Jumps in Rank and Expected Returns: Introducing Varying Cross-sectional Risk*

Gloria González-Rivera
Department of Economics
University of California, Riverside

Tae-Hwy Lee
Department of Economics
University of California, Riverside

Santosh Mishra
Department of Economics
Oregon State University

Abstract

We propose an extension of the meaning of volatility by introducing a measure, namely the **V**arying **C**ross-sectional **R**isk (VCR), that is based on a ranking of returns. VCR is defined as the conditional probability of a sharp jump in the position of an asset return within the cross-sectional return distribution of the assets that constitute the market, which is represented by the Standard and Poor's 500 Index (SP500). We model the *joint* dynamics of the cross-sectional position and the asset return by analyzing (1) the *marginal* probability distribution of a sharp jump in the cross-sectional position within the context of a duration model, and (2) the probability distribution of the asset return *conditional* on a jump, for which we specify different dynamics in returns depending upon whether or not a jump has taken place. As a result, the marginal probability distribution of returns is a mixture of distributions. The performance of our model is assessed in an out-of-sample exercise. We design a set of trading rules that are evaluated according to their profitability and riskiness. A trading rule based on our VCR model is dominant providing superior mean trading returns and accurate Value-at-Risk forecasts.

Key words: Cross-sectional position, Duration, Leptokurtosis, Momentum, Mixture of normals, Nonlinearity, Risk, Trading rule, VaR, VCR.

JEL Classification: C3, C5, G0.

*Corresponding author: Gloria González-Rivera, Department of Economics, University of California, Riverside, CA 92521-0427, email gloria.gonzalez@ucr.edu, tel. (951) 827-1470, fax (951) 827-5685. We thank Peter Chung, Valentina Corradi, René García, Oscar Jordà, Canlin Li, David Mayers, Vassilis Polimenis, Eric Renault, and seminar participants at Krannert School of Management of Purdue University, Anderson Graduate School of Management of UCR, the 2003 Econometric Society European Meeting (Stockholm), and the 2003 NBER/NSF Time Series Conference (Chicago).

1 Introduction

Economists, investors, regulators, decision makers at large face uncertainty in a daily basis. While most of us share an intuitive notion on the meaning of uncertainty, which involves future events and their probability of occurrence, the measurement of uncertainty depends on who you are and what you do. For instance, financial economists and econometricians for most part equate uncertainty with risk, risk with volatility, and volatility with variance (or any other measure of dispersion). Investors and regulators are not only concerned with measures of volatility but also they monitor the tails of the probability distribution of returns. Regulators worry about catastrophic or large losses that can jeopardize the health of the financial system of the economy. Decision theorists deal with volatility measures only in particular instances since the variance may not be a sufficient statistic to summarize the risk faced by an agent, and often they claim that an ordinal measure of risk, i.e. the construction of a ranking of assets, may be sufficient for an agent to make a rational choice under uncertainty (Granger, 2002).

In this paper we put forward a different measure of volatility that combines the views of the variance advocates with those of the ranking advocates. For financial assets, our conception of risk is based on an ordinal measure (the cross-sectional ranking of an asset in relation to its peers) and on a cardinal measure (the univariate conditional volatility of the asset). A graphical introduction to our main idea is contained in Figure 1, where we present a stylized description of the problem that we aim to analyze. Let $y_{i,t}$ be the asset excess return of firm i at time t, and $z_{i,t}$ the cross-sectional position (percentile ranking) of this return in relation to all other assets that constitute the market. For every t, we observe the realizations of all asset returns, and for each asset, we assign a cross sectional position $z_{i,t}$. For one period to the next, the cross-sectional position changes. In Figure 1, we draw a histogram to represent the ranking of realized asset returns, which is time-varying. Our objective is to model the dynamics of the cross-sectional position jointly with the dynamics of the asset return. To illustrate the different dynamics of $y_{i,t}$ and $z_{i,t}$, we choose four points in time. Consider the movements of $y_{i,t}$ and $z_{i,t}$ going from t_1 to t_4. We observe that from t_1 to t_2, the market overall has gone down as well as the return and the cross-sectional position of asset i, $y_{i,t_1} > y_{i,t_2}$ and $z_{i,t_1} > z_{i,t_2}$. However, from t_2 to t_3, the asset return has decreased $y_{i,t_2} > y_{i,t_3}$ but its cross-sectional position has improved $z_{i,t_2} < z_{i,t_3}$. In relation to its peers, asset i is a better performer. The opposite happens on going from t_3 to t_4. The overall market is going up; for asset i, the return increases $y_{i,t_3} < y_{i,t_4}$ but its cross-sectional position is unchanged $z_{i,t_3} = z_{i,t_4}$. In this case, the asset, even though is doing well, comparatively speaking it could have done better.

We are interested in this notion of relative risk. Note that the time series $y_{i,t}$ conveys univariate information about asset i but the time series $z_{i,t}$ implicitly conveys information about the full market and, in this sense, it is a multivariate measure.

We set our problem as to model the joint distribution of the return and the probability of a (sharp) jump ($J_{i,t}$) in the cross-sectional position of the asset $z_{i,t}$, i.e. $f(y_{i,t}, J_{i,t}|\mathcal{F}_{t-1})$ where $\mathcal{F}_{t-1}$ is an information set up to time $t-1$. Since $f(y_{i,t}, J_{i,t}|\mathcal{F}_{t-1}) = f_1(J_{i,t}|\mathcal{F}_{t-1})f_2(y_{i,t}|J_{i,t}, \mathcal{F}_{t-1})$, our task will be accomplished by modelling the marginal distribution of the jump and the conditional distribution of the return. The former provides the conditional probability of jumping cross-sectional positions, which we call the time-**V**arying **C**ross-sectional **R**isk (VCR). It is straightforward to understand that VCR is time-varying because it depends on an information set that changes over time, and that it conveys cross-sectional information because it depends on the position of the asset in relation to its peers. We say that is risk because the probability of a sharp jump in cross-sectional positions is an assessment of the chances of being a winner or a loser within the available set of assets in the market.

On modelling $f_1(J_{i,t}|\mathcal{F}_{t-1})$, our paper also connects with the recent literature in microstructure of financial markets and duration analysis (Engle and Russell, 1998). This line of research aims to model events (e.g., trades) and waiting times between events. The traditional question in duration analysis is what is the expected length of time between two events given some information set. In this paper, the event is the jump in the cross-sectional position of the asset return, however, when we model the expected duration between jumps or its mirror image –the conditional probability of the jump–, our analysis is performed in calendar time as in Hamilton and Jordà (2002). Given some information set, the question of interest is what is the likelihood that tomorrow there will be a sharp change in the position of this firm in relation to the cross-sectional distribution of returns. This calendar time approach is necessary because asset returns are reported in calendar time (days, weeks, etc.) and it has the advantage of incorporating other information available in calendar time.

Furthermore, the modelling of $f_1(J_{i,t}|\mathcal{F}_{t-1})$ also connects with the financial literature on "momentum". At time t, a typical momentum strategy consists of buying stocks that have performed well and selling those that have performed poorly over the previous 3 or 12 months, and holding this portfolio over the next 3 to 12 month periods. Jegadeesh and Titman (1993), among many others, have documented that this strategy generates positive returns over the holding period. From a statistical point of view, this strategy implies that there is a continuation of patterns, i.e. those stocks that have performed well (poorly) in the recent past keep performing well (poorly) in the near future, that is winners (losers) continue to be winners (losers). It is natural to think that this type of persistence or time patterns must be a target for statistical modeling and may be exploited from a forecasting point of view. If we model the conditional probability of a sharp jump in cross sectional position and calculate its inverse, we are assessing the conditional expected duration of a high (low) position in the cross-sectional ranking of assets.

The ultimate justification of any time series model is its forecasting ability. The model that we propose is highly nonlinear and its performance is assessed in an out-of-sample exercise within the context of investment decision making. We consider two criteria. In the first, we deal with an investor whose interest is to maximize profits of a portfolio long in stocks. The second criterion is to consider an investor who worries about potential large losses and wishes to add a Value-at-Risk evaluation to her trading strategy. Profitability and riskiness are the two coordinates in the mind of the investor. We design a set of trading rules that will be compared in the two aforementioned criteria. The statistical comparison is performed within the framework of White's (2000) reality check, which controls for potential data snooping biases. A trading rule that exploits the one-step ahead forecast of the cross-sectional positions of asset returns will be shown to be clearly superior to other rules based on more standard models.

The organization of the paper is as follows. In section 2, we provide our strategy for the joint modelling of asset returns and jumps in the cross-sectional position. We present the estimation results for a sample of weekly returns of those SP500 firms that have survived for the last ten years. In section 3, we assess the out-of-sample performance of our model. We explain the trading rules, loss functions, and the statistical framework to compare different trading rules. Finally, in section 4 we conclude.

2 Cross-sectional position and expected returns

In this section, we propose a bivariate model of *expected returns* and *jumps* in the ranking of a given asset within the cross-sectional distribution of asset returns.

Let $y_{i,t}$ be the return of the i^{th} firm at time t, and $\{y_{i,t}\}_{i=1}^{M}$ be the collection of asset returns of the M firms that constitute the *market* at time t. For every time period, we order the asset returns from the smallest to the largest, and we define $z_{i,t}$, the *cross-sectional position* (or *percentile rank*) of the i^{th} firm within the market, as the percentage of firms that have a return less than or equal to the return of the i^{th} firm. We write

$$z_{i,t} \equiv M^{-1} \sum_{j=1}^{M} \mathbf{1}(y_{j,t} \leq y_{i,t}), \tag{1}$$

where $\mathbf{1}(\cdot)$ is the indicator function, and for M large, $z_{i,t} \in (0, 1]$ In Figure 1, $z_{i,t}$ is the shaded area of the cross-sectional histogram of returns. We say that a (sharp) *jump* in the cross-sectional position of the i^{th} firm has occurred when there is a minimum (upwards or downwards) movement of 0.5 in the cross-sectional position of the return of the i^{th} firm. We define such a jump as a binary variable

$$J_{i,t} \equiv \mathbf{1}(|z_{i,t} - z_{i,t-1}| \geq 0.5). \tag{2}$$

The choice of the magnitude of the jump is not arbitrary. The sharpest jump that we could consider is 0.5. In every time period, we need to allow for the possibility of jumping, either up or down, in the following period regardless of the present position of the asset. For instance, if we choose a jump greater than 0.5, say 0.7, and $z_{i,t} = 0.4$, then the probability of jumping up or down in the next time period is zero. Note that the defined jump does not imply that the return will be above or below the median. As an example, if $z_{i,t-1} = 0.4$ and $z_{i,t} = 0.6$, then $J_{i,t} = 0$ but the return at time t will be above the cross-sectional median of returns. However, if $J_{i,t} = 1$, then the asset return has moved either above or below the median. Note that an upward (downward) jump implies neither a higher (lower) return, nor a larger (smaller) variance. This is so because the cross-sectional position is the result of the interaction of the relative movements of all individual assets in the market. Moreover our definition of jump is different from the jump process that can be defined in the univariate return process (e.g., Aït-Sahalia 2003). Our jump is a jump in cross-sectional positions, which implicitly depends on numerous univariate return processes.

