Empirical Modeling of Extreme Events from Return–Volume Time Series in Stock Market

Peter Bühlmann *
Department of Statistics
University of California Berkeley
Berkeley, CA 94720-3860

June 1996

Abstract

We propose the discretization of real-valued financial time series into few ordinal values and use non-linear likelihood modeling for sparse Markov chains within the framework of generalized linear models for categorical time series.

We analyze daily return and volume data and estimate the probability structure of the process of extreme lower, extreme upper and the complementary usual events. Knowing the whole probability law of such ordinal-valued vector processes of extreme events of return and volume allows us to quantify non-linear associations. In particular, we find a (new kind of) asymmetry in the return–volume relationship which is a partial answer to a research issue given by Karpoff (1987).

We also propose a simple prediction algorithm which is based on an empirically selected model.

Keywords: Asymmetric volume–return relation; Finance; Generalized linear model; Markov chain; Non-linear association; Prediction

*Research supported in part by the Swiss National Science Foundation.

1 Introduction

Many of the time series occurring in finance suffer from the fact that there are no simple mechanistic models which substantially reduce the complexity of the data and still explain it well. Either the models are too simple to explain the nature of financial phenomena or the model is too complex so that over-fitting occurs. Over-fitting results in high variability of estimates and reproducibility in a similar situation cannot be satisfactorily achieved.

If we believe in an intrinsic high complexity of the financial nature, we seem to be forced to work with models having many parameters and hence with estimates being highly variable. We propose one way out of this impasse by the following question: Why modeling the whole structure of an observed financial time series, rather than modeling only some structure, such as extreme events? We propose here a discretization of a real-valued financial time series into only three ordinal categories, corresponding to extreme lower, extreme upper and the complementary (usual) events. This results in a huge reduction of the number of parameters for a model and still allows to explore questions about extreme events, such as unusual values or large positive or negative increments.

We exemplify the method by analyzing daily data of return from the Dow Jones index and volume from the New York Stock Exchange (NYSE). For analyzing the three events (lower extreme, upper extreme and usual) we use a likelihood modeling approach based on a higher order Markovian assumption, and model cumulative probabilities within the framework of generalized linear models with lagged variables treated as factors. The word 'linear' can be misleading, our model class is very general and broad, since we are treating lagged variables as factors. This then includes processes similar to arbitrary finite order Markov chains. By putting a natural hierarchy on such models we typically model sparse rather than full Markov chains, which allows to avoid the curse of dimensionality, cf. section 3.1. Such models are very flexible, one can also include continuous covariates and explanatory exogenous factors for describing the dynamics. For a good overview in the independent set-up, cf. McCullagh and Nelder (1989). Within this model class, the selection of a model can be supported by considering measures of predictive power such as Akaike's information criterion (AIC), cf. Akaike (1973). Having selected a model, we also propose a simple prediction algorithm.

By discretizing we pay a price by restricting the focus to extreme events (and the complementary usual events). On the other hand we gain a lot in the process of empirical model searching and fitting. Moreover, our conclusions are probabilistic statements rather than the often used correlation which measures only linear association.

We obtain by our method fully probabilistic interpretations for the relationship of extreme events of return from the Dow Jones and volume from the NYSE. This gives new insight into the structure of financial markets. In particular, our empirical results are an answer to a research question posed by Karpoff (1987): for our data, the volume–return relationship, at least for extreme events, is in a new sense asymmetric.

This paper is organized as follows. In section 2 we describe the data set, in section 3 we explain the modeling and prediction techniques, in section 4 we report our empirical findings for the analyzed data set, in section 5 we draw some conclusions and in section 6 we briefly outline some more mathematical and computational details.

2 The data

The data is about daily return of the Dow Jones index and daily volume of the New York Stock Exchange (NYSE). The daily measurements are from the period July 1962 through June 1988, corresponding to 6430 days [2]. What we term 'volume' is the standardized aggregate turnover on the NYSE,

$$\mathrm{Vol}_t = log(F_t) - log(\sum_{s=t-100}^{t-1} F_s/100),$$

where F_s equals the fraction of shares traded on day s.
The return is defined as

$$\mathrm{Ret}_t = log(D_t) - log(D_{t-1}),$$

where D_t is the index of the Dow Jones on day t. This is the same data with the same standardization as in Weigend and LeBaron (1994). Figure 1 shows these standardized financial time series; the time t=6253, where return $\mathrm{Ret}_t < -0.2$ corresponds to the 1987 crash. Besides this special structure around the crash, both series look quite stationary.

3 Modeling

Since mechanistic theories for volume or return are not broadly accepted, we want to use a modeling approach which is in a nonparametric spirit. Our models in sections 3.3 and 3.4 can easily be extended to include some usual parametric parts, yielding then a kind of semiparametric model.

3.1 The curse of dimensionality for full Markov chains

The most natural model for a stationary process $\{X_t\}_{t\in\mathbb{Z}}$ ($X_t \in \mathbb{R}$), assuming no particular underlying mechanistic system, is maybe a full Markov chain of order p, i.e.,

$$\mathbb{P}[X_t \le x | X_{t-1}, X_{t-2}, \ldots] = \mathbb{P}[X_t \le x | X_{t-1}, \ldots, X_{t-p}],\ x \in \mathbb{R},\ 1 \le p < \infty.$$

This implies the only implicit assumption, namely that the process $\{X_t\}_{t\in\mathbb{Z}}$ has a finite memory of length p.

Such models are very general and usually too complex. Even if X_t would only take values in a finite space $\mathcal{S}$ of cardinality $|\mathcal{S}|$, these models involve $|\mathcal{S}|^{p+1} - 1$ free parameters to estimate. For example, $|\mathcal{S}| = 5, p = 5$ yields 15624 free parameters, which is prohibitive! Such an explosion of parameters corresponds to a space which is very hard to cover with data points. This phenomenon is called the curse of dimensionality.

We learn that full Markov chains with values in a space $\mathcal{S}$ with about $|\mathcal{S}| \ge 4$ do often not yield a big reduction of the complexity in the data, unless the sample size is huge.

[2] This data is publicly available via the internet:
http://ssdc.ucsd.edu/ssdc/NYSE.Date.Day.Return.Volume.Vola.text

3.2 Discretizing into ordinal values

Given a single realization $\mathrm{Ret}_1, \ldots, \mathrm{Ret}_n$ from the stationary (which we assume), financial time series of returns, we discretize the data into a small number of ordinal values yielding $R_1, \ldots, R_n$. We propose here

$$R_t = \begin{cases} 1, & \text{if } \mathrm{Ret}_t \le c_1 \\ 2, & \text{if } c_1 < \mathrm{Ret}_t \le c_2 \\ 3, & \text{if } \mathrm{Ret}_t > c_2 \end{cases}, \tag{3.1}$$

where $c_1,\ c_2$ are the 2.5% and 97.5% sample quantiles of $\mathrm{Ret}_1, \ldots, \mathrm{Ret}_n$.
Thus, $R_t \neq 2$ describes an extreme event with expected occurrence being about 5%. Analogously as in (3.1) we can discretize volume yielding $V_1, \ldots, V_n$. For the data described in section 2, the discretized series of daily return and daily volume are shown in Figure 2. The more extreme activities can be exploited a little bit by eye: for the series of returns around time points 3000, 5000 and 6200, extreme events are clustered together. This clustering is also visible from the original series of returns in Figure 1. Some clustering for the volume series is also but less clearly visible.

Ordinal data is generally referred to quantities whose values are categories falling on an ordinal scale, cf. McCullagh and Nelder (1989, Ch. 5). Other discretizations are possible, for example into binary variables. However, the number of different ordinal values should be kept small in order to avoid the curse of dimensionality.

The discretization in (3.1) reduces already the complexity of the data. Modeling such ordinal data is now an easier task than modeling the original financial time series. Of course, we can now only say something about 'usual events', described by $R_t = 2$, and extreme events, described by $R_t = 1$ or $R_t = 3$. By paying a price to consider only 'usual' and two kinds of 'extreme events', we have a much better chance to find good models for such ordinal data. We address in section 3.6 the issue about loosing information by discretizing and introduce there a state space type model for the real-valued time series of return and volume. In the sequel we consider only ordinal-valued time series.

3.3 GLM for individual ordinal-valued time series

Generalized linear models (GLM) can be used to model ordinal time series data. For data as in (3.1), the idea is to model the complementary cumulative probabilities $\mathbb{P}[R_t \ge j]$ $(j = 2, 3)$ as a function of the history $R_{t-1}, \ldots, R_{t-p}$ for some $1 \le p < \infty$.

In the independent set-up, a link function $g : (0,1) \to \mathbb{R}$ is specified and then transformed probabilities are modeled as

$$g(\mathbb{P}[R_t \ge j | x]) = \theta_j + x'\alpha,\ j = 2, 3;\ t = 1, \ldots, n, \tag{3.2}$$

where x is a $p \times 1$ explanatory vector, α a $p \times 1$ parameter vector and $\theta_2 \ge \theta_3$ parameters in $\mathbb{R}$.
Typical link functions g are the logit-, probit- and complementary loglog-function. For an overview, see McCullagh and Nelder (1989, Ch. 5).

The model (3.2) for independent data can be adapted to the time series case. We follow the idea of likelihood modeling, based on the history of the process. Assuming the Markov property of order p for $\{R_t\}_{t\in\mathbb{Z}}$, the conditional log-likelihood function, given $R_1, \ldots, R_p$, can be written as

$$\ell(\beta; R) = \sum_{t=p+1}^{n} log(p_\beta(R_t|R_{t-1}, \ldots, R_{t-p})), \tag{3.3}$$

where $p_\beta(R_t|R_{t-1}, \ldots, R_{t-p})$ is the conditional distribution of R_t given by some parametric model with parameters β, see formula (3.4) and (3.5). As explanatory variables x we thus choose the lagged observations $R_{t-1}, \ldots, R_{t-p}$. We treat here these lagged variables as factors, each of them having in our case 3 levels.

Treating the lagged variables as factors is a crucial distinction which brings us away from pure linear modeling. In the sequel, we denote by $\mathcal{R}_{t-1}, \ldots, \mathcal{R}_{t-p}$ the factors corresponding to the lagged variables $R_{t-1}, \ldots, R_{t-p}$. We are also using the notation as in cross-classification, cf. McCullagh and Nelder (1989, Ch. 3.4). For example, we write $\mathcal{R}_{t-i}$ for the main effect, $\mathcal{R}_{t-i}.\mathcal{R}_{t-j}$ for the second order interaction and $\mathcal{R}_{t-i} * \mathcal{R}_{t-j} = \mathcal{R}_{t-i} + \mathcal{R}_{t-j} + \mathcal{R}_{t-i}.\mathcal{R}_{t-j}$ for the full (saturated) second order model; an analogous notation is used for three or more terms.

