Jumps in Cross-Sectional Rank and Expected Returns: A Mixture Model[*]

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 a new non-linear time series model of expected returns. The novelty lies on the investigation of the dynamics of the cross-sectional rank of asset returns. We model the *joint* dynamics of a sharp jump in the cross-sectional rank and the asset return by analyzing (1) the *marginal* probability distribution of a jump in the cross-sectional rank 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 depending upon whether or not a jump has taken place. As a result, the expected returns are generated by a mixture of normal distributions weighted by the probability of jumping. The model is estimated for the weekly returns of the constituents of the SP500 index from 1990 to 2000, and its performance is assessed in an out-of-sample exercise from 2001 to 2005. Based on the one-step ahead forecast of the mixture model we propose a trading rule, which is evaluated according to several forecast evaluation criteria and compared to eighteen alternative trading rules. We find that the proposed trading strategy is the dominant rule by providing superior risk-adjusted mean trading returns and accurate Value-at-Risk forecasts.

Key Words: Cross-sectional rank, Duration, Leptokurtosis, Mixture of normals, Nonlinearity, Risk, Sharpe ratio, Trading rule, VaR.

JEL Classification: C3, C5, G0.

---

1 Introduction

We propose a new non-linear model of expected returns that combines time series information with cross-sectional information on firm-specific returns. Though this general purpose is not novel, for instance the empirical testing of classical asset pricing models such as CAPM and APT is carried out combining time series and cross-sectional information, the introduction of cross-sectional information embedded in the cross-sectional ranking of returns is new. Our objective is the development of a time series model for the cross-sectional rankings that will be helpful to forecast expected returns, which in turn will be the basis of some trading strategies.

Let $y_{i,t}$ be the return of firm i 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 each time t, we order the asset returns from the smallest to the largest, and we define $z_{i,t}$, the *Varying Cross-sectional Rank* (VCR) of firm i within the market, as the proportion of firms that have a return less than or equal to the return of firm i. 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, $z_{i,t} \in (0,1]$.

A graphical introduction to our main idea is contained in Figure 1, in which $z_{i,t}$ is the shaded area of the cross-sectional histogram of returns. A stylized description of the problem that we aim to analyze is as follows. For one period to the next, the VCR changes. We draw a histogram to represent the VCR of realized asset returns, which is time-varying. Our objective is to model the dynamics of the VCR $z_{i,t}$ jointly with the dynamics of the asset return $y_{i,t}$. 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 overall market has gone down as well as the return and the VCR 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 VCR 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 VCR is unchanged $z_{i,t_3} = z_{i,t_4}$. In this case the asset, even though is a good performer, comparatively speaking, has been outperformed by its peers. 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. Our interest is the jointly time series modeling of the return $y_{i,t}$ and the VCR $z_{i,t}$.

The important question is whether there is time dependence in $z_{i,t}$. In our preliminary investi-

gation, we have observed that, though there is no linear dependence of the ARMA type, there are some assets that tend to stick to the upper or lower ranks (broadly defined) of the cross-sectional distribution of returns while some others tend to move around very frequently. It seems that there is some time persistence in rank (high or low) though not of the linear type. In addition, from a cross-sectional perspective, the rank is a dependent variable. Any movement, small or large, in an asset return will affect not only the rank of this asset but also the rank of the remaining assets. Only in the exceptional and improbable case that all asset returns move in the same direction and by more or less similar amount, the overall ranking will not be affected. Because of this high cross-sectional dependence, we consider that small movements in the asset ranking will not contain significant information and that we should focus in the large movements in ranking, which most likely are the result of news in the overall market and/or of news concerning a particular asset. For these reasons, we analyze the variability in VCR rather than the VCR itself. Focusing on large rank movements, we define, at time t, a sharp jump as a binary variable that takes the value one when the there is a minimum (upward or downward) movement of 0.5 in the ranking of asset i , and zero otherwise:

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

A jump of this magnitude brings the asset return above or below the median of the cross-sectional distribution of returns. Note that our notion of jumps differs from the more traditional meaning of the word in the context of continuous-time modelling of the univariate return process. A jump in the cross-sectional rank implicitly depends on numerous univariate return processes.

Now, the analytical problem becomes to model the joint distribution of the return $y_{i,t}$ and the jump $J_{i,t}$, i.e. $f(y_{i,t}, J_{i,t}|\mathcal{F}_{t-1})$ where $\mathcal{F}_{t-1}$ is the 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 conditioning on the jump. As $J_{i,t}$ is a Bernoulli variable, the marginal distribution of the jump is $f_1(J_{i,t}|\mathcal{F}_{t-1}) = p_{i,t}^{J_{i,t}}(1 - p_{i,t})^{(1-J_{i,t})}$ where $p_{i,t} \equiv \Pr(J_{i,t} = 1|\mathcal{F}_{t-1})$ is the conditional probability of a jump in the cross-sectional ranks. We model $p_{i,t}$ within the context of a dynamic duration model specified in calendar time as in Hamilton and Jordà (2002). Though in event time the traditional question in duration analysis would be to analyze the expected length of time between two jumps, in calendar time the question will focus on the likelihood of a sharp jump in the rank of a given asset. The calendar time approach is necessary for our analysis because asset returns are reported in calendar time (days, weeks, etc.) and it has the advantage of incorporating any other available information

2

also reported in calendar time.

The important implication of modeling $f(y_{i,t}, J_{i,t}|\mathcal{F}_{t-1}) = f_1(J_{i,t}|\mathcal{F}_{t-1})f_2(y_{i,t}|J_{i,t}, \mathcal{F}_{t-1})$ is that the marginal return distribution, say $g(y_{i,t}|\mathcal{F}_{t-1})$, is a mixture distribution where the mixture weights are given by $p_{i,t}$. The result is therefore a highly nonlinear model for the expected return $E(y_{i,t+1}|\mathcal{F}_t)$. We assess the performance of the nonlinear model in an out-of-sample forecasting exercise within the context of investment decision making. We consider two types of criteria. In the first type, we deal with an investor whose interest is to maximize profits and risk-adjusted profits of a portfolio long in stocks. The second type criteria is to consider an investor who worries about potential large losses and wishes to add a Value-at-Risk (VaR) evaluation to her trading strategy. Based on the one-step ahead forecast of returns provided by our mixture model, we design a trading rule that will be compared to a set of different trading rules within the framework of White (2000). It will be shown that the proposed trading rule is superior to the other trading rules.

The organization of the paper is as follows. In section 2, we provide the joint modelling of asset returns and jumps in VCR. We present the estimation results for the weekly returns of the constituents of the SP500 index. In section 3, we assess the out-of-sample performance of our model. We explain the trading rules, forecast evaluation criteria, and the statistical framework to compare different trading rules. Finally, in section 4 we conclude.[1]

2 Jumps in Cross-sectional Rank and Asset Returns

In this section, our purpose is to build an econometric model that combines time information and cross-sectional information about firm-specific returns. First we discuss the model for the cross-sectional jump $f_1(J_{i,t}|\mathcal{F}_{t-1})$ and then the model for the returns $f_2(y_{i,t}|J_{i,t}, \mathcal{F}_{t-1})$.

A natural starting point to incorporate the rank in our empirical analysis is to understand its most basic time series properties. In our sample of weekly returns for the constituents of the SP500 index, which will be explained in more detail in the following sections, we have found that there is no linear dependence either in $z_{i,t}$ or in $(z_{i,t} - \bar{z}_t)^2$ and consequently, further modeling of the ARMA-GARCH type of $z_{i,t}$ have yielded insignificant parameter estimates.[2] It might be possible to find a

---

[1] Due to constraints on the size of the manuscript, we also provide a supplementary appendix that contains intermediate results and further discussion about the methodology used in the paper.

