Regime switching model for financial data: Empirical risk analysis

Khaled Salhi [a,b,c,d], Madalina Deaconu [a,b,c], Antoine Lejay [a,b,c,*],
Nicolas Champagnat [a,b,c], Nicolas Navet [e]

[a] Université de Lorraine, Institut Elie Cartan de Lorraine, UMR 7502, Vandoeuvre-lès-Nancy, F-54506, France
[b] CNRS, Institut Elie Cartan de Lorraine, UMR 7502, Vandoeuvre-lès-Nancy, F-54506, France
[c] Inria, Villers-lès-Nancy, F-54600, France
[d] Alphability, Vandoeuvre-lś-Nancy, F-54500, France
[e] FSTC/CSC Research Unit/Lassy lab., Université du Luxembourg, L-1359, Luxembourg

HIGHLIGHTS

- This paper introduces a regime switching model for Value-at-Risk estimation.
- Hidden Markov models and extreme value theory are combined into a hybrid model.
- The regime switching model is applied to real data NYSE Euronext stocks.
- Classifying data in two states permits a fast detection of regime switching.
- This new model increases predictive performance of VaR forecasting.

ARTICLE INFO

ABSTRACT

This paper constructs a regime switching model for the univariate Value-at-Risk estimation. Extreme value theory (EVT) and hidden Markov models (HMM) are combined to estimate a hybrid model that takes volatility clustering into account. In the first stage, HMM is used to classify data in crisis and steady periods, while in the second stage, EVT is applied to the previously classified data to rub out the delay between regime switching and their detection. This new model is applied to prices of numerous stocks exchanged on NYSE Euronext Paris over the period 2001–2011. We focus on daily returns for which calibration has to be done on a small dataset. The relative performance of the regime switching model is benchmarked against other well-known modeling techniques, such as stable, power laws and GARCH models. The empirical results show that the regime switching model increases predictive performance of financial forecasting according to the number of violations and tail-loss tests. This suggests that the regime switching model is a robust forecasting variant of power laws model while remaining practical to implement the VaR measurement.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

The Value-at-Risk (VaR) is one of the main risk indicators for management of financial portfolios [1]. It is the threshold above which a loss over a chosen time horizon occurs with at most a given level of confidence.

---

* Corresponding author at: Inria, Villers-lès-Nancy, F-54600, France.
E-mail addresses: Khaled.Salhi@inria.fr (K. Salhi), Madalina.Deaconu@inria.fr (M. Deaconu), Antoine.Lejay@inria.fr (A. Lejay), Nicolas.Champagnat@inria.fr (N. Champagnat), nicolas.navet@uni.lu (N. Navet).

The VaR may be estimated either by parametric or non-parametric techniques. The non-parametric ones use only the empirical distributions (historical, resampling) without fitting a model. Due to the small amount of available data, it does not provide an accurate estimate for the probability of extreme events. The parametric approach overrides partly the problem induced by the lack of data by fitting the parameters of a model on historical data and computing afterwards the VaR, either by analytical or numerical methods.

The RiskMetric methodology [2] is widely used to estimate the risk associated to a portfolio by establishing quantitative relations between variations of the risk indicator with respect to risk factors (stocks and prices of derivatives for example). Nowadays this methodology incorporates heavy tail distributions, but it was initially developed in the Gaussian context, which is still prevalent in risk management and enforced by Basel Accord [3].

However, actual regulations and standard procedures for computing the VaR, based mainly on the Gaussian world, have been invalidated by many studies (see *e.g.* Refs. [4,5]) as they strongly underestimate the extreme events observed in the market.

The successive financial crises since 1987 led to a greater attention in modeling tail behavior of the induced returns distributions, and to the use of Extreme Value Theory (EVT) as a central concept in Risk Management [6].

Neither single model nor statistical methodology is acknowledged as standard for dealing with heavy-tails. In this study, we construct a variant of power law tails of the distributions by taking into consideration the regime switching between crisis and steady periods. We compare our model to several known models (stable, Generalized Pareto tails of distribution and GARCH). We consider heavy tail distributions for the losses L with distribution function F, typically, generalized Pareto (power laws):

$$1 - F(x) = \mathbb{P}(L \geqslant x) = \frac{\ell(x)}{x^{\alpha}}, \tag{1}$$

where ℓ is a function of slow variation and α is the *tail index* which summarizes the heaviness of the tail distribution and characterizes also the existence of moments.

In our approach, we aim to benefit from the stability of power law models in the VaR forecasting and the detection of volatility clustering given by conditional models. We assume the existence of two states: crisis and steady, and we classify data in two regimes using hidden Markov models (HMM). Then, the power law tail distribution is estimated from the past crisis and steady periods. The model gives more weight to the current regime, which reduces under-estimations and over-estimations at the beginning of crisis and steady periods.

We focus here only on univariate distributions. The case of several stocks leads to higher complexity, as the notion of VaR itself is not properly defined (see *e.g.* Ref. [7]). The multivariate case will be subject of further studies.

Our models of risk are intensively tested on historical market data from NYSE Euronext Paris. We focus only on daily returns of single stock prices, which is the practical situation of small investors and for which the problem of model calibration is the most difficult because of the small amount of data.

Outline. In Section 2, we give an overview of heavy-tailed models used in finance, the markets on which they have been applied and the estimators for power laws. In Section 3, we introduce our methodology for estimating the parameters and performing the backtesting. Finally, in Section 4, we introduce our dataset and discuss the conclusions we draw from the study over a selection of 56 stocks from NYSE Euronext Paris. Perspectives and conclusions are then highlighted in Section 5.

2. Empirical evidence for heavy tails

2.1. Evidences for the power law in financial markets

It is widely acknowledged that prices and returns of stocks obey to general laws usually called "stylized facts" [8]. Skewness and heavy-tail are the two main properties of observed prices which are not verified by the Black&Scholes models.

Although stable distributions have been proposed since the 60's [9,10] and an alternative model (finite variance subordinated log-normal distributions) has been proposed in 1973 by P. Clark [11], the systematic use of Extreme Value Theory (EVT) is recent, as the crash of 1987 urged for a better understanding of large losses. For early occurrences of the use of EVT focusing only on the tail distributions, let us cite [12–14].

Developed markets have been investigated as well, mainly through market indices: S&P 500, Dow Jones and Nasdaq [15], German DAX Stocks [16], Australian ASX-ALL [17], Nikkei and Eurostoxx 50 [18], etc.