We work here only with the logit link function $g(\mu) = logit(\mu) = exp(\mu)/(1 + exp(\mu))$ which has a probabilistic interpretation with the logistic distribution for an underlying latent variable, namely for the original value Ret_t of the financial time series. Also, we usually look $p = 5$ lags back into the past. Inspired from (3.2), our model for the conditional distributions $p_\beta(R_t|R_{t-1}, \ldots, R_{t-p})$ then becomes

$$logit(\mathbb{P}[R_t \geq j]) = \theta_j + \mathcal{R}_{t-1} * \mathcal{R}_{t-2} * \mathcal{R}_{t-3} * \mathcal{R}_{t-4} * \mathcal{R}_{t-5},\ j = 2, 3;\ t = 1, \ldots, n, \tag{3.4}$$

or some sub-model thereof, such as

$$\begin{aligned} logit(\mathbb{P}[R_t \geq j]) &= \theta_j + \sum_{i=1}^{5} \mathcal{R}_{t-i} + \sum_{i=2}^{5} \mathcal{R}_{t-1}.\mathcal{R}_{t-i} + \sum_{i=3}^{5} \mathcal{R}_{t-2}.\mathcal{R}_{t-i} \\ &+ \sum_{i=4}^{5} \mathcal{R}_{t-3}.\mathcal{R}_{t-i} + \mathcal{R}_{t-4}.\mathcal{R}_{t-5},\ j = 2, 3;\ t = 1, \ldots, n. \end{aligned} \tag{3.5}$$

In both models (3.4) and (3.5), we have $\theta_2 \geq \theta_3 \in \mathbb{R}$.

We then estimate intercepts θ_2, θ_3 and the main and interaction effects (all abbreviated by a parameter vector β) by maximum likelihood. Maximization of the log-likelihood function in (3.3) with $p = 5$ can be achieved by an iteratively reweighted least squares algorithm, cf. McCullagh and Nelder (1989, Ch. 2.5).

The full (saturated) model (3.4) is similar to a full Markov chain of order 5, as described in section 3.1. Model (3.4) has 245 free parameters. Model (3.5) exhibits some hierarchical structure: the main effects and all second-order interactions. Model (3.5) has 52 free parameters. By looking at the number of free parameters, it should become clear that models (3.4) and (3.5) are quite different from classical linear AR-type models which would include only as many free autoregressive parameters as lagged variables.

With models (3.4) or (3.5) (or other sub-models of (3.4)), we can quantify probabilities for extreme small or extreme large events of our series of returns. Also, having a model we can do prediction for usual and two kinds of extreme events.

3.4 Joint GLM for two ordinal-valued time series

There is a substantial interest to quantify the dependence or association between return and volume, cf. Karpoff (1987). Most of the previous work was done by looking at cross-correlations, a measure for linear dependence. We will develop here joint likelihood modeling, resulting in estimates for the joint probability of discretized return and discretized volume. In other words, we will track down the ultimate aim in this set-up, namely the knowledge of the whole probability structure for usual and extreme events of return and volume.

As in (3.1) we discretize return and volume into 3 ordinal values, yielding $R_1, \ldots, R_n$ for discretized return and $V_1, \ldots, V_n$ for discretized volume. We study first the relationship of return given volume, according to a Wall Street saying that 'it takes volume to make prices move'. This means that R_t should depend on V_t for the same time point t.

For likelihood modeling, we again assume the Markov property of order p for the ordinal-vector process $\{(R_t, V_t)'\}_{t \in \mathbb{Z}}$. The log-likelihood function, conditional on $R_1, \ldots, R_p$, $V_1, \ldots, V_p$, then becomes

$$\begin{aligned}
\ell(\beta; R, V) &= \sum_{t=p+1}^{n} log(p_\beta(R_t, V_t | R_{t-1}, V_{t-1}, \ldots, R_{t-p}, V_{t-p})) \\
&= \sum_{t=p+1}^{n} log(p_{\beta^R}(R_t | V_t, R_{t-1}, V_{t-1}, \ldots, R_{t-p}, V_{t-p})) \\
&+ \sum_{t=p+1}^{n} log(p_{\beta^V}(V_t | R_{t-1}, V_{t-1}, \ldots, R_{t-p}, V_{t-p})) \\
&= \ell(\beta^R; R, V) + \ell(\beta^V; R, V),
\end{aligned} \tag{3.6}$$

where the $p_\beta(.|.)$'s are the conditional probabilities given by some parametric model with parameters β as in formula (3.7).

We thus introduce an explanatory factor $\mathcal{V}_t$ with 3 levels corresponding to the variable V_t. Additionally, the factors for the lagged variables are denoted by $\mathcal{R}_{t-i}$ and $\mathcal{V}_{t-i}$ (i=1,...,5). Similarly as in models (3.4) and (3.5) we model

$$\begin{aligned}
logit(\mathbb{P}[R_t \geq j]) &= \theta_j^R + \mathcal{V}_t + \sum_{i=1}^{5} \mathcal{R}_{t-i} + \sum_{i=2}^{5} \mathcal{R}_{t-1}.\mathcal{R}_{t-i} + \sum_{i=3}^{5} \mathcal{R}_{t-2}.\mathcal{R}_{t-i} \\
&+ \sum_{i=4}^{5} \mathcal{R}_{t-3}.\mathcal{R}_{t-i} + \mathcal{R}_{t-4}.\mathcal{R}_{t-5}, \\
logit(\mathbb{P}[V_t \geq j]) &= \theta_j^V + \sum_{i=1}^{5} \mathcal{V}_{t-i} + \sum_{i=1}^{5} \mathcal{R}_{t-i}, \; j = 2, 3; \; t = 1, \ldots, n,
\end{aligned} \tag{3.7}$$

where $\theta_2^R \geq \theta_3^R$, $\theta_2^V \geq \theta_3^V$.

Of course, other models than (3.7) are possible. In particular, the model for $\mathbb{P}[V_t \geq j]$ could be thought to be without the factors $\mathcal{R}_{t-i}$ $(i = 1, \ldots, 5)$. Model (3.7) has for the R_t-part the same structure as model (3.5), except that we include now in addition the no lagged cross-dependence factor $\mathcal{V}_t$. This model (3.7) exhibits again a hierarchical structure, involving 76 free parameters.

The factorization of the likelihood function in (3.6) which yields the separation of the parameters β^R from β^V, has the computational advantage, that the maximum likelihood estimates can be found by maximizing each of the two parts in the log-likelihood function separately. Note, that putting a logit-link function as in (3.7) on each of the terms $p_{\beta^R}(R_t|V_t, R_{t-1}, V_{t-1}, \ldots, R_{t-p}, V_{t-p})$ and $p_{\beta^V}(V_t|R_{t-1}, V_{t-1}, \ldots, R_{t-p}, V_{t-p})$ in the log-likelihood function in (3.6) is different than using a logit-link function for the joint distribution $p_\beta(R_t, V_t|R_{t-1}, V_{t-1}, \ldots, R_{t-p}, V_{t-p})$.

With model (3.7) we can quantify joint, and hence conditional and marginal probabilities for usual and extreme events of return and volume. Moreover, prediction can be done based on the past of both series.

3.5 Prediction of return

Though prediction of extreme events in a financial time series seems to be an extremely difficult task, we analyze here the prediction power of our model (3.7) for forecasting return. The one-step ahead prediction can be realized by the following algorithm.

1. Given $R_1, \ldots, R_n, V_1, \ldots, V_n$, compute the estimates

$$\hat{\mathbb{P}}[R_{t+1} = j|V_{t+1}, R_t, V_t, \ldots, R_{t-4}, V_{t-4}],$$
$$\hat{\mathbb{P}}[V_{t+1} = j|R_t, V_t, \ldots, R_{t-4}, V_{t-4}],\ j = 1, 2, 3,$$

according to the estimated parameters in model (3.7).

2. Predict

$$\hat{V}_{n+1} = \text{argmax}_j \hat{\mathbb{P}}[V_{n+1} = j|R_n, V_n, \ldots, R_{n-4}, V_{n-4}],$$

which is the MAP (maximum probability) estimator.

3. Predict

$$\hat{R}_{n+1} = \text{argmax}_j \hat{\mathbb{P}}[R_{n+1} = j|\hat{V}_{n+1}, R_n, V_n, \ldots, R_{n-4}, V_{n-4}],$$

which is the MAP estimator, with the predictor $\hat{V}_{n+1}$ from step 2 plugged in for the unknown value V_{n+1}.

For predicting return, both quantities $\hat{\mathbb{P}}[R_{n+1} = j|\hat{V}_{n+1}, R_n, V_n, \ldots, R_{n-4}, V_{n-4}]$ ($j = 1, 2, 3$) and $\hat{R}_{n+1}$ are of interest, the first being a kind of prediction density and the second being the MAP predictor itself.

3.6 Loosing information by discretizing?

We briefly address the issue of modeling and predicting without using directly the full information of the real-valued financial time series at the beginning. When taking the view of being only interested in extreme events (and their remaining complementary part), the question to be asked is whether the discretized time series, as given by (3.1) is still a sufficient statistic for the problem to analyze.

A simplified situation is given by the following state space type model,

$$\begin{aligned}&\mathbb{P}[R_t = j, V_t = k | R_{t-1}, V_{t-1}, \ldots] = f_{j,k}(R_{t-1}, V_{t-1}, \ldots, R_{t-p}, V_{t-p})\\ &(\mathrm{Ret}_t, \mathrm{Vol}_t)' = g(R_t, V_t, R_{t-1}, V_{t-1}, \ldots, R_{t-q}, V_{t-q}, Z_t)\\ &1 \le p, q < \infty;\ j, k = 1, 2, 3;\ t \in \mathbb{Z}, \end{aligned} \tag{3.8}$$

where $\{Z_t\}_{t\in\mathbb{Z}}$ is a stationary sequence, independent of $\{(R_t, V_t)'\}_{t\in\mathbb{Z}}$ and $g = (g_1, g_2)'$ is an $\mathbb{R}^2$-valued function, compatible with the discretization operation in (3.1), e.g., $g_1(1, v_t, r_{t-1}, v_{t-1}, \ldots, r_{t-q}, v_{t-q}, z_t) \le c_1$ for all $v_t, r_{t-1}, v_{t-1}, \ldots, r_{t-q}, v_{t-q}, z_t$, where c_1 is defined in (3.1).

