Analyzing and Exploiting Asymmetries in the News Impact Curve

Markus Haas          Stefan Mittnik
Marc S. Paolella               Sven C. Steudee

First version: June 2006
Current version: June 2006

This research has been carried out within the NCCR FINRISK projects on “Behavioural and Evolutionary Finance” and “Financial Econometrics for Risk Management”
Analyzing and Exploiting Asymmetries in the News Impact Curve

Markus Haas\(^a\) Stefan Mittnik\(^{a,b,c}\) Marc S. Paolella\(^d\)† Sven C. Steude\(^e\)

\(^a\)Department of Statistics, University of Munich, Germany
\(^b\)Center for Financial Studies, Frankfurt, Germany
\(^c\)Ifo Institute for Economic Research, Munich, Germany
\(^d\)Swiss Banking Institute, University of Zurich, Switzerland
\(^e\)Institute for Empirical Research in Economics, University of Zurich, Switzerland

June 2006

Abstract

The recently proposed class of MixN–GARCH models, which couple a mixed normal distributional structure with linked GARCH–type dynamics, has been shown to offer a plausible decomposition of the contributions to volatility, as well as admirable out–of–sample forecasting performance, for financial asset returns. A general feature of these processes are constant mixing weights of the component densities, which leads to tractable stationarity conditions, but may not be a realistic assumption in general. This paper relaxes this assumption by considering different specifications of time–varying mixing weights for the MixN–GARCH model class. In particular, by relating current weights to past returns via sigmoid response functions, an empirically more realistic representation of Engle and Ng’s (1993) news impact curve with an asymmetric impact of unexpected return shocks on future volatility is obtained, and large gains in terms of in–sample fit and out–of–sample VaR forecasting performance can be realized.

\textit{JEL classification:} C22; C51; G10

\textit{Keywords—}GARCH, news impact curve, leverage effect, down-market effect, time–varying weights, Value–at–Risk.

*Mittnik conducted part of the research while visiting the Department of Economics at Washington University in St. Louis with a grant from the Fulbright Commission. Part of the research of Paolella and Steude has been carried out within the National Centre of Competence in Research “Financial Valuation and Risk Management” (NCCR FINRISK), which is a research program supported by the Swiss National Science Foundation. The authors are grateful to Gerd Hansen and Helmut Herwartz for discussions on a previous version of this paper, and also the numerous comments and suggestions from the participants at the March 2006 Ausschuss für Ökonometrie meeting of the German Verein für Socialpolitik.

†Corresponding author: paolella@isb.unizh.ch
1 Introduction

The use of a mixed normal distribution for modeling the unconditional distribution of asset returns is very effective, and has been considered by numerous authors, including Fama (1965), Kon (1984) and Tucker and Pond (1988). 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 behaviour.” 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, Mittnik and Paolella (2005) 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 studied by Alexander and Lazar (2004, 2006) and Haas, Mittnik and Paolella (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 modeled, 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 defining aspect 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. This new class allows the mixing weights to be functions of their previous values and of past returns.
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 form, are smooth monotonic functions bounded between zero and one (as applied, for example, in neural networks). Indeed, the concept of time–varying parameters is well established in econometrics, and is now widely used in a variety of empirical finance contexts. For example, Anderson, Nam, and Vahid (1999) introduce a class of asymmetric nonlinear smooth transition GARCH models, Dahlhaus and Subba Rao (2003) investigate the behavior of ARCH models with time–varying coefficients, and Medeiros and Veiga (2004) use flexible coefficient GARCH dynamics to model multiple regimes in volatility. Time–varying parameters have also already been found useful in a context related to ours: Vlaar and Palm (1993) proposed a time–varying jump intensity, while Beine and Laurent (2003) have implemented a normal mixture model with time–varying jump probabilities, used for the estimation of volatility shocks.

While in the constant weight MixN–GARCH model 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.

One prominent model in the asymmetric GARCH world is Nelson’s (1991) exponential GARCH, or EGARCH, which (besides the aforementioned asymmetry) allows big shocks to have a greater impact on volatility estimates than in standard GARCH models. 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 2.2, while an empirical comparison with our approach is provided in Section 6.