Stable distributions and processes exhibit generalized Pareto tail distributions as in (1). Yet, choosing this model imposes tail index $\alpha < 2$. This parameter is difficult to estimate, especially when close to 2 [19]. Many critical studies (see *e.g.* Refs. [19,20]) show that the tail index is greater than 2, and around 3 for short terms returns.

To estimate the generalized Pareto distribution, let us consider that the return X at a given time satisfies (1) and that n successive returns $(X_1, \ldots, X_n)$ are independent or at least stationary. It is a crucial and complex problem to estimate α and $\ell(x)$ written in a parametric or semi-parametric form (for example, $\ell(x) = C$ or $\ell(x) = C_1 + C_2 x^{-\beta} + o(x^{-\beta})$), as well as the threshold x_0 above which the previous expressions for $\ell(x)$ are valid. For studying the tail of the distribution of X, we use *order statistics* $(X_{(1)}, \ldots, X_{(n)})$ of $(X_1, \ldots, X_n)$ with $X_{(1)} \leq X_{(2)} \leq \cdots b \leq X_{(n)}$.

A large family of estimators is available but none of them supersedes the others. For our purposes, we adopt the *Hill estimator* defined by

$$H_{k,n} = \frac{1}{k} \sum_{j=1}^{k} \ln X_{(n-j+1)} - \ln X_{(n-k)}$$

as an estimator of $\gamma = 1/\alpha$. There are several ways to interpret this estimator (maximum likelihood, least squares…). The main difficulty for its implementation consists in choosing the optimal index k. We refer to the book [21] and references therein.

2.2. Commonly used models for heavy tails

Many studies have looked for alternatives to power laws modeling. There is a huge literature on this subject [22], and we give here the main lines with a focus on univariate returns. Regarding stochastic processes in continuous time, various authors considered jump diffusion models [23], variance Gamma processes [24] and subordinated processes [11], and SDE whose invariant distributions are fat-tailed [25,26] or present moments explosions [27, Chapter 7].

Chronological series play also a very important role in the development of such financial models for stock prices. A large class of models assumes that the returns are solutions to equations of the form $r_{t+1} = \mu r_t + \sigma_t \varepsilon_t$, where σ_t is itself described by an equation of similar form, leading to General Autoregressive Conditional Heteroskedasticity (GARCH) models which capture clustering effects in volatility. Here, the innovation ε_t is a noise, that could be Gaussian or follow other distributions like Student's t distributions [28]. A quantitative manifestation of volatility clustering is that, while returns themselves are uncorrelated, absolute returns $|r_t|$ or their squares display a positive, significant and slowly decaying autocorrelation function $\mathrm{corr}(|r_t|, |r_{t+\tau}|) > 0$ for τ ranging from a few minutes to several weeks. However, in terms of extreme values, several studies, in different financial markets, concluded that the quantile forecasting performance of power law models is better than that of GARCH type models [29–32]. GARCH models yield volatile quantile forecasts, while power laws lead to more stable ones. A detailed examination of the VaR forecasts from these two classes of models proved that wild swings observed in the GARCH VaR predictions are more an artifact of the GARCH model, rather than the underlying data [33].

Finally, another approach consists in separating the tail of the distribution from its bulk through hybrid models. Several approaches may be found in Ref. [34]. Furthermore, mixtures of models may also lead to heavy tails [35].

3. Framework and methodology

3.1. Log returns and Value-at-Risk

Let $R_t = \ln(S_{t+1}/S_t)$ be the (log-)return, or more simply the return, at time t where S_t is the price of a stock at time t. We call *losses* the values $L_t = -R_t$. The daily $\mathrm{VaR}_t(a)$ at level $a \in (0, 1)$ is defined by

$$1 - a = \mathbb{P}(R_t < \mathrm{VaR}_t(a)) \tag{2}$$

which represents the quantile q_{1-a}^R at level $(1 - a)$ of the returns distribution (or the quantile q_a^L at level a of the losses distribution).

After having fixed a class of parametric models, the practical computation of the VaR consists in calibrating the parameters for the common distribution of R or L (actually only its tail) and computing the quantile q_{1-a}^R or q_a^L.

3.2. Models

We consider four classes of parametric univariate models for the returns or losses:

(I) The **stable distribution for the returns** defined by its characteristic function

$$\phi_X(t) = \begin{cases} \exp\left[i\mu t - \sigma^\alpha |t|^\alpha \left(1 - i\beta \,\mathrm{sign}(t) \tan\left(\frac{\pi\alpha}{2}\right) \right) \right] & \text{if } \alpha \neq 1, \\ \exp\left[i\mu t - \sigma |t| \left(1 + \frac{2}{\pi} i\beta \,\mathrm{sign}(t) \ln|t| \right) \right] & \text{if } \alpha = 1, \end{cases} \tag{3}$$

where $\alpha \in (0, 2)$ is the *tail index*, $\beta \in (-1, 1)$ is the *skewness*, $\sigma \geqslant 0$ is the *scale parameter* and $\mu \in \mathbb{R}$ the *location parameter* (see *e.g.* Refs. [36,37]). Stable distributions arise naturally as universal classes when looking to limit theorems.

(II) The **Pareto distribution for tails of the losses** gives for the distribution function of the returns,

$$\mathbb{P}(L_t \geqslant x) = C x^{-\alpha} \quad \text{for } x \geqslant x_0. \tag{4}$$

This model does not make any supplementary assumption on the bulk of the distribution of the returns.

Except for $\alpha = 2$, where stable distributions are Gaussian, the parameter α of stable laws corresponds to their Pareto tail index, with $\ell(x)$ converging to a constant [36, Theorem 1.12]. Pareto distributions offer a wider variety of fat tails by removing the constraint $\alpha < 2$.

(III) The **GARCH**-*t* **model**. We consider the GARCH (1, 1) model for the time series of financial returns. More precisely, suppose (R_t) satisfies the following model:

$$R_t = \mu + \varepsilon_t = \mu + \sigma_t Z_t, \tag{5}$$

$$\sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2, \tag{6}$$

where (Z_t) is a sequence of i.i.d. Student-*t* innovations. Our choice for Student-*t* innovations is based on many studies showing that Student-*t* innovations give the best GARCH fitting performance [28].