By allowing quite general processes $\{Z_t\}_{t\in\mathbb{Z}}$ and functions g, this model includes very complicated dynamics for the $\mathbb{R}^2$-valued series $\{(\mathrm{Ret}_t, \mathrm{Vol}_t)'\}_{t\in\mathbb{Z}}$. The dynamics for $\{(R_t, V_t)'\}_{t\in\mathbb{Z}}$ can be motivated by the description of the market, that extreme events are only triggered by previous extreme events and not influenced by other characteristics of $\{(\mathrm{Ret}_t, \mathrm{Vol}_t)'\}_{t\in\mathbb{Z}}$. This corresponds to a self-generating trade: extreme events act as local generators, influencing the next few outcomes of return and volume. For such a model, the conditional probability law

$$\mathcal{L}(\mathrm{Ret}_1, \ldots, \mathrm{Ret}_n, \mathrm{Vol}_1, \ldots, \mathrm{Vol}_n | R_1, \ldots, R_n, V_1, \ldots, V_n)$$

does not depend on the functions $f_{j,k}$ describing the dynamics of $\{(R_t, V_t)'\}_{t\in\mathbb{Z}}$.

This is the analogon to sufficiency, saying that the ordinal-valued observed time series $\{(R_t, V_t)'\}_{t=1}^n$ contains the full information about the functions $f_{j,k}$. Thus, for learning about extreme events, namely learning about the functions $f_{j,k}$, we would not gain by using the real-valued data $\{(\mathrm{Ret}_t,\ \mathrm{Vol}_t)'\}_{t=1}^n$.

Granger (1992) has proposed switching regime models for $\{\mathrm{Ret}_t\}_{t\in\mathbb{Z}}$. Our model (3.8) involves some kind of non-lagged switching variables $(R_t, V_t, R_{t-1}, V_{t-1}, \ldots, R_{t-q}, V_{t-q}, Z_t)'$, some of them being as extreme events of interest themselves. Our model (3.8) allows a quite simple empirical model search for R_t and V_t, but generally not for Ret_t and Vol_t. This is a main advantage over switching regime models, where the empirical selection of regimes, explanatory variables and specification of their functional form are very hard to do.

Even though the model (3.8) might be too simple, we believe that under certain circumstances an analysis can be based on discretized observations without loosing much relevant information. We have also tried a version of model (3.7) including real-valued observations,

$$\begin{aligned} logit(\mathbb{P}[R_t \ge j]) &= \theta_j^R + \mathcal{V}_t + \sum_{i=1}^{5} \mathcal{R}_{t-i} + \sum_{i=2}^{5} \mathcal{R}_{t-1}.\mathcal{R}_{t-i} + \sum_{i=3}^{5} \mathcal{R}_{t-2}.\mathcal{R}_{t-i}\\ &+ \sum_{i=4}^{5} \mathcal{R}_{t-3}.\mathcal{R}_{t-i} + \mathcal{R}_{t-4}.\mathcal{R}_{t-5} + \sum_{i=1}^{5} \gamma_i R_{t-i},\\ logit(\mathbb{P}[V_t \ge j]) &= \theta_j^V + \sum_{i=1}^{5} \mathcal{V}_{t-i} + \sum_{i=1}^{5} \mathcal{R}_{t-i} + \sum_{i=1}^{5} \delta_i R_{t-i} + \sum_{i=1}^{5} \lambda_i V_{t-i},\ j = 2, 3, \end{aligned}$$

resulting in 15 additional parameters.

For our data set in section 2, this model had less predictive power than model (3.7). In addition we tried to include not only linear but also higher order polynomials as explanatory real-valued variables. Again, such models were slightly worse than (3.7) in terms of prediction.

More work is needed concerning the issue of discretization. The operation in (3.1) works well for our particular data set. We enjoy the advantage of having a reduction in the complexity of the data, which makes the empirical modeling part a much easier task, still getting satisfactory answers.

4 Empirical results

We report here our results for the data described in section 2 by using the models from sections 3.3 and 3.4, respectively.

4.1 Results for individual series of return

For individual series, we only report our findings for returns, which is often the more interesting quantity. The conceptual modeling for the volume series is the same.

Model (3.4) is with 245 free parameters already quite complex and we do not try it. Instead, we want to see how well some simpler models like (3.5) explain the return data. As a measure for selecting between competing models we use Akaike's information criterion (AIC) (see section 6),

$$AIC = -2\ell(\hat{\beta}; R) + 2(\text{number of free parameters}),$$

where $\ell(\hat{\beta}; R)$ is the log-likelihood function evaluated at its maximizer $\hat{\beta}$.
Based on the log-likelihood function in (3.3) we tried the models,

(Mret1) $logit(\mathbb{P}[R_t \geq j]) = \theta_j + \sum_{i=1}^{5} \mathcal{R}_{t-i}$

(Mret2) $logit(\mathbb{P}[R_t \geq j]) = \theta_j + \sum_{i=1}^{5} \mathcal{R}_{t-i} + \sum_{i=2}^{5} \mathcal{R}_{t-1}.\mathcal{R}_{t-i}$

(Mret3) $logit(\mathbb{P}[R_t \geq j]) = \theta_j + \sum_{i=1}^{5} \mathcal{R}_{t-i} + \sum_{i=2}^{5} \mathcal{R}_{t-1}.\mathcal{R}_{t-i} + \sum_{i=3}^{5} \mathcal{R}_{t-2}.\mathcal{R}_{t-i}$

(Mret4) as given in (3.5)

(Mret5) $logit(\mathbb{P}[R_t \geq j]) = \theta_j + \sum_{i=1}^{5} \mathcal{R}_{t-i} + \sum_{i=2}^{5} \mathcal{R}_{t-1}.\mathcal{R}_{t-i} + \sum_{i=3}^{5} \mathcal{R}_{t-2}.\mathcal{R}_{t-i} + \sum_{i=4}^{5} \mathcal{R}_{t-3}.\mathcal{R}_{t-i} + \mathcal{R}_{t-4}.\mathcal{R}_{t-5} + \mathcal{R}_{t-1}.\mathcal{R}_{t-2}.\mathcal{R}_{t-3}$

Note that we have here a nested sequence of models (Mret1) $\subseteq$ (Mret2) $\subseteq \ldots \subseteq$ (Mret5). All of these models are hierarchical, in (Mret2), (Mret3) and (Mret5) in the sense that factor $\mathcal{R}_{t-1}$ is assumed to be more important than $\mathcal{R}_{t-2}$ and so on. The values of the AIC statistic are given by the following table.

	(Mret1)	(Mret2)	(Mret3)	(Mret4)	(Mret5)
AIC	2978.5	2978.4	2947.1	2908.9	–

In model (Mret5), the third order interaction was not estimable and hence we do not consider this model anymore. By looking at the AIC statistic, we thus select model (Mret4) which is model (3.5).

In the following we give now the more detailed analysis of model (Mret4). The reduction in the deviance (see section 6) from an independence model to our current model is 196.8, achieved by 50 additional parameters, i.e.,

$$2\ell(\hat{\beta}_{Mret4}; R) - 2\ell(\hat{\beta}_0; R) = 196.8,$$

where $\hat{\beta}_0$ corresponds to the model $logit(\mathbb{P}[R_t \geq j]) = \theta_j,\ j = 2, 3$.
Therefore, our model (Mret4) seems highly significant. The coefficient estimates for $\theta_2,\ \theta_3$ and for the different effects are given in Table 1 in the Appendix. Corresponding t-values for such individual parameters depend on the unknown underlying stochastic process and are not easily available in a way, being robust against model misspecification, see section 6. A possible way to do nonparametric, model-free inference is given by resampling techniques, see section 6.

In Figure 3 we show the fitted probabilities

$$\hat{\mathbb{P}}[R_t = j | R_{t-1}, \ldots, R_{t-5}]\ (j = 1, 2, 3).$$

The activity for extreme events around time points 3000, 5000 and 6200 is clearly visible. Moreover, we have a quantification for usual and the two kinds of extreme returns, namely the probabilities for these events. In Figure 4 we show two graphical tools to check the goodness of fit of our model (3.5) for the return series. Based on the fitted probabilities in Figure 3 we compute

$$\hat{\mathbf{E}}[R_t | R_{t-1}, \ldots, R_{t-5}] = \sum_{j=1}^{3} j \hat{\mathbb{P}}[R_t = j | R_{t-1}, \ldots, R_{t-5}],$$

and plot it, together with the observed data, against the time axis t. The technique of uniform residuals is explained in section 6. From Figure 4 we do not see any severe lacks of our model and we have some confidence in it.

4.2 Results for joint modeling of return and volume

Besides the model (3.7) we consider some other competing models, all being logit-models for the conditional probabilities of $p_{\beta^R}(R_t | V_t, R_{t-1}, V_{t-1}, \ldots, R_{t-p}, V_{t-p})$ and $p_{\beta^V}(V_t | R_{t-1}, V_{t-1}, \ldots, R_{t-p}, V_{t-p})$ in the decoupled log-likelihood function in (3.6). It is then sufficient to find good models for $logit(\mathbb{P}[R_t \geq j])$ and $logit(\mathbb{P}[V_t \geq j]$, $j = 2, 3$. Our candidates are as follows.

(JMret1) $logit(\mathbb{P}[R_t \geq j]) = \theta_j^R + \mathcal{V}_t + \sum_{i=1}^{5} \mathcal{R}_{t-i} + \sum_{i=2}^{5} \mathcal{R}_{t-1}.\mathcal{R}_{t-i}$

(JMret2) $logit(\mathbb{P}[R_t \geq j]) = \theta_j^R + \mathcal{V}_t + \sum_{i=1}^{5} \mathcal{R}_{t-i} + \sum_{i=1}^{5} \mathcal{V}_{t-i} + \sum_{i=2}^{5} \mathcal{R}_{t-1}.\mathcal{R}_{t-i}$

(JMret3) as given in the first part of (3.7)

(JMvol1) $logit(\mathbb{P}[V_t \geq j]) = \theta_j^V + \sum_{i=1}^{5} \mathcal{V}_{t-i}$

(JMvol2) $logit(\mathbb{P}[V_t \geq j]) = \theta_j^V + \sum_{i=1}^{5} \mathcal{V}_{t-i} + \mathcal{V}_{t-1}.\mathcal{V}_{t-2}$

(JMvol3) as given in the second part of (3.7)

(JMvol4) $logit(\mathbb{P}[V_t \geq j]) = \theta_j^V + \sum_{i=1}^5 \mathcal{V}_{t-i} + \sum_{i=1}^5 \mathcal{R}_{t-i} + \mathcal{R}_{t-1}.\mathcal{R}_{t-2}$

(JMvol5) $logit(\mathbb{P}[V_t \geq j]) = \theta_j^V + \sum_{i=1}^5 \mathcal{V}_{t-i} + \sum_{i=1}^5 \mathcal{R}_{t-i} + \sum_{i=2}^5 \mathcal{R}_{t-1}.\mathcal{R}_{t-i}$

This is not a nested sequence of models, there are many possibilities for competing models. The variety of different models is actually prohibitively large. Here, we make an intuitively reasonable (and small) search, we again use the AIC statistic (see section 6) for selecting models. The following Table gives the numerical values.