The remainder of this paper is as follows. Section 2 briefly introduces the original (constant parameter) MixN–GARCH model. Section 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 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 5, while Section 6 briefly summarizes our findings for other financial return series. Section 7 provides
concluding remarks and some ideas for future research.

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).

2.1 Mixed Normal GARCH

The mixed normal GARCH model, denoted MixN-GARCH, has recently been proposed by Haas, Mittnik, and Paolella (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

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

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

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

with the latter being relevant to the discussion in Section 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,

$$r_t | \mathcal{F}_{t-1} \sim \text{MixN}(\lambda_t, \mu_t, \sigma_t^{(2)}),$$

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

$$\sigma_t^{(2)} = \alpha_0 + \sum_{i=1}^{r} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{s} \beta_j \sigma_{t-j}^{(2)},$$

where $\alpha_i = (\alpha_{i1}, \alpha_{i2}, \ldots, \alpha_{ik})'$, $i = 0, \ldots, r$, are $k \times 1$ vectors; $\beta_j, j = 1, \ldots, s$, are diagonal $k \times k$ matrices\footnote{The full–matrix specification is considered in Haas, Mittnik and Paolella (2004a), though, as discussed there and confirmed with other data sets, the diagonal restriction is usually favored in empirical applications.} with the component–specific persistence parameters on the main diagonal; and
the innovation, $\epsilon_t$, is

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

(5)

Haas, Mittnik and Paolella (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 (3)–(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 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) 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

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

(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) arises when only a subset of the component variances gathered in the vector $\sigma_t^{(2)}$ 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 (1), (3) and (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 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.
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 (3) and (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

\[
\sigma^2_{jt} = \alpha_0 + \alpha_1(\epsilon_{t-1} - \theta_j)^2 + \beta_j \sigma^2_{j,t-1}, \quad j = 1, \ldots, k,
\]

where the \(\theta_j\)'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 \(\sigma^2_{jt}\) 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

\[
\sigma^2_{jt} = \alpha_0 + \alpha_1 \epsilon^2_{t-1} + \theta_j d_{t-1} \epsilon^2_{t-1} + \beta_j \sigma^2_{j,t-1}, \quad j = 1, \ldots, k,
\]

where \(d_{t-1} = 1\) if \(\epsilon_{t-1} < 0\) and \(d_{t-1} = 0\) otherwise. As in (7), a positive \(\theta_j\) implies that \(\sigma^2_{jt}\) reacts more intensely to negative shocks than to positive shocks.

3 Time–Varying Weights

The idea of modeling 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 modeling 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., Bekaert and Gray, 1998; Neely, 1999; and Haas, Mittnik and Mizrach, 2005). The conditional densities of such mixture models exhibit an enormous flexibility. For example, as illustrated by Haas, Mittnik and Mizrach (2005) 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.
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 $\epsilon_t$'s defined in (5).

### 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

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

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

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

and which we denote as TV($u, v, w$)MixN($k, g$), where $u, v$ and $w$ are the orders of the lagged $\epsilon_t$, lagged $\lambda_j$, and lagged $|\epsilon_t|^d$, respectively. For chosen values of $u, v, w$, the parameters $\gamma_{0j}$, $d$, and $\gamma_{mj}$, $\kappa_{mj}$, and $\delta_{mj}$, $m = 1, \ldots, k - 1$, are jointly estimated with the other model parameters in the MixN–GARCH process (4) and (6), via conditional maximum likelihood. 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, $\epsilon_{t-1}$, via sigmoid functions with two parameters: one for the location, $\gamma_{0j}$, and one for the scale, $\gamma_{1j}$, $j = 1, \ldots, k - 1$. Expressions (9) and (10) reduce to

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

\footnote{Wong and Li (2001) provide a general discussion of a mixture autoregressive model with time–varying mixing weights, where the mixing weights are logistic transformations of lagged endogenous and/or exogenous variables, including derivation of an EM algorithm for parameter estimation. However, with GARCH effects in the component variances, the EM algorithm possesses no advantages compared to standard numerical optimization methods, both in terms of stability as well as elegance.}
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 (11), and let component two have the larger variance. We have

\[
\lambda_t = \frac{e^{\gamma_0 + \gamma_1 \epsilon_{t-1}}}{1 + e^{\gamma_0 + \gamma_1 \epsilon_{t-1}}},
\]

and

\[
d\lambda_t/d\epsilon_{t-1} = \gamma_1 \lambda_t (1 - \lambda_t),
\]

so that, from (2), the overall variance at time \( t \) decreases with increasing \( \epsilon_{t-1} \) if \( \gamma_1 > 0 \). This pattern is indeed observed when the model is applied to stock market data.