(IV) The **regime switching model**. The VaR forecasting under the assumption of Pareto distribution shows a clustering of under-estimations (respectively over-estimations) at the beginning of each period of large (respectively small) fluctuations (see Fig. 3). The interpretation is that, after a regime switching, we estimate the VaR with data from the other regime. Therefore, we construct a regime switching model that considers a mixture of Pareto distribution. These regime switching models suppose the existence of a state process that is at the origin of the returns. This process is generally chosen as being a Markov chain. Many authors use logistic functions of lagged endogenous variables [38,39], probabilistic functions [40] or Hamilton filter [41] to construct the state chain. These methods are based on an a priori knowledge of the transition probabilities. In our methodology, we are looking to use the a priori knowledge as less as possible. So, we introduce a Hidden Markov Model (HMM) that represents an unsupervised learning procedure from data. We suppose that the return R_t at time t depends on a hidden state X_t, which can take 2 values c and s corresponding to crisis ($X_t = c$, large fluctuations) or steady ($X_t = s$, small fluctuations) periods.

HMM has been first proposed by L.E. Baum and his co-authors [42,43] in the late 60's. These models assume the existence of a non-observed variable that is the source of observations, and try to estimate the hidden variable. This hidden variable X_t is supposed to be a Markov chain. The model is fully characterized by the parameters $M = (\rho, Q, \psi)$ of the HMM:

(i) the initial law ρ of X:

$$\rho(x) = \mathbb{P}(X_0 = x), \quad x \in \{c, s\},$$

(ii) the transition matrix Q of X:

$$Q(x, x') = \mathbb{P}(X_{t+1} = x' | X_t = x), \quad x, x' \in \{c, s\}, \ \forall t \geqslant 0,$$

(iii) the emission kernel of R given X:

$$\psi(x, dy) = \mathbb{P}(R_t \in dy | X_t = x), \quad x \in \{c, s\}, y \in \mathbb{R}, \ \forall t \geqslant 0,$$

where $\psi(x, \cdot)$ is a probability measure on $\mathbb{R}$.

The **model of losses tails** is then given by

$$\mathbb{P}(L_t \geqslant x) = \mathbb{P}(L_t \geqslant x | X_t = c)\mathbb{P}(X_t = c) + \mathbb{P}(L_t \geqslant x | X_t = s)\mathbb{P}(X_t = s)$$

$$= C_c x^{-\alpha_c} \mathbb{P}(X_t = c) + C_s : x^{-\alpha_s} \ (1 - \mathbb{P}(X_t = c)) \tag{7}$$

for $x \geqslant x_0$, where (C_c, α_c) and (C_s, α_s) are respectively the Pareto distribution parameters of crisis and steady losses. The parameters $M = (\rho, Q, \psi)$, $C_c, \alpha_c, C_s, \alpha_s$ and $\mathbb{P}(X_t = c)$ are estimated from data as explained in the next section.

3.3. Parameters estimation and computation of the VaR

We make the hypothesis that the series of stock prices $(S_t)_t$ is stationary over the time (models I and II), $(\sigma_t)_t$ is time homogeneous (model III), or the HMM $(X_t, R_t)_t$ is time homogeneous (model IV). For the first two models, we adopt a moving window approach with a window size of 252 days (one year of data). For instance, the window is placed between the first and 252nd days and a given quantile is forecast for the 253rd day. Next, the window is slid one step forward to forecast quantiles for the 254th, 255th, ..., last days. The motivation behind the moving window technique is to capture dynamic time-varying characteristics of the data in different time periods and to emulate the situation where a small investor wants to forecast the VaR for the next day. The GARCH approach is not based on a moving window but uses all the available data up to the day on which forecasts are generated. This approach is preferable since the detection of volatility clustering requires more data. The last model uses a window composed from 252 crisis data and 252 steady periods data as explained in (IV). We proceed as follows:

(I) **Stable distribution**. For the stable distribution, we use for the returns the McCulloch method [44] as implemented in the R library fBasics on the data $(R_1, \ldots, R_n)$ to estimate the four parameters α, β, σ and μ and to compute the quantile through a direct estimation of the distribution function.

(II) **Pareto distribution**. We use a slight modification of the Hill estimator. Let $(L_{(1)}, \ldots, L_{(n)})$ be the increasing order statistics of $(L_1, \ldots, L_n)$ whose common distribution is assumed to satisfy $\mathbb{P}(L_1 \geqslant x) = Cx^{-\alpha}$ for $x \geqslant x_0$.

Then, if one writes

$$\ln L_{(i)} = -\gamma \ln \left(\frac{n + 1 - i}{n + 1} \right) + K + \varepsilon_i, \tag{8}$$

where $\gamma = 1/\alpha$ and $K = \gamma \ln C$, the noise ε_i is small for n and i large enough. Plotting $\ln L_{(i)}$ as a function of $-\ln((n + 1 - i)/(n + 1))$ gives the *Pareto plot* which should be close to linear for large i. The Hill estimator computes the slope

of this graph using weighted least squares. For more stability, we remove the highest values. We set $d_n = \lfloor 0.95 \times n \rfloor$, $u_n = \lfloor 0.99 \times n \rfloor$ and we use $\hat{\gamma}$ as an estimator of $\gamma = \alpha^{-1}$:

$$\hat{\gamma} = -\frac{\sum\limits_{i=d_n}^{u_n} \ln(L_{(i)}) \cdot \ln\left(\frac{n+1-i}{n+1}\right)}{\sum\limits_{i=d_n}^{u_n} \left(\ln\left(\frac{n+1-i}{n+1}\right)\right)^2}.$$

For better numerical stability, we borrow ideas from I. Weissman [45] to estimate the constant C: for $w \in (0, 1)$ close to 1 (we fix $w = 0.90$), the constant C and the threshold x_0 in (4) are estimated by

$$\hat{C} = L_{\lfloor nw \rfloor}^{\hat{\alpha}}(1 - w) \quad \text{and} \quad \hat{x}_0 = L_{\lfloor nw \rfloor}. \tag{9}$$

The rationale of this approximation is that $L_{\lfloor nw \rfloor} = \hat{q}_w^L$ is an approximation of the quantile q_w^L of $(L_1, \ldots, L_n)$. The quantile of the losses at any level $a \geqslant w$ is then approximated by

$$\hat{q}_a^L = \left(\frac{\hat{C}}{1-a}\right)^{\hat{\gamma}} = L_{\lfloor nw \rfloor} \left(\frac{1-w}{1-a}\right)^{\hat{\gamma}}.$$

(III) **GARCH model**. For returns, we fit a GARCH Student-t model by using the R library `rugarch` over an expanding window to compute the quantile through a direct estimation of the distribution function.