[2] For only eight firms in our estimation sample the p-value of the Box-Pierce-Ljung Q(4) for $z_{i,t}$ is smaller than 5%. Within the 466 firms, the average p-value is 0.48, the minimum is 0.001 and the maximum is 0.99. For $(z_{i,t} - \bar{z}_t)^2$ the minimum p-value of the LM tests for ARCH(4) is 0.06, the maximum 0.98 and the average p-value is 0.41. These results seem to indicate that there is no statistical evidence of linear dependence in the first and second moments of

non-linear model for $z_{i,t}$ but we will not pursue this route as we do not have much guidance from any theoretical or empirical model of the cross-sectional ranking in the present literature. Instead, as we mention in the introduction, we have observed that while some assets tend to stick to the upper or lower ranks (above or below the median) some others tend to move around very frequently. For example in the empirical section we find that the stocks in the information/technology sector tend to jump more often as compared to the stocks in the utility sector. Our objective is to model the persistence/dynamics of the VCR in a non-linear fashion, by introducing a duration variable that measures how long the asset return remains in the upper or lower ranks.

For these reasons, in (2) we define a sharp jump as a binary variable that takes the value one when the there is a minimum (upward or downward) movement of 0.5 in the ranking of asset i on going from time $t-1$ to t, and zero otherwise. The choice of the magnitude of the jump is not arbitrary. The sharpest jump that we can consider is 0.5. In every time period, we need to allow for the possibility of a jump, either up or down, in the following period regardless of the present rank 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 rank is the result of the interaction of the relative movements of all individual assets in the market.

Our objective is to model the joint conditional 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 may contain the past realizations of returns, jumps, and VCRs. To simplify notation, we drop the subindex i, but in the following analysis it 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

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

the VCR.

Let us call $\mathcal{L}_1(\theta_1) = \sum_{t=1}^{T} \ln f_1(J_t|\mathcal{F}_{t-1};\theta_1)$ and $\mathcal{L}_2(\theta_2) = \sum_{t=1}^{T} \ln f_2(y_t|J_t,\mathcal{F}_{t-1};\theta_2)$. The maximization of the joint log-likelihood function can be achieved by maximizing $\mathcal{L}_1(\theta_1)$ and $\mathcal{L}_2(\theta_2)$ separately without loss of efficiency by assuming that the parameter vectors θ_1 and θ_2 are "variation free" in the sense of Engle *et al.* (1983).

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

On modelling $f_1(J_{i,t}|\mathcal{F}_{t-1})$, our paper also connects with the 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 question in duration analysis is the length of time between two events given some information set. In this paper, the event is the jump in VCR of the asset return. When we model the expected duration between jumps (or its mirror image $p_{i,t}$), our analysis is performed in calendar time, as in Hamilton and Jordà (2002), to allow the analysis of information that is reported in a calendar basis.

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 it is customary in duration models. The question of interest is, what is the probability of a jump at time t in the VCR of the asset return of firm i 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 is the conditional probability of a jump. From (3), we note that p_t 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 VCR of the asset return in relation to the other firms in the market. Furthermore, because $J_t = 1$, p_t assesses the possibility of being in the upper ranks (winner) or in the lower ranks (loser) of the cross-sectional distribution of asset returns.

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 rank is

given by $\Psi_{N(t)} = \sum_{j=1}^{\infty} j(1-p_t)^{j-1} p_t = p_t^{-1}.$[3] 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. The ACH model is a calendar-time version of the autoregressive conditional duration (ACD) of Engle and Russell (1998). In both ACD and ACH models, the expected duration is a linear function of lag durations. However as the ACD model is set up in event time, there are some difficulties on how to introduce information that arrives between events. This is not the case in the ACH model because the set-up is in calendar time. In the ACD model, the forecasting object is the expected time between events; in the ACH model, the objective is to forecast the probability that the event will happen tomorrow given the information known up to today. 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, as we mentioned above, working in calendar time has the advantage that we can incorporate information that becomes available between jumps and can affect the probability of a jump in future periods. We specify the 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 VCRs and past returns. Given a sample of observed jumps in VCR, the log-likelihood function for $\theta_1 = (\alpha', \beta', \delta')'$ is

$$\mathcal{L}_1(\theta_1) = \sum_{t=1}^{T} \ln f_1(J_t | \mathcal{F}_{t-1}; \theta_1) = \sum_{t=1}^{T} [J_t \ln p_t(\theta_1) + (1 - J_t) \ln(1 - p_t(\theta_1))]. \tag{6}$$

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

We assume that the return to asset i may behave differently depending upon the occurrence of a jump. We distinguish between active and passive stocks depending on the reasons why the jump in the asset ranking has occurred. A sharp jump in the VCR may happen because: (i) the asset has had an unusual return (an active movement in the asset ranking), (ii) the overall market may have moved (up or down) in the opposite direction from the asset return (a passive movement in the asset ranking), and (iii) a combination of active and passive movements. If a jump has occurred,

---

[3]Note that $\sum_{j=0}^{\infty}(1-p_t)^j = p_t^{-1}$. Differentiating with respect to p_t yields $\sum_{j=0}^{\infty} -j(1-p_t)^{j-1} = -p_t^{-2}$. Multiplying by $-p_t$ gives $\sum_{j=0}^{\infty} j(1-p_t)^{j-1} p_t = p_t^{-1}$ and thus $\sum_{j=1}^{\infty} j(1-p_t)^{j-1} p_t = p_t^{-1}$

the return was pushed either towards the lower tail or upper tail of the cross-sectional distribution of returns, and in relation to the market, this asset becomes either a loser or a winner. However, when $J_t = 1$, being a winner (loser) does not imply extraordinary positive (negative) returns; in addition, as an asset return can be in the top (bottom) of the cross-sectional distribution of returns and be negative (positive). Likewise, an asset may not have experienced any jump $J_t = 0$, and experience a large return, positive or negative. The modelling of two potential different states (whether a jump has occurred or not) will permit to differentiate whether the conditional expected return is driven by active or/and passive movements in the asset ranking in conjunction with its own return dynamics.

A priori, we hypothesize different dynamics in these two states. A general specification is

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

where $\mu_{j,t} = \int y_{t+1} \cdot f_2(y_t|J_t = j, \mathcal{F}_{t-1}; \theta_2) dy_{t+1}$ and $\sigma_{j,t}^2 = \int y_{t+1}^2 \cdot f_2(y_t|J_t = j, \mathcal{F}_{t-1}; \theta_2) dy_{t+1} - \mu_{j,t}^2$ ($j = 1, 0$). Whether these two states are present in the data is an empirical question and it will be answered through statistical testing. The normal density in each state is an assumption which may be modified depending upon the performance of the model. For instance, a thick-tailed density or a skewed density may be needed if there is remaining kurtosis or skewness left in the data. Standard diagnostic tests will shed further light on the need to modify the chosen density.