To illustrate the sigmoid model, Figure 1 shows the relation between \( \epsilon_{t-1} \) and the first mixing weight, \( \lambda_{1t} \), for the daily returns on the NASDAQ index from 1971 to 2001, for a time–varying weight AR(3)–MixN–GARCH(1, 1) model with two component densities. In this example, having a second component with a higher 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 1](image)

**Figure 1:** Estimated relation (from joint ML estimation of the full model) between lagged innovation at time \( t - 1, \epsilon_{t-1} \), and the first mixing weight at time \( t, \lambda_{1t} \), for continuously compounded daily NASDAQ returns from 1971 to 2001 for the TV(1) model (11) with \( k = 2 \) mixture components (and with an AR(3) term for the mean and \( r = s = 1 \)); the sigmoid function has location parameter \( \hat{\gamma}_{01} = 1.501 \) and scale parameter \( \hat{\gamma}_{11} = 0.488 \).

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 \( \epsilon_{t-1} \) and \( \epsilon_{t-2} \). 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.
weights.

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

\[
\lambda_{jt} = \frac{\exp\{\gamma_{0j} + \gamma_{1j} (\epsilon_{t-1} + \epsilon_{t-2})\}}{1 + \sum_{j=1}^{k-1} \exp\{\gamma_{0j} + \gamma_{1j} (\epsilon_{t-1} + \epsilon_{t-2})\}}, \quad j = 1, \ldots, k-1,
\]

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 \( \epsilon_{t-2} \), albeit in a restricted fashion which constrains the impact of \( \epsilon_{t-1} \) and \( \epsilon_{t-2} \) to be identical. Anticipating our empirical results below, this specification is preferred for the NASDAQ data over competing models, based on the Bayesian information criteria (BIC).

### 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.\(^3\)

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 given by

\[
\lambda_{1t}(\epsilon_{t-1}) = \begin{cases} 
  b_1, & \text{if } \epsilon_{t-1} < \theta_1, \\
  b_2, & \text{if } \theta_1 \leq \epsilon_{t-1} < \theta_2, \\
  b_3, & \text{if } \theta_2 \leq \epsilon_{t-1} < \theta_3, \\
  \vdots \\
  b_{m-1}, & \text{if } \theta_{m-2} \leq \epsilon_{t-1} < \theta_{m-1}, \\
  b_m, & \text{if } \epsilon_{t-1} \geq \theta_{m-1},
\end{cases}
\]

where, as usual, \( \lambda_{2t} = 1 - \lambda_{1t} \) and \( 0 \leq b_i \leq 1, \quad i = 1, \ldots, m \) (but the constraint \( b_i < b_j, \quad i < j \), is not imposed). To simplify matters, the boundaries are constructed such that the difference between consecutive \( \theta_i \) are equal. It is worth emphasizing that the \( \theta_i \) need to be determined before the estimation is carried out, but the \( \theta_i \) depend on the domain of \( \lambda_{1t}(\epsilon_{t-1}) \)—which is unknown before the estimation because the \( \epsilon_t \) result from the filtering of the estimated model. To circumvent this problem and determine an approximate range of the \( \epsilon_t \), we first estimate the

---

\(^3\)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.
model with constant weights and then, based on the range of the filtered innovations, specify appropriate values for $\theta_1$ and $\theta_m$. 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 5.1 below discusses the estimation results for this model.

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 rational 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.

The leverage effect can be calculated as the correlation of lagged returns and a measure of future volatility; see, for example, Cont (2001), who defines the leverage effect of lag $\tau$ 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 $t$. Figure 2 shows the estimated leverage effect, as a function of $\tau$, for the daily NASDAQ index returns from the years 1971 to 2001. We see that the correlation starts negative and tends towards zero for increasing lags, which is in line with the results of Bouchaud *et al.* (2001), who find a decay time for the leverage effect for stock indices of about 50 days.