(IV) **Regime switching model**. The procedure of classification contains two steps. In the first step, starting from the sequence of observations R, we determine the parameters $M = (\rho, Q, \psi)$ of the HMM (learning problem). Then, we construct the states' sequence X from the observations sequence and the model's parameters (recognition problem).

The solution of the learning problem is obtained by likelihood maximization. No tractable algorithm is known for solving this problem exactly, but a local maximum likelihood can be derived efficiently using a type of Expectation–maximization algorithms known as Baum–Welch algorithm [43], implemented in the R library `RHmm`. We give the results of this algorithm, with the returns of BNP stock as a typical example of observed sequence:

- The estimated initial law ρ of X is $\rho(c) = 4.97 \cdot 10^{-05}$ and $\rho(s) = 1 - \rho(c)$, meaning that the first data corresponds to a steady period with high probability.

- The estimated transition matrix Q of X is:
$$\begin{pmatrix} Q(c, c) & Q(c, s) \\ Q(s, c) & Q(s, s) \end{pmatrix} = \begin{pmatrix} 0.979 & 0.021 \\ 0.007 & 0.993 \end{pmatrix}.$$
In particular, the probabilities of remaining in crisis or in steady periods are close to 1, which reflects the tendency of the financial market to remain in the same state. In addition, the probability of moving from crisis to steady state is three times higher than from steady to crisis, which reflects the tendency to have shorter crisis periods than steady periods in the market.

- The Baum–Welsh algorithm assumes Gaussian emission distribution, characterized by the conditional mean and variance given the hidden state:
$$\begin{pmatrix} \overline{\psi(c, R)} & \text{Var}(\psi(c, R)) \\ \overline{\psi(s, R)} & \text{Var}(\psi(s, R)) \end{pmatrix} = \begin{pmatrix} -0.12 & 0.18 \\ 0.05 & 0.02 \end{pmatrix} \times 10^{-2},$$
where, $\overline{\cdot}$ and Var denote the mean and variance of a probability distribution. Our choice of the hidden states denomination was driven by the fact that the variance of returns with the crisis state is 3 times higher than the variance of returns with the steady state.

These results are consistent with the qualitative properties of crisis in a financial market. Similar results were found when applying the HMM to other market stocks.

Once the model parameters are determined, we solve the recognition problem by the Viterbi algorithm [46], a dynamic programming algorithm for finding the most likely sequence of hidden states corresponding to a sequence of observed events. In fact, the algorithm uses a global maximum a posteriori estimation of states to maximize the probabilities $\delta_t(i)$ of being in state i at time t knowing the partial sequence $R_1, \ldots, R_t$. Fig. 1 describes the state sequence built for the BNP stock. The crisis periods are in red and the steady periods are in blue.

This information of crisis periods is integrated in the estimation of return distribution as follows.

- We set $t_0 = 1500$ for getting enough data for the HMM estimation. For fixed $t \geqslant t_0$, the VaR at time t is given by q_a^L as follows:
  (a) Estimate the HMM parameters $M^{1:t-1} = \left(\rho^{1:t-1}, Q^{1:t-1}, \psi^{1:t-1}\right)$ from the return sample $(R_1, \ldots, R_{t-1})$.
  (b) Estimate the sequence of hidden states $(X_1, \ldots, X_{t-1})$.
  (c) Estimate the power law parameters (C_c, α_c) from the last 252 returns of crisis periods of the sample $(R_1, \ldots, R_{t-1})$.
  (d) Estimate power law parameters (C_s, α_s) from the last 252 returns of steady periods of the sample $(R_1, \ldots, R_{t-1})$.
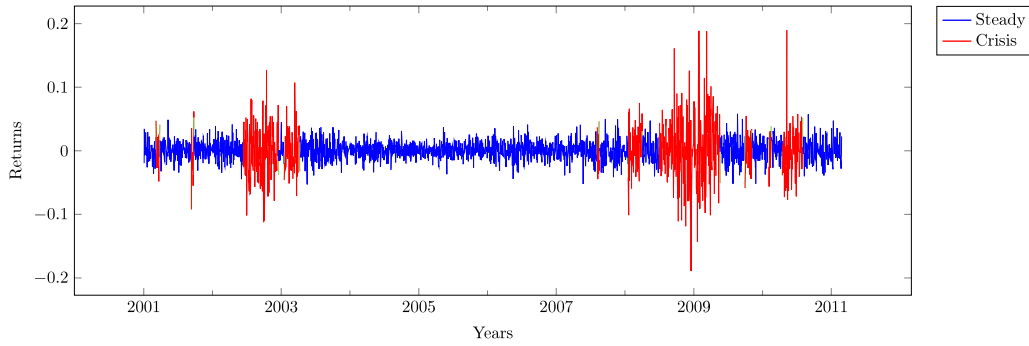
Fig. 1. Classification and detection of crisis and steady periods for the BNP stock. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

(e) With

$$p_t = \mathbb{P}(X_t = c) = \mathbb{P}(X_t = c|X_{t-1} = c)\mathbb{P}(X_{t-1} = c) + \mathbb{P}(X_t = c|X_{t-1} = s)\mathbb{P}(X_{t-1} = s)$$
$$= Q^{0:t-1}(c, c)\,\delta_{t-1}^{0:t-1}(c) + Q^{0:t-1}(s, c)\,\delta_{t-1}^{0:t-1}(s),$$

the quantile of the losses at level a is obtained with

$$1 - a = p_t\, C_c\, x^{-\alpha_c} + (1 - p_t)\, C_s\, x^{-\alpha_s}.$$

This equation is then solved by using Newton–Raphson method.

3.4. Backtesting

The backtesting procedure consists in comparing the successive estimated VaR by moving window approach with the actual returns [47]. A violation occurs when the actual return R_t is smaller than $\text{VaR}_t(a)$ estimated from the previous data. The *violation ratio* is defined as the total number of violations, over the total number of one-period forecasts. When the VaR is estimated at $(1 - a)$th quantile, the expected violation ratio should be $q = 1 - a$. A higher violation ratio implies an under-estimation of the VaR. Conversely, a lower ratio implies an over-estimation of the VaR by the underlying model.