	(JMret1)	(JMret2)	(JMret3)
AIC	2960.6	2963.8	2889.5

	(JMvol1)	(JMvol2)	(JMvol3)	(JMvol4)	(JMvol5)
AIC	2542.4	2544.5	2446.4	2448.9	2454.3

By looking at the AIC statistic, we thus select model (JMret3) and (JMvol3) which is model (3.7). Lagged cross-dependence of R_t from V_{t-i} $(i > 0)$ as in model (JMret2) seems non-relevant, which is a consistent finding with the result in Rogalski (1978).

In the following we give now the more detailed analysis of the joint model (JMret3)&(JMvol3). The reduction in the deviance (see section 6) from an independence model to our current model is 812.3, achieved by 74 additional parameters. Therefore, our model (JMret3)&(JMvol3) seems highly significant. The coefficient estimates for θ_j^R, θ_j^V $(j = 2, 3)$ and for the different effects are given in Table 2 in the Appendix. Corresponding t-values for such individual parameters, which are asymptotically correct and robust against model-misspecification, are again not easily available and a possible way to do nonparametric, model-free inference is given by resampling techniques, see section 6.

We are particularly interested to assess some measures of uncertainty for the estimated factor $\mathcal{V}_t$. The difference in the deviance (see section 6) from model (Mret3)&(JMvol3) without the factor $\mathcal{V}_t$ to the model (JMret3)&(JMvol3) with the factor $\mathcal{V}_t$ is 23.4, achieved by 2 additional parameters, which appears substantial. Asymptotically correct nonparametric, model-free confidence intervals can be constructed with the aid of the blockwise bootstrap, see section 6. By using 100 blockwise bootstrap [3] replicates we obtain the 99%-confidence intervals

$$[-1.452, -0.114] \text{ for the coefficient } \mathcal{V}_{0,1},$$
$$[-0.800, -0.149] \text{ for the coefficient } \mathcal{V}_{0,2},$$

where $\mathcal{V}_{0,i}$ denotes the coefficient of the factor $\mathcal{V}_t$ on level $i \in \{1, 2\}$ and the constraint that summing over all levels equals zero is implicitly made.
Both confidence intervals do not contain zero, so that the factor $\mathcal{V}_t$ is significant. The standard errors, estimated nonparametrically by the blockwise bootstrap are

$$S.E.(\hat{\mathcal{V}}_{0,1}) = 0.253,$$
$$S.E.(\hat{\mathcal{V}}_{0,2}) = 0.148.$$

Our significance results are obtained in a model-free way, which emphasizes the 'true relevance' of the variable V_t in explaining the variable R_t, given the history $\mathcal{H}_{t-1} = (R_{t-1}, V_{t-1}, \ldots)$.

[3] We used the blocklength $\ell = 38 \approx 2n^{1/3}$.

In Figure 5 we show the fitted probabilities

$$\hat{\mathbb{P}}[R_t = j | V_t = k, R_{t-1}, V_{t-1}, \ldots, R_{t-5}, V_{t-5}],\ j, k = 1, 2, 3.$$

From the coefficient estimates in Table 2, we get

$$\mathrm{argmax}_k \hat{\mathbb{P}}[R_t = j | V_t = k, R_{t-1}, V_{t-1}, \ldots, R_{t-5}, V_{t-5}] = j,\ j = 1, 3,\ \text{for all } t, \tag{4.9}$$

saying that the same extreme events are more likely to be present in R_t if they are already present in V_t. We refer to this as a positive association, in the same spirit as a positive correlation. But note that we have here an association in terms of conditional probabilities, whereas correlation measures only a linear association. There is a lot of discussion in the literature about association between return and volume, mainly about correlation. For an overview see Karpoff (1987).

For comparison with Figure 3 we show in Figure 6 the fitted probabilities

$$\hat{\mathbb{P}}[R_t = j | V_t, R_{t-1}, V_{t-1}, \ldots, R_{t-5}, V_{t-5}],\ j = 1, 2, 3.$$

In Figure 3 we did not include the information about volume. Comparing the Figures 3 and 6 is hard to do by eye, the absolute value of actual differences can be as large as 0.19. This is consistent with the result above that volume at time t is significant for explaining return at time t.

In Figure 7 we check the goodness of fit of our model (JMret3)&(JMvol3) for the return series. Based on the fitted probabilities in Figure 6 we compute

$$\hat{\mathbb{E}}[R_t | V_t, R_{t-1}, V_{t-1}, \ldots, R_{t-5}, V_{t-5}] = \sum_{j=1}^{3} j \hat{\mathbb{P}}[R_t = j | V_t, R_{t-1}, V_{t-1}, \ldots, R_{t-5}, V_{t-5}],$$

and plot it, together with the observed events, against the time axis t. We also show a plot of uniform residuals, see section 6. From Figure 7 we do not see any severe lacks of our model.

4.3 Karpoff's asymmetric volume–return relationship

We are also able to answer in some sense Karpoff's first research issue (Karpoff, 1987, Sec. VI). He poses the question whether the volume–price relationship is asymmetric. Loosely speaking the question is, whether the volume is negatively correlated with negative price changes and positively correlated with positive price changes. Clearly, a linear function could not describe such a relation and this question can only be answered by analyzing non-linear relations, as we do. The possible association is here directed, namely studying volume as a function of price changes. This is the inverse relation of our analysis exhibited in Figures 5 and 6. We take the point of view that the relation of V_t given R_t should be studied as a function of the history $\mathcal{H}_{t-1} = (R_{t-1}, V_{t-1}, \ldots)$. Only such a view allows to quantify the dependence of V_t from R_t after the effect of the history $\mathcal{H}_{t-1}$ has been removed. This seems much more natural than studying the effect of R_t on V_t without regarding the history $\mathcal{H}_{t-1}$. With our model for the joint distribution for R_t, V_t given the history of the first five lagged variables, we are able to calculate

$$\hat{\mathbb{P}}[V_t = j | R_t = k, V_{t-1}, R_{t-1}, \ldots, V_{t-5}, R_{t-5}],\ j, k = 1, 2, 3.$$

In Figure 8 we plot these probabilities against the time axis t. In Figure 9 we give a magnification for the above probabilities with $j = 1$; $k = 1, 3$. The Figures 8 and 9 suggest that a similar formula as in (4.9) is no longer true. By exact computation we get the following,

$$\text{argmax}_k \hat{\mathbb{P}}[V_t = j | R_t = k, V_{t-1}, R_{t-1}, \ldots, V_{t-5}, R_{t-5}] \neq j, \; j = 1, 3, \text{ for some } t. \quad (4.10)$$

The volume–return relationship is asymmetric in the sense of formula (4.10). In particular, as suggested by Figure 9, the conditional probability for an extreme negative event of volume ($V_t = 1$) is often more likely when the return at the corresponding time point is extremely high ($R_t = 3$) than extremely low ($R_t = 1$). Suggested by Figure 8 and exactly computed from the coefficients estimates in Table 2, Formula (4.10) sharpens informally to,

$$\begin{aligned}
&\text{argmax}_{k \in \{1,3\}} \hat{\mathbb{P}}[V_t = 1 | R_t = k, V_{t-1}, R_{t-1}, \ldots, V_{t-5}, R_{t-5}] = 3, \text{ for `quite many' } t, \\
&\text{argmax}_{k \in \{1,3\}} \hat{\mathbb{P}}[V_t = 3 | R_t = k, V_{t-1}, R_{t-1}, \ldots, V_{t-5}, R_{t-5}] = 3, \text{ for `almost all' } t.
\end{aligned}$$

This gives a positive answer in probabilistic terms to Karpoff's question. Our analysis allows to quantify the volume–return relation of extreme events.

4.4 Results for prediction

For the prediction purpose, we fit our model (3.7) for the first $n = 6130$ observations and use then this estimated model for predicting the last 300 remaining values of the return series. To be precise, we are doing here one-step ahead predictions as described in section 3.5, but our estimated model will always be the same based on the first 6130 observations. Figure 10 shows the prediction probabilities

$$\hat{\mathbb{P}}[R_{n+1} = j | \hat{V}_{n+1}, R_n, V_n, \ldots, R_{n-4}, V_{n-4}], \; j = 1, 2, 3; \; n = 6130, \ldots, 6429.$$

These predicted probabilities clearly reflect some of the exceptional behavior of the return series around time points 6250-6300, which correspond to the times around the 1987 crash. In Figure 11 we show the MAP predictor

$$\hat{R}_{n+1}, \; n = 6130, \ldots, 6429$$

and the actual values which occurred. Only rarely, our MAP rule predicts an extreme event: 4 out of 5 extreme event predictions are correct. This shows that our algorithm is not really making magnificent predictions, but we still can gain, in that 80% of the few extreme value predictions are correct. Note also that the MAP predictor from step 3 in section 3.5 is conservative. For example, if we could gain a lot by making a correct extreme event prediction, an alternative predictor might be defined as

$$\hat{R}_{n+1} = \begin{cases} 1, & \text{if } \hat{\mathbb{P}}_1 > d_1 \hat{\mathbb{P}}_2 \text{ and } \hat{\mathbb{P}}_1 \geq \hat{\mathbb{P}}_3 \\ 3, & \text{if } \hat{\mathbb{P}}_3 > d_2 \hat{\mathbb{P}}_2 \text{ and } \hat{\mathbb{P}}_3 > \hat{\mathbb{P}}_1 \\ 2, & \text{otherwise} \end{cases},$$

where $\hat{\mathbb{P}}_j = \hat{\mathbb{P}}[R_{n+1} = j | \hat{V}_{n+1}, R_n, V_n, \ldots, R_{n-4}, V_{n-4}]$, and d_1, d_2 could depend on some cost function. This then would yield more progressive and risky predictions.

Other explanatory quantities might improve the power for predicting extreme values of return. LeBaron (1990) has used volatility (and not volume) as an additional explanatory variable, acting as a switching variable. Our modeling and prediction approach could be extended in a straightforward way to include other variables.

5 Conclusions

We have presented an approach for modeling extreme events in financial time series. The discretization into three ordinal categories and the likelihood modeling using the framework of generalized linear models with lagged factor variables seem to be an innovative and new idea in the field of finance.

Our models have the attractive feature to draw direct probabilistic statements, here given for joint, marginal and conditional distributions of return of the Dow Jones and volume of the NYSE at time t, always given the history $t-1, t-2, \ldots$. Unlike the frequently used correlation for measuring association, our estimated (joint) probabilities are not restricted to linear associations. Correlations between changes of volume and price (or absolute value of price) were often found to be weak, cf. Crouch (1970a, 1970b), Rogalski (1978). But this might be due to a substantial non-linear association which cannot be picked up by the correlation measure. Our models are an attempt to describe the whole probability structure of the vector time series of extreme events for return and volume. Given the history, volume at time t appears to be significant for explaining the conditional probability of return at time t. This conclusion can be drawn from the coefficients of the factor $\mathcal{V}_t$ which are significant, as well as from the reduction of the AIC statistic 2908.9 of model (Mret4) to 2889.5 of (JMret3), which is remarkable. Moreover, volume at times $t-i$, $(i > 0)$ seems unimportant for explaining the conditional probability of return at time t, indicating an independence of return and lagged volume. A related result is given in Rogalski (1978).

