RESEARCH

Swiss Finance Institute Research Paper Series N°21-37

Disasters, Large Drawdowns, and Long-term Asset Management

Eric Jondeau
University of Lausanne and Swiss Finance Institute

Alexandre Pauli
University of Lausanne and Swiss Finance Institute

Disasters, Large Drawdowns, and Long-term Asset Management

Eric Jondeau[a] and Alexandre Pauli[b]

June 2021

Abstract

Long-term investors are often reluctant to invest in assets or strategies that can suffer from large drawdowns. A major challenge for such investors is to gain access to predictions of large drawdowns in order to precisely design strategies minimizing these drawdowns. In this paper, we describe a multivariate Markov-switching model framework that allows us to predict large drawdowns. We provide evidence that three regimes are necessary to capture the negative trends in expected returns that generate large drawdowns, and we correctly predict conditional drawdowns. In addition, investment strategies based on these models outperform model-free strategies based on the empirical distribution of drawdowns. These results hold within and out of the sample.

Keywords: Large drawdowns, Stock-market returns, Markov-switching model, Portfolio allocation model.

JEL Classification: C53, G15.

[a]Swiss Finance Institute and University of Lausanne, Faculty of Business and Economics (HEC Lausanne), CH 1015 Lausanne, Switzerland. E-mail: eric.jondeau@unil.ch.

[b]Swiss Finance Institute and University of Lausanne, Faculty of Business and Economics (HEC Lausanne), CH 1015 Lausanne, Switzerland. E-mail: alexandre.pauli@unil.ch.

1 Introduction

In recent decades, the recurrence of disasters (severe macroeconomic and financial crises, climate change, the COVID-19 pandemic) has raised the issue of protecting investors' portfolios from large market drawdowns. In the first quarter of 2020, the 34% decline in the U.S. equity market was one of the five largest one-quarter drawdowns in the last century. These large drawdowns are of particular importance for long-term asset managers, such as insurers, pension funds, or sovereign funds. Some asset managers are forced to take organizational measures when their portfolio experiences a large drawdown. For this reason, they may be willing to pay a premium to avoid such large losses or include a drawdown objective in their portfolio optimization process. However, modeling and predicting the temporal evolution of large drawdowns are particularly challenging because such drawdowns usually develop over relatively long periods of time, a time scale that is not consistent with most standard financial econometric models. For this reason, the literature mostly relies on nonparametric (or model-free) measures to predict large drawdowns but does not address the modeling issue.

In this paper, we address both modeling and prediction issues. We describe a model for financial returns based on regime switches that enables us to predict large market drawdowns. We demonstrate that three regimes are necessary to capture the long-term dynamics of financial returns. With such a model, a long-term asset manager could easily implement an investment strategy that minimizes the expected value of a large drawdown measure. We also provide empirical evidence that investing according to this strategy effectively allows investors to reduce their losses in large market downturns compared with using standard GARCH-type models, which fail to predict large drawdowns. This strategy also outperforms the nonparametric approach based on the historical distribution of large drawdowns.

Rare disasters were initially introduced in macroeconomic and financial models to analyze asset-pricing puzzles (Rietz, 1988; Barro, 2006; Barro and Ursua, 2008; Barro, 2009).[1] Dis-

[1] This approach has then been extended fruitfully by several authors by allowing time variability and predictability with respect to the frequency or severity of disasters (Gabaix, 2012; Gourio, 2012, 2013; Wachter, 2013; Hasler and Marfe, 2016, among others). See Tsai and Wachter (2015) for a review of rare disaster frameworks in asset pricing.

asters were typically viewed as one-period shocks to the economy, corresponding to wars and economic crises. More recently, similar approaches were adopted to analyze the consequences in financial markets of disasters arising from climate change (Karydas and Xepapadeas, 2019; Barnett et al., 2020; Engle et al., 2020) or from the COVID-19 pandemic (Pagano et al., 2020).