To construct the confidence interval (CI) for the violation ratio, the sum of I_t of one-period forecasts, where I_t takes 1 if there is a violation otherwise 0, follows a binomial distribution as we assume it is the sum of independent random variables. An exact CI at $100 \times (1 - \theta)\%$ is given by

$$\left[\frac{1}{1 + \frac{n-k+1}{k}F_{2(n-k+1),2k}(1 - \theta/2)}, \frac{\frac{k+1}{n-k}F_{2(k+1),2(n-k)}(1 - \theta/2)}{1 + \frac{k+1}{n-k}F_{2(k+1),2(n-k)}(1 - \theta/2)} \right],$$

where k is the success number, n the sample size, and $F_{v_1,v_2}(p)$ the inverse of the quantile at level $p \in [0, 1]$ of the Fisher F-distribution with degree of freedoms v_1 and v_2 [48,49].

In order to test our assumption of independence of the I_t, we use the unconditional coverage (UC) and conditional coverage (CC) tests proposed by Christoffersen [50]. The CC tests jointly independence and correct coverage. It combines the UC test and a test of independence. The likelihood ratio (LR) of the CC test is given by [50]:

$$LR_{cc} = LR_{uc} + LR_{ind}$$
$$= -2\ln\left[q^N(1-q)^{T-N}\right] + 2\ln\left[(1-\pi_{01})^{n_{00}}\pi_{01}^{n_{01}}(1-\pi_{11})^{n_{10}}\pi_{11}^{n_{11}}\right] \sim \chi_{(2)}^2 \qquad (10)$$

where T is the number of observations, N is the number of violations, a is the confidence level, n_{ij} is the number of observations with value i followed by j, and $\pi_{ij} = \mathbb{P}(I_t = j|I_{t-1} = i)$ is the probability value. The first part of (9) is the LR of the unconditional coverage test (LR_{uc}), and the second part of (9) is the LR of the independence test (LR_{ind}). In the case of $n_{11} = 0$, the LR_{cc} will be limited to the first-order Markov likelihood as given in [51]:

$$LR_{cc} = -2\ln\left[q^N(1-q)^{T-N}\right] + 2\ln\left[(1-\pi_{01})^{n_{00}}\pi_{01}^{n_{01}}\right] \sim \chi_{(2)}^2. \qquad (11)$$

In our numerical study (Section 4), we estimate a VaR at 1% confidence level, as in Basel II requirements [3]. VaR forecasting performances of our models are compared by the violation ratio, the exact confidence interval test and the UC and CC Christoffersen tests.
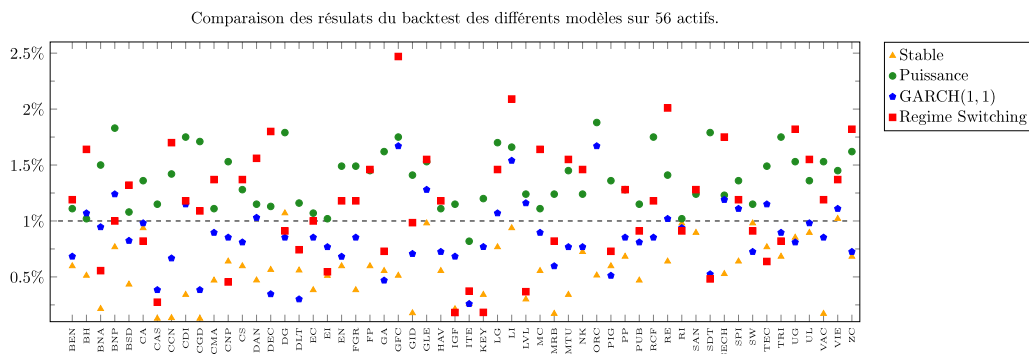
Fig. 2. Violation ratios of 56 Euronext Paris stocks under the four estimation models.

4. Discussion: empirical results and backtesting analysis

4.1. The dataset: stocks from Euronext Paris

The financial instruments considered in the experiments are stocks exchanged on NYSE Euronext Paris. The market data are provided by eSignal (Interactive Data), and, in the following, the stocks are identified by their eSignal symbol. Out of all the stocks listed on the Euronext Paris exchange in February 2011 (more than 600), we selected those having quotations throughout all the period ranging from January 2001 till February 2011 (more than 11 years). This leads us with a subset of 56 stocks including some of the most liquid stocks making up the CAC40 index. The list of these assets can be found in Table 1. In the following experiments, the time series considered are the returns of the end-of-day closing prices of the selected stocks and the parameters of the returns distributions are estimated on a sample as explained in Section 3.3. The choice of a window length set to one year, in models I and II, is a trade-off between the need to have enough data to make good calibration and to include recent crises and the increased risk of departure from the stationarity hypothesis with larger datasets. Regarding the stationarity, it is difficult to draw a clear cut conclusion from Dickey–Fuller and Kwiatkowski–Phillips–Schmidt–Shin tests on unit root and stationarity tests [52,53]. In our experiments, we are not able to identify other sample sizes that would consistently outperform one year with regard to the VaR backtesting or stationarity measures.

4.2. Discussion

In this section, the relative forecasting performances of the regime switching and benchmarked models are illustrated. We apply the methods of estimation of the Value-at-Risk and of backtesting described in Section 3 on the dataset described above. We plot for all stocks their prices, historical volatility computed from one-year data over a moving window, and VaR computed under our four models. Our comparisons are based on a backtesting for a single level of confidence one level of confidence 1% of the VaR. The performance of each model is given by an average of its results for the selected stocks.

Fig. 2 gives the violation ratios for the four models on the 56 selected stocks. Clearly the stable model over-estimates the VaR and the power law model under-estimates the VaR. GARCH and regime switching models results fluctuate around the target value 1%. In term of average of violation ratios, the regime switching model is the closest to 1% with 1.12%, followed by 0.86% for the GARCH model, then the power law model with 1.37% and the stable model with 0.55%. The number of stocks for which the backtest results in [0.90%, 1.10%] is 8 for the regime switching, 8 for the GARCH, 5 for the power law and 7 for the stable model.