The log-likelihood function $\mathcal{L}_2(\theta_2) = \sum_{t=1}^{T} \ln f_2(y_t|J_t, \mathcal{F}_{t-1}; \theta_2)$ is

$$\mathcal{L}_2(\theta_2) = \sum_{t=1}^{T} \ln \left[\frac{J_t}{\sqrt{2\pi\sigma_{1,t}^2}} \exp\left\{ -\frac{1}{2}\left(\frac{y_t - \mu_{1,t}}{\sigma_{1,t}}\right)^2 \right\} + \frac{1 - J_t}{\sqrt{2\pi\sigma_{0,t}^2}} \exp\left\{ -\frac{1}{2}\left(\frac{y_t - \mu_{0,t}}{\sigma_{0,t}}\right)^2 \right\} \right],$$

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

2.3   A mixture model for expected returns

If the two proposed states are granted in the data, the marginal density function of the asset return must be a mixture of two normal density functions where the mixture weights are given by the probability of jumping p_t:

$$\begin{aligned} g(y_t|\mathcal{F}_{t-1}; \theta) &\equiv \sum_{J_t=0}^{1} f(y_t, J_t|\mathcal{F}_{t-1}; \theta) \\ &= \sum_{J_t=0}^{1} f_1(J_t|\mathcal{F}_{t-1}; \theta_1) f_2(y_t|J_t, \mathcal{F}_{t-1}; \theta_2) \\ &= p_t \cdot f_2(y_t|J_t = 1, \mathcal{F}_{t-1}; \theta_2) + (1 - p_t) \cdot f_2(y_t|J_t = 0, \mathcal{F}_{t-1}; \theta_2), \end{aligned} \tag{8}$$

as $f_1(J_t|\mathcal{F}_{t-1};\theta_1) = p_t^{J_t}(1-p_t)^{(1-J_t)}$. Therefore, the one-step ahead forecast of the return is

$$E(y_{t+1}|\mathcal{F}_t;\theta) = \int y_{t+1} \cdot g(y_{t+1}|\mathcal{F}_t;\theta)dy_{t+1} = p_{t+1}(\theta_1) \cdot \mu_{1,t+1}(\theta_2) + (1 - p_{t+1}(\theta_1)) \cdot \mu_{0,t+1}(\theta_2). \quad (9)$$

The expected return is a function of the probability of jumping p_t, 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 December 27, 2005 for all the constituents of the SP500 index. The full sample is split into the estimation sample that runs for the first eleven years, from January 1, 1990 to December 27, 2000 (with $R = 573$ weeks), and the prediction sample from January 2, 2001 to December 27, 2005 (with $P = 260$ weeks). The set of firms in the SP500 index is not fixed; over time new firms are coming into the index and some others drop out. Consequently, the sample size for each individual firm may differ depending on the date in which it became a member of the SP500. For instance, Microsoft is a constituent of the SP500 for all our sample period but Yahoo, which was created years after 1990, enters in the SP500 in April 1996. The sample size for the in-sample estimation analysis varies across firms with a maximum in-sample size of 573 weekly returns. The only restriction that we impose in the estimation sample arises from the need of having a minimum of observations to carry out the model estimation. We require at least four years of data (208 weekly returns) prior to December 27, 2000. This requirement constrains the in-sample analysis to 466 companies. However, new firms will show up during the prediction sample and some other will drop out. We also account for these changes in the constituents of the SP500 in the out-of-sample forecasting.

In the supplementary appendix (Table S.1), we summarize the unconditional moments (mean, standard deviation, skewness, and kurtosis) of all 500 firms in the estimation sample. The frequency distribution of the unconditional mean is unimodal with a weekly mean return of 0.029%. For the unconditional standard deviation, the median value is 5.25%. The coefficient of skewness is predominantly negative with a median value of -0.12. All the firms have a large coefficient of kurtosis with a median value of 10.34. We calculate the Box-Pierce-Ljung statistics up to the fourth order to test for autocorrelation in returns and we find mild autocorrelation for about one-third of the firms. However, the Box-Pierce-Ljung test up to the fourth order to test for autocorrelation in squared returns indicates strong dependence in second moments for all the firms in the SP500 index.

8

2.4.1 Estimation Results for $f_1(J_t|\mathcal{F}_{t-1}; \theta_1)$

For 466 firms, we fit a conditional duration model as in (4) and (5). The information set consists of past durations, past returns and past VCRs : $\{D_{N(t)-j}, y_{t-j}, z_{t-j}; \ j = 1, 2, \ldots\}$. 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 $\theta_1 \equiv (\alpha', \beta', \delta')'$. Based on standard model selection criteria (t-statistics and log-likelihood ratio tests), we select the following final specification

$$
\begin{aligned}
p_t &= [\Psi_{N(t-1)} + \delta' X_{t-1}]^{-1} \\
\Psi_{N(t)} &= \alpha D_{N(t)-1} + \beta \Psi_{N(t)-1} \\
\delta' X_{t-1} &= \delta_1 + \delta_2 y_{t-1} \mathbf{1}(z_{t-1} \leq 0.5) + \delta_3 y_{t-1} \mathbf{1}(z_{t-1} > 0.5)
\end{aligned}
\tag{10}
$$

The conditional duration model is an ACH(1,1). There is a nonlinear effect of the predetermined variables on duration. The effect of past returns on duration depends on whether the VCR of the asset is above or below the median.

In the supplementary appendix (Table S.2), we report the cross-sectional frequency distributions of the estimates $\hat{\theta}_1 \equiv (\hat{\alpha}, \hat{\beta}, \hat{\delta}_1, \hat{\delta}_2, \hat{\delta}_3)'$ for all the 466 firms in the estimation sample. All the parameters are statistically significant at the customary 5% level. For $\hat{\alpha}$, the median is 0.34 with 90% of the firms having an $\hat{\alpha}$ below 0.47. For $\hat{\beta}$, its frequency distribution is highly skewed to the right with a median of 0.07 and with 90% of the firms having a $\hat{\beta}$ below 0.30. The median $\hat{\alpha} + \hat{\beta}$ is 0.42 and for 90% of the firms, the $\hat{\alpha} + \hat{\beta}$ is below 0.67. The estimates $\hat{\delta}_2$ and $\hat{\delta}_3$ have mostly opposite signs, the former is predominantly positive and the latter is predominantly negative. The effect of $\hat{\delta}_2$ and $\hat{\delta}_3$ in expected duration depends on the interaction between the VCR and the sign of the return. There are four possible scenarios. For instance, for most of the cases when $\hat{\delta}_2$ is positive and $\hat{\delta}_3$ is negative, if the past asset return is positive and below (above) the median market return, its expected duration is longer (shorter) and the probability of a 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 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 0.57 with 90% of the firms having a $\hat{\delta}_2$ below 1.67. For $\hat{\delta}_3$, the median value is -0.64 with 90% of the firms having a $\hat{\delta}_3$ above -1.83. Roughly speaking, for a representative firm with median parameter estimates, the expected duration is approximately

between 3 and 4 weeks, and since $E(p_t) \geq [E(\Psi_{N(t-1)} + \delta' X_{t-1})]^{-1}$, a lower bound for the expected probability of a jump is between 25 and 33%.

In Table 1A we 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.[4] The largest share corresponds to the Consumer Goods sector with 25.2% of the SP500 companies, and the Information Technology sector with 17.7% of the firms. The smallest share corresponds to the Energy sector with 5.5% of the firms. *Ceteris paribus,* the larger $\hat{\alpha} + \hat{\beta}$, the longer the expected duration and the lower the probability of a jump. The Information Technology sector has the smallest $\hat{\alpha} + \hat{\beta} = 0.33$ while the Utilities sector has the largest $\hat{\alpha} + \hat{\beta} = 0.50$, indicating that the former has a higher tendency to move from the lower cross-sectional ranks to the upper and vice versa. On the contrary, the Utilities sector is relatively more stable. In the columns labeled $\hat{\delta}_2$ and $\hat{\delta}_3$, we report the median impact of the calendar variables on the probability of a jump. As we mention above, the median $\hat{\delta}_2$ is strictly positive, and the median $\hat{\delta}_3$ is strictly negative and larger in magnitude than $\hat{\delta}_2$ for all sectors but one (Finance). Most importantly, it is the joint effect of $(\hat{\alpha}, \hat{\beta}, \hat{\delta}_2, \hat{\delta}_3)$, which is summarized in the column labeled $\hat{p}_t$ the median probability of a jump in each sector. Not surprisingly, the Information Technology sector has the largest median probability with $\hat{p}_t = 0.45$, which means that about every 2.5 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 $\hat{p}_t = 0.13$, which implies jumps every 7.5 weeks approximately.