Our objective is to model the conditional joint probability density function of returns and jumps $f(y_{i,t}, J_{i,t}|\mathcal{F}_{t-1}; \theta)$, where $\mathcal{F}_{t-1}$ is the information set up to time $t-1$, which contains past realizations of returns, jumps, and cross-sectional positions. To simplify notation, we drop the subindex i but in the following analysis should be understood that the proposed modelling is performed for every single firm in the market. We factor the joint probability density function as the product of the marginal density of the jump and the conditional density of the return

$$f(y_t, J_t|\mathcal{F}_{t-1}; \theta) = f_1(J_t|\mathcal{F}_{t-1}; \theta_1) f_2(y_t|J_t, \mathcal{F}_{t-1}; \theta_2),$$

where $\theta = (\theta_1' \ \theta_2')'$. For a sample $\{y_t, J_t\}_{t=1}^T$, the joint log-likelihood function is

$$\sum_{t=1}^{T} \log f(y_t, J_t | \mathcal{F}_{t-1}; \theta) = \sum_{t=1}^{T} \log f_1(J_t | \mathcal{F}_{t-1}; \theta_1) + \sum_{t=1}^{T} \log f_2(y_t | J_t, \mathcal{F}_{t-1}; \theta_2).$$

Let us call $L_1(\theta_1) = \sum_{t=1}^{T} \log f_1(J_t|\mathcal{F}_{t-1}; \theta_1)$ and $L_2(\theta_2) = \sum_{t=1}^{T} \log f_2(y_t|J_t, \mathcal{F}_{t-1}; \theta_2)$. The maximization of the joint log-likelihood function can be achieved by maximizing $L_1(\theta_1)$ and $L_2(\theta_2)$ separately without loss of efficiency if the parameter vectors θ_1 and θ_2 are "variation free" in the sense of Engle *et al.* (1983) as is the case in our set-up below.

2.1 Modelling the cross-sectional jump $f_1(J_t|\mathcal{F}_{t-1}; \theta_1)$

In order to model the conditional probability of jumping, we define a counting process $N(t)$ as the cumulative number of jumps up to time t, that is, $N(t) = \sum_{n=1}^{t} J_n$. This is a non-decreasing step function that is discontinuous to the right and to the left and for which $N(0) = 0$. Associated with this counting process, we define a duration variable $D_{N(t)}$ as the number of periods between two jumps. Note that because our interest is to model the jump jointly with returns and these are recorded in a calendar basis (daily, weekly, monthly, etc.), the duration variable needs to be defined in calendar time instead of event time as is customary in duration models. The question of interest is, what is the probability of a sharp jump at time t in the cross-sectional position of the i^{th} firm asset return given all available information up to time $t-1$? This is the conditional hazard rate p_t

$$p_t \equiv \Pr(J_t = 1|\mathcal{F}_{t-1}) = \Pr(N(t) > N(t-1)|\mathcal{F}_{t-1}), \tag{3}$$

which we call the time-*varying cross-sectional risk* (VCR). From (3), we note that VCR is time-varying because it depends on the information set $\mathcal{F}_{t-1}$, and it is cross-sectional because J_t depends on the positioning of the asset return in relation to the other firms in the market. Furthermore, because $J_t = 1$, VCR assesses the possibility of being in the upper tail (winner) or in the lower tail (loser) of the cross-sectional distribution of asset returns. In this sense, we called this measure "risk".

It is easy to see that the probability of jumping and duration must have an inverse relationship. If the probability of jumping is high, the expected duration must be short, and vice versa. Let $\Psi_{N(t)}$ be the expected duration. The expected duration until the next jump in the cross-sectional position is given by $\Psi_{N(t-1)} = \sum_{j=1}^{\infty} j(1-p_t)^{j-1} p_t = p_t^{-1}$. Consequently, to model (3), it suffices to model the expected duration and compute its inverse. Following Hamilton and Jordà (2002), we specify an autoregressive conditional hazard (ACH) model.[1] A general ACH model is specified as

$$\Psi_{N(t)} = \sum_{j=1}^{m} \alpha_j D_{N(t)-j} + \sum_{j=1}^{r} \beta_j \Psi_{N(t)-j}. \tag{4}$$

Since p_t is a probability, it must be bounded between zero and one. This implies that the conditional duration must have a lower bound of one. Furthermore, working in calendar time has the advantage that we can incorporate information that becomes available between jumps and can affect the probability of jumping in future periods. We can write a general conditional hazard rate as

$$p_t = [\Psi_{N(t-1)} + \delta' X_{t-1}]^{-1}, \tag{5}$$

where X_{t-1} is a vector of relevant calendar time variables such as past cross-sectional positions and past returns.

The log-likelihood function $L_1(\theta_1) = \sum_{t=1}^{T} \log f_1(J_t|\mathcal{F}_{t-1};\theta_1)$ corresponding to a sample of observed jumps in the cross-sectional position is

$$L_1(\theta_1) = \sum_{t=1}^{T} \left[J_t \log p_t(\theta_1) + (1 - J_t)\log(1 - p_t(\theta_1))\right], \tag{6}$$

where $\theta_1 = (\alpha', \beta', \delta')'$ is the parameter vector for which the log-likelihood function is maximized.

2.2 Modelling the return conditional on jumps $f_2(y_t|J_t, \mathcal{F}_{t-1}; \theta_2)$

We assume that the return to the i^{th} firm asset may behave differently depending upon the occurrence of a sharp jump. If a sharp jump has occurred, the return was pushed either towards the lower tail or upper tail of the cross-sectional distribution of returns. In relation to the market, this asset becomes either a loser or a winner. *A priori,* one may expect different dynamics in these two states. A general specification is

$$f_2(y_t|J_t, \mathcal{F}_{t-1}; \theta_2) = \begin{cases} N(\mu_{1,t},\ \sigma^2_{1,t}) & \text{if } J_t = 1, \\ N(\mu_{0,t},\ \sigma^2_{0,t}) & \text{if } J_t = 0, \end{cases} \tag{7}$$

where μ_t is the conditional mean and σ^2_t is the conditional variance, potentially different depending upon the existence of a jump. The information set consists of past returns and cross-sectional positions, i.e., $\mathcal{F}_{t-1} = \{y_{t-1}, y_{t-2}, \ldots, z_{t-1}, z_{t-2}, \ldots\}$.

We should note that knowing whether the jump has occurred does not imply that the return has to come from a skewed probability density function. One may be tempted to think that a left

[1] The ACH model is a calendar-time version of the autoregressive conditional duration (ACD) of Engle and Russell (1998).

skewed (right skewed) distribution of returns would be appropriate when a downward (upward) jump has taken place. This is not the case. Remember that whether a jump happens is a function of the state of the market. One may think of a bullish market where a downward jump in the cross-sectional position of a particular asset could be associated with a high return (in the right tail of the return distribution), and on the contrary, in a bearish market an upward jump may be associated with a low return (in the left tail of the return distribution). Consequently, our assumption of symmetry (normality) is justified whether the jump is upward or downward. However, more complex distributions can be assumed.

The log-likelihood function $L_2(\theta_2) = \sum_{t=1}^{T} \log f_2(y_t|J_t, \mathcal{F}_{t-1}; \theta_2)$ is

$$L_2(\theta_2) = \sum_{t=1}^{T} \log \frac{1}{\sqrt{2\pi}} \left[\frac{J_t}{\sigma_{1,t}} \exp \frac{-(y_t - \mu_{1,t})^2}{\sigma_{1,t}^2} + \frac{1 - J_t}{\sigma_{0,t}} \exp \frac{-(y_t - \mu_{0,t})^2}{\sigma_{0,t}^2} \right],$$

where θ_2 includes all parameters in the conditional means and conditional variances.

2.3 Modelling the expected return

The marginal density function of returns is a mixture of normal density functions where the weights in the mixture are given by the probability of jumping p_t, the VCR measure:

$$\begin{aligned} g(y_t|\mathcal{F}_{t-1}) &\equiv \sum_{J_t=0}^{1} f(y_t, J_t|\mathcal{F}_{t-1}; \theta) \\ &= \sum_{J_t=0}^{1} f_1(J_t|\mathcal{F}_{t-1}; \theta_1) f_2(y_t|J_t, \mathcal{F}_{t-1}; \theta_2) \\ &= p_t\ f_2(y_t|J_t = 1, \mathcal{F}_{t-1}; \theta_2) + (1 - p_t)\ f_2(y_t|J_t = 0, \mathcal{F}_{t-1}; \theta_2) \\ &= \begin{cases} N(\mu_{1,t},\ \sigma_{1,t}^2) & \text{with probability } p_t, \\ N(\mu_{0,t},\ \sigma_{0,t}^2) & \text{with probability } (1 - p_t). \end{cases} \end{aligned} \tag{8}$$

According to (8), the one-step ahead forecast of the return is

$$E(y_{t+1}|\mathcal{F}_t) = \int y_{t+1} \cdot g(y_{t+1}|\mathcal{F}_t) dy_{t+1} = p_{t+1}\ \mu_{1,t+1} + (1 - p_{t+1})\ \mu_{0,t+1}. \tag{9}$$

The expected return is a function of the VCR, which is a nonlinear function of the information set as shown in (5). Hence the expected returns are nonlinear functions of the information set, even in a simple case where $\mu_{1,t}$ and $\mu_{0,t}$ are linear.

2.4 Estimation Results

We collect the weekly returns, from January 1, 1990 to August 31, 2001, for all the firms in the SP500 index that have survived for the last ten years (343 firms). The firms that have not survived during this period are usually very volatile firms and tend to reside in the tails of the cross-sectional distribution of asset returns, thus they will not affect substantially the cross-sectional positions of the surviving firms. The total number of observations is 599. In Table 1, we summarize the unconditional moments (mean, standard deviation, coefficient of skewness, and coefficient of kurtosis) for the 343 firms. The frequency distribution of the unconditional mean seems to be bimodal with two well defined groups of firms, a cluster with a negative mean return of approximately -0.25%, and another with a positive mean return of 0.25%. For the unconditional standard deviation, the median value is 4.37. The median coefficient of skewness is 0.01, with most of the firms exhibiting moderate to low asymmetry. All the firms have a large coefficient of kurtosis with a median value is 5.23. We calculate the Box-Pierce statistic to test for up to fourth order autocorrelation in returns and we find mild autocorrelation for about one-third of the firms. However, the Box-Pierce test for up to fourth order autocorrelation in squared returns indicates strong dependence in second moments for all the firms in the SP500 index.