Due to the small amount of data available for the backtest, the confidence intervals (CI) are quite large. The average size of the CI is 0.39% for the stable distribution, 0.60% for the power law distribution, 0.98% for the GARCH model and 0.80% for the regime switching model. So it is not possible to give affirmative conclusions for a single stock, but the backtest results over the 56 stocks in Table 1 indicate that the regime switching model gives the best results in terms of backtest. With the regime switching model, 49 out of the 56 exact-CI contain the target value 1%. It is the case for only 46, 43 and 36 out of the 56 with the GARCH, power law and stable models respectively. The unconditional and conditional coverage tests give better results for the GARCH followed by the regime switching model. The results of UC and CC tests are given in Table 1.

We can interpret this result as follows: the tail distribution in the stable law is too fat (since its tail index α is always smaller than 2). This is confirmed by the fact that the estimated tail index for stable distribution is close to 2 for all stocks in our dataset. On the other hand, the tail of the power law distribution estimated on the full dataset is not fat enough to describe Euronext Paris market during crisis periods. The regime switching model computes a combination of tail power law for crisis and steady periods which produces a fatter tail distribution.

Fig. 3 shows the quantile estimations on our typical example BNP Paribas stock. As stable and power law models do not take into consideration the regime switching, they suffer from a delay in the estimation of the quantile. The beginning of periods of small (resp. large) fluctuations is characterized by an over-estimation (resp. under-estimation) of the VaR.

Table 1
Exact CI backtesting (EBT) and unconditional (UC) and conditional coverage (UC) tests over 56 Euronext Paris stocks.

Stock	Stable			Power			GARCH(1, 1)			Regime switch		
	EBT	UC	CC	EBT	UC	CC	EBT	UC	CC	EBT	UC	CC
BEN	★	□	★	★	★	★	★	★	★	★	★	★
BH	★	□	□	★	★	★	★	★	★	★	★	★
BNA	□	□	□	★	□	□	★	★	★	★	★	★
BNP	★	★	★	□	□	□	★	★	★	★	★	★
BSD	□	□	□	★	★	★	★	★	★	★	★	★
CA	★	★	□	★	★	★	★	★	★	★	★	★
CAS	□	□	□	★	★	★	□	□	□	□	□	□
CCN	□	□	□	★	★	★	★	★	★	★	□	★
CDI	□	□	□	□	□	□	★	★	□	★	★	★
CGD	□	□	□	□	□	□	□	□	□	★	★	□
CMA	□	□	□	★	★	★	★	★	★	★	★	□
CNP	★	★	□	★	□	□	★	★	★	★	□	★
CS	★	★	★	★	★	★	★	★	★	★	★	★
DAN	□	□	□	★	★	★	★	★	★	★	★	□
DEC	★	□	★	★	★	★	□	□	□	★	□	□
DG	★	★	★	□	□	□	★	★	★	★	★	★
DLT	★	□	★	★	★	★	□	□	□	★	★	★
EC	□	□	□	★	★	★	★	★	★	★	★	★
EI	★	□	□	★	★	★	★	★	★	★	★	★
EN	★	★	★	★	□	★	★	★	★	★	★	★
FGR	□	□	□	★	□	□	★	★	★	★	★	★
FP	★	★	★	★	□	★	—	—	—	★	★	★
GA	★	□	□	□	□	□	□	□	★	★	★	★
GFC	★	□	□	□	□	□	□	□	□	□	□	□
GID	□	□	□	★	★	□	★	★	★	★	★	★
GLE	★	★	□	★	□	□	★	★	★	★	★	□
HAV	★	□	★	★	★	★	★	★	★	★	★	★
IGF	□	□	□	★	★	★	★	★	★	□	□	□
ITE	□	□	□	★	★	★	□	□	□	★	□	★
KEY	□	□	□	★	★	★	★	★	★	□	□	□
LG	★	★	★	□	□	□	★	★	★	★	★	□
LI	★	★	★	□	□	□	★	★	★	□	□	□
LVL	□	□	□	★	★	□	★	★	★	★	□	★
MC	★	★	★	★	★	★	★	★	★	★	★	★
MRB	□	□	□	★	★	□	★	★	★	★	★	□
MTU	□	□	□	★	□	★	★	★	★	★	★	★
NK	★	★	★	★	★	★	★	★	★	★	★	★
ORC	★	□	□	□	□	□	□	□	□	□	□	□
PIG	★	□	□	★	★	□	★	□	★	★	★	★
PP	★	★	★	★	★	★	★	★	★	★	★	★
PUB	★	□	□	★	★	★	★	★	★	★	★	★
RCF	□	□	□	□	□	□	★	★	★	★	★	★
RE	★	★	★	★	★	★	★	★	★	□	□	□
RI	★	★	★	★	★	★	★	★	★	★	★	★
SAN	★	★	★	★	★	★	—	—	—	★	★	★
SDT	□	□	□	□	□	□	★	★	★	★	★	★
SECH	★	□	□	★	★	★	★	★	★	★	□	□
SPI	★	★	□	★	★	□	★	★	★	★	★	★
SW	★	★	★	★	★	★	★	★	★	★	★	★
TEC	★	★	★	★	□	□	★	★	★	★	★	★
TRI	★	★	★	□	□	□	★	★	★	★	★	★
UG	★	★	★	★	□	□	★	★	★	★	□	□
UL	★	★	★	★	★	★	★	★	★	★	★	★
VAC	□	□	□	★	□	□	★	★	★	★	★	★
VIE	★	★	□	★	□	□	★	★	★	★	★	★
ZC	★	★	★	□	□	□	★	★	★	★	□	□
Total	36	24	23	43	32	30	46	45	46	49	41	39

★ Fail to reject, □ Rejection, — Not converging.