2.4.2 Estimation Results for $f_2(y_t | J_t, \mathcal{F}_{t-1}; \theta_2)$

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. The selected specification of (7) is

$$f_2(y_t | J_t, \mathcal{F}_{t-1}; \theta_2) = \begin{cases} N(\mu_{1t}, \sigma_{1t}^2) & \text{if } J_t = 1 \\ N(\mu_{0t}, \sigma_{0t}^2) & \text{if } J_t = 0 \end{cases}, \tag{11}$$

$$\mu_{1t} \equiv E(y_t | \mathcal{F}_{t-1}, J_t = 1) = \nu_1 + \gamma_1 y_{t-1} + \eta_1 z_{t-1},$$

$$\mu_{0t} \equiv E(y_t | \mathcal{F}_{t-1}, J_t = 0) = \nu_0 + \gamma_0 y_{t-1} + \eta_0 z_{t-1},$$

$$\sigma_{1t}^2 = \sigma_{0t}^2 = \sigma_t^2 = E(\epsilon_t^2 | \mathcal{F}_{t-1}, J_t) = \omega + \rho \epsilon_{t-1}^2 + \tau \sigma_{t-1}^2,$$

---

[4]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.

10

where $\epsilon_{t-1} = (y_{t-1} - \mu_{1,t-1})J_{t-1} + (y_{t-1} - \mu_{0,t-1})(1 - J_{t-1})$ and $\theta_2 = (\nu_1, \gamma_1, \eta_1, \nu_0, \gamma_0, \eta_0, \omega, \rho, \tau)'$. We arrive to this specification by sequentially implementing a battery of likelihood ratio tests. We aim to gather statistical evidence for or against the mixture of normals that we propose in (7). The first null hypothesis of interest states the same dynamics in the conditional mean for both states, i.e. $H_0^1 : \nu_1 = \nu_0, \gamma_1 = \gamma_0, \eta_1 = \eta_0$, where in both the restricted and the unrestricted models we assume a constant conditional variance, $\sigma_{1t}^2 = \sigma_{0t}^2 = \sigma^2$. This null is rejected very strongly for all the firms in the SP500 index.[5] Next, we relax the assumption of constant variance across states and write a second null hypothesis as $H_0^2 : \nu_1 = \nu_0, \gamma_1 = \gamma_0, \eta_1 = \eta_0$, where, in both the restricted and the unrestricted models, we assume $\sigma_{1t}^2 = \sigma_{0t}^2 = \sigma_t^2$ with σ_t^2 specified as in (11). For all firms, we reject again very strongly this null hypothesis and hence, we conclude that there is statistical evidence in favor of different dynamics in the conditional mean across states jointly with a time-varying conditional variance. Following the rejection of H_0^1 and H_0^2, we test for equal conditional variances maintaining the nonlinearity in the conditional mean, i.e. $\mu_{1t} \neq \mu_{0t}$ as in (11), $H_0^3 : \sigma_{1t}^2 = \sigma_{0t}^2$. In the unrestricted model, the conditional mean is specified as in (11) and the conditional variances follow a GARCH(1,1) process with different parameters depending on $J_t = 1$ or $J_t = 0$. We fail to reject the third hypothesis to finally entertain a model as in (11). Within this model, the well known unconditional leptokurtosis of asset returns is explained by a location-mixture of normals with time-varying conditional variances.

The estimation results for the 466 firms are summarized in the supplementary appendix (Table S.3), where we report the cross-sectional frequency distributions of the parameters estimates in the conditional mean and conditional variance. All parameters are statistically significant at the 5% level. When we consider asset returns for which a jump has taken place, the marginal impact of past returns, $\hat{\gamma}_1$, is predominantly negative, with a median value of -0.44, though there is wide range of values $(-1.64, 0.52)$. The negative sign is expected for "active" stocks in which the jump in ranking is mainly associated with a reversal in its past returns, as opposed to "passive" stocks in which the jump is mainly associated with movements in other asset returns. The marginal effect of past VCRs, $\hat{\eta}_1$, is also predominantly negative, with a median value of -0.38 and a wide range of $(-2.15, 1.54)$. For a given asset, and keeping everything else equal, a negative (positive) sign seems to indicate that the jump in VCR is on average associated with a decrease (increase) in expected returns. Eventually the combination of both effects will determine the expected return. We can say that on average an asset for which $\hat{\gamma}_1$ and $\hat{\eta}_1$ are set to the median values, $|\hat{\gamma}_1| = 0.44 > |\hat{\eta}_1| = 0.38$,

---

[5] We do not report all the testing results for the 466 firms but they are available upon request.

should be considered an "active" stock. When there is no jump, the marginal effect of past returns, $\hat{\gamma}_0$, is predominantly positive with a median value of 0.36 such that positive (negative) returns are followed by positive (negative) returns. As for the marginal effect of past VCRs, $\hat{\eta}_0$, it seems to be small with a median value of 0.03 and of either sign.

In Table 1B, we report the median values of the parameter estimates for every industrial sector represented in the SP500 index. We observe that $\hat{\gamma}_1$ and $\hat{\eta}_1$ are both negative and, for most sectors, $|\hat{\gamma}_1| > |\hat{\eta}_1|$. In the no-jump state, $\hat{\gamma}_0$ and $\hat{\eta}_0$ are both mostly positive with $\hat{\gamma}_0$ significantly larger than $\hat{\eta}_0$. The model for the conditional variance 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 0.04 and a median value for $\hat{\tau}$ of 0.89. A leverage effect in the conditional variance does not seem to be warranted as the different specifications of the conditional mean across states take care of potential asymmetries in returns.

We run standard diagnostic checks in the standardized residuals corresponding to model (11). A summary of these tests over the 466 firms follows:

	residual mean	residual std.dev.	residual skewness	residual kurtosis	p-value of $Q(4)$	p-value of $Q^2(4)$
mean	0.00	1.00	0.01	3.02	0.61	0.89
median	0.00	0.99	0.01	3.01	0.67	0.99
max	0.00	1.19	0.36	4.57	0.99	1.00
min	0.00	0.82	0.00	2.66	0.00	0.12

The specification (11) passes standard diagnostic checks for model adequacy, which provide strong support for the mixture of normals. The model performs extremely well on modelling the reported skewness and kurtosis of the unconditional distribution (Table S.1), delivering standardized residuals that are symmetric and have a kurtosis of 3. The p-values of the Box-Pierce-Ljung $Q(4)$ and McLeod-Li $Q^2(4)$ statistics of order four for the standardized residuals and squared standardized residuals are above the 5% significance level for all the firms but one, concluding that there is not significant linear dependence left in the data.[6] However, given the nonlinearity of the model, a more drastic check on the adequacy of the model is to assess its out-of-sample forecasting performance, which we analyze in the following section.

---

[6] As pointed out by Hong and Lee (2003) and Chen (2007), the asymptotic null distributions of the Box-Pierce-Ljung and McLeod-Li statistics will be affected by the estimation of the standardized residuals. Therefore the p-values reported here should be interpreted with caution and taken just as a rough guide for diagnostic checking.

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 of the returns based on the VCR model in Equations (10) and (11). This trading strategy will be called *VCR-Mixture 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. However, given that large profits can be generated at the expense of engaging in high risk strategies, we account for this possibility by assessing excess asset returns per unit of risk. To this end, we consider two objective functions, one based on the traditional Sharpe ratio (SR) and the other based on a modified Sharpe ratio (MSR). In the second scenario, we consider a situation where the investor wishes to assess potential large losses by adding a Value-at-Risk (VaR) evaluation of her trading strategy.