[Table 1 about here]

2.4.1 The duration model

For 343 firms, we fit a conditional duration model as in (4) and (5). The information set consists of past durations, past returns and past cross-sectional positions : $\{D_{N(t)-j}, y_{t-j}, z_{t-j}\}$ for $j = 1, 2, \ldots$. The duration time series for every firm is characterized by clustering – long (short) durations are followed by long (short) durations – and, consequently the specification of an ACH model may be warranted. We maximize the log-likelihood function (6) with respect to the parameter vector $\theta_1 = (\alpha', \beta', \delta')'$. We estimate different lag structures (linear and nonlinear) of the information set and, based on standard model selection criteria (t-statistics and log-likelihood ratio tests), we obtain the following final specification

$$
\begin{aligned}
p_t &= [\Psi_{N(t-1)} + \delta^{'} X_{t-1}]^{-1}, \\
\Psi_{N(t)} &= \alpha D_{N(t)-1} + \beta \Psi_{N(t)-1}, \\
\delta^{'} X_{t-1} &= \delta_1 + \delta_2 y_{t-1} \mathbf{1}(z_{t-1} \leq 0.5) + \delta_3 y_{t-1} \mathbf{1}(z_{t-1} > 0.5) + \delta_4 z_{t-1}.
\end{aligned}
$$

The conditional duration model is an ACH(1,1) with persistence $\alpha + \beta$. There is a nonlinear effect of the predetermined variables on duration. The effect of past returns on duration depends on whether the cross-sectional position of the asset is below or above the median. In Table 2, we report the cross-sectional frequency distributions of the estimates $\hat{\theta}_1 = (\hat{\alpha}, \hat{\beta}, \hat{\delta}^{'})'$ for all the 343 firms. All the parameters are statistically significant at the customary 5% level with the exception of δ_4, for which do not report its frequency distribution.

[Table 2 about here]

For $\hat{\alpha}$, the median is 0.36 with 90% of the firms having an $\hat{\alpha}$ below 0.48. For $\hat{\beta}$, its frequency distribution is highly skewed to the right with a median of 0.06 and with 90% of the firms having a $\hat{\beta}$ below 0.25. The median persistence is 0.45 and for 90% of the firms, the persistence is below 0.63. The estimates $\hat{\delta}_2$ and $\hat{\delta}_3$ have opposite signs, the former is positive and the latter is negative with $\hat{\delta}_2$ being larger in magnitude than $\hat{\delta}_3$. The effect of $\hat{\delta}_2$ and $\hat{\delta}_3$ in expected duration depends on the interaction between the cross-sectional position and the sign of the return. There are four possible scenarios. For instance, when the past asset return is positive and below (above) the median market return, its expected duration is longer (shorter) and the probability of a sharp jump is smaller (larger), other things equal. On the contrary, when the past asset return is negative and below (above) the median market return, its expected duration is shorter (longer) and the probability of a sharp jump is larger (smaller), other things equal. Both $\hat{\delta}_2$ and $\hat{\delta}_3$ have a very skewed cross-sectional frequency distributions. For $\hat{\delta}_2$, the median value is 0.15 with 90% of the firms having a $\hat{\delta}_2$ below 0.55. For $\hat{\delta}_3$, the median value is -0.11 with 90% of the firms having a $\hat{\delta}_3$ above -0.45. Roughly speaking, for a representative firm with median parameter estimates, the expected duration is between 4 and 5 weeks, and since $E(p_t) \geq [E(\Psi_{N(t-1)} + \delta' X_{t-1})]^{-1}$ by Jensen's inequality, a lower bound for the expected VCR is between 0.20 and 0.25.

In the last section of Table 2, we also report the median estimates of the parameters of the duration model for the industrial sectors that are represented in the SP500 index. There are ten sectors in the index, which have been reduced to eight. [2] The largest shares correspond to the Consumer Goods sector with 28% of the SP500 companies, and the Information Technology sector with 21% of the firms. The smallest share corresponds to the Energy sector with 5% of the firms. The column denoted as $\hat{\alpha} + \hat{\beta}$ provides the median persistence in duration for each sector. A high persistence implies longer expected durations, thus a lower probability of a sharp jump. The Information Technology sector with $\hat{\alpha} + \hat{\beta} = 0.40$ has the lowest persistence and the Utilities sector with $\hat{\alpha} + \hat{\beta} = 0.55$ has the highest persistence. The former is relatively riskier than the former according to the VCR measure. In the columns labeled $\hat{\delta}_2$ and $\hat{\delta}_3$, we report the median impact

[2] Consumer Discretionary Product and Consumer Staple Product are combined to form one category, namely Consumer Goods. Information Technology and Telecommunication Services are merged into one group, namely Information Technology.

of the calendar variables on the probability of a jump. *Ceteris paribus,* a low absolute value of the estimate denotes a lower tendency of the asset return to move towards the median of the market. As the volatile stocks are usually found in the tails of the cross-sectional distribution of the market, a low absolute value of the estimate implies a tendency to remain in the lower or upper tail of the cross-sectional distribution. We find that the Information Technology sector has the smallest absolute values of $\hat{\delta}_2$ and $\hat{\delta}_3$ and the Utilities sector the largest, indicating that the latter is a median performer compared to the former which has a tendency to move towards the tails of the cross-sectional distribution of returns. The joint effect of duration persistence and impact of the calendar variables is summarized in the column labeled $\hat{p}_t$, which is the median probability of a sharp jump in each sector. Not surprisingly, the Information Technology sector has the largest median probability with $\hat{p}_t = 0.30$, which means that about every three weeks these stocks jump from the top (bottom) to the bottom (top) of the cross-sectional distribution of the market. On the other side of the spectrum, we have the Utilities sector with the smallest median probability $\hat{p}_t = 0.11$, which implies jumps every nine weeks approximately.

2.4.2 The nonlinear model for conditional returns

We proceed to estimate (7). Since this model is already nonlinear, we restrict the specification of the conditional mean and conditional variance in each state ($J_t = 1$ or $J_t = 0$) to parsimonious linear functions of the information set. After customary specification tests, the preferred specification of the conditional first two moments is

$$\begin{aligned} \mu_{1,t} &\equiv E(y_t|\mathcal{F}_{t-1}, J_t = 1) = \nu_1 + \gamma_1 y_{t-1} + \eta_1 z_{t-1}, \\ \mu_{0,t} &\equiv E(y_t|\mathcal{F}_{t-1}, J_t = 0) = \nu_0 + \gamma_0 y_{t-1} + \eta_0 z_{t-1}, \\ \sigma^2_{1,t} &= \sigma^2_{0,t} = \sigma^2_t = E(\varepsilon^2_t|\mathcal{F}_{t-1}, J_t) = \omega + \rho\varepsilon^2_{t-1} + \tau\sigma^2_{t-1}, \end{aligned} \tag{10}$$

where $\varepsilon_{t-1} \equiv (y_{t-1} - \mu_{1,t-1})J_{t-1} + (y_{t-1} - \mu_{0,t-1})(1 - J_{t-1})$. We arrive to this specification by implementing a battery of likelihood ratio tests for the two null hypotheses:

$$\mathbb{H}^1 : \mu_{1,t} = \mu_{0,t} \quad \text{and} \quad \mathbb{H}^2 : \sigma^2_{1,t} = \sigma^2_{0,t}.$$

The first null hypothesis $\mathbb{H}^1$ is rejected very strongly for all the firms in the SP500 index indicating that there are different dynamics in the conditional mean across states. For the second null hypothesis $\mathbb{H}^2$, we fail to reject.[3] So, the marginal density of the return is a location-mixture of

[3] We do not report all the testing results for the 343 firms but they are available upon request.

normals. Within our model (10), the well known unconditional leptokurtosis of asset returns is explained by the mixture of distributions.

The estimation results for the 343 firms are summarized in Table 3. We report the cross-sectional frequency distributions of the parameters estimates of the conditional mean and conditional variance. The majority of the parameters are statistically significant at the 5% level.

[Table 3 about here]

When we consider asset returns for which a sharp jump has taken place, the impact of past returns, $\hat{\gamma}_1$, is predominantly negative (in 75% of the firms $\hat{\gamma}_1 < 0$). The effect of past cross-sectional positions, $\hat{\eta}_1$, is clearly negative for all the firms. These signs are expected. Consider an asset for which a sharp jump has taken place ($J_t = 1$), and whose past return has been going down ($\Delta y_{t-1} < 0$). A movement down in past returns implies that the likelihood of a sharp jump increases. Given that we are considering an asset for which a sharp jump is happening, we should expect that the most likely direction of the jump is up, thus increasing the present expected return. The same type or argument goes through when we consider movements in the cross-sectional position. When there is no jump, the marginal effect of past returns, $\hat{\gamma}_0$, on expected returns could be positive or negative across the firms with a median value of 0.15. On the contrary, the marginal effect of past cross-sectional positions, $\hat{\eta}_0$, is clearly positive. This means that asset returns who move up in the cross-sectional ranking of firms, but who have not experienced a sharp jump, tend to have an increase in their expected returns, other things equal. For individual firms we observe that $|\hat{\gamma}_1| > |\hat{\gamma}_0|$ and $|\hat{\eta}_1| > \hat{\eta}_0$, which is consistent with the notion that, most of the time, sharp jumps in cross-sectional positions must be associated with large movements in expected returns. The median value of $\hat{\gamma}_1$ is -1.11 compared to the median of $\hat{\gamma}_0$ that is 0.15; and the median of $\hat{\eta}_1$ is -0.61 compared to the median of $\hat{\eta}_0$ that is 0.37. We also report the median estimates for every industrial sector represented in the SP500 index. The most salient feature is the behavior of the Utilities sector. The median estimate of $\hat{\gamma}_1(\hat{\gamma}_0)$ is very small (large) compared to that of other sectors meaning that jumps in the Utilities sector are not very frequent corroborating the results of the duration model. The median estimates of $\hat{\eta}_1$ and $\hat{\eta}_0$ are approximately of the same magnitude and sign for all the sectors although the estimates for the Utilities sector are the most extreme ones.

In Table 3, we also report the estimates of the parameters of the conditional variance. The model is a standard GARCH(1,1). The persistence is measured by $\hat{\rho} + \hat{\tau}$. The median persistence is 0.93, with a median value for $\hat{\rho}$ of 0.06 and a median value for $\hat{\tau}$ of 0.87. A leverage effect in the conditional variance does seem to be warranted, the different specifications of the conditional mean across states take care of potential asymmetries in returns. We run standard diagnostic checks such as the Box-Pierce statistics for autocorrelation in residuals, squared residuals, and standardized squared residuals and we conclude that the residuals, standardized residuals, and standardized squared residuals seem to be white noise. The specification (10) passes standard diagnostic checks for model adequacy. However, given the nonlinearity of the model, a more drastic check on the validity of the model is the assessment of its forecasting performance, which we analyze in the following section.

3 Out-of-sample evaluation of the VCR model

In this section we assess the performance of the proposed VCR model within the context of investment decision making. We consider two major scenarios. First, we deal with an investor whose interest is to maximize profits from trading stocks. We assume that her trading strategy – what to buy, what to sell – depends on the forecast returns based on the VCR model in (9) and (10). This trading strategy will be called *VCR-Mixture Position Trading Rule* and it is based on the one-step ahead forecast of individual asset returns based on the VCR model. The superiority of the proposed specification depends on its potential ability to generate larger profits than those obtained with more standard models. In the second scenario, in addition to the profits (returns), the investor wishes to assess potential large losses by adding a Value-at-Risk evaluation of her trading strategy. In this case, the modelling of the conditional variance becomes also relevant. Both scenarios provide an out-of-sample evaluation of the VCR model.