However, by detecting the volatility clustering, the GARCH model takes into consideration the regime switching and overcomes this drawback of under-estimation and over-estimation. Our new HMM approach aims to give a simpler model that combines the advantage of power law and GARCH models. We consider crisis and steady periods and we combine their tail power law distributions. This permits at the beginning of large fluctuations (crisis periods) to reduce the VaR threshold due to the additional term computed from previous crisis periods. Similarly, at the beginning of small fluctuations (steady periods), it permits to raise the VaR threshold by adding a term computed from previous steady periods. The consideration
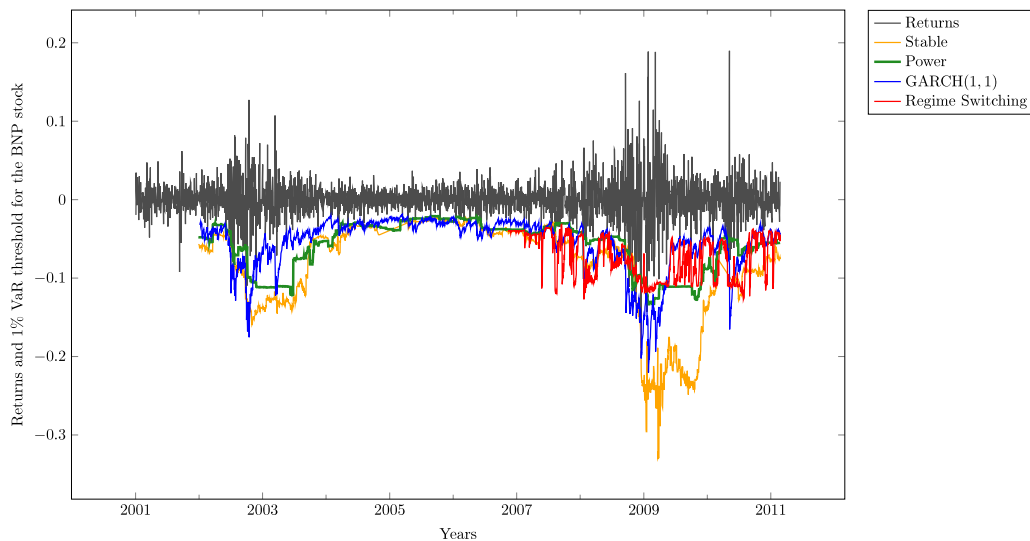
Fig. 3. Comparison of 1% VaR models for the BNP stock.

of two states in our model allows to reduce the delay of regime switching detection in the quantile estimation while keeping the model simple.

5. Conclusions and perspectives

Value-at-Risk (VaR) and conditional volatility models have become common tools for financial forecasting. However, conditional volatility models cannot capture extreme movements, as these models are based on past volatility rather than the extreme observations. On the other hand, extreme value theory models can capture extreme movements and forecasting performance of these models are better than that of conventional volatility models such as GARCH. In this paper, hidden Markov models and EVT are combined to construct a variant of power laws model that forecasts the extreme observations taking into consideration both the data clustering and the regime switching.

The main contribution of this paper is to propose a combined EVT model and compare the predictive performance of this model with conventional models. This hybrid model is tested on real data of 56 stocks exchanged on NYSE Euronext Paris. The relative performance of regime switching model is benchmarked against stable, power laws and GARCH-t models. Our regime switching model gives an average violation ratio on 56 stocks closer to 1% than the other models, and the model is statistically significant for all Christoffersen [50] tail-loss tests for most of the stocks. The regime switching model has the advantage of a simple power law model in the threshold estimation. In addition, it reduces the violation clustering observed in stable and power law models by taking into account the crisis and steady periods. However, it shows a fluctuation over and under the target value 1% in the backtest results characterizing almost all conditional models.

Our results suggest further study by classifying data with a long-memory model of hidden states to have more stability in the VaR forecasting. Co-movements and correlations between stocks in terms of crisis periods can be studied to identify global crisis periods or classes of stocks that have the same behavior. Such clustering would permit to refine the model for each class of stocks and further increase the performance of VaR forecasting.

Acknowledgments

References

[1] P. Jorion, Value at Risk: The New Benchmark for Managing Financial Risk, third ed., McGraw-Hill, 2006.
[2] J. Mina, J.Y. Xiao, Return to RiskMetrics: The Evolution of a Standard RiskMetrics techreport, 2001.
[3] Supervision, Basel Committee on Banking S. Bank of International Settlements International Convergence of Capital Measurement and Capital Standards—A Revised Framework 06 2004.
[4] J. Berkowitz, J. O'Brien, How accurate are value-at-risk models at commercial banks? J. Finance 57 (3) (2002) 1093–1111.
[5] C. Alexander, E. Sheedy, Developing a stress testing framework based on market risk models, J. Bank. Finance 32 (10) (2008) 2220–2236.
[6] F.X. Diebold, T. Schuermann, J.D. Stroughair, Pitfalls and opportunities in the use of extreme value theory in risk management, J. Risk Finance 1 (2) (2000) 30–35.
[7] R. Serfling, Quantile functions for multivariate analysis: approaches and applications, Stat. Neerl. 56 (2) (2002) 214–232.
[8] R. Cont, Empirical properties of asset returns: stylized facts and statistical issues, Quant. Finance 1 (2001) 223–236.
[9] B. Mandelbrot, The variation of certain speculative prices, J. Bus. XXXVI (1963) 392–417.