![Figure 2: Estimated leverage effect, $L(\tau) = \text{Corr}(r_{t-\tau}, r_t^2)$, of daily NASDAQ returns from February 1971 to June 2001.](image)
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 3 shows the leverage effect for the same data set of NASDAQ returns as in Figure 2, 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 2. However, for some of the sub–samples, there is no leverage effect at all. For example, in the upper right plot in Figure 3, which covers the years 1975 to 1979, the correlations hover around zero for all depicted lags.

**Figure 3:** Same as in Figure 2 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 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 that it is less pronounced and less volatile for increasing lags.

Figure 4: 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 $r_{t-\tau}$ is re-calculated every day using the most recent 1000 observations.

A sigmoid structure can be used to account for the asymmetric impact of unexpected shocks on future volatility. From (6), $\epsilon_t = r_t - \mu_t$, where $\epsilon_t$ can be interpreted as a measure of news (Engle and Ng, 1993). A negative $\epsilon_t$ indicates the arrival of bad news (negative shock) while a positive $\epsilon_t$ indicates good news (positive shock). As mentioned above, from Figure 1, in which the first mixing weight of a two-component TV(1) model is shown, it is clear that negative shocks increase the weight of the “wilder” component (component two) while positive shocks decrease its weight. Thus, a negative shock increases predictable volatility more than a positive shock with the same size.

Figure 5 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 we do not know the unconditional variance in the time–varying setting, we have taken the sample mean in Figure 5 instead. The NIC of TV(1) clearly reveals an asymmetric behavior, while for the original MixN–GARCH model, the
NIC is symmetric and centered at $\epsilon_{t-1} = 0$.\footnote{Another source of asymmetry in the NIC (which 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).}

**Figure 5:** 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 2. Both models share two components and an AR$(3)$–MixN–GARCH$(1, 1)$ process.

## 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 $t$.

### 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 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$(3)$–MixN–GARCH$(1, 1)$ structure with $\Psi_1$ in (4) diagonal. Parameter estimates are obtained by numerically maximizing the
Table 1 compares the likelihood value and the two likelihood–based standard information criteria AIC and BIC for the different time–varying structures discussed in Section 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 information criteria and likelihood value are shown.

As can be seen from Table 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 \(\epsilon_{t-1}\) and \(\epsilon_{t-2}\), while the AIC favors TV(2), which allows for different coefficients of the two lagged innovations.

We now turn to the results when using the non–parametric setting introduced in Section 3.2. After some experimenting, a reasonable choice of \(m\) in (13) is between 5 and 10 for the NASDAQ data set. The resulting fit using \(m = 6\) is shown in Figure 6 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 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 6, though in general, we have noticed that the choice of starting values can lead to local likelihood maxima. Figure 7 is similar, but uses \(m = 8\).

To help understand where the shape of the piecewise linear function is most important, Figures 6 and 7 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.

A potentially new stylized fact which might deserve future investigation emerges from Figures 6 and 7, 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 6 and 7 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 5 for the basic sigmoid structure) but sharply increases for very large positive innovations, which is in line with

---

5This 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.
the findings of Linton and Mammen (2003), 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).

The sigmoid structures relating the mixture components also affect the variance of the models. The component variances differ between the constant-weight MixN–GARCH model and its time–varying counterpart. The reason is that the sigmoid structures capture part of the volatility process, i.e., when markets go down (up), less weight is given to the low (high) volatility components. Figure 8 shows the evolution of the square root of the variances of the MixN(2, 2) and the TV(1)(2, 2) model. We see that the volatility spikes of the time-varying weight model are always smaller than the constant weight model. Also the mean and median volatility over the whole period illustrates that part of the volatility process is captured by the sigmoid structures. For the two components of the TV(1)(2, 2) the means of the volatilities are 0.80 and 2.53 compared to 0.85 and 2.68 for the MixN(2, 2). The results for the medians are qualitatively the same.