We proceed as follows. For each firm i in the market (343 firms), we compute the one-step ahead forecast $\hat{y}_{i,t+1}$ of the return as in (9). The sequence of one-step ahead forecasts is obtained with a "rolling" sample. For a sample size of T and with the first R observations, we estimate the parameters of the model $\hat{\theta}_R$ and compute the one-step ahead forecast $\hat{y}_{i,R+1}(\hat{\theta}_R)$. Next, using observations 2 to $R+1$, we estimate the model again to obtain $\hat{\theta}_{R+1}$ and calculate the one-step ahead forecast $\hat{y}_{i,R+2}(\hat{\theta}_{R+1})$. We keep rolling the sample one observation at a time until we reach $T-1$, to obtain $\hat{\theta}_{T-1}$ and the last one-step forecast $\hat{y}_{i,T}(\hat{\theta}_{T-1})$. Based on the forecasted returns, the investor predicts the cross-sectional position of all assets in relation to the overall market, that is,

$$\hat{z}_{i,t+1} = M^{-1} \sum_{j=1}^{M} \mathbf{1}(\hat{y}_{j,t+1} \leq \hat{y}_{i,t+1}), \qquad t = R, \ldots, T-1,$$

and buys the top K performing assets if their return is above the risk-free rate. In every subsequent out-of-sample period ($t = R, \ldots, T-1$), the investor revises her portfolio, selling the assets that fall out of the top performers and buying the ones that rise to the top, and she computes the one-period portfolio return

$$\pi_{t+1} = K^{-1} \sum_{j=1}^{M} y_{j,t+1} \cdot \mathbf{1}(\hat{z}_{j,t+1} \geq z_{t+1}^{K}), \qquad t = R, \ldots, T-1,$$

where z_{t+1}^{K} is the cutoff cross-sectional position to select the K best performing stocks such that $\sum_{j=1}^{M} \mathbf{1}(\hat{z}_{j,t+1} \geq z_{t+1}^{K}) = K$. We form a portfolio with the top 1% performers in the SP500 index. Every asset in the portfolio is weighted equally.

3.1 Competing trading rules

To evaluate the out-of-sample performance of the VCR model, we compare it with that of various competing models.

A simple alternative to the VCR-mixture rule may be constructed by imposing $\mathbb{H}^1 : \mu_{1,t} = \mu_{0,t}$ in (10). This trading rule assesses the importance of the nonlinearity in expected returns. It will be called *Position Trading Rule* because it takes into account the cross-sectional position of an asset while it ignores the nonlinearity of the model (mixture of normal densities) for expected returns. The one-step ahead forecast for every asset in the market is obtained from a linear specification of the conditional mean where the regressors are past returns and past cross-sectional positions. As in the previous rule, the ordinal position is predicted and a rolling sample scheme is used to obtain the sequence of one-step ahead forecasts. The investor follows the same strategy as before buying the top five performing assets and revising her portfolio in every period.

The third trading rule is a buy-and-hold strategy of the market portfolio. At the beginning of the forecasting interval, the investor buys the SP500 index and holds it until the end of the interval. At any given t, the one-period portfolio return is $\pi_t = y_{m,t}$ where $y_{m,t}$ is the return to the SP500 index. This strategy will be called *Buy-and-Hold-the-Market Trading Rule.*

The fourth trading rule is driven by the market efficiency hypothesis. We call this rule the *Random Walk Hypothesis.* If stock prices follow a random walk, the best predictor of price is the previous price, and the best forecast for the return of any given asset is zero. Hence $\pi_t = 0$ for any t and any asset.

In summary, these four trading rules aim to assess the predictability of stock returns: the VCR-Mixture Position Trading Rule claims that stock returns are non-linearly predictable, the Position Trading Rule claims that stock returns are linearly predictable, the Random Walk Hypothesis claims that stock returns are non-predictable, and the Buy-and-Hold-the-Market Trading Rule claims that actively managed portfolios have no advantage over passively index investing.

3.1.1 Technical trading rules

In addition to the above four models, we also consider four classes of *technical* trading rules considered by Sullivan, Timmermann and White (1999). All the trading rules are based on the SP500 index. For each of the four technical trading rules below, we consider four parameterization.

Filter-Rule(x): If the weekly closing price of a particular security moves up at least x per cent, buy and hold the security until its price moves down at least x per cent from a subsequent high, at which time simultaneously sell and go short. The short position is maintained until the weekly closing price rises at least x per cent above a subsequent low at which time one covers and buys. The neutral position is obtained by liquidating a long position when the price decreases y percent from the previous high, and covering a short position when the price increases y percent from the previous low. We apply one of the rules of Sullivan *et al.* (1999) to define subsequent high (low). A subsequent high (low) is the highest (lowest) closing price achieved while holding a particular long (short) position. We also allow for the holding of the asset for c weeks ignoring any signals generated from the market. We consider $\{x : 0.05, 0.10, 0.20, 0.50\}$, $y = 0.5x$, and $c = 1$.

Moving-Average-Rule(l, s): This rule involves going long (short) when the short period moving average (s) rises above (falls below) the long period moving average (l). Its idea is to smooth out the series and locate the initiation of trend (when s penetrates l). We consider four sets of local moving averages with $\{(l, s) : (10, 2), (20, 2), (10, 4), (20, 4)\}$, a fixed percentage band filter to rule out false signals with the band $b = 0.05$ for all cases, and $c = 1$ as for the filter rule.

Channel-Break-Out-Rule(n, x): A channel is said to occur when the high over the previous n time periods is within x percent of the low over the previous n time periods, not including the current price. The strategy is to buy when the closing price exceeds the channel, and to go short when the price moves below the channel. Long and short positions are held for a fixed number of days, $c = 1$. A fixed percentage band, $b = 0.05$, is applied to the channel as a filter. We consider the four sets of parameters $\{(n, x) : (4, 0.05), (10, 0.05), (4, 0.10), (10, 0.10)\}$.

Support-and-Resistance-Rule(n): Buy when the closing price exceeds the maximum price over the previous n time periods, and sell when the closing price is less than the minimum price over the previous n time periods. We consider $\{n : 2, 4, 8, 16\}$, and a fixed percentage band $b = 0.05$.

3.2 Loss functions: mean trading return and VaR

The simplest evaluation criterion is to compute the return of each trading strategy over the forecast sample $(R+1, T)$. There are $P \equiv T - R$ periods in this interval. For every trading rule we compute the "mean trading return",

$$MTR = P^{-1} \sum_{t=R}^{T-1} \pi_{t+1}.$$

The rule that provides the largest MTR would be a preferred trading strategy. However, with each trading rule we choose different portfolios that may have different levels of risk. The investor while pursuing a high MTR may also wish to control for catastrophic events maintaining a minimum amount of capital to cushion against excessive losses. Consequently, each trading rule would be evaluated according to their ability to allocate the optimal amount of capital for unlikely events, rendering a Value-at-Risk evaluation criterion necessary.[4]

Consider a portfolio of assets whose realized return is given by π_{t+1}. We are interested in $VaR^{\alpha}_{t+1}(\theta)$, the one-step ahead Value-at-Risk forecast of π_{t+1} at a given nominal tail coverage probability α. This is defined as the conditional quantile such that $\Pr[\pi_{t+1} \leq VaR^{\alpha}_{t+1}(\theta)|\mathcal{F}_t] = \alpha$. For Position Trading Rule and Buy-and-Hold-the-Market Trading Rule, for which we assume a location-scale distribution of π_{t+1}, the forecast of the portfolio VaR can be estimated as $VaR^{\alpha}_{t+1}(\hat{\theta}_t) = \mu^{\pi}_{t+1}(\hat{\theta}_t) + \Phi^{-1}_{t+1}(\alpha)\sigma^{\pi}_{t+1}(\hat{\theta}_t)$, where $\mu^{\pi}_{t+1}(\hat{\theta}_t), \sigma^{\pi}_{t+1}(\hat{\theta}_t)$ are the forecasts of the portfolio return and conditional standard deviation, $\Phi_{t+1}(\cdot)$ is the conditional cumulative distribution function of the standardized portfolio return, and $\hat{\theta}_t$ is the parameter vector estimated with information up to time t. For VCR-Mixture Position Trading Rule, where we are interested in the VaR of a portfolio of K asset, each one following a mixture of conditional normal distributions, the computation of the VaR is not straightforward because a mixture of normals does not belong to the location-scale family. We implement the analytical Monte Carlo method of Wang (2001).

[4]While we consider all the trading rules with MTR (see Table 4), we consider only three rules (VCR-Mixture Position Trading Rule, Position Trading Rule, and Buy-and-Hold-the-Market Trading Rule) with VaR-based loss functions (see Table 5). The reasons are because the selection of assets to form the portfolio is indeterminate for the Random Walk Hypothesis, and because the VaR calculation for all the technical trading rules, which are based on the SP500 index, is the same as that of the Buy-and-Hold-the-Market Trading Rule.

We evaluate the trading rules according to three VaR-based loss functions. The first loss function aims to minimize the amount of capital to put aside (that is required to protect the investor against a large negative return), the second loss function assesses which trading rule provides the correct predicted tail coverage probability, and the third loss function is based on quantile estimation and it evaluates which trading rule provides the best estimate of the VaR.

The first loss function V_1 sets the mean predicted "minimum required capital", $MRC^{\alpha}_{t+1}(\hat{\theta}_t)$,

$$V_1 \equiv P^{-1} \sum_{t=R}^{T-1} MRC^{\alpha}_{t+1}(\hat{\theta}_t) \simeq P^{-1} \sum_{t=R}^{T-1} VaR^{\alpha}_{t+1}(\hat{\theta}_t).$$

A formula for MRC^{α} as a function of VaR^{α} with $\alpha = 0.01$ is set by the Bassel Accord. See Jorion (2000, p. 65). We approximate the formula by setting $MRC^{\alpha} \simeq VaR^{\alpha}$. Over the forecast period, the trading rule that provides the lowest amount of capital to put aside will be preferred.