3.1 VCR-Mixture Trading Rule

We proceed as follows. For each firm i in the market (466 firms), we compute the one-step ahead forecast of the return as in (9)

$$\hat{y}_{i,t+1}(\hat{\theta}_t) = p_{t+1}(\hat{\theta}_{1,t}) \cdot \hat{\mu}_{1,t+1}(\hat{\theta}_{2,t}) + (1 - p_{t+1}(\hat{\theta}_{1,t})) \cdot \hat{\mu}_{0,t+1}(\hat{\theta}_{2,t}), \quad t = R, \dots, T-1, \qquad (12)$$

where $\hat{\theta} = (\hat{\theta}_1', \hat{\theta}_2')'$, $\hat{\theta}_1 = (\hat{\alpha}, \hat{\beta}, \hat{\delta}_1, \hat{\delta}_2, \hat{\delta}_3)'$, and $\hat{\theta}_2 = (\hat{\nu}_1, \hat{\gamma}_1, \hat{\eta}_1, \hat{\nu}_0, \hat{\gamma}_0, \hat{\eta}_0, \hat{\omega}, \hat{\rho}, \hat{\tau})'$.[7] Based on the forecasted returns $\{\hat{y}_{i,t+1}(\hat{\theta}_t)\}_{t=R}^{T-1}$, the investor predicts the VCR of all assets in relation to the overall market, that is,

$$\hat{z}_{i,t+1} = M^{-1} \sum_{j=1}^{M} \mathbf{1}(\hat{y}_{j,t+1} \leq \hat{y}_{i,t+1}), \quad t = R, \dots, T-1, \qquad (13)$$

and buys the top K performing assets if their return is above the risk-free rate. In every subsequent out-of-sample period ($t = R, \dots, T-1$), 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

---

[7]The sequence of one-step ahead forecasts is obtained with a "rolling" sample of size R. 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 $\hat{y}_{i,T}(\hat{\theta}_{T-1})$.

portfolio return

$$\pi_{t+1} = K^{-1} \sum_{j=1}^{M} y_{j,t+1} \cdot \mathbf{1}(\hat{z}_{j,t+1} \geq z_{t+1}^{K}), \quad t = R, \ldots, T-1, \tag{14}$$

where z_{t+1}^{K} is the cutoff cross-sectional rank 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% ($K = 5$ stocks) performers in the SP500 index. Every asset in the portfolio is weighted equally.[8]

The out-of-sample forecast runs from January 2, 2001 through December 27, 2005. By the very nature of the SP500 index, over time some firms are added and some drop, consequently the index needs to be updated periodically. During the forecasting period, there were 105 companies that were added to the index replacing the same number of companies in the process. To avoid any survivorship bias, at every point in time in the out-of-sample analysis we deal with 500 companies. We consider the dates in which the companies are added and deleted from the SP500 index. If a company is deleted on a given week, it will no longer be included in the out-of-sample forecast evaluation from that week onwards. Similarly, if a new company is added on a given week, we estimate the duration and expected returns models for this company and we will include it in the forecast evaluation from that week onwards. At any point in time, we face the possibility of choosing, by the prediction of the VCR-Mixture Trading Rule, a stock for the optimal portfolio that may drop from the SP500 index in the next period. However, in our analysis, we have not encountered such a possibility and the one-period-ahead predicted optimal portfolio always contains companies that are constituents of the SP500 index in the following period.

3.2 Competing trading rules

To evaluate the out-of-sample performance of the VCR model, we compare it with that of various competing models. The second trading rule is a simple alternative to the VCR-Mixture rule which is constructed by imposing $\mathbb{H}_0^1 : \mu_{1,t} = \mu_{0,t} = \mu_t$ in (11). This trading rule has no mixture and thus enable to assess the importance of the nonlinearity in the VCR-Mixture Trading Rule. It will be called *VCR Trading Rule* because μ_t depends on the lagged VCR z_{t-1} of an asset while it ignores the mixture of two normal densities. 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 VCRs. As in the previous rule, the rolling sample scheme is used to obtain the sequence of

---

[8]We also carried out the analysis for the top 2% (10 stocks) performers. As expected, the realized profit was smaller but the risk-weighted profit (captured by the Sharpe ratio) did not change significantly. Thus, the qualitative nature of our conclusions remains the same.

one-step ahead forecasts $\hat{y}_{i,t+1}$. The ordinal rank is predicted by (13) and the investor follows the same strategy of (14) as before by 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.*

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

In addition to the above three models, we also consider four classes of *technical* trading rules considered by Sullivan, Timmermann and White (1999): *Filter-Rule, Moving-Average-Rule, Channel-Break-Out-Rule, and Support-and-Resistance-Rule.* All of these four trading rules are based on the SP500 index and they can be considered rules that exploit the momentum in returns. For each of the four technical trading rules, we consider four parameterizations. See the supplementary appendix for further explanation.

3.3   Forecast evaluation criteria

The first type of evaluation criterion is to compute the return of each trading strategy over the forecasting 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. We also correct MTR according to the level of risk of the chosen portfolio. We consider two broad types of risk-corrected criteria: one is based on the Sharpe ratio, and the other based on VaR calculations to manage catastrophic losses. The criterion based on the traditional Sharpe ratio is given by the excess return per unit of risk measured by the standard deviation of the selected portfolio

$$SR = P^{-1} \sum_{t=R}^{T-1} \frac{(\pi_{t+1} - r_{f,t+1})}{\sigma_{t+1}^{\pi}(\hat{\theta}_t)},$$

15

where $r_{f,t+1}$ is the risk free rate. A modified Sharpe ratio (MSR) can also be implemented when the excess return is measured per unit of VaR. It is defined as

$$MSR = P^{-1} \sum_{t=R}^{T-1} \frac{(\pi_{t+1} - r_{f,t+1})}{VaR_{t+1}^{\alpha}(\hat{\theta}_t)},$$

where $VaR_{t+1}^{\alpha}(\hat{\theta}_t)$ is the one-step ahead VaR forecast of π_{t+1} at a given nominal tail coverage probability α. The above three evaluation criteria, MTR, SR, MSR, are to be maximized.

The second type of evaluation criteria will be useful to an investor who wishes to control for catastrophic events by maintaining a minimum amount of capital to cushion against excessive losses. We would like to evaluate each trading rule according to their ability to allocate the optimal amount of capital for unlikely events. For this purpose a Value-at-Risk evaluation criterion is useful. Consider a portfolio of assets whose realized return is given by π_{t+1}. We are interested in $VaR_{t+1}^{\alpha}(\theta)$, 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 $\Pr[\pi_{t+1} \leq VaR_{t+1}^{\alpha}(\theta)|\mathcal{F}_t] = \alpha$. For the VCR Trading Rule and the Buy-and-Hold-the-Market Trading Rule, where we assume a location-scale distribution of π_{t+1}, the forecast of the portfolio VaR can be estimated as $VaR_{t+1}^{\alpha}(\hat{\theta}_t) = \mu_{t+1}^{\pi}(\hat{\theta}_t) + \Phi_{t+1}^{-1}(\alpha)\sigma_{t+1}^{\pi}(\hat{\theta}_t)$, where $\mu_{t+1}^{\pi}(\hat{\theta}_t), \sigma_{t+1}^{\pi}(\hat{\theta}_t)$ are the forecasts of the portfolio return and conditional standard deviation respectively, $\Phi_{t+1}(\cdot)$ 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 the VCR-Mixture 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), which is described in some detail in the supplementary appendix.

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 the tick function which evaluates which trading rule provides the best quantile forecast.