Disasters and the associated financial market crashes exhibit properties that differ from the standard measures of extreme losses usually used in asset management. In this context, the risk related to extreme losses is in general described by measures of downside risk such as Value-at-Risk, expected shortfall, or tail risk (Ang and Bekaert, 2002; Ang et al., 2006; Bollerslev and Todorov, 2011; Kelly and Jiang, 2014). These downside risk measures complement or replace the portfolio variance as an investment criterion to mitigate the effects of large losses.[2] These measures, however, correspond to acute shocks, whereas large drawdowns develop over relatively long periods (one quarter or longer). One difficulty in dealing with large drawdowns is that they are long-term path dependent, a property that most statistical models in finance fail to describe.

The literature defines several concepts of large drawdowns. A widely-accepted measure is the period maximum drawdown (MDD), which corresponds to the largest loss from peak to trough in a given period. Investment strategies based on MDD are analyzed by Grossman and Zhou (1993) in a continuous-time framework or by Reveiz and Leon (2008), who use nonlinear optimization techniques to solve the MDD optimization problem. Chekhlov et al. (2005) define the concept of conditional drawdown (CDD), which corresponds to an average of the largest drawdowns in a given period and includes the average drawdown (ADD) and the MDD as limiting cases. More recently, Goldberg and Mahmoud (2016) define the conditional expected drawdown (CED), which corresponds to an average of the largest period MDD over a long sample. These measures have interesting properties, as CDD and CED are convex measures of risk and as such can be reduced by portfolio diversification.

[2]For instance, Alexander and Baptista (2004) consider a mean-variance criterion with Value-at-Risk or expected shortfall constraints, which correspond to large one-period losses.

Most of these papers use nonparametric (or model-free) approaches to measure large drawdowns. In particular, Chekhlov et al. (2005) simulate scenarios based on historical data to predict next-period drawdowns. This approach is likely to work well for the sample but may not be well-suited when conditions vary through time, such as in climate change, which may increase the probability of occurrence of disasters. Another difficulty when working with large drawdowns is that most econometric models fail to capture large losses accumulating over relatively long periods. For instance, GARCH models, even with nonnormal distributions, are not able to generate the negative trend that we observe in financial markets during crises. Consequently, reproducing large drawdowns similar to those observed in the 1999 dotcom crisis, the 2008 subprime crisis, or the 2020 COVID-19 pandemic becomes highly challenging with these models. In contrast, Markov-Switching (MS) models can generate large losses associated with a crisis because they allow for different drifts in market returns across regimes. When a given return process enters a bear regime, it accumulates negative values for relatively long periods of time. Papers describing these models for financial returns often overlook this property because their main focus is on the implied nonnormality (as in Guidolin and Timmermann, 2008) or the volatility dynamics (as in Klaassen, 2002, or Haas et al., 2004). Peng and Kim (2020) recently proposed an MS model with two regimes for measuring large drawdowns over a 10-day horizon. We provide evidence that three regimes are needed for long horizons to capture the occurrence of financial crises and the ensuing drawdowns.

Our objective is to assess the ability of MS models to predict large drawdowns and to evaluate the out-of-sample performance of the competing models. For this reason, we use a long sample of daily returns of small caps and large caps. The sample covers more than 90 years, of which the last 30 years are used for the out-of-sample evaluation.[3] We estimate multivariate MS-GARCH models with one, two, or three regimes and different distributions for the innovation process. We then run two investment exercises. In the first exercise (full-sample analysis), we estimate the models over the full sample and perform a unique,

[3]Such a long sample would not be necessary in practice to estimate MS models and to use them for predicting subsequent large drawdowns.

in-sample allocation that corresponds to the average characteristics of the data. For a given model, we simulate drawdown-based risk measures with horizons ranging from one to four quarters and determine the optimal allocation for an investor minimizing the drawdown risk of the portfolio. Although this exercise may suffer from a forward-looking bias, it provides evidence that predictions based on three-regime models perform as well as predictions based on the model-free approach in sample. As this approach relies on the empirical distribution of large drawdowns, it provides an unbiased estimate of their true (unknown) distribution in sample. In the second exercise (out-of-sample analysis), we estimate the various models over subsamples and allocate the portfolio for the subsequent period. This out-of-sample experiment is rather time-consuming, but it enables us to mimic an investor's allocation process in real time. In addition, it would penalize overparameterized models and therefore mitigate the winner's curve problem (Hansen, 2009).