Using such estimates of the whole probability structure, we are able to give an answer to the first research issue proposed by Karpoff (1987): the volume–return relationship, at least for extreme events, appears to be asymmetric. This asymmetry has been found after the effect of the history has been removed, thus establishing a probabilistic result about the sample path of the financial process, rather than only for one time point, such as 'today'. Beaver (1968) suggests that returns correspond to changes in the expectation of the market as a whole, whereas volumes correspond to changes in the expectation of the individual investor. Thus, the asymmetry in (4.10) and its following formulas imply: given the history, an extreme negative change in the expectation of the individual investor is often more likely under a given extreme positive than negative change in the expectation of the whole market.

Our models also yield a natural prediction algorithm. Its predictive power is neither magnificent nor impractically poor.

6 Mathematical and computational remarks

GLM, deviance and AIC

The modeling approach as described in sections 3.3 and 3.4 is an extension of cumulative logits models for independent observations to the dependent Markovian case. References for the independent case are Agresti (1990), McCullagh and Nelder (1989) and for the dependent case Fahrmeir and Tutz (1994, Chs. 6.1, 8.2-8.3), Brillinger (1996).

The deviance is a measure for goodness of fit, which is useful for non-normal and non-linear models. A model (M) of interest is compared to the baseline model (M_0), which has as many parameters as observations. Given data $\mathbf{X} = (X_1, \ldots, X_n)$, we denote by $\ell(\hat{\beta}_M; \mathbf{X})$ the log-likelihood function evaluated at the MLE estimate $\hat{\beta}_M$ under model (M). The deviance is defined as

$$D_M = 2(\ell(\hat{\beta}_{M_0}; \mathbf{X}) - \ell(\hat{\beta}_M; \mathbf{X})),$$

which is two times the log likelihood ratio statistic. Having nested models $(M_1) \subseteq (M_2)$, the difference of deviances is defined as

$$\Delta D_{M_1 \subseteq M_2} = D_{M_1} - D_{M_2} = 2(\ell(\hat{\beta}_{M_2}; \mathbf{X}) - \ell(\hat{\beta}_{M_1}; \mathbf{X})),$$

which measures the significance of model (M_2) relative to its sub-model (M_1). For a more detailed treatment, cf. McCullagh and Nelder (1989, Ch. 2.3).

Comparing the predictive power of different models can be done with Akaike's information criterion (AIC). The goal is to minimize

$$AIC = -2\ell(\hat{\beta}_M; \mathbf{X}) + 2(\text{number of free parameters in model } M),$$

where the minimization is done over different models M of consideration.
Since the MLE estimate $\hat{\beta}_M$ is the maximizer of $\ell(.; \mathbf{X})$ for model M, the first term on the right hand side of the AIC statistic decreases for larger models, whereas the second term acts as a penalty for large models. See Akaike (1973).

Model-free inference and bootstrap

Assessing correct asymptotic standard errors, asymptotic confidence intervals or to construct asymptotically correct tests is completely different (and inherently more difficult) from the set-up with independent data. With dependent data, techniques for estimating variances from the independent case are asymptotically valid, only if the (GLM-) model we work with is correct. But this is a rather stringent assumption and we often cannot trust it. Our model might be good for approximating the 'truth', but assessing accuracy of such approximations or testing if the model approximates well should be done in a model-free fashion. To get model-free, nonparametric variance or distribution estimates, the idea of bootstrapping can be used. Efron's (1979) classical bootstrap will fail, since we are dealing here with dependent data. One has to rely on techniques adapted to the dependent case. A model free resampling scheme is given by the blockwise bootstrap proposed in Künsch (1989) and further developed in Bühlmann (1994).

Uniform residuals

The plot of uniform residuals was developed by Brillinger and Preisler (1983). Suppose X is an ordinal-valued variable with $\mathbb{P}[X = 1] = p_1$, $\mathbb{P}[X = 2] = p_2$, $\mathbb{P}[X = 3] = 1 - p_1 - p_2$. Then, the variable

$$Z = \begin{cases} \text{Uniform on } (0, p_1], & \text{if } X = 1 \\ \text{Uniform on } (p_1, p_1 + p_2], & \text{if } X = 2 \\ \text{Uniform on } (p_1 + p_2, 1), & \text{if } X = 3 \end{cases}$$

has a Uniform distribution on $(0, 1)$. This can then be used to plot such a variable $Z = Z(x)$ versus different points on the x-axis with now $p_i = p_i(x)$ depending on x.

Computations

The computations can all be carried out by a statistical package which is able to fit cumulative logits models as in (3.3)-(3.7). In particular, since the log-likelihood function in (3.6) separates into two parts which are modeled individually, both terms $\ell(\beta^R; R, V)$ and $\ell(\beta^V; R, V)$ can be maximized separately and no multivariate techniques are necessary. All computations and graphics have been done with S-Plus, cf. Becker et al. (1988), Chambers and Hastie (1992). The cumulative logit model has been fitted with the additional function *logist* from a special library called *logist*.

Appendix

The Tables 1 and 2 give the MLE estimates of our analysis. Identifiability constraints are chosen such that summing over all levels of a factor equals zero. Denote by $\hat{\theta}_j$ the estimates for θ_j $(j = 2, 3)$, by $\hat{\mathcal{R}}_{i,j}$ the estimates for the factor $\mathcal{R}_{t-i}$ $(i = 1, \ldots, 5;\ j = 1, 2)$ and by $\hat{\mathcal{R}}_{i,j;i',j'}$ the estimates for the factor $\mathcal{R}_{t-i}.\mathcal{R}_{t-i'}$ $(i, i' = 1, \ldots, 5;\ j, j' = 1, 2)$. Analogously for the factors $\mathcal{V}_{t-i}$ $(i = 0, 1, \ldots, 5)$ and interaction factors.

Acknowledgments: I would like to thank David Brillinger for many helpful comments and discussions.

References

[1] Agresti, A. (1990). Categorical Data Analysis. Wiley, New York.

[2] Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In 2nd Int. Symposium on Information Theory, Eds. B.N. Petrov and F. Csáki, pp. 267-281. Akadémia Kiado, Budapest.

[3] Beaver , W.V. (1968). The information content of annual earnings announcements. Empirical Research in Accounting: Selected Studies. Supplement to J. of Accounting Research **6** 67-92.

[4] Becker, R.A., Chambers, J.M. and Wilks, A.R. (1988). The New S Language. Wadsworth, Pacific Grove.

[5] Brillinger, D.R. (1996). An analysis of an ordinal–valued time series. In Athens Conference on Applied Probability and Time Series Analysis, Vol. II: Time Series Analysis. In Memory of E.J. Hannan. Lect. Notes in Statistics **115**. Springer-Verlag, New York.

[6] Brillinger, D.R. and Preisler, H.K. (1983). Maximum likelihood estimation in a latent variable problem. In Studies in Econometrics, Time Series and Multivariate Statistics, Eds. S. Karlin et al., pp. 31-65. Academic, New York.

[7] Bühlmann, P. (1994). Blockwise bootstrapped empirical processes for stationary sequences. Annals of Statistics **22** 995-1012.

[8] Chambers, J.M. and Hastie, T.J. (1992). Statistical Models in S. Wadsworth, Pacific Grove.

[9] Crouch, R.L. (1970a). A nonlinear test of the random-walk hypothesis. American Economic Rev. **60** 199-202.

[10] Crouch, R.L. (1970b). The volume of transactions and price changes on the New York Stock Exchange. Financial Analysts Journal **26** 104-109.

[11] Efron, B. (1979). Bootstrap methods: another look at the jackknife. Annals of Statistics **7** 1-26.

[12] Fahrmeir, L. and Tutz, G. (1994). Multivariate Statistical Modelling based on Generalized Linear Models. Springer, New York.

[13] Granger, C.W.J. (1992). Forecasting stock market prices: lessons for forecasters. Internat. J. of Forecasting **8** 3-13.

[14] Karpoff, J.M. (1987). The relation between price changes and trading volume: a survey. J. of Financial and Quantitative Analysis **22** 109-126.

[15] Künsch, H.R. (1989). The jackknife and the bootstrap for general stationary observations. Annals of Statistics **17** 1217-1241.

[16] LeBaron, B. (1990). Forecasting improvements using a volatility index. Working paper, Economics Dept. University of Wisconsin, Madison.

[17] McCullagh, P. and Nelder, J.A. (1989). Generalized Linear Models. 2nd Edition. Chapman and Hall, London.

[18] Rogalski, R.J. (1978). The dependence of prices and volume. The Rev. of Economics and Statistics **36** 268-274.

[19] Weigend, A. S. and LeBaron, B. (1994). Evaluating Neural Network Predictors by Bootstrapping. In Proc. of the International Conference on Neural Information Processing (ICONIP'94), pp. 1207-1212. Seoul, Korea.