The second loss function V_2 aims to choose the trading rule that minimizes the difference between the nominal and the empirical lower tail probability. It is an out-of-sample evaluation criterion based on the likelihood ratio statistic of Christoffersen (1998). Over the forecast interval $(R+1, T)$, consider the following counts $n_1 = \sum_{t=R}^{T-1} \mathbf{1}(\pi_{t+1} < VaR^{\alpha}_{t+1}(\hat{\theta}_t))$ and $n_0 = P - n_1$. If the VaR^{α} has been correctly forecasted, n_0 must be $P \times (1-\alpha)$ and n_1 equals to $P \times \alpha$. In fact, the predictive likelihood function of α given a sample $\left\{\mathbf{1}(\pi_{t+1} < VaR^{\alpha}_{t+1}(\hat{\theta}_t))\right\}_{t=R}^{T-1}$ is $L(\alpha) = (1-\alpha)^{n_0}\alpha^{n_1}$ and the maximum likelihood estimator of α is $\hat{\alpha} = n_1/P$. If we were to test for the null hypothesis that $E[\mathbf{1}(\pi_{t+1} \leq VaR^{\alpha}_{t+1}(\hat{\theta}_t))] = \alpha$, the likelihood ratio test $2(\log L(\hat{\alpha}) - \log L(\alpha))$ would be a suitable statistic. The loss function V_2 is based on this statistic, and it is a distance measure between α and $\hat{\alpha}$. A trading rule that minimizes V_2 will be preferred.

$$\begin{aligned} V_2 &\equiv P^{-1}\left[2\log L(\hat{\alpha}) - 2\log L(\alpha)\right] \\ &= P^{-1}\sum_{t=R}^{T-1} 2\left[\mathbf{1}(\pi_{t+1} < VaR^{\alpha}_{t+1}(\hat{\theta}_t))\log\frac{\hat{\alpha}}{\alpha} + \mathbf{1}(\pi_{t+1} > VaR^{\alpha}_{t+1}(\hat{\theta}_t))\log\frac{1-\hat{\alpha}}{1-\alpha}\right]. \end{aligned}$$

The third loss function V_3 chooses the trading rule that minimizes the objective function used in quantile estimation (Koenker and Bassett, 1978)

$$V_3 \equiv P^{-1}\sum_{t=R}^{T-1}(\pi_{t+1} - VaR^{\alpha}_{t+1}(\hat{\theta}_t))\left[\alpha - \mathbf{1}(\pi_{t+1} < VaR^{\alpha}_{t+1}(\hat{\theta}_t))\right].$$

The trading rule that provides the smallest V_3 is preferred because it indicates a better goodness of fit.

3.3 Reality check

The question of interest is, among the twenty trading rules, which one is the best? Each rule produces different forecasts that are evaluated according to the four loss functions introduced in the previous subsection: $-MTR, V_1, V_2$, and V_3.[5] The best trading rule is the one that provides the minimum loss. However, for every loss function, how can we tell when the difference among the losses produced by each trading rule is statistically significant? Furthermore, is a pairwise comparison among the trading rules sufficient? Note that all trading rules are based on the same data, and that their forecasts are not independent. We need a statistical procedure to assess whether the difference among the losses is significant while, at the same time, taking into account any forecast dependence across trading rules and controlling for potential biases due to data snooping. This procedure is the "reality check" proposed by White (2000). Suppose that we choose one of the trading rules as a benchmark. We aim to compare the loss of the remaining trading rules to that of the benchmark. We formulate a null hypothesis that the trading rule with the smallest loss is no better than the benchmark rule. If we reject the null hypothesis, there is at least one competing trading rule that produces smaller loss than the benchmark. A brief sketch of the formal testing procedure follows.

Let l be the number of competing trading rules ($k = 1, \ldots, l$) to compare with the benchmark rule (indexed as $k = 0$). For each trading rule k, one-step predictions are to be made for P periods from $R + 1$ through T using a rolling sample, as explained in the previous sections. Consider a generic loss function $L(Y, \theta)$ where Y consists of variables in the information set. In our case, we have four loss functions: $-MTR, V_1, V_2$, and V_3. The best trading rule is the one that minimizes the expected loss. We test a hypothesis about an $l \times 1$ vector of moments, $E(\mathbf{f})$, where $\mathbf{f} \equiv \mathbf{f}(Y, \theta)$ is an $l \times 1$ vector with the k^{th} element $f_k = L_0(Y, \theta) - L_k(Y, \theta)$, $\theta \equiv \text{plim } \hat{\theta}_T$, and $L_0(\cdot, \cdot)$ is the loss under the benchmark rule, $L_k(\cdot, \cdot)$ is the loss provided by the k^{th} trading rule. A test for a hypothesis on $E(\mathbf{f})$ may be based on the $l \times 1$ statistic $\bar{\mathbf{f}} \equiv P^{-1} \sum_{t=R}^{T-1} \hat{\mathbf{f}}_{t+1}$, where $\hat{\mathbf{f}}_{t+1} \equiv \mathbf{f}(Y_{t+1}, \hat{\theta}_t)$.

Our interest is to compare all the trading rules with a benchmark. An appropriate null hypothesis is that all the trading rules are no better than a benchmark, i.e., $\mathbb{H}_0 : \max_{1 \le k \le l} E(f_k) \le 0$. This is a multiple hypothesis, the intersection of the one-sided individual hypotheses $E(f_k) \le 0,\ k = 1, \ldots ., l$. The alternative is that $\mathbb{H}_0$ is false, that is, the best trading rule is superior to the benchmark. If the null hypothesis is rejected, there must be at least one trading rule for which $E(f_k)$

[5] We write a negative sign in the mean trading return ($-MTR$) so as to minimize this function as well as those based on VaR calculations.

is positive. Suppose that $\sqrt{P}(\bar{\mathbf{f}} - E(\mathbf{f})) \xrightarrow{d} N(0, \Omega)$ as $P(T) \to \infty$ when $T \to \infty$, for Ω positive semi-definite. White's (2000) test statistic for $\mathbb{H}_0$ is formed as $\bar{V} \equiv \max_{1 \leq k \leq l} \sqrt{P}\bar{f}_k$, which converges in distribution to $\max_{1 \leq k \leq l} G_k$ under $\mathbb{H}_0$, where the limit random vector $G = (G_1, \ldots, G_l)'$ is $N(0, \Omega)$. However, as the null limiting distribution of $\max_{1 \leq k \leq l} G_k$ is unknown, White (2000, Theorem 2.3) shows that the distribution of $\sqrt{P}(\bar{\mathbf{f}}^* - \bar{\mathbf{f}})$ converges to that of $\sqrt{P}(\bar{\mathbf{f}} - E(\mathbf{f}))$, where $\bar{\mathbf{f}}^*$ is obtained from the stationary bootstrap of Politis and Romano (1994). By the continuous mapping theorem this result extends to the maximal element of the vector $\sqrt{P}(\bar{\mathbf{f}}^* - \bar{\mathbf{f}})$ so that the empirical distribution of $\bar{V}^* = \max_{1 \leq k \leq l} \sqrt{P}(\bar{f}_k^* - \bar{f}_k)$ is used to compute the p-value of $\bar{V}$ (White, 2000, Corollary 2.4). This p-value is called the "Reality Check p-value".

3.4 Evaluation of trading rules

The out-of-sample performance of the aforementioned trading rules is provided in Tables 4 and 5. In Table 4, the trading rules are evaluated according to the MTR function, and in Table 5 according to the three VaR-based loss functions. In both cases, the in-sample size for the rolling estimation is $R = 300$, and the out-of-sample forecast horizon is $P = 299$. The stationary bootstrap (with smoothing parameter 0.25,which corresponds to the mean block length of 4) is implemented with 1000 bootstrap resamples.[6] In the first column of each table, we report the benchmark trading rule to which the remaining rules will be compared.

[Table 4 about here]

In Table 4, we report the value of MTR function for each trading strategy. The VCR-Mixture Position Trading Rule produces a mean trading return of 0.264 that is twice as much as the next most favorable rule, which is the Position Trading Rule. This one and the Buy-and-Hold-the-Market Rule produce similar results. We also find that all the technical trading rules are clearly dominated by the VCR-Mixture Position rule, even though some of the Filter Rules and one of the Support-and-Resistance Rule produce similar mean trading returns to those of the Position Trading Rule and the Buy-and-Hold-the-Market Trading Rule. The statistical difference among the rules is assessed with the White procedure. In each row, a benchmark rule is compared with all the remaining $l = 19$ rules. When the VCR-Mixture-Position Rule is the benchmark, the reality check p-value is 1.000, indicating that it is not dominated by any of the other rules. When any other rule is taken as the benchmark, the reality check p-value is less than 8%.

[6] We experiment with different smoothing parameters 0.50 and 0.75 and we obtain similar results.

Following upon some of the criticisms of the profitability of momentum strategies, the superior MTR of the VCR-Mixture Position Trading Rule may be the result of forming portfolios that are very risky and consequently, the profits we observe are just due to a compensation for risk. We have calculated the beta of each stock through a CAPM-type time series regression and, over each period of the forecasting interval (299 periods), we have computed the average beta of the selected portfolio of winner stocks. Since we are dealing with SP500 companies, the average beta of the stocks in the index is about one but there are stocks with betas as high as 4.9. For the winner portfolio, in one-third of the forecasting periods, the average beta is less than 1.1; in half of the periods, the average beta is less than 1.4; and in about one-third of the periods is greater than 1.8. Hence, the VCR-Mixture Position Trading Rule tends to pick up stocks over the full spectrum of risk. As for the industrial nature of the five winner stocks that form the portfolio, we find that the Information Technology sector (the riskiest according to the VCR model) is chosen 16% of the time and the Utilities sector (the least risky) is chosen 10% of the time.[7] Thus, if our strategy were based solely on the choice of risky asset, we would have expected to choose Information Technology stocks more frequently.

[Table 5 about here]

In Table 5, we report the out-of-sample performance of the three trading rules evaluated according to the loss functions V_1, V_2, and V_3, for two tail nominal probabilities $\alpha = 1\%$ and $\alpha = 5\%$. The results for $\alpha = 1\%$ and 5% are virtually identical for all the three loss functions. With respect to V_1, the Position Trading Rule seems to dominate statistically the remaining two rules providing the least amount of required capital. However, when we consider V_2, the same rule performs very poorly because it estimates the tail coverage rate of 7.2% at a nominal rate of 1%, and 14.0% at a nominal rate of 5%. Thus, the smallest MRC of the Position Trading Rule comes at the expense of a high tail failure rate. The VCR-Mixture Position Trading Rule delivers the best tail coverage, estimating a tail coverage probability of 1.3% at a nominal rate of 1% and 4.3% at a nominal rate of 5%. The White p-value confirms that this is the dominant rule in V_2. With respect to the loss function V_3, the reality check p-values asserts that, once more, the VCR-Mixture Position Trading Rule dominates the alternative two strategies.

[7] Given that some sectors are more represented than others in the SP500 index, we have adjusted the number of times a firm is chosen from a particular sector by dividing by the share of the sector in the SP500 and normalizing the resulting numbers to 1.

In addition to the four loss functions reported in Tables 4 and 5, we have also computed the mean squared forecast errors (MSFE) of the returns, $P^{-1}\sum_{t=R}^{T-1} M^{-1}\sum_{i=1}^{M}(\hat{y}_{i,t+1}-y_{i,t+1})^2$, and the MSFE of the cross-sectional positions, $P^{-1}\sum_{t=R}^{T-1} M^{-1}\sum_{i=1}^{M}(\hat{z}_{i,t+1}-z_{i,t+1})^2$. Based on these MSFE losses, we compare the VCR-Mixture Position Trading Rule and the Position Trading Rule. With the Position Trading Rule as the benchmark, the reality check p-values for both of the MSFE loss functions are 0.000, clearly indicating that the VCR-Mixture Position Trading Rule is better than the Position Trading Rule and thus underlying the importance of nonlinearity in the conditional mean of the return process due to the VCR-weighted mixture.