5.2 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 \( \xi \) is calculated as \( \hat{F}_{t|t-1}^{\mathcal{M}}(\text{VaR}_t(\xi)) = \xi \), where \( \hat{F}_{t|t-1}^{\mathcal{M}} \) 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, 1998; Kuester, Mittnik and Paolella, 2005; 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 2 for the 1%, 2.5%, 5% and 10% \( \xi \)-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, Mittnik and Paolella, 2005, and the references therein), which might be more costly to financial institutions than a slight overestimation of risk.

To further illustrate this point, the results for a spectrum of VaR levels up to \( \xi = 10\% \) can
Figure 8: Volatility evolution for the two components of the MixN(2, 2) (upper panel) and the TV(1)(2, 2) (lower panel) for the NASDAQ index.

<table>
<thead>
<tr>
<th>Model</th>
<th>VaR(1%)</th>
<th>VaR(2.5%)</th>
<th>VaR(5%)</th>
<th>VaR(10%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MixN(2, 2)</td>
<td>0.91</td>
<td>2.86</td>
<td>5.78</td>
<td>11.14</td>
</tr>
<tr>
<td>MixN(3, 2)</td>
<td>1.29</td>
<td>2.86</td>
<td>5.66</td>
<td>10.43</td>
</tr>
<tr>
<td>MixN(3, 3)</td>
<td>1.18</td>
<td>2.93</td>
<td>5.55</td>
<td>10.19</td>
</tr>
<tr>
<td>TV(1)(2, 2)</td>
<td>0.94</td>
<td>2.25</td>
<td>5.46</td>
<td>10.45</td>
</tr>
<tr>
<td>TV(1)(3, 2)</td>
<td>1.27</td>
<td>2.47</td>
<td>5.37</td>
<td>9.96</td>
</tr>
<tr>
<td>TV(1)(3, 3)</td>
<td>1.05</td>
<td>2.44</td>
<td>5.16</td>
<td>10.19</td>
</tr>
</tbody>
</table>

Table 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.

be seen from Figure 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.

Table 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 corre-
sponding 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.

<table>
<thead>
<tr>
<th>Model</th>
<th>MAD(5%)</th>
<th>MSD(5%)</th>
<th>MAD(10%)</th>
<th>MSD(10%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MixN(2, 2)</td>
<td>0.28</td>
<td>0.12</td>
<td>0.56</td>
<td>0.42</td>
</tr>
<tr>
<td>MixN(3, 2)</td>
<td>0.36</td>
<td>0.15</td>
<td>0.46</td>
<td>0.24</td>
</tr>
<tr>
<td>MixN(3, 3)</td>
<td>0.27</td>
<td>0.09</td>
<td>0.30</td>
<td>0.11</td>
</tr>
<tr>
<td>TV(1)(2, 2)</td>
<td>0.15</td>
<td>0.03</td>
<td>0.30</td>
<td>0.12</td>
</tr>
<tr>
<td>TV(1)(3, 2)</td>
<td>0.16</td>
<td>0.04</td>
<td>0.20</td>
<td>0.05</td>
</tr>
<tr>
<td>TV(1)(3, 3)</td>
<td>0.08</td>
<td>0.01</td>
<td>0.08</td>
<td>0.01</td>
</tr>
</tbody>
</table>

Table 3: The mean of the absolute deviation (MAD) and mean squared deviation (MSD) of the empirical from the theoretical tail probability (“Deviation” in Figure 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.

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

**Table 4:** Same as in Table 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.
Lazar (2005) the results concerning their forecasting performance can be seen in Table 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.

### 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 (9) and (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.

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 $\epsilon_{t-2}$ 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. Even 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 $\tau$. 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.