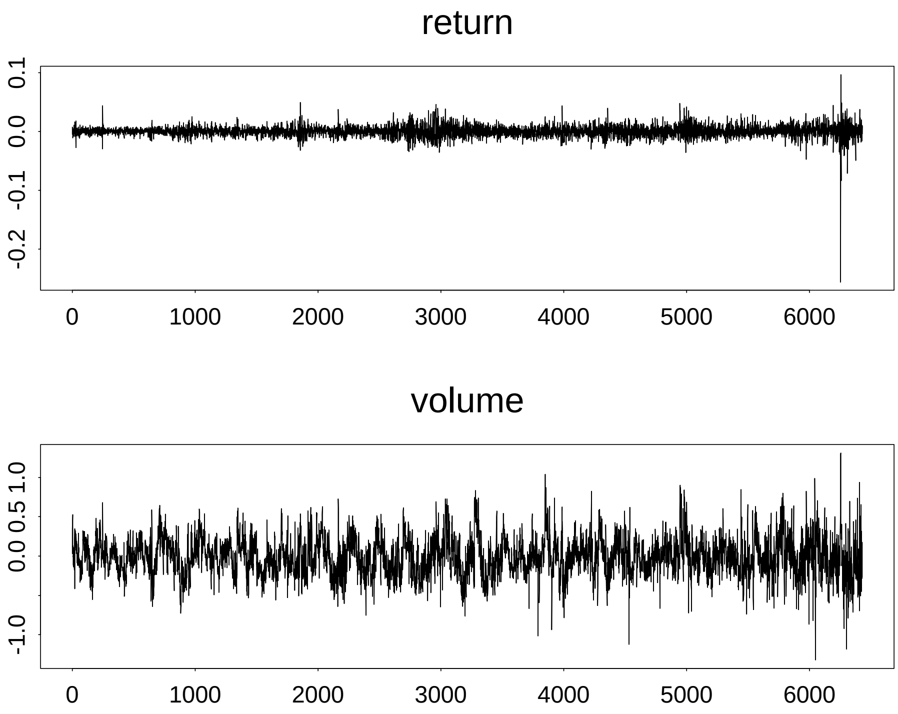

Figure 1: Daily return and volume

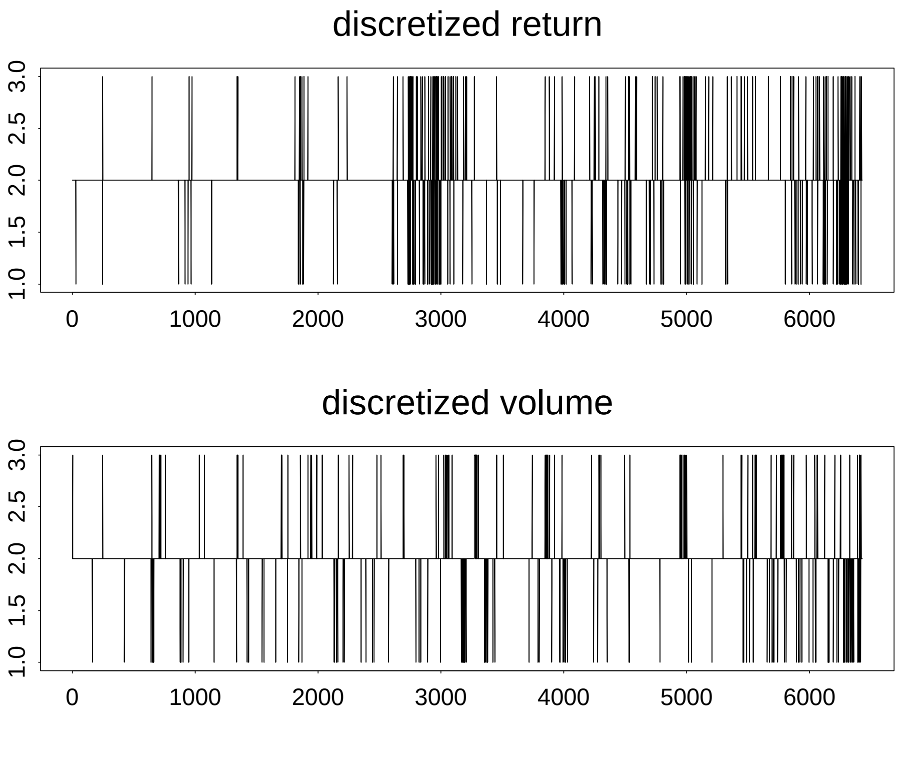

Figure 2: Discretized daily return and daily volume

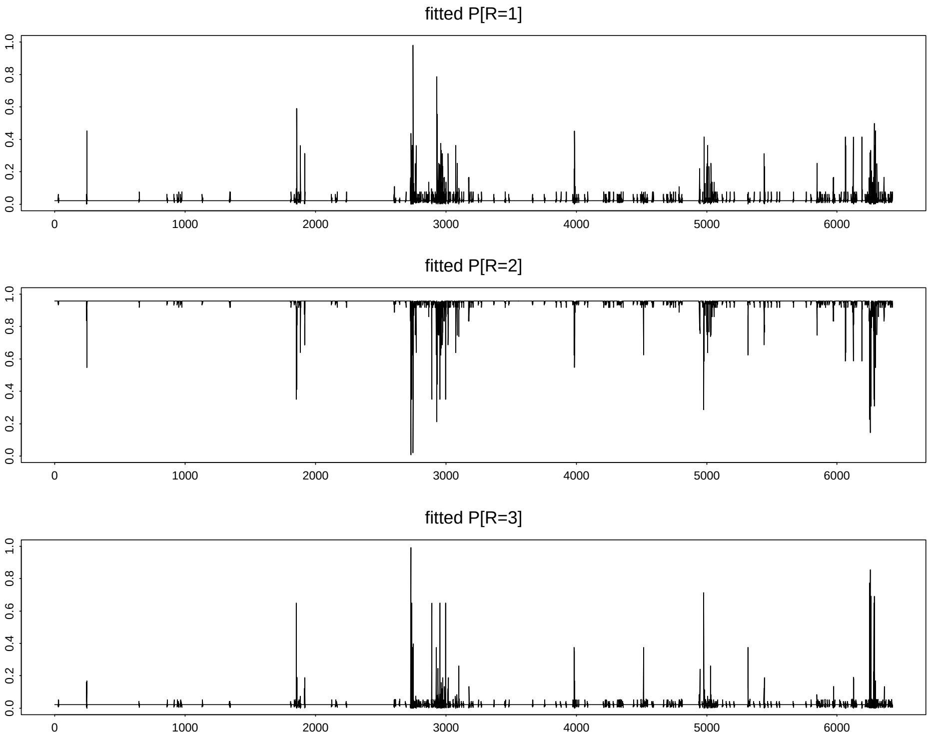

Figure 3: Fitted probabilities for return

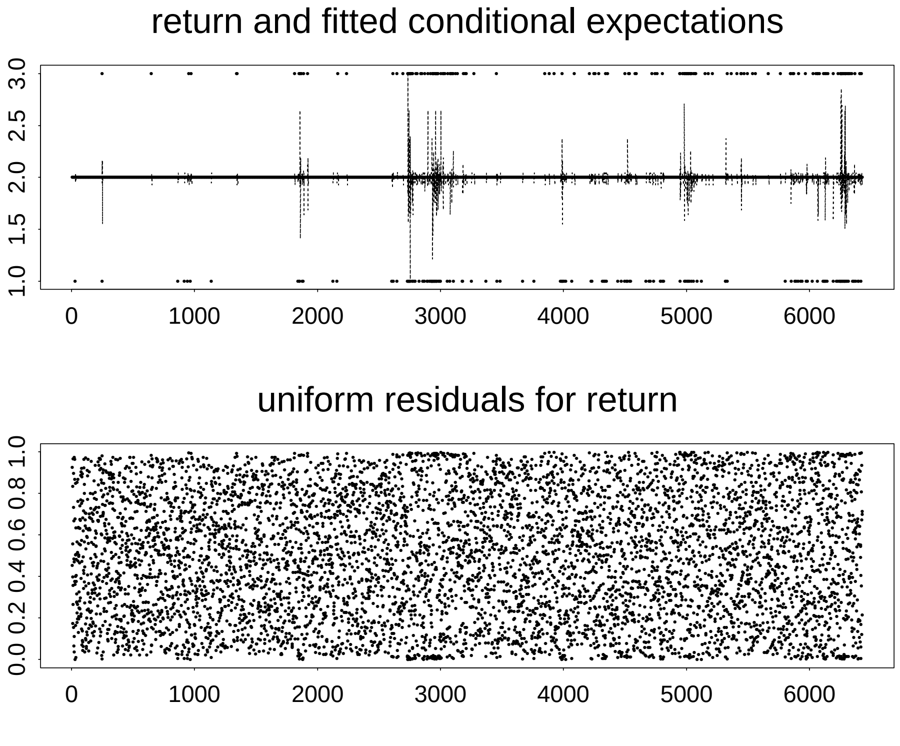

Figure 4: Real (dots) and expected returns, and uniform residuals for return

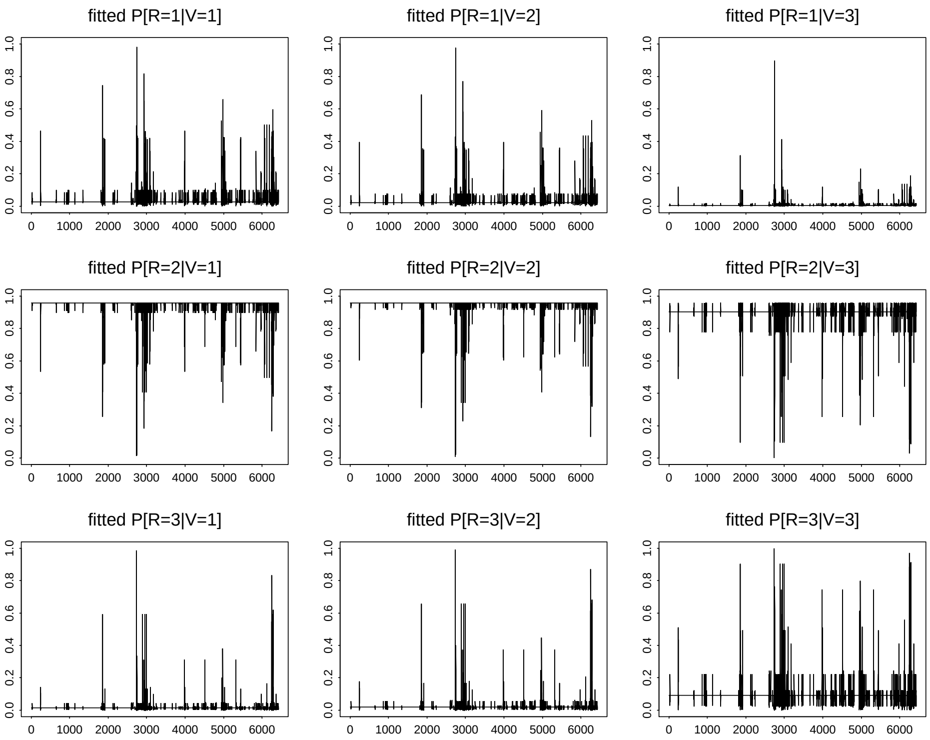

Figure 5: Fitted probabilities for return given all levels of volume

