk. *Journal of Empirical Finance*, 12:339–352.

Bühlmann, P. and McNeil, A. J. (2002). An Algorithm for Nonparametric GARCH Modelling. *Journal of Computational Statistics & Data Analysis*, 40(4):665–683.

Burnham, K. P. and Anderson, D. (2003). *Model Selection and Multi-Model Inference*. Springer–Verlag, New York, second edition.

Cheng, T.-C. (2005). Robust regression diagnostics with data transformations. *Computational Statistics and Data Analysis*, 49(3):875–891.

Christie, A. (1982). The Stochastic Behavior of Common Stock Variances- Value, Leverage, and Interest Rate Effects. *Journal of Financial Economic Theory*, 10:407–432.

Christoffersen, P. F. and Pelletier, D. (2004). Backtesting Value–at–Risk: A Duration–Based Approach. *Journal of Financial Econometrics*, 2:84–108.

Christoffersen, P. (1998). Evaluating Interval Forecasts. *International Economic Review*, 39(4):841–862.

Cont, R. (2001). Empirical Properties of Asset Returns: Stylized Facts and Statistical Issues. *Quantitative Finance*, 1:223–236.

Danielsson, J. and Morimoto, Y. (2000). Forecasting Extreme Financial Risk: A Critical Analysis of Practical Methods for the Japanese Market. *Monetary and Economic Studies*, 18(2):25–48.

Diebold, F. X., Gunther, T. A., and Tay, A. S. (1998). Evaluating Density Forecasts with Applications to Financial Risk Management. *International Economic Review*, 39(4):863–883.

Ding, Z., Granger, C. W., and Engle, R. F. (1993). A Long Memory Property of Stock Market Returns and a New Model. *Journal of Empirical Finance*, 1:83–106.

Doganoglu, T. and Mittnik, S. (1998). An Approximation Procedure for Asymmetric Stable Paretian Densities. Unpublished manuscript, Institute of Statistics and Econometrics, Christian Albrechts University at Kiel, Germany.

Dowd, K. (2002). *Measuring Market Risk*. John Wiley & Sons, chichester.

Engle, R. F. and Ng, V. K. (1993). Measuring and Testing the Impact of News on Volatility. *The Journal of Finance*, 48:1749–1778.

Engle, R. F. (1982). Autoregressive Conditional Heterorskedsticity with Estimates of the Variance of UK Inflation. *Econometrica*, 50(4):987–1007.

Engle, R. F. (1990). Stock Volatility and the Crash of '87: Discussion. *Review of Financial Studies*, 3:103–106.

Fama, E. F. (1965). The Behavior of Stock Market Prices. *Journal of Business*, 38:34–105.

Figlewski, S. and Wang, X. (2000). Is the "Leverage Effect" a Leverage Effect? Working paper series 00-37, New York University.

Giot, P. and Laurent, S. (2004). Modelling Daily Value–at–Risk Using Realized Volatility and ARCH Type Models. *Journal of Empirical Finance*, 11:379–398.

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.

Gouriéroux, C. (1997). *ARCH Models and Fiancial Applications.* Springer, New York.

Granger, C. W. J. (2005). The Past and Future of Empirical Finance: Some Personal Comments. *Journal of Econometrics*, 129:35–40.

Haas, M., Mittnik, S., and Mizrach, B. (2006). Assessing Central Bank Credibility During the EMS Crises: Comparing Option and Spot Market-Based Forecasts. *Journal of Financial Stability*, 2:28–54.

Haas, M., Mittnik, S., Paolella, M. S., and Steude, S.-C. (2005). Analyzing and Exploiting Asymmetries in the News Impact Curve. NCCR Working Paper Series: Financial Valuation and Risk Management, No. 256, University of Zurich.

Haas, M., Mittnik, S., and Paolella, M. S. (2004a). Mixed Normal Conditional Heteroskedasticity. *Journal of Financial Econometrics*, 2(2):211–250.

Haas, M., Mittnik, S., and Paolella, M. S. (2004b). A New Approach to Markov Switching GARCH Models. *Journal of Financial Econometrics*, 2(4):493–530.

Hadi, A. S. and Luceno, A. (1997). Maximum trimmed likelihood estimators: a unified approach, examples, and algorithms. *Computational Statistics and Data Analysis*, 25:251–272.

Hamilton, J. D. (1989). A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. *Econometrica*, 57:357–384.

Hartz, C., Mittnik, S., and Paolella, M. S. (2006). Accurate Value–at–Risk Forecasting Based on the Normal–GARCH Model. *Computational Statistics & Data Analysis*, 51:2295–2312.

Harvey, C. R. and Siddique, A. (1999). Autoregressive Conditional Skewness. *Journal of Financial and Quantitative Analysis*, 34(4):465–487.

Kim, D. and Kon, S. J. (1994). Alternative Models for the Conditional Heteroscedasticity of Stock Returns. *Journal of Business*, 67(4):563–598.

Kim, T.-H. and White, H. (2004). On More Robust Estimation of Skewness and Kurtosis. *Finance Research Letters*, 1:56–73.