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

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

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 VaR-based 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 the binary variable $\mathbf{1}(\pi_{t+1} < VaR_{t+1}^\alpha(\hat{\theta}_t))$. Over the forecasting period, consider the following counts $n_1 = \sum_{t=R}^{T-1} \mathbf{1}(\pi_{t+1} < VaR_{t+1}^\alpha(\hat{\theta}_t))$ and $n_0 = \sum_{t=R}^{T-1} \mathbf{1}(\pi_{t+1} \geq VaR_{t+1}^\alpha(\hat{\theta}_t))$. Note that $P = T - R = n_0 + n_1$. If the VaR^α has been correctly forecasted, it must be that $n_1 = P \times \alpha$ and $n_0 = P \times (1 - \alpha)$. The predictive log-likelihood function of α given a sample $\left\{ \mathbf{1}(\pi_{t+1} < VaR_{t+1}^\alpha(\hat{\theta}_t)) \right\}_{t=R}^{T-1}$ is $\mathcal{L}(\alpha) = \ln\left(\alpha^{n_1}(1-\alpha)^{n_0}\right)$ and the maximum likelihood estimator of α is $\hat{\alpha} = n_1/P$. If we were to test for the null hypothesis that $E[\mathbf{1}(\pi_{t+1} \leq VaR_{t+1}^\alpha(\hat{\theta}_t))] = \alpha$, the likelihood ratio test $-2(\mathcal{L}(\alpha) - \mathcal{L}(\hat{\alpha}))$ would be a suitable statistic. The loss function V_2 is based on this statistic, as it is a distance measure between α and $\hat{\alpha}$. A trading rule that minimizes V_2 will be preferred.

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

The third VaR-based loss function V_3 is the tick function used in quantile estimation (Koenker and Bassett, 1978)

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

The trading rule that provides the smallest V_3 is preferred.

The question of interest is, among a set of trading rules, which one is the best? Each rule produces different forecasts that are evaluated according to the six objective functions introduced above. The best trading rule is the one that provides the maximum of MTR, SR, MSR and the minimum of V_1, V_2, and V_3. To compare the trading rules in terms of the six criteria we use the "reality check" proposed by White (2000) and modified by Hansen (2005). Given a benchmark trading rule, we aim to compare the values of the evaluation function produced by the other trading rules to that of the benchmark. We formulate a null hypothesis that says that the best of the competing trading rules is no better than the benchmark rule. If we reject the null hypothesis, there is at least one competing trading rule that produces a better value than the benchmark. A brief sketch of the formal testing procedure is provided in the supplementary appendix.

3.4 Evaluation of trading rules

The out-of-sample performance of the aforementioned trading rules is provided in Table 2. In Panel A, the trading rules are evaluated according to the MTR function; in Panel B, according to SR and MSR; and in Panel C, according to the three VaR-based loss functions. In all cases, the maximum in-sample size for the rolling estimation is $R = 573$, and the out-of-sample forecasting period has $P = 260$ weeks. The stationary bootstrap of Politis and Romano (1994) with the bootstrap smoothing parameter 0.25 (corresponding to the mean block length of 4) is implemented with 1000 bootstrap resamples. In the first column of each table, we report the benchmark trading rule to which the remaining rules will be compared.

In Panel A, we report the value of MTR for each trading strategy. The VCR-Mixture Trading Rule produces a weekly mean trading return of 0.93% that is significantly more than the next most favorable rule, which is the Buy-and-Hold-the-Market Trading Rule with a weekly mean return of -0.02%. We also find that all the technical trading rules are clearly dominated by the VCR-Mixture rule. In each row, a benchmark rule is compared with all the remaining 18 rules. When the VCR-Mixture Rule is the benchmark, the White's reality check p-value is 1.000 and Hansen's p-value is 0.956, indicating that it is not dominated by any of the other 18 trading rules. When any other rule is taken as a benchmark, these reality check p-values are less than 1%, indicating that they are easily dominated.

Following upon some of the criticisms of the profitability of momentum strategies, the superior MTR of the VCR-Mixture 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. To assess the return-risk trade off, we implement the Sharpe ratio and a modified version of it. In Panel B, the largest SR (largest mean return per unit of standard deviation) is provided by the VCR-Mixture rule with a weekly return of 0.22%, which is lower than the mean return provided by the same rule under the MTR criterion, but still a dominant return when compared to the mean returns provided by the VCR Trading Rule and Buy-and-Hold-the-Market Trading Rule. The White p-value is 1.000 and Hansen's p-value is 0.954 indicating the superiority of the VCR-Mixture rule. The results for the Modified Sharpe ratio (MSR, mean return per unit of VaR) are qualitatively identical. MSR is the largest for the VCR-Mixture rule with a weekly return of 0.793 (for the 1% VaR) and 0.881 (for the 5% VaR). In both cases the White's p-values are 1.000 and Hansen's p-values are above 0.90.

In Panel C, we report the out-of-sample performance of the three trading rules evaluated accord-

ing to the VaR-based loss functions V_1, V_2, and V_3, for $\alpha = 0.01, 0.05$. The results for $\alpha = 0.01, 0.05$ are virtually identical for all the three loss functions. With respect to V_1, the VCR-Mixture Trading Rule dominates the other two rules as it provides the least amount of required capital. When we consider V_2, the same results hold. The VCR-Mixture rule delivers the best tail coverage by estimating a tail coverage probability of $\hat{\alpha} = 0.013$ at a nominal rate $\alpha = 0.01$, and $\hat{\alpha} = 0.05$ at a nominal rate of $\alpha = 0.05$. Finally, with respect to the tick function V_3, the reality check p-values indicate that neither the Buy-and-Hold-the-Market strategy nor the VCR-Mixture strategy are dominated.

In addition to the six forecast criteria reported in Table 2, 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 VCRs, $P^{-1} \sum_{t=R}^{T-1} M^{-1} \sum_{i=1}^{M} (\hat{z}_{i,t+1} - z_{i,t+1})^2$. Based on these MSFE losses, we compare the VCR-Mixture Trading Rule and the VCR Trading Rule. With the VCR 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 Trading Rule is better than the VCR Trading Rule and thus the importance of nonlinearity in the conditional mean of the return process due to the weighted mixture of normal densities.

3.5 Betas, average cumulative profits, and transaction costs

We have calculated the beta of the selected portfolios, through a CAPM-type time series regression, over each period of the forecasting interval ($P = 260$ weeks). For the winner portfolio, in 70% of the forecasting periods the average beta of the portfolio chosen by the VCR-Mixture Trading Rule is greater than 1, and in 14% of the periods is greater than 1.5. In Figure 2, we plot the evolution of the portfolio beta over the forecasting period for the VCR-Mixture Trading Rule and the VCR Trading Rule. There is no observable pattern. The VCR-Mixture Trading Rule tends to pick up stocks over the *full* spectrum of risk with portfolio betas as low as 0.18 and as high as 1.79.

In Figure 3, we plot the average cumulative profits over the forecasting period from the three aforementioned trading rules. With the exception of a few periods at the beginning of the forecasting sample, the weekly average return provided by the VCR-Mixture Trading Rule clearly dominates the other two rules.

Since the VCR-Mixture Trading Rule and the VCR Trading Rule involve dynamic adjustments of the portfolio in every week of the forecasting sample, it is natural to ask whether the transaction costs involved will render these strategies unprofitable. Given that the Buy-and-Hold-the-Market

strategy does not incur into any transaction cost, we calculate the threshold transaction cost associated with the VCR-Mixture strategy such that an investor is indifferent between the VCR-Mixture portfolio or holding the SP500 index. We assume that the transaction cost is a certain percentage (c_t) of the price paid to buy or sell a certain portfolio. Let P_{it} and r_{pt} be the price and excess return of the VCR-Mixture portfolio at time t respectively, and let r_{mt} be the excess return obtained by holding the SP500 index. The *threshold* (or *break-even*) transaction cost c_t is obtained by solving the equation that equates returns at every time period t, that is, $\ln\left[P_{i,t}(1-c_t)\right] - \ln\left[P_{i,t-1}(1+c_t)\right] = r_{mt}$. Hence, $c_t = \frac{1-\exp(D_{it})}{1+\exp(D_{it})}$, where $D_{it} \equiv r_{mt} - r_{pt}$. The average and the median values of c_t over the out-of-sample forecasting period is 17% and 36% respectively. By any industry standards these costs are exorbitant. Thus, transaction costs are not a deterrent to the implementation of the VCR-Mixture Trading Rule.