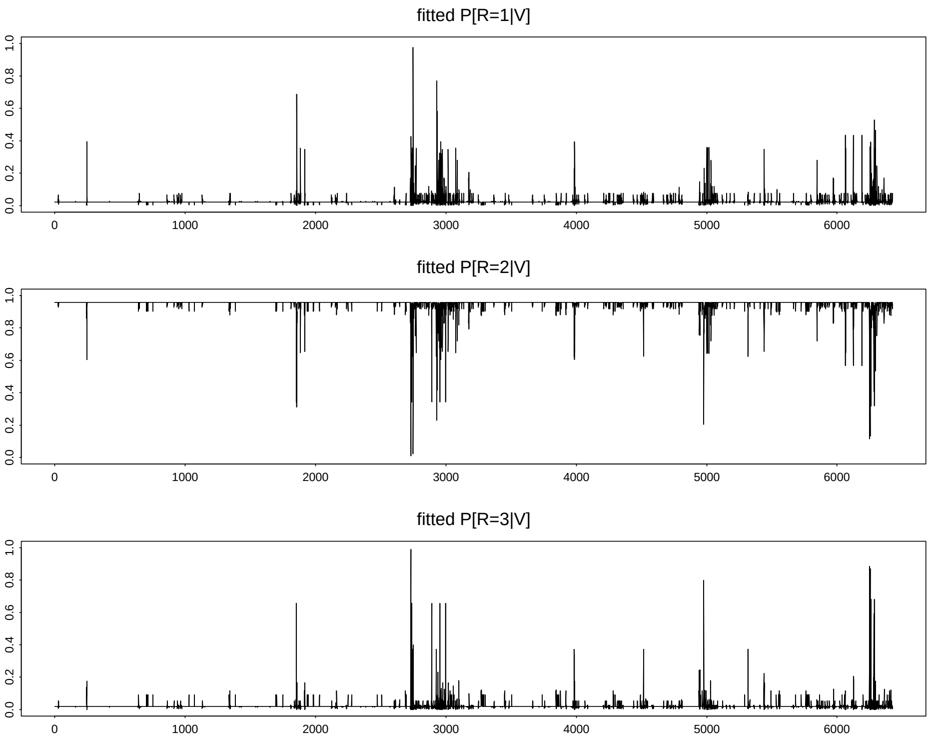

Figure 6: Fitted probabilities for return given volume

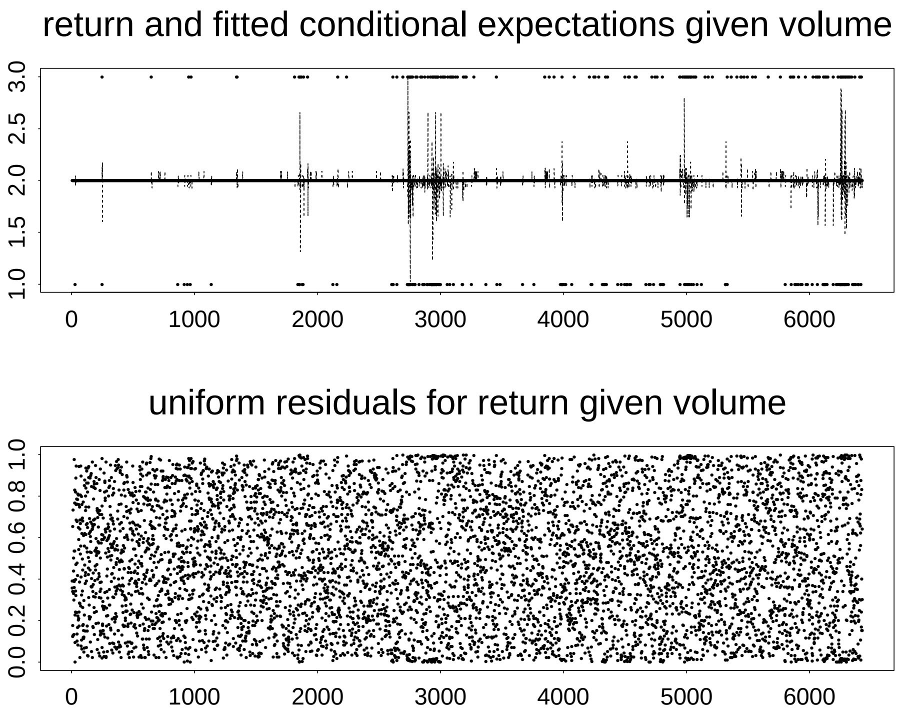

Figure 7: Real (dots) and expected returns given volume, and uniform residuals for return given volume

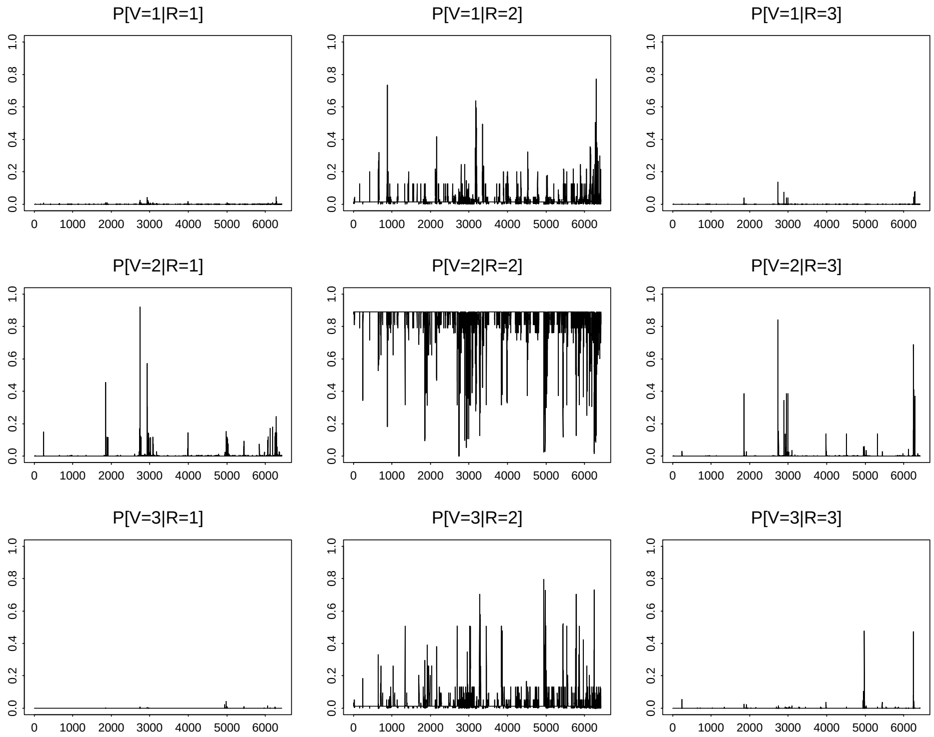

Figure 8: Fitted probabilities for volume given all levels of return

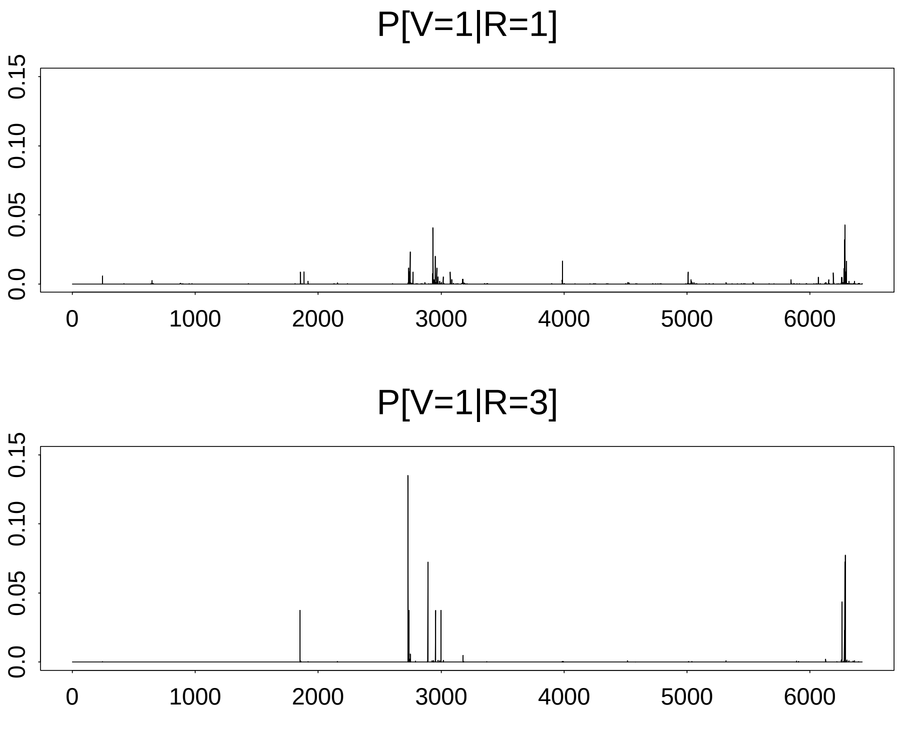

Figure 9: Fitted probabilities for extremely small volume given the extreme levels of return