[10] E.F. Fama, Risk, return, and equilibrium, J. Polit. Econ. 791 (1) (1971) 30–55.
[11] P.K. Clark, A subordinated stochastic process model with finite variance for speculative prices, Econometrica (1973) 135–155.
[12] D.W. Jansen, C.G. de Vries, On the frequency of large stock returns: Putting booms and busts into perspective, Rev. Econ. Stat. (1991) 18–24.
[13] M. Loretan, P. Philips, Testing the covariance stationarity of heavy-tailed time series: An overview of the theory with applications to several financial datasets, J. Empir. Finance 1 (2) (1994) 211–248.
[14] F. Longin, The asymptotic distribution of extreme stock market returns, J. Bus. 69 (3) (1996) 383–408.
[15] Y. Malevergne, V. Pisarenko, D. Sornette, On the power of generalized extreme value (GEV) and generalized Pareto distribution (GPD) estimators for empirical distributions of stock returns, Appl. Financ. Econ. 16 (3) (2006) 271–289.
[16] T. Doganoglu, C. Hartz, S. Mittnik, Portfolio optimization when risk factors are conditionally varying and heavy tailed, Comput. Econ. 29 (3–4) (2007) 333–354.
[17] A. Singh, D. Allen, P. Robert, Extreme market risk and extreme value theory, Math. comput. Simul. 94 (2013) 310–328.
[18] M. Gilli, E. Këllezi, An application of extreme value theory for measuring financial risk, Comput. Econ. 27 (2–3) (2006) 207–228.
[19] W.H. DuMouchel, Estimating the stable index α in order to measure tail thickness: A critique, Ann. Stat. 11 (4) (1983) 1019–1031.
[20] R. Weron, Levy-stable distributions revisited: tail index >2 does not exclude the Levy-stable regime, Internat. J. Modern Phys. C 12 (2) (2001) 209–223.
[21] J. Beirlant, Y. Goegebeur, J. Teugels, J. Segers, Statistics of Extremes. Theory and Applications, in: Wiley Series in Probability and Statistics, John Wiley & Sons Ltd., 2004, With contributions from Daniel De Waal and Chris Ferro.
[22] S. Rachev, Handbook of Heavy Tailed Distributions in Finance, vol. 1, North-Holland, 2003.
[23] R. Cont, P. Tankov, Financial Modelling with Jump Processes, in: Chapman & Hall/CRC Financial Mathematics Series, Chapman & Hall/CRC, Boca Raton, FL, 2004.
[24] R. Marfè, A generalized variance gamma process for financial applications, Quant. Finance 12 (1) (2012) 75–87.
[25] Y. Nagahara, Non-Gaussian distribution for stock returns and related stochastic differential equation, Financ. Eng. Japan. Mark. 3 (2) (1996) 121–149.
[26] W.T. Shaw, M. Schofield, A model of returns for the post-credit-crunch reality: hybrid Brownian motion with price feedback, Quant. Finance (2012) 1–24.
[27] A. Gulisashvili, Analytically Tractable Stochastic Stock Price Models, in: Springer Finance, Springer, Heidelberg, 2012.
[28] J.D. Curto, J.C. Pinto, G.N. Tavares, Modeling stock markets' volatility using GARCH models with Normal, Student's t and stable Paretian distributions, Statist. Papers 50 (2) (2007) 311–321.
[29] R. Gençay, F. Selçuk, A. Ulugülyağci, High volatility, thick tails and extreme value theory in value-at-risk estimation, Insurance Math. Econom. (ISSN: 0167-6687) 33 (2) (2003) 337–356. Papers presented at the 6th IME Conference, Lisbon, 15-17 July 2002.
[30] R. Gençay, F. Selçuk, Extreme value theory and Value-at-Risk: Relative performance in emerging markets, Int. J. Forecast. 20 (2) (2004) 287–303.
[31] A. Cifter, Value-at-risk estimation with wavelet-based extreme value theory: Evidence from emerging markets, Physica A (ISSN: 0378-4371) 390 (12) (2011) 2356–2367.
[32] K. Kuester, S. Mittnik, M.S. Paolella, Value-at-risk prediction: A comparison of alternative strategies, J. Financ. Econ. 4 (1) (2006) 53–89.
[33] J. Danielsson, Y. Morimoto, Forecasting extreme financial risk: a critical analysis of practical methods for the Japanese market, Monet. Econ. Stud. 18 (2) (2000).
[34] C. Scarrott, A. MacDonald, A review of extreme value threshold estimation and uncertainty quantication, REVSTAT 10 (1) (2012) 33–60.
[35] S.A. Broda, M. Haas, J. Krause, M.S. Paolella, S.C. Steude, Stable mixture GARCH models, J. Econometrics 17 (2013) 292–316.
[36] J. Nolan, Stable Distributions: Models for Heavy Tailed Data, Birkhauser, Boston, 2013.
[37] H. Fofack, J.P. Nolan, Tail behavior, modes and other characteristics of stable distributions, Extremes 2 (1) (1999) 39–58.
[38] A. Ang, G. Bekaert, Regime switches in interest rates, J. Bus. Econ. Stat. (ISSN: 07350015) 20 (2) (2002) 163–182.
[39] J.D. Hamilton, Specification testing in Markov-switching time-series models, J. Econometrics (ISSN: 0304-4076) 70 (1) (1996) 127–157.
[40] H. Schaller, S.V. Norden, Regime switching in stock market returns, Appl. Financ. Econ. 7 (2) (1997) 177–191. eprint: http://dx.doi.org/10.1080/096031097333745.
[41] M. Billio, L. Pelizzon, Value-at-Risk: a multivariate switching regime approach, J. Empir. Finance (ISSN: 0927-5398) 7 (5) (2000) 531–554.
[42] B. Baum, P. Petrie, Statistical inference for probabilistic functions of finite state Markov chains, Ann. Math. Statist. (ISSN: 0003-4851) 37 (1966) 1554–1563.
[43] L.E. Baum, T. Petrie, G. Soules, N. Weiss, A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains, Ann. Math. Statist. (ISSN: 0003-4851) 41 (1970) 164–171.
[44] J.H. McCulloch, Measuring tail thickness to estimate the stable index α: a critique, J. Bus. Econom. Statist. 15 (1) (1997) 74–81.
[45] I. Weissman, Estimation of parameters and large quantiles based on the k largest observations, J. Amer. Statist. Assoc. 73 (364) (1978) 812–815.
[46] A. Viterbi, Error bounds for convolutional codes and an asymptotically optimum decoding algorithm, IEEE Trans. Inform. Theory (ISSN: 0018-9448) 13 (2) (1967) 260–269.
[47] S.D. Campbell, A review of backtesting and backtesting procedures techreport, 2007, pp. 1–17.
[48] C.R. Blyth, Approximate binomial confidence limits, J. Amer. Statist. Assoc. 81 (395) (1986) 843–855.
[49] C.R. Blyth, Correction: Approximate binomial confidence limits [J. Amer. Statist. Assoc. 81 (1986), no. 395, 843–855], J. Amer. Statist. Assoc. 84 (406) (1989) 636.
[50] P.F. Christoffersen, Evaluating interval forecasts, Internat. Econom. Rev. (ISSN: 0020-6598) 39 (4) (1998) 841–862. Symposium on Forecasting and Empirical Methods in Macroeconomics and Finance.
[51] P. Christoffersen, D. Pelletier, Backtesting value-at-risk: A duration-based approach, J. Financ. Econ. 2 (1) (2004) 84–108. http://jfec.oxfordjournals.org/content/2/1/84.abstract.
[52] D.A. Dickey, W.A. Fuller, Distribution of the estimators for autoregressive time series with a unit root, J. Amer. Statist. Assoc. 74 (366) (1979) 427–431. part 1.
[53] D. Kwiatkowski, P. Phillips, P. Schmidt, Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? J. Econometrics 54 (1992) 159–178.