4 Conclusions

Uncertainty, risk, and volatility aim to describe random events faced by economic agents, to which we attach a probability of occurrence. Within this general context, we have added a new perspective to the meaning of volatility. We have departed from the more classical meaning of volatility, measured by a variance (or any other measure of dispersion), and we have added an ordinal measure, the cross-sectional position of an asset return in relation to its peers. The meaning of volatility that we put forward is a combination of a time-varying variance and a time-varying probability of jumping positions in the cross-sectional return distribution of the assets that constitute the market. The latter conveys a notion of interdependence among assets, implicitly revealing market information, and in this sense, it has a multivariate flavor. Our task has been to investigate the relevance of this approach.

For individual assets, we have modelled the *joint* dynamics of the cross-sectional position and the asset return by analyzing (1) the *marginal* probability distribution of a sharp jump in the cross-sectional position, and (2) the probability distribution of the asset return *conditional* on a jump. The former is conducted within the context of a duration model, and the latter assumes that there are different dynamics in returns depending upon whether or not a jump has occurred. We have estimated and tested the proposed models with weekly returns of those SP500 corporations that have survived in the index from January 1, 1990 to August 31, 2001. The estimation results for the 343 firms are diverse but, broadly speaking, we have found that the expected probability of jumping increases when the firm's cross-sectional position is either at the very top or the very bottom of the cross-sectional distribution of returns, hence extreme positions tend to be shorter lived than intermediate ones. For a representative firm such as one with median parameter estimates, we have calculated that the expected duration is between 4 and 5 weeks implying a minimum expected probability of a sharp jump of 20-25%. Furthermore, we found that the expected return is a function of past cross-sectional positions and that there are different dynamics when the return is either at extreme positions (top or bottom of the cross-sectional distribution) or at intermediate positions.

From an investor's point of view, the most relevant question is how useful this model is. We have designed an out-of-sample exercise where we have judged the adequacy of our model in two dimensions: profitability and risk monitoring. Twenty different trading rules, some model-based and some technical rules, are compared within the statistical framework of White's reality check, which controls for potential biases due to data snooping. The VCR-Mixture Position Trading Rule based on the one-step ahead forecast of our proposed model dominates all standard alternative trading rules. It provides superior mean trading returns and accurate Value-at-Risk forecasts.

We summarize this research by underlining our main contribution. The conditional probability of jumping cross-sectional positions is forecastable. This is possible because there is a persistence on return performance. That is, assets that perform well (poorly) in the past, keep performing well (poorly) in the near future. We have called this probability Varying Cross-sectional Risk because it provides an assessment of the chances of being a winner or a looser within the available set of assets.

References

Aït-Sahalia, Yacine (2003), "Disentangling Diffusion from Jumps", forthcoming in *the Journal of Financial Economics.*

Christoffersen, Peter F. (1998), "Evaluating Interval Forecasts," *International Economic Review*, 39, 841-862.

Engle, Robert F., David F. Hendry, and J.-F. Richard (1983), "Exogeneity", *Econometrica*, 51, 277-304.

Engle, Robert F. and Jeffrey R. Russell (1998), "Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data," *Econometrica,* 66, 1127-1162.

Granger, Clive W. J. (2002), "Some Comments on Risk", *Journal of Applied Econometrics*, 17(5), 447-456.

Hamilton, James D. and Oscar Jordà (2002), "A Model of the Federal Funds Target," *Journal of Political Economy,* 110, 1135-1167.

Jegadeesh, Narasimhan and Sheridan Titman (1993), "Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency", *Journal of Finance,* 48, 65-91.

Jorion, Philippe (2000), *Value at Risk*, 2ed., McGraw-Hill.

Koenker, Roger and Gilbert Bassett, Jr. (1978), "Regression Quantiles", *Econometrica*, 46(1), 33-50.

Lintner, John (1965), "The Valuation of Risky Assets and the Selection of Risky Investment in Stock Portfolios and Capital Budget", *Review of Economics and Statistics,* 47, 13-37.

Markowitz, Harry (1952), "Portfolio Selection", *Journal of Finance,* 7, 77-99.

Politis, Dimitris N. and Joseph P. Romano (1994), "The Stationary Bootstrap", *Journal of American Statistical Association*, 89, 1303-1313.

Sharpe, William (1964), "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk", *Journal of Finance,* 19, 425-442.

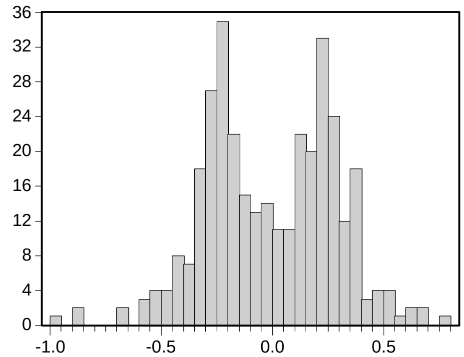
Sullivan, Ryan, Allan Timmermann, and Halbert White (1999), "Data Snooping, Technical Trading Rule Performance, and the Bootstrap," *Journal of Finance*, 54, 1647-1692.

Wang, Jin (2001), "Generating Daily Changes in Market Variables Using a Multivariate Mixture of Normal Distributions", *Proceedings of the 2001 Winter Simulation Conference*, B.A., Peters, J.S. Smith, D.J. Medeiros, and M.W. Roher, eds., 283-289.

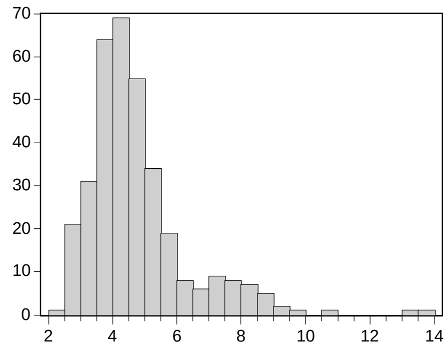
White, Halbert (2000), "A Reality Check for Data Snooping", *Econometrica*, 68(5), 1097-1126.

Table 1
Descriptive Statistics of Weekly Returns of the SP500 firms
January 1, 1990-August 31, 2001

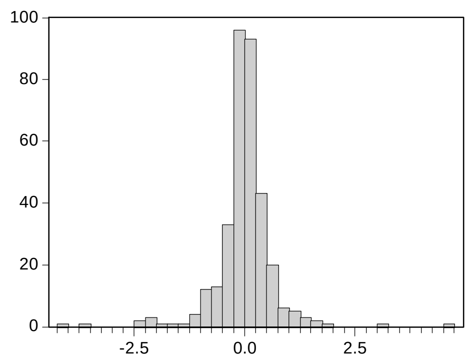
Cross-sectional frequency distribution of unconditional moments



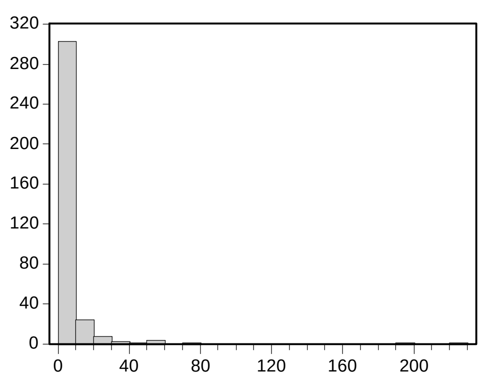
Series: MEAN	
Sample 1 343	
Observations 343	
Mean	-0.007470
Median	-0.016200
Maximum	0.771030
Minimum	-0.964030
Std. Dev.	0.291280
Skewness	-0.134563
Kurtosis	2.735350
Jarque-Bera	2.036118
Probability	0.361296



Series: STD. DEVIATION	
Sample 1 343	
Observations 343	
Mean	4.734877
Median	4.372920
Maximum	13.53324
Minimum	2.475010
Std. Dev.	1.583924
Skewness	1.911303
Kurtosis	8.657083
Jarque-Bera	666.2048
Probability	0.000000



Series: SKEWNESS	
Sample 1 343	
Observations 343	
Mean	-0.009417
Median	0.014850
Maximum	4.706040
Minimum	-4.086720
Std. Dev.	0.681469
Skewness	-0.173610
Kurtosis	17.02490
Jarque-Bera	2812.864
Probability	0.000000



Series: KURTOSIS	
Sample 1 343	
Observations 343	
Mean	8.475258
Median	5.228110
Maximum	229.9880
Minimum	3.116470
Std. Dev.	17.33199
Skewness	10.07482
Kurtosis	117.6397
Jarque-Bera	193627.3
Probability	0.000000

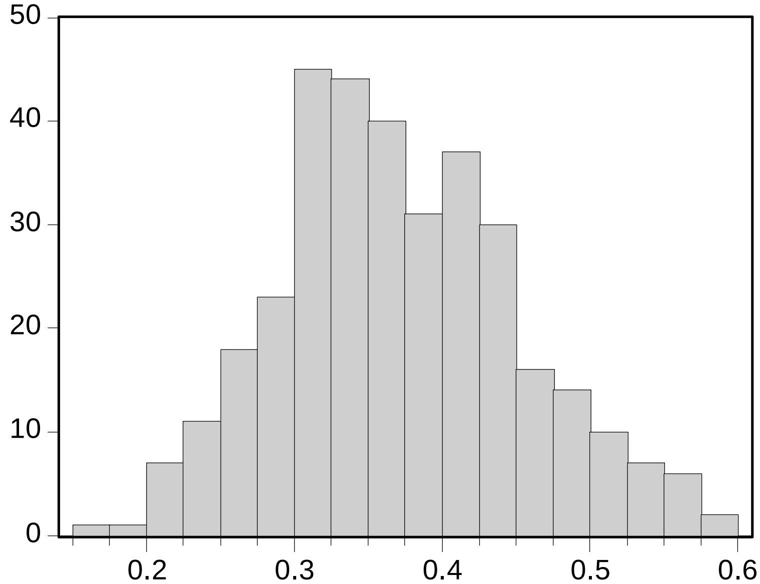
Table 2

Cross-sectional frequency distribution of the estimates of the duration model

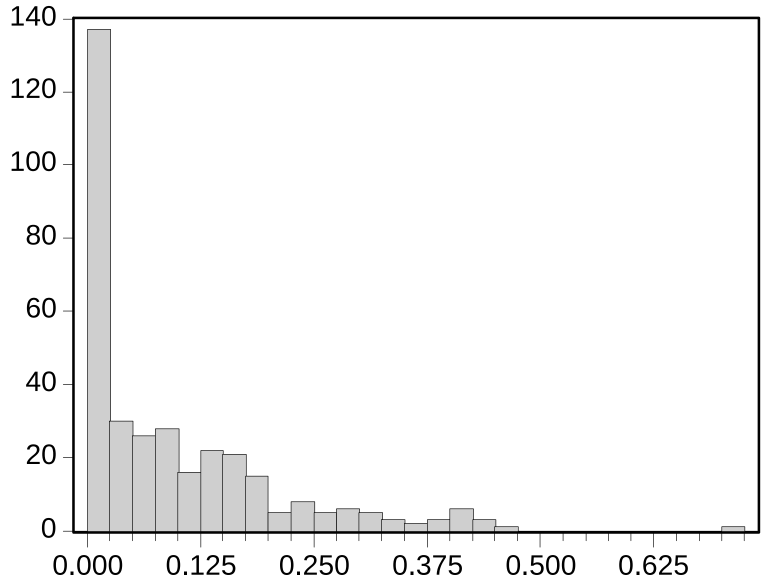
$$p_t = \Pr(J_t = 1 \mid F_{t-1}) = [y_{N(t-1)} + d'X_{t-1}]^{-1}$$

$$y_{N(t)} = a\ D_{N(t)-1} + b y_{N(t)-1}$$

$$d'X_{t-1} = d_1 + d_2 y_{t-1} 1(z_{t-1} \leq 0.5) + d_3 y_{t-1} 1(z_{t-1} > 0.5) + d_4 z_{t-1}$$