As the leverage effect presumably does not have to be modeled with all the mixture components, we also suggest considering a more general model class (with $k > 2$) in which not all components exhibit a sigmoid structure. It would be interesting to see if there is a general pat-
<table>
<thead>
<tr>
<th>Data</th>
<th>Model</th>
<th>MAD(1%)</th>
<th>MSD(1%)</th>
<th>MAD(2.5%)</th>
<th>MSD(2.5%)</th>
<th>MAD(5%)</th>
<th>MSD(5%)</th>
<th>MAD(10%)</th>
<th>MSD(10%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DAX</td>
<td>MixN(2,2)</td>
<td>0.14</td>
<td>0.02</td>
<td>0.14</td>
<td>0.02</td>
<td>0.19</td>
<td>0.04</td>
<td>0.20</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>TV(1)(2,2)</td>
<td>0.29</td>
<td>0.10</td>
<td>0.38</td>
<td>0.16</td>
<td>0.36</td>
<td>0.14</td>
<td>0.27</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td>AMixN(1)(2)</td>
<td>0.18</td>
<td>0.04</td>
<td>0.19</td>
<td>0.04</td>
<td>0.13</td>
<td>0.03</td>
<td>0.20</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>AMixN(2)(2)</td>
<td>0.24</td>
<td>0.07</td>
<td>0.29</td>
<td>0.09</td>
<td>0.26</td>
<td>0.08</td>
<td>0.19</td>
<td>0.05</td>
</tr>
<tr>
<td>Dow</td>
<td>MixN(2,2)</td>
<td>0.17</td>
<td>0.03</td>
<td>0.08</td>
<td>0.01</td>
<td>0.10</td>
<td>0.02</td>
<td>0.13</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>TV(1)(2,2)</td>
<td>0.06</td>
<td>0.00</td>
<td>0.12</td>
<td>0.02</td>
<td>0.15</td>
<td>0.03</td>
<td>0.19</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>AMixN(1)(2)</td>
<td>0.28</td>
<td>0.10</td>
<td>0.43</td>
<td>0.21</td>
<td>0.44</td>
<td>0.21</td>
<td>0.28</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td>AMixN(2)(2)</td>
<td>0.29</td>
<td>0.10</td>
<td>0.48</td>
<td>0.26</td>
<td>0.44</td>
<td>0.22</td>
<td>0.30</td>
<td>0.13</td>
</tr>
<tr>
<td>Nikkei</td>
<td>MixN(2,2)</td>
<td>0.06</td>
<td>0.00</td>
<td>0.11</td>
<td>0.02</td>
<td>0.23</td>
<td>0.08</td>
<td>0.41</td>
<td>0.21</td>
</tr>
<tr>
<td></td>
<td>TV(1)(2,2)</td>
<td>0.21</td>
<td>0.05</td>
<td>0.21</td>
<td>0.05</td>
<td>0.36</td>
<td>0.16</td>
<td>0.44</td>
<td>0.23</td>
</tr>
<tr>
<td></td>
<td>AMixN(1)(2)</td>
<td>0.28</td>
<td>0.10</td>
<td>0.56</td>
<td>0.38</td>
<td>0.73</td>
<td>0.60</td>
<td>0.63</td>
<td>0.47</td>
</tr>
<tr>
<td></td>
<td>AMixN(2)(2)</td>
<td>0.27</td>
<td>0.09</td>
<td>0.57</td>
<td>0.40</td>
<td>0.74</td>
<td>0.62</td>
<td>0.64</td>
<td>0.48</td>
</tr>
<tr>
<td>S&amp;P</td>
<td>MixN(2,2)</td>
<td>0.28</td>
<td>0.09</td>
<td>0.27</td>
<td>0.08</td>
<td>0.18</td>
<td>0.05</td>
<td>0.12</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>TV(1)(2,2)</td>
<td>0.10</td>
<td>0.01</td>
<td>0.23</td>
<td>0.07</td>
<td>0.32</td>
<td>0.12</td>
<td>0.25</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>AMixN(1)(2)</td>
<td>0.30</td>
<td>0.11</td>
<td>0.42</td>
<td>0.20</td>
<td>0.41</td>
<td>0.18</td>
<td>0.29</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>AMixN(2)(2)</td>
<td>0.29</td>
<td>0.10</td>
<td>0.43</td>
<td>0.21</td>
<td>0.50</td>
<td>0.27</td>
<td>0.40</td>
<td>0.20</td>
</tr>
</tbody>
</table>

Table 5: Same as in Table 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.
tern concerning leverage and components such that, for example, the high volatility component is always more capable of modeling the leverage effect than lower volatility components.

Finally, in addition to, or instead of, relating current mixing weights to the past innovations and past mixing weights as suggested in (10), it might be advantageous to consider use of the conditional variance, skewness or kurtosis.

References


Figure 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 \(F_\xi\) 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.