We find that models with three regimes provide the best predictions of large drawdowns in the in-sample and out-of-sample experiments. When we allocate the investor's wealth by minimizing the predicted large drawdowns in the subsequent period, the ex post drawdown of the portfolio is in most cases lower with the three-regime model than it is with the GARCH model or the two-regime model. In the three-regime model, one of the regimes has large negative expected returns, which allows us to generate large windfalls consistent with actual crises. In contrast, standard GARCH models do not generate such large drawdowns. Investors allocating their portfolio using a three-regime model tend to be invested in long large caps and short small caps, an allocation that turns out to be the right choice for reducing subsequent large drawdowns. Our results also demonstrate that the three-regime model usually outperforms the model-free approach out of sample. We also compare the performances of models based on normal or Student's t innovations. We find that the three-regime model with Student's t innovations systematically outperforms the model with normal innovations.

The remainder of the paper is organized as follows. In Section 2, we summarize the definition and construction of the large drawdown measures and the investor's decision problem.

In Section 3, we briefly present the multivariate MS-GARCH model and explain how to use these models to predict large drawdowns and construct a portfolio that minimizes these drawdowns. Section 4 describes the data. Section 5 exposes the empirical results with respect to the asset allocation problem when we consider a unique, in-sample allocation based on the model's estimation over the full sample. Section 6 discusses the empirical findings on the out-of-sample asset allocation problem. Section 7 concludes the paper.

2 Large Drawdowns and the Portfolio Problem

2.1 Definitions

The maximum drawdown is the most widely-used concept of large drawdowns. It represents the maximum loss from peak to trough over a given time period. As MDD increases for longer time series, it is customary to measure MDD over a given fixed time period of one quarter or one year, for example. This section defines the various concepts of large drawdowns that we will analyze in the remainder of the paper.

Let $p_t = (p_{t,1}, \cdots, p_{t,H})$ be a sample path associated with the stochastic process P with a continuous and strictly increasing distribution, where $p_{t,h}$ denotes the log-price on day h in period t (e.g., a given quarter or year).[4] The drawdown within this path on day h corresponds to the return loss between the last peak and the current price:

$$DD_{t,h} = \max_{1 \le j \le h} p_{t,j} - p_{t,h}.$$

We denote the drawdowns within the sample path as $DD_t = (DD_{t,1}, \cdots, DD_{t,H})'$. The maximum drawdown is defined as the largest drawdown:

$$MDD_t = \max_{1 \le h \le H} DD_{t,h}.$$

[4]For simplicity, we assume here a fixed number of days H per period. In the empirical analysis, we will use the actual number of days per period.

This measure suffers from limitations: It is an extreme statistic and as such it is highly sensitive to outliers. For this reason, it may not be well-suited for use in an optimization process. In addition, it is not a coherent measure of risk, according to the definition by Artzner et al. (1999).

Chekhlov et al. (2003, 2005) define the conditional drawdown (CDD) as the average of the largest drawdowns in a given period exceeding a quantile of the drawdown distribution, which mitigates the impact of outliers. For a probability θ, CDD is given by the average of the worst $(1-\theta)\times 100\%$ drawdowns. Formally, we define the drawdown threshold $Th_\theta(DD_t)$ as the θ-quantile of the drawdown distribution:

$$Th_\theta(DD_t) = \inf\{s \mid \Pr(DD_t > s) \leq 1-\theta\}.$$

CDD corresponds to the tail conditional expectation, i.e., the average of the drawdowns above the threshold:

$$CDD_{\theta,t} = E_t[DD_t \mid DD_t > Th_\theta(DD_t)], \tag{1}$$

where the expectation is applicable over the sample path. When $\theta = 0$, $CDD_{\theta,t}$ is equal to $E_t[DD_t]$, defined as the average drawdown (ADD_t). When $\theta \to 1$, $CDD_{\theta,t}$ coincides with MDD_t. For a given sample path p_t, we have $ADD_t \leq CDD_{\theta,t} \leq MDD_t$.