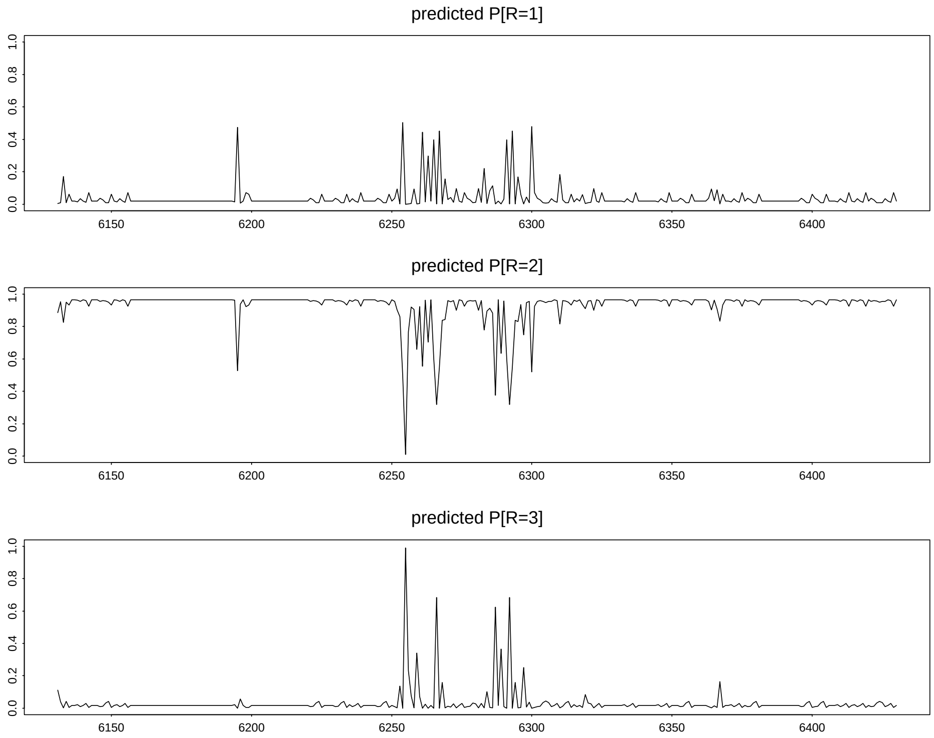

Figure 10: Predicted probabilities for return

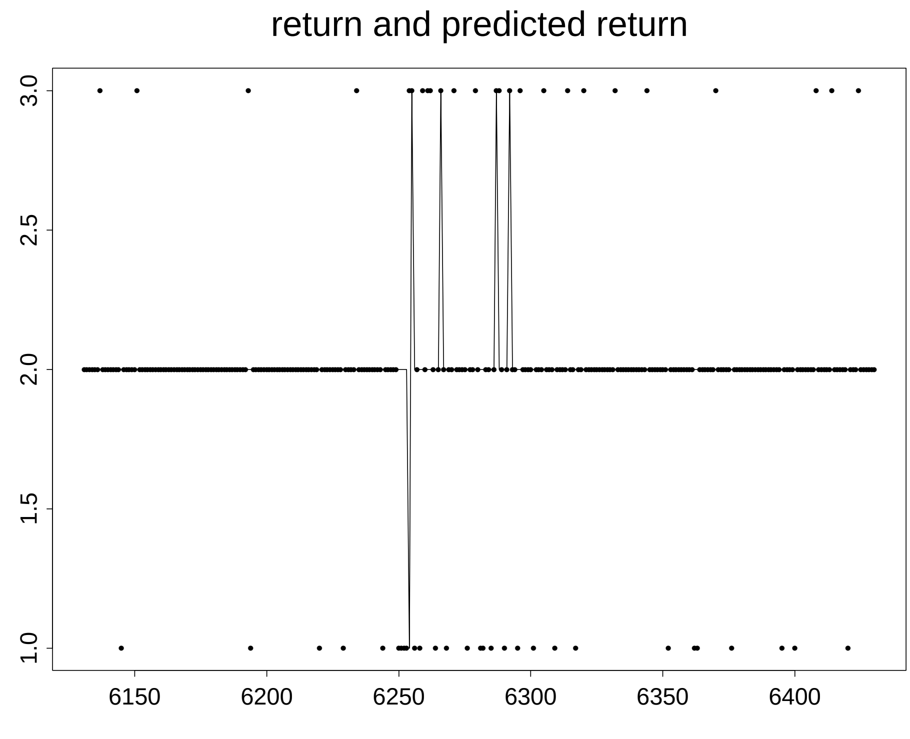

Figure 11: Real returns (dots) and MAP predictions (line) for returns

$\hat{\theta}_2 = 3.80$	$\hat{\theta}_3 = -3.85$		
$\hat{\mathcal{R}}_{1,1} = -0.94$	$\hat{\mathcal{R}}_{1,2} = -0.47$		
$\hat{\mathcal{R}}_{2,1} = 2.72$	$\hat{\mathcal{R}}_{2,2} = -0.69$		
$\hat{\mathcal{R}}_{3,1} = -0.35$	$\hat{\mathcal{R}}_{3,2} = -0.07$		
$\hat{\mathcal{R}}_{4,1} = 0.54$	$\hat{\mathcal{R}}_{4,2} = 0.48$		
$\hat{\mathcal{R}}_{5,1} = -0.33$	$\hat{\mathcal{R}}_{5,2} = 0.33$		
$\hat{\mathcal{R}}_{1,1;2,1} = -0.09$	$\hat{\mathcal{R}}_{1,2;2,1} = -0.36$	$\hat{\mathcal{R}}_{1,1;2,2} = 0.98$	$\hat{\mathcal{R}}_{1,2;2,2} = 0.09$
$\hat{\mathcal{R}}_{1,1;3,1} = 1.48$	$\hat{\mathcal{R}}_{1,2;3,1} = -0.24$	$\hat{\mathcal{R}}_{1,1;3,2} = -0.96$	$\hat{\mathcal{R}}_{1,2;3,2} = 0.84$
$\hat{\mathcal{R}}_{1,1;4,1} = -0.41$	$\hat{\mathcal{R}}_{1,2;4,1} = 0.73$	$\hat{\mathcal{R}}_{1,1;4,2} = 0.01$	$\hat{\mathcal{R}}_{1,2;4,2} = -0.16$
$\hat{\mathcal{R}}_{1,1;5,1} = -0.35$	$\hat{\mathcal{R}}_{1,2;5,1} = -0.45$	$\hat{\mathcal{R}}_{1,1;5,2} = -0.06$	$\hat{\mathcal{R}}_{1,2;5,2} = -0.14$
$\hat{\mathcal{R}}_{2,1;3,1} = -0.53$	$\hat{\mathcal{R}}_{2,2;3,1} = 0.36$	$\hat{\mathcal{R}}_{2,1;3,2} = -0.01$	$\hat{\mathcal{R}}_{2,2;3,2} = 0.15$
$\hat{\mathcal{R}}_{2,1;4,1} = 0.43$	$\hat{\mathcal{R}}_{2,2;4,1} = -0.61$	$\hat{\mathcal{R}}_{2,1;4,2} = -0.47$	$\hat{\mathcal{R}}_{2,2;4,2} = 0.03$
$\hat{\mathcal{R}}_{2,1;5,1} = 1.95$	$\hat{\mathcal{R}}_{2,2;5,1} = -0.81$	$\hat{\mathcal{R}}_{2,1;5,2} = -2.08$	$\hat{\mathcal{R}}_{2,2;5,2} = 0.69$
$\hat{\mathcal{R}}_{3,1;4,1} = -2.03$	$\hat{\mathcal{R}}_{3,2;4,1} = -0.06$	$\hat{\mathcal{R}}_{3,1;4,2} = 1.19$	$\hat{\mathcal{R}}_{3,2;4,2} = -0.82$
$\hat{\mathcal{R}}_{3,1;5,1} = -0.54$	$\hat{\mathcal{R}}_{3,2;5,1} = 1.00$	$\hat{\mathcal{R}}_{3,1;5,2} = -0.26$	$\hat{\mathcal{R}}_{3,2;5,2} = -0.29$
$\hat{\mathcal{R}}_{4,1;5,1} = 2.11$	$\hat{\mathcal{R}}_{4,2;5,1} = 0.62$	$\hat{\mathcal{R}}_{4,1;5,2} = -0.15$	$\hat{\mathcal{R}}_{4,2;5,2} = 0.09$

Table 1: MLE estimates for model (3.5)

$\hat{\theta}_2 = 4.03$	$\hat{\theta}_3 = -3.68$		
$\hat{\mathcal{V}}_{0,1} = -0.71$	$\hat{\mathcal{V}}_{0,2} = -0.43$		
$\hat{\mathcal{R}}_{1,1} = -0.80$	$\hat{\mathcal{R}}_{1,2} = -0.34$		
$\hat{\mathcal{R}}_{2,1} = 2.73$	$\hat{\mathcal{R}}_{2,2} = -0.59$		
$\hat{\mathcal{R}}_{3,1} = -0.35$	$\hat{\mathcal{R}}_{3,2} = 0.04$		
$\hat{\mathcal{R}}_{4,1} = 0.54$	$\hat{\mathcal{R}}_{4,2} = 0.46$		
$\hat{\mathcal{R}}_{5,1} = -0.31$	$\hat{\mathcal{R}}_{5,2} = 0.37$		
$\hat{\mathcal{R}}_{1,1;2,1} = -0.20$	$\hat{\mathcal{R}}_{1,2;2,1} = -0.44$	$\hat{\mathcal{R}}_{1,1;2,2} = 0.97$	$\hat{\mathcal{R}}_{1,2;2,2} = 0.07$
$\hat{\mathcal{R}}_{1,1;3,1} = 1.45$	$\hat{\mathcal{R}}_{1,2;3,1} = -0.23$	$\hat{\mathcal{R}}_{1,1;3,2} = -0.99$	$\hat{\mathcal{R}}_{1,2;3,2} = 0.85$
$\hat{\mathcal{R}}_{1,1;4,1} = -0.48$	$\hat{\mathcal{R}}_{1,2;4,1} = 0.69$	$\hat{\mathcal{R}}_{1,1;4,2} = 0.04$	$\hat{\mathcal{R}}_{1,2;4,2} = -0.10$
$\hat{\mathcal{R}}_{1,1;5,1} = -0.39$	$\hat{\mathcal{R}}_{1,2;5,1} = -0.45$	$\hat{\mathcal{R}}_{1,1;5,2} = -0.10$	$\hat{\mathcal{R}}_{1,2;5,2} = -0.17$
$\hat{\mathcal{R}}_{2,1;3,1} = -0.58$	$\hat{\mathcal{R}}_{2,2;3,1} = 0.36$	$\hat{\mathcal{R}}_{2,1;3,2} = 0.04$	$\hat{\mathcal{R}}_{2,2;3,2} = 0.08$
$\hat{\mathcal{R}}_{2,1;4,1} = 0.34$	$\hat{\mathcal{R}}_{2,2;4,1} = -0.57$	$\hat{\mathcal{R}}_{2,1;4,2} = -0.39$	$\hat{\mathcal{R}}_{2,2;4,2} = 0.00$
$\hat{\mathcal{R}}_{2,1;5,1} = 1.87$	$\hat{\mathcal{R}}_{2,2;5,1} = -0.84$	$\hat{\mathcal{R}}_{2,1;5,2} = -2.11$	$\hat{\mathcal{R}}_{2,2;5,2} = 0.71$
$\hat{\mathcal{R}}_{3,1;4,1} = -2.15$	$\hat{\mathcal{R}}_{3,2;4,1} = -0.03$	$\hat{\mathcal{R}}_{3,1;4,2} = 1.22$	$\hat{\mathcal{R}}_{3,2;4,2} = -0.85$
$\hat{\mathcal{R}}_{3,1;5,1} = -0.69$	$\hat{\mathcal{R}}_{3,2;5,1} = 1.06$	$\hat{\mathcal{R}}_{3,1;5,2} = -0.21$	$\hat{\mathcal{R}}_{3,2;5,2} = -0.34$
$\hat{\mathcal{R}}_{4,1;5,1} = 1.99$	$\hat{\mathcal{R}}_{4,2;5,1} = 0.60$	$\hat{\mathcal{R}}_{4,1;5,2} = -0.11$	$\hat{\mathcal{R}}_{4,2;5,2} = 0.08$

$\hat{\theta}_2 = 4.15$	$\hat{\theta}_3 = -4.30$
$\hat{\mathcal{V}}_{1,1} = -2.26$	$\hat{\mathcal{V}}_{1,2} = 0.02$
$\hat{\mathcal{V}}_{2,1} = -0.80$	$\hat{\mathcal{V}}_{2,2} = -0.24$
$\hat{\mathcal{V}}_{3,1} = -0.83$	$\hat{\mathcal{V}}_{3,2} = -0.05$
$\hat{\mathcal{V}}_{4,1} = -1.17$	$\hat{\mathcal{V}}_{4,2} = 0.04$
$\hat{\mathcal{V}}_{5,1} = -0.64$	$\hat{\mathcal{V}}_{5,2} = 0.04$
$\hat{\mathcal{R}}_{1,1} = 0.16$	$\hat{\mathcal{R}}_{1,2} = -1.41$
$\hat{\mathcal{R}}_{2,1} = -0.45$	$\hat{\mathcal{R}}_{2,2} = 0.30$
$\hat{\mathcal{R}}_{3,1} = -0.52$	$\hat{\mathcal{R}}_{3,2} = 0.35$
$\hat{\mathcal{R}}_{4,1} = -0.05$	$\hat{\mathcal{R}}_{4,2} = 0.35$
$\hat{\mathcal{R}}_{5,1} = -0.57$	$\hat{\mathcal{R}}_{5,2} = 0.56$

Table 2: MLE estimates for model (3.7)