4 Conclusion

We have proposed a trading rule that is based on a non-linear time series model for expected returns. The novelty of the proposed modeling lies on the investigation of the dynamics of the cross-sectional rank of asset returns, which is conducted within the context of a duration model. For the weekly returns of the constituents of the SP500 index, we have modelled the *joint* dynamics of the cross-sectional rank and the asset return by analyzing (1) the *marginal* probability distribution of a jump in the cross-sectional rank, and (2) the probability distribution of the asset return *conditional* on a jump. As a result, we claim that the expected return is generated by a mixture of normal densities weighted by the probability of jumping. Though the proposed model passes a battery of standard diagnostics and given the non-linear nature of the model we judge the adequacy of our specification within a forecasting exercise. Based on the one-step ahead forecast of the mixture model, we have designed a trading rule that is evaluated, over the forecasting sample, within the context of several evaluation functions: trading return, risk-adjusted trading return (Sharpe ratios), and VaR-based loss functions. When we compare our trading rule based on the proposed VCR-mixture model, against eighteen other different trading rules (some model-based and some technical), we find that in all cases considered it is the dominant rule.

References

Chen, Y.-T. (2007), "A Unified Approach to Standardized-Residuals-Based Correlation Tests for GARCH-type Models", *Journal of Applied Econometrics,* forthcoming.

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.

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

Hansen, Peter R. (2005), "A Test for Superior Predictive Ability," *Journal of Business and Economic Statistics,* 23, 365-380

Hong, Yongmiao and Tae-Hwy Lee (2003), "Diagnostic Checking for Adequacy of Nonlinear Time Series Models", *Econometric Theory*, 19(6), 1065-1121.

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

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

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

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. In-Sample Estimation

Panel A
Median Values of the Parameter Estimates of q_1
in the ACH Model for $f_1(J_t \mid \mathfrak{I}_{t-1}; q_1)$ in Equation (10)

Industry Sectors in the SP500 Index	% of firms	$\hat{a} + \hat{b}$	$\hat{d}_2$	$\hat{d}_3$	$\hat{p}_t$
Consumer Goods	25.2	0.456	0.542	-0.573	0.257
Energy	5.5	0.359	0.769	-0.840	0.347
Finance	16.5	0.444	0.959	-0.488	0.257
Health Care	11.2	0.387	0.501	-0.672	0.326
Industrials	11.2	0.451	0.586	-0.946	0.238
Information Technology	17.7	0.334	0.318	-0.413	0.447
Material	6.4	0.437	0.587	-0.690	0.290
Utilities	6.4	0.497	0.851	-1.101	0.132
All sectors	100.0	0.422	0.571	-0.641	0.267

Panel B
Median Values of the Parameter Estimates of q_2
in the Model for $f_2(y_t \mid J_t, \mathfrak{I}_{t-1}; q_2)$ in Equation (11)

Industry Sectors in the SP500 Index	% of firms	$\hat{g}_1$	$\hat{g}_0$	$\hbar_1$	$\hbar_0$	$\hat{r} + \hat{t}$
Consumer Goods	25.2	-0.445	0.345	-0.375	-0.010	0.909
Energy	5.5	-0.591	0.384	-0.252	0.022	0.940
Finance	16.5	-0.422	0.347	-0.369	0.058	0.972
Health Care	11.2	-0.319	0.365	-0.658	0.031	0.899
Industrials	11.2	-0.415	0.338	-0.358	0.056	0.885
Information Technology	17.7	-0.419	0.480	-0.422	0.015	0.920
Material	6.4	-0.444	0.353	-0.473	-0.036	0.965
Utilities	6.4	-0.575	0.136	-0.069	0.208	0.909
All sectors	100.0	-0.440	0.358	-0.378	0.029	0.925

TABLE 2. Out-of-Sample Evaluation of Trading Rules

Panel A
Mean Trading Return (MTR)

	MTR	White p-value	Hansen p-value
VCR-Mixture Rule	0.926	1.000	0.956
VCR Rule	-0.297	0.001	0.000
Buy-and-Hold-the-Market Rule	-0.019	0.001	0.000
Filter-Rule (0.05)	-0.052	0.001	0.000
Filter-Rule (0.10)	-0.032	0.001	0.000
Filter-Rule (0.20)	-0.023	0.009	0.004
Filter-Rule (0.50)	-0.081	0.002	0.001
Moving-Average Rule (10, 2)	-0.034	0.003	0.000
Moving-Average Rule (20, 2)	-0.076	0.001	0.000
Moving-Average Rule (10, 4)	-0.049	0.001	0.000
Moving-Average Rule (20, 4)	-0.047	0.001	0.000
Channel-Break-Out Rule (4, 0.05)	-0.055	0.001	0.000
Channel-Break-Out Rule (10, 0.05)	-0.076	0.000	0.000
Channel-Break-Out Rule (4, 0.10)	-0.089	0.000	0.000
Channel-Break-Out Rule (10, 0.10)	-0.060	0.001	0.000
Support-and-Resistance Rule (2)	-0.124	0.002	0.000
Support-and-Resistance Rule (4)	-0.054	0.001	0.000
Support-and-Resistance Rule (8)	-0.080	0.004	0.001
Support-and-Resistance Rule (16)	-0.082	0.001	0.000

Panel B
Sharpe Ratio (SR) and Modified Sharpe Ratio (MSR)

		White p-value	Hansen p-value
Sharpe Ratio	*SR*		
VCR-Mixture Rule	0.218	1.000	0.954
VCR Rule	0.076	0.020	0.121
Buy-and-Hold-the-Market Rule	0.019	0.022	0.109
Modified Sharpe Ratio with $a = 0.01$	*MSR*		
VCR-Mixture Rule	0.793	1.000	0.923
VCR Rule	0.030	0.000	0.000
Buy-and-Hold-the-Market Rule	0.008	0.000	0.000
Modified Sharpe Ratio with $a = 0.05$	*MSR*		
VCR-Mixture Rule	0.881	1.000	0.912
VCR Rule	0.039	0.000	0.000
Buy-and-Hold-the-Market Rule	0.010	0.000	0.000

Three VaR Based-Loss Functions

	V_1	White p-value	Hansen p-value	V_2	$\hat{a}$	White p-value	Hansen p-value	V_3	White p-value	Hansen p-value
VaR with $a = 0.01$										
VCR-Mixture Rule	2.748	1.000	0.879	0.002	0.013	1.000	0.999	0.070	0.832	0.712
VCR Rule	9.172	0.000	0.000	0.089	0.085	0.000	0.000	0.478	0.000	0.000
Buy-and-Hold-the-Market	5.467	0.000	0.000	0.005	0.023	0.000	0.000	0.090	0.618	0.418
VaR with $a = 0.05$										
VCR-Mixture Rule	1.724	1.000	0.945	0.000	0.050	1.000	0.929	0.255	0.259	0.145
VCR Rule	6.973	0.000	0.000	0.049	0.138	0.002	0.001	0.604	0.000	0.000
Buy-and-Hold-the-Market	4.103	0.000	0.000	0.000	0.050	1.000	0.929	0.121	0.973	0.861

Notes: The out-of-sample period is $P=260$ and the in-sample period is $R=573$. In each row of the panels, we report the values of the forecast evaluation functions together with the reality-check p-values of White (2000) and Hansen (2005). The null hypothesis is that no other trading rules are better than the selected benchmark. A large reality-check p-value indicates that the null hypothesis cannot be rejected. In Panel A, MTR represents the profit accrued from the respective trading rules. When the VCR-Mixture Rule is the benchmark, the reality check p-value 1.000 means that this benchmark rule is not dominated by any of the other 18 trading rules. In Panel B, SR represents the average profit per unit of standard deviation accrued from the respective trading rules. MSR represents the average profit scaled by VaR with different tail probabilities (0.01 or 0.05). When the VCR-Mixture Rule is the benchmark, the reality check p-value is again 1.000 which means this benchmark rule is not dominated by any of the other two trading rules. In Panel C, V_1 , V_2 , and V_3 represent the three VaR-based loss functions (MRC, coverage failure rate, and the tick loss). $\hat{a}$ denotes the empirical failure rate at the nominal rate a . When the VCR-Mixture Rule is the benchmark, the p-values are very large for all V_1 , V_2 and V_3 , implying that this benchmark trading rule is not dominated by the other two trading rules.