Klaster, M. A. and Knot, K. H. W. (2002). Toward an Econometric Target Zone Model with Endogenous Devaluation Risk for a Small Open Economy. *Economic Modelling*, 19:509–529.

Knight, J. and (editors), S. S. (2001). *Return Distributions in Finance*. Butterworth-Heinemann, Oxford.

Knittel, C. R. and Roberts, M. R. (2001). An Empirical Examination of Deregulated Electricity Prices. Working papers sereis of the program on workable energy regulation (power), University of California Energy Institute.

Kon, S. J. (1984). Models of Stock Returns: A Comparison. *The Journal of Finance*, 39:147–165.

Kuester, K., Mittnik, S., and Paolella, M. S. (2005). Value–at–Risk Prediction: A Comparison of Alternative Strategies. Submitted.

Kuester, K., Mittnik, S., and Paolella, M. S. (2006). Value–at–Risk Prediction: A Comparison of Alternative Strategies. *Journal of Financial Econometrics*, 4:53–89.

Lanne, M. and Saikkonen, P. (2003). Modeling the U.S. Short–Term Interest Rate by Mixture Autoregressive Processes. *Journal of Financial Econometrics*, 1:96–125.

Linton, O. and Mammen, E. (2005). Estimating Semiparametric ARCH(∞) Models by Kernal Smoothing Methods. *Econometrica*, 73:771–83.

Liu, S. and Brorsen, B. W. (1995). Maximum Likelihood Estimation of a GARCH-Stable Model. *Journal of Applied Econometrics*, 10:273–285.

Mandelbrot, B. (1963). The Variation of Certain Speculative Prices. *Journal of Business*, 36:394–419.

Markatou, M. (2000). Mixture Models, Robustness, and the Weighted Likelihood Methodology. *Biometrics*, 56:483–486.

McCulloch, J. H. (1985). Interest-Risk Sensitive Deposit Insurance Premia: Stable ACH Estimates. *Journal of Banking and Finance*, 9:137–156.

McCulloch, J. H. (1996). Financial Applications of Stable Distributions. In Maddala, G. S. and Rao, C. R., editors, *Handbook of Statistics*, volume 14. Elsevier Science.

McCulloch, J. H. (1998). Numerical Approximation of the Symmetric Stable Distribution and Density. In Adler, R. J., Feldman, R. E., and Taqqu, M. S., editors, *A Practical Guide to Heavy Tails*, pages 489–499. Birkhäuser, Boston.

McDonald, J. B. (1996). Probability Distributions for Financial Models. In Maddala, G. and Rao, C., editors, *Handbook of Statistics*, volume 14. Elsevier Science.

McNeil, A. J. and Frey, R. (2000). Estimation of Tail-related Risk Measures for Heteroscedastic Financial Time Series: An Extreme Value Approach. *Journal of Empirical Finance*, 7(3):271–300.

Mittnik, S., Doganoglu, T., and Chenyao, D. (1999). Computing the Probability Density Function of the Stable Paretian Distribution. *Mathematical and Computer Modelling*, 29:235–240.

Mittnik, S., Paolella, M. S., and Rachev, S. T. (1998a). A Tail Estimator for the Index of the Stable Paretian Distribution. *Communications in Statistics–Theory and Methods*, 27(5):1239–1262.

Mittnik, S., Paolella, M. S., and Rachev, S. T. (1998b). Unconditional and Conditional Distributional Models for the Nikkei Index. *Asia-Pacific Financial Markets*, 5(2):99–128.

Mittnik, S., Paolella, M. S., and Rachev, S. T. (2000). Diagnosing and Treating the Fat Tails in Financial Returns Data. *Journal of Empirical Finance*, 7:389–416.

Mittnik, S., Paolella, M. S., and Rachev, S. T. (2002). Stationarity of Stable Power–GARCH Processes. *Journal of Econometrics*, 106:97–107.

Mittnik, S. and Paolella, M. S. (2000). Conditional Density and Value-at-Risk Prediction of Asian Currency Exchange Rates. *Journal of Forecasting*, 19:313–333.

Mittnik, S. and Paolella, M. S. (2003). Prediction of Financial Downside Risk with Heavy Tailed Conditional Distributions. In Rachev, S. T., editor, *Handbook of Heavy Tailed Distributions in Finance*. Elsevier Science, Amsterdam.

Mittnik, S. and Rachev, S. (1993a). Modeling Asset Returns With Alternative Stable Models. *Econometric Reviews*, 12:261–330.

Mittnik, S. and Rachev, S. (1993b). Reply to Comments on 'Modeling Asset Returns With Alternative Stable Models' and Some Extensions. *Econometric Reviews*, 12:347–389.

Modigliani, F. and Miller, M. (1958). The Cost of Capital, Corporation Finance, and the Theory of Investment. *American Economic Review*, 48:261–297.

Neely, C. J. (1999). Target Zones and Conditional Volatility: The Role of Realignments. *Journal of Empirical Finance*, 6:177–192.