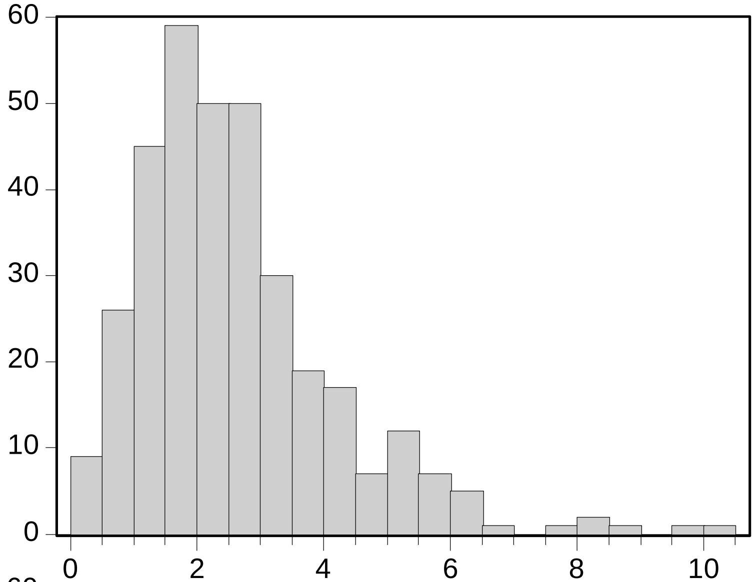
Series: ALPHA	
Sample 1 343	
Observations 343	
Mean	0.371292
Median	0.364000
90% percentile	0.485000
Maximum	0.598000
Minimum	0.167000
Std. Dev.	0.081061
Skewness	0.309177
Kurtosis	2.814036
Jarque-Bera	5.958832
Probability	0.050823



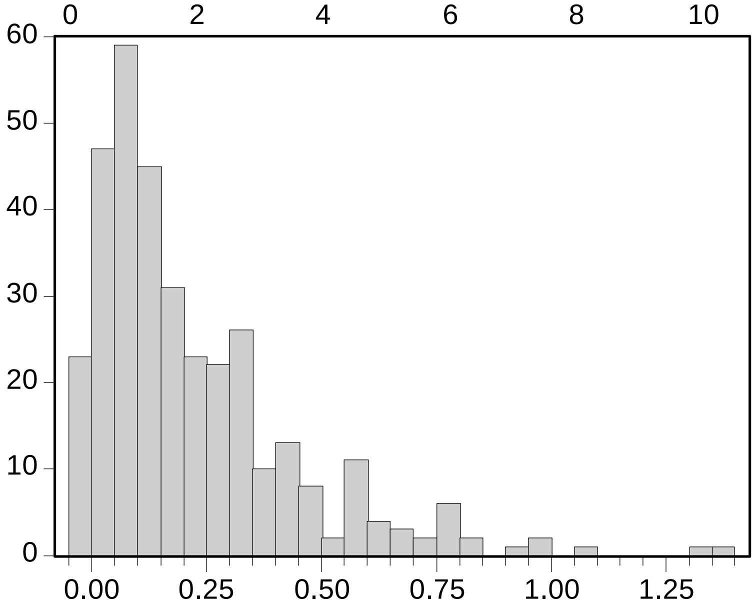
Series: BETA	
Sample 1 343	
Observations 343	
Mean	0.098172
Median	0.056000
90% percentile	0.252000
Maximum	0.724000
Minimum	0.010000
Std. Dev.	0.110893
Skewness	1.780148
Kurtosis	6.773379
Jarque-Bera	384.6472
Probability	0.000000

Table 2 (cont.)

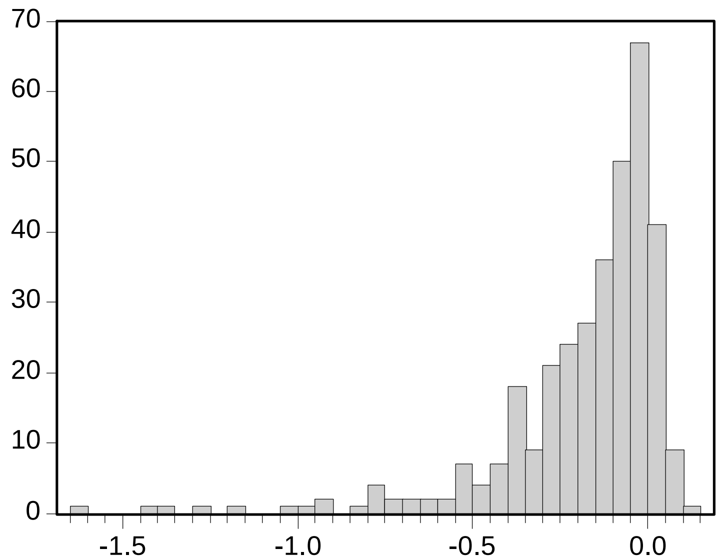
Cross-sectional frequency distribution of the estimates of the duration model



Series: DELTA1	
Sample 1 343	
Observations 343	
Mean	2.621921
Median	2.325000
90%percentile	4.757000
Maximum	10.23000
Minimum	0.008000
Std. Dev.	1.580073
Skewness	1.480903
Kurtosis	6.378431
Jarque-Bera	288.4929
Probability	0.000000



Series: DELTA2	
Sample 1 343	
Observations 343	
Mean	0.218213
Median	0.146000
90% percentile	0.551000
Maximum	1.369000
Minimum	-0.027000
Std. Dev.	0.223472
Skewness	1.804511
Kurtosis	7.277704
Jarque-Bera	447.6692
Probability	0.000000



Series: DELTA3	
Sample 1 343	
Observations 343	
Mean	-0.181584
Median	-0.105000
10%percentile	-0.446000
Maximum	0.135000
Minimum	-1.617000
Std. Dev.	0.244567
Skewness	-2.509015
Kurtosis	11.44026
Jarque-Bera	1377.982
Probability	0.000000

Table 2 (cont.)

Duration Model

Median parameter estimates for the industry sectors represented in the SP500 index

$$p_t = \Pr(J_t = 1 | F_{t-1}) = [y_{N(t-1)} + d' X_{t-1}]^{-1}$$
$$y_{N(t)} = a\, D_{N(t)-1} + b y_{N(t)-1}$$
$$d' X_{t-1} = d_1 + d_2 y_{t-1} 1(z_{t-1} \leq 0.5) + d_3 y_{t-1} 1(z_{t-1} > 0.5) + d_4 z_{t-1}$$

Sector	% of firms	$\hat{a} + \hat{b}$	$\hat{d}_2$	$\hat{d}_3$	$\hat{p}_t$
Consumer Goods	28.3	0.455	0.162	-0.107	0.221
Energy	5.2	0.434	0.178	-0.198	0.239
Finance	12.5	0.457	0.221	-0.083	0.211
Health Care	7.0	0.444	0.103	-0.075	0.251
Industrials	12.8	0.455	0.193	-0.132	0.205
Information Technology	20.7	0.399	0.081	-0.056	0.301
Material	6.7	0.484	0.120	-0.128	0.223
Utilities	6.7	0.548	0.214	-0.416	0.109
All sectors	100.0	0.455	0.146	-0.105	0.220

Table 3
Cross-sectional frequency distribution of the estimates of the nonlinear model for expected returns

$$f_2(y_t \mid J_t, F_{t-1}; q_2) = \begin{cases} N(m_{1,t}, s_{1,t}{}^2) & if \quad J_t = 1 \\ N(m_{o,t}, s_{o,t}{}^2) & if \quad J_t = 0 \end{cases}$$

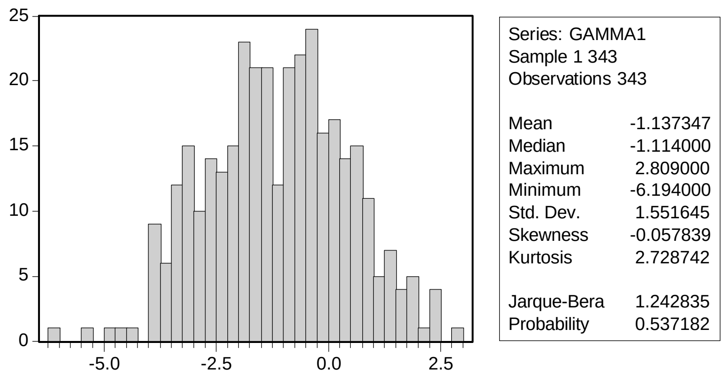
$$m_{1,t} = n_1 + g_1 y_{t-1} + h_1 z_{t-1}$$

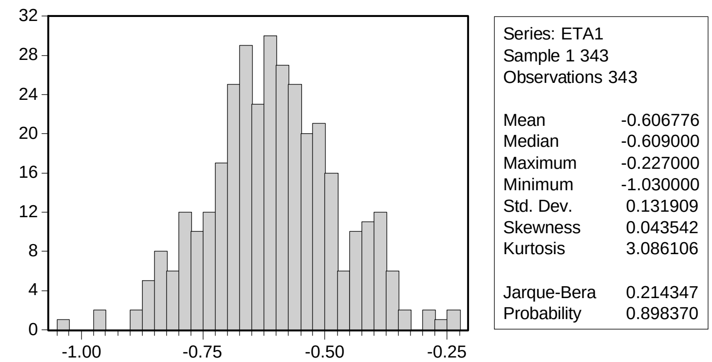
$$m_{0,t} = n_0 + g_0 y_{t-1} + h_0 z_{t-1}$$