FIGURE 1
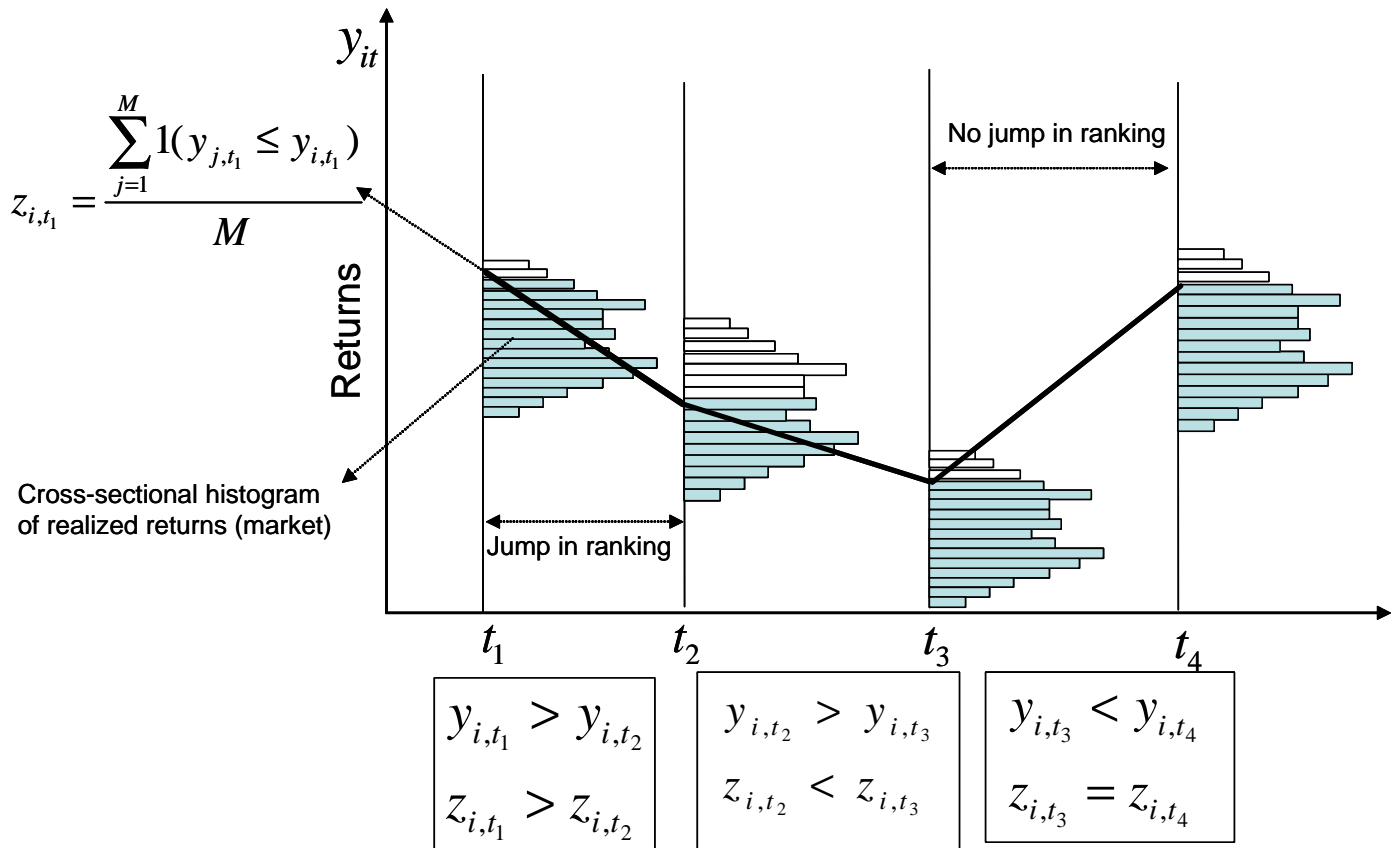**Stylized Description of the Modeling Problem**

$$z_{i,t_1} = \frac{\sum_{j=1}^{M} 1(y_{j,t_1} \leq y_{i,t_1})}{M}$$

y_{it}

Returns

No jump in ranking

Cross-sectional histogram
of realized returns (market)

Jump in ranking

t_1 t_2 t_3 t_4

$y_{i,t_1} > y_{i,t_2}$

$z_{i,t_1} > z_{i,t_2}$

$y_{i,t_2} > y_{i,t_3}$

$z_{i,t_2} < z_{i,t_3}$

$y_{i,t_3} < y_{i,t_4}$

$z_{i,t_3} = z_{i,t_4}$

FIGURE 2
Beta of the Winner Portfolio
Forecasting sample: January 2, 2001 to December 27, 2005
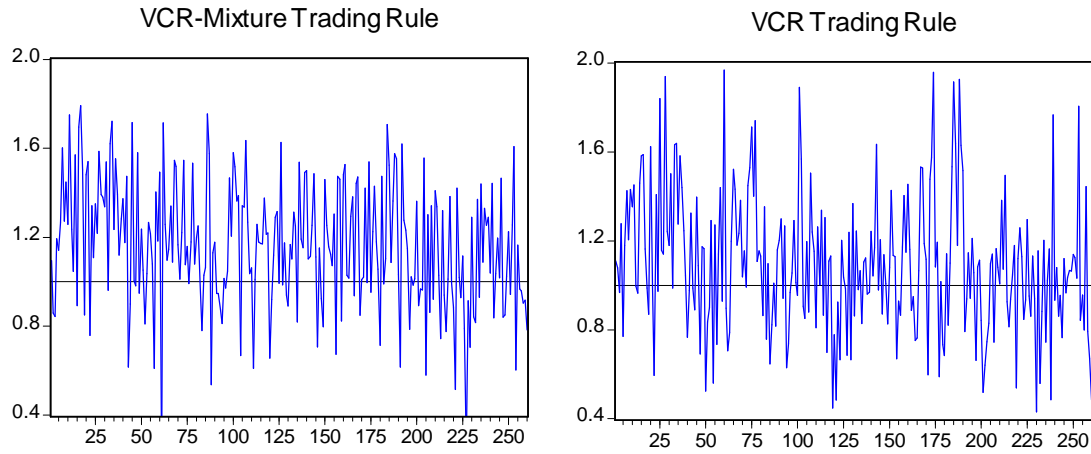
VCR-Mixture Trading Rule

VCR Trading Rule

FIGURE 3
Average Cumulative Percentage Profits
Forecasting sample: January 2, 2001 to December 27, 2005

VCR-Mixture Position Trading Rule
Position Trading Rule
Buy-and-Hold Trading Rule