Neftci, S. (2000). Value at Risk Calculation, Extreme Events, and Tail Estimation. *Journal of Derivatives*, 7(3):23–37.

Nelson, D. B. and Foster, D. B. (1994). Asymptotic Filtering Theory For Univariate ARCH Models. *Econometrica*, 62:1–41.

Palm, F. C. and Vlaar, P. J. G. (1997). Simple Diagnostic Procedures for Modeling Financial Time Series. *Allgemeines Statistisches Archiv*, 81:85–101.

Palm, F. C. (1996). GARCH Models of Volatility. In Maddala, G. and Rao, C., editors, *Handbook of Statistics*, volume 14, pages 209–240. Elsevier Science, Amsterdam.

Palm, F. C. (1997). GARCH Models of Volatility. In Maddala, G. and Rao, C., editors, *Handbook of Statistics*, volume 14. Elsevier Science.

Panorska, A., Mittnik, S., and Rachev, S. (1995). Stable GARCH Models for Financial Time Series. *Applied Mathematics Letters*, 8(5):33–37.

Paolella, M. S. (1997). Generalized GARCH Models with Nonnormal Distributions. Working Paper 111, Institute of Statistics and Econometrics, University of Kiel.

Paolella, M. S. (2001). Testing the Stable Paretian Assumption. *Mathematical and Computer Modelling*, 34:1095–1112.

Platen, E., West, J., and Breymann, W. (2004). Intraday Empirical Analysis of Electricity Price Behaviour. Research paper 140, Quantitative Finance Research Centre, University of Technoloty Sydney.

Rachev, S. T. (editor) (2003). *Handbook of Heavy Tailed Distributions in Finance*. North-Holland.

Rachev, S. T. and Mittnik, S. (2000). *Stable Paretian Models in Finance*. Wiley & Sons, Chichester.

Rachev, S., Kim, J.-R., and Mittnik, S. (1997). Econometric Modelling in the Presence of Heavy-Tailed Innovations. *Communications in Statistics – Stochastic Models*, 13:841–886.

Rachev, S., Kim, J.-R., and Mittnik, S. (1999). Stable Paretian econometrics part I and II. *The Mathematical Scientist*, 24:24–55 and 113–127.

Richter, M. and Soerensen, C. (2002). Stochastic Volatility and Seasonality in Commodity Futures and Options: The Case of Soybeans. Working paper, Department of Finance, Copenhagen Business School.

Rockinger, M. and Jondeau, E. (2002). Entropy Densities with an Application to Autoregressive Conditional Skewness and Kurtosis. *Journal of Econometrics*, 106:119–142.

Samanidou, E., Zschischang, E., Stauffer, D., and Lux, T. (2002). Microscopic Models of Financial Markets. In Schweitzer, F., editor, *Microscopic Models of Economic Dynamics*. Springer, Berlin.

Samorodnitsky, G. and Taqqu, M. (1994). *Stable Non-Gaussian Random Processes, Stochastic Models with Infinite Variance*. Chapman & Hall, London.

Schwert, W. G. (1989). Why Does Stock Market Volatility Change Over Time? *Journal of Finance*, 44:1115–1153.

Sentana, E. (1995). Quadratic ARCH Models. *Review of Economic Studies*, 62:639–661.

Tasche, D. (2002). Expected Shortfall and Beyond. *Journal of Banking and Finance*, 26(7):1519–1533.

Taylor, S. (1986). *Modelling Financial Time Series*. Wiley, New York.

Tay, A. S. and Wallis, K. F. (2000). Density Forecasting: A Survey. *Journal of Forecasting*, 19:235–254.

Tucker, A. L. and Pond, L. (1988). The Probability Distribution of Foreign Exchange Price Changes: Tests of Candidate Processes. *Review of Economics and Statistics*, 70:638–647.

Vigfusson, R. (1997). Switching between Chartists and Fundamentalists: A Markov Regime–Switching Approach. *International Journal of Finance and Economics*, 2:291–305.

Vlaar, P. J. G. and Palm, F. C. (1993). The Message in Weekly Exchange Rates in the European Monetary System: Mean Reversion, Conditional Heteroscedasticity, and Jumps. *Journal of Business and Economic Statistics*, 11(3):351–360.

Wong, C. S. and Li, W. K. (2000). On a Mixture Autoregressive Model. *Journal of the Royal Statistical Society*, 62:95–115.

Wong, C. S. and Li, W. K. (2001). On a Logistic Mixture Autoregressive Model. *Biometrika*, 88:833–846.

Curriculum Vitae

Gymnasium Heinrich Heine, Germany, Heikendorf	1988-1997
Military Service, Germany, Eutin	1997-1998
Christian Albrechts University, Germany, Kiel -Studies in Quantitative Economics and Business Administration	1998-2003
University of Zurich, Switzerland, Zurich -Doctoral Studies in Economics	2004-2007
J.P. Morgan Markets, Great Britain, London -Equity Derivatives Group	since 2007