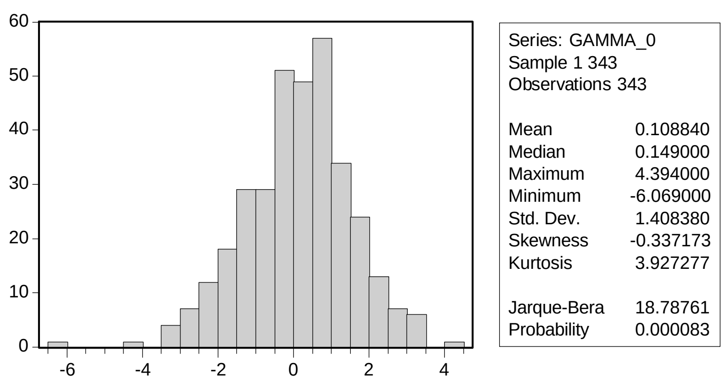
$$s^2{}_t = w + re^2{}_{t-1} + ts^2{}_{t-1}$$

$$with \quad e_{t-1} = (y_{t-1} - m_{1,t-1})J_{t-1} + (y_{t-1} - m_{0,t-1})(1 - J_{t-1})$$

Conditional mean parameter estimates







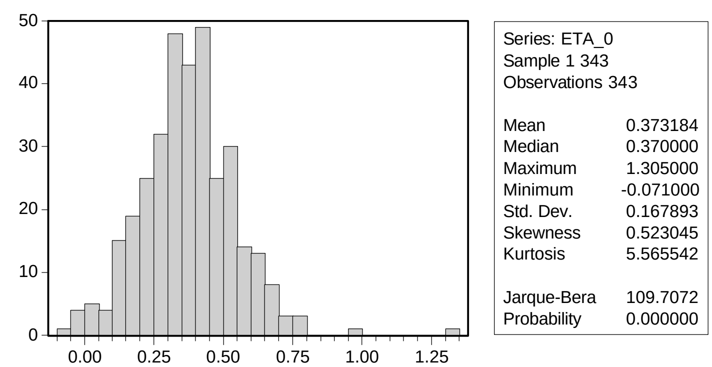
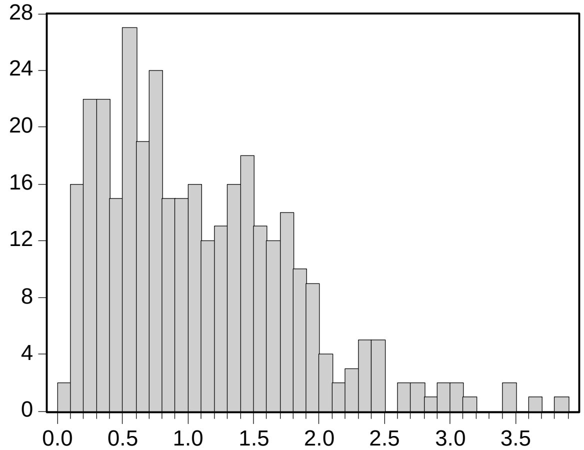
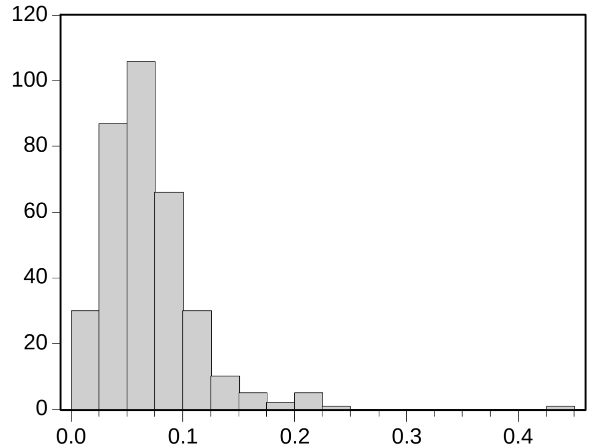


Table 3 (cont.)

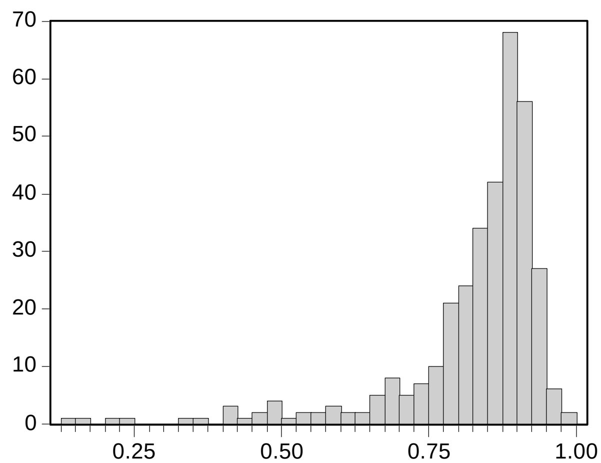
Conditional variance parameter estimates



Series: OMEGA	
Sample 1 343	
Observations 343	
Mean	1.099802
Median	0.956000
Maximum	3.847000
Minimum	0.072000
Std. Dev.	0.724246
Skewness	0.951112
Kurtosis	3.860637
Jarque-Bera	62.29958
Probability	0.000000



Series: RHO	
Sample 1 343	
Observations 343	
Mean	0.068586
Median	0.063000
Maximum	0.438000
Minimum	0.010000
Std. Dev.	0.043179
Skewness	2.707263
Kurtosis	19.34593
Jarque-Bera	4237.573
Probability	0.000000



Series: TAU	
Sample 1 343	
Observations 343	
Mean	0.825708
Median	0.866000
Maximum	0.980000
Minimum	0.126000
Std. Dev.	0.131893
Skewness	-2.476561
Kurtosis	10.27886
Jarque-Bera	1107.822
Probability	0.000000

Table 3 (cont.)

Expected Returns Model

Median parameter estimates for the industry sectors represented in the SP500 index

$$f_2(y_t \mid J_t, F_{t-1}; q_2) = \begin{cases} N(m_{1,t}, s_{1,t}^{\ 2}) & if \quad J_t = 1 \\ N(m_{o,t}, s_{o,t}^{\ 2}) & if \quad J_t = 0 \end{cases}$$

$$m_{1,t} = n_1 + g_1 y_{t-1} + h_1 z_{t-1}$$

$$m_{0,t} = n_0 + g_0 y_{t-1} + h_0 z_{t-1}$$

$$s^2_t = w + re^2_{t-1} + ts^2_{t-1}$$

$$with \quad e_{t-1} = (y_{t-1} - m_{1,t-1})J_{t-1} + (y_{t-1} - m_{0,t-1})(1 - J_{t-1})$$

Sector	% of firms	$\hat{g}_1$	$\hat{g}_0$	$\hat{h}_1$	$\hat{h}_0$	$\hat{r} + \hat{t}$
Consumer Goods	28.3	-1.389	-0.073	-0.585	0.378	0.922
Energy	5.2	-1.025	0.145	-0.629	0.363	0.953
Finance	12.5	-0.660	0.519	-0.661	0.356	0.942
Health Care	7.0	-1.801	-0.121	-0.611	0.395	0.921
Industrials	12.8	-1.430	-0.166	-0.596	0.372	0.953
Information Technology	20.7	-1.405	0.196	-0.595	0.444	0.928
Material	6.7	-1.069	0.557	-0.601	0.349	0.921
Utilities	6.7	0.116	0.987	-0.729	0.110	0.953
All sectors	100.0	-1.110	0.149	-0.609	0.370	0.930

Table 4

Out-of-sample evaluation of the trading rules
Loss function: Mean Trading Return

Benchmark trading rule	MTR	White p-value
VCR-Mixture Position Rule	0.264	1.000
Position Rule	0.131	0.079
Buy-and-Hold-the-Market Rule	0.115	0.072
Random Walk Hypothesis	0.000	0.000
Filter-Rule (0.05)	0.062	0.008
Filter-Rule (0.10)	0.057	0.008
Filter-Rule (0.20)	0.119	0.077
Filter-Rule (0.50)	0.109	0.067
Moving-Average Rule (10, 2)	0.024	0.000
Moving-Average Rule (20, 2)	-0.036	0.000
Moving-Average Rule (10, 4)	0.006	0.001
Moving-Average Rule (20, 4)	-0.009	0.000
Channel-Break-Out Rule (4, 0.05)	0.029	0.001
Channel-Break-Out Rule (10, 0.05)	0.086	0.006
Channel-Break-Out Rule (4, 0.10)	-0.006	0.000
Channel-Break-Out Rule (10, 0.10)	0.040	0.002
Support-and-Resistance Rule (2)	0.124	0.040
Support-and-Resistance Rule (4)	0.006	0.000
Support-and-Resistance Rule (8)	0.040	0.000
Support-and-Resistance Rule (16)	0.072	0.004

Notes: The mean trading return (MTR) represents the profit accrued from the respective trading rules. The out-of-sample horizon is P=299 and the in-sample horizon is R=300. In each row, we present the trading rule selected as the benchmark with its corresponding MTR and the White reality check p-value for testing the null hypothesis that the remaining trading rules are not any better than the benchmark rule. A large reality check p-value indicates that the null hypothesis cannot be rejected. For example, when the VCR-Mixture Position Rule is the benchmark, the reality check p-value 1.000 means that this benchmark rule is not dominated by any of the other 19 alternative trading rules.

Table 5

Out-of-sample evaluation of the trading rules
Loss function: Value-at-Risk calculations

Panel 1. $a = 0.01$

Benchmark trading rule	V_1	White p-value	V_2	$\hat{a}$	White p-value	V_3	White p-value
VCR-Mixture Position	1.952	0.000	0.007	0.013	1.000	0.025	0.996
Position	0.980	1.000	0.097	0.072	0.000	0.038	0.101
Buy-and-Hold-the-Market	2.351	0.000	0.040	0.029	0.000	0.040	0.042

Panel 2. $a = 0.05$

Benchmark trading rule	V_1	White p-value	V_2	$\hat{a}$	White p-value	V_3	White p-value
VCR-Mixture Position	1.395	0.000	0.019	0.043	1.000	0.073	0.998
Position	0.565	1.000	0.171	0.140	0.002	0.101	0.034
Buy-and-Hold-the-Market	1.625	0.000	0.055	0.081	0.000	0.141	0.000

Notes: The out-of-sample period is P=299 and the in-sample period is R=300. V_1 , V_2 , and V_3 represent the values of the three VaR-based loss functions: MRC, coverage failure rate, and goodness of fit in quantile estimation. $\hat{a}$ denotes the empirical failure rate at the nominal rate a . The White reality check p-value corresponds to the test with a null hypothesis that the remaining trading rules are not any better than the benchmark rule. A large reality check p-value indicates that the null hypothesis cannot be rejected. For example, when the VCR-Mixture Position Trading Rule is the benchmark, the p-values are very large in terms of V_2 and V_3 , implying that this benchmark trading rule is not dominated by the two alternative trading rules. In terms of V_1, the reality check p-value is 0.0 and thus it is dominated by the alternative rules.

Figure 1
Stylized description of the modelling problem