Finally, Goldberg and Mahmoud (2015, 2016) introduce the concept of conditional expected drawdown (CED) as the average of the MDDs exceeding a quantile of the MDD distribution. For a probability $\tilde{\theta}$, $CED_{\tilde{\theta}}$ corresponds to the average of the worst $(1-\tilde{\theta})\times 100\%$ maximum drawdowns. Consequently, the threshold $Th_{\tilde{\theta}}(MDD)$ is determined by the $\tilde{\theta}$-quantile of the MDD distribution:

$$Th_{\tilde{\theta}}(MDD) = \inf\{s \mid \Pr(MDD > s) \leq 1-\tilde{\theta}\}$$

and the $CED_{\tilde{\theta}}$ is therefore given by

$$CED_{\tilde{\theta}} = E[MDD \mid MDD > Th_{\tilde{\theta}}(MDD)], \tag{2}$$

where the expectation is taken over the full sample. When $\tilde{\theta} = 0$, $CED_{\tilde{\theta}}$ is equal to $E[MDD]$.

Therefore, CDD and CED are comparable in nature to the well-known expected shortfall of the return distribution, except that they are obtained as the tail mean of the drawdown distribution and the tail mean of the maximum drawdown distribution, respectively. Importantly, as shown by Chekhlov et al. (2005) and Goldberg and Mahmoud (2015), CDD and CED satisfy the properties of deviation measures, i.e., positivity, shift invariance, positive homogeneity, and convexity. Convexity of a measure of risk implies that this measure can be reduced by diversification and used in quantitative optimization. As a consequence, if an investor minimizes this measure, the minimum, if it exists, is a global minimum.

Our subsequent analysis will focus on the four large drawdown measures: ADD, CDD, MDD, and CED. An essential feature of these measures is that drawdowns can occur over different time frames.[5] To cope with this feature, we consider three different horizons H, corresponding to one quarter, two quarters, and four quarters.

2.2 Empirical Measures of Large Drawdowns

Evaluating the performance of an asset or a portfolio of assets in terms of its ex post drawdown requires empirical measurement. We match the horizon of the drawdown measures to a long-term investor's horizon H. For instance, an asset manager rebalancing the portfolio every quarter is interested in large drawdowns occurring throughout the quarter. Therefore, the full sample is divided into T nonoverlapping subsamples of length H, with the sequence of log-prices in subsample t given by $p_t = (p_{t,1}, \cdots, p_{t,H})$ for $t = 1, \cdots, T$. For each subsample, we compute the vector of drawdowns $DD_t = (DD_{t,1}, \cdots, DD_{t,H})$.

[5] A drawdown may span over a few days (as for the COVID-19 crisis, with a 36.4% drawdown in 24 days) or over a window of more than a year (as for the subprime crisis, with a 59% drawdown in 355 days).

We obtain the drawdown-based measures for each subsample as follows. The period ADD is simply given by the sample mean of the drawdowns: $ADD_t = \frac{1}{H}\sum_{h=1}^{H} DD_{t,h}$, and the MDD is given by the maximum drawdown of the subsample: $MDD_t = \max_{1\leq h\leq H} DD_{t,h}$. CDD is calculated as the average of the drawdowns over the θ-quantile:

$$CDD_{\theta,t} = \frac{1}{(1-\theta)H}\sum_{h=1}^{H} DD_{t,h}\ I_{(DD_{t,h}>Th_{\theta,t})}, \tag{3}$$

where $Th_{\theta,t} = \inf\{s \mid \frac{1}{H}\sum_{h=1}^{H} I_{(DD_{t,h}>s)} \leq 1-\theta\}$.

CED is based on the distribution of period MDDs over the full sample: we collect the maximum drawdowns over all the subsamples $MDD = (MDD_1, \cdots, MDD_T)$ and the sample CED corresponds to the average of the worst $(1-\tilde{\theta})\times 100\%$ MDDs over the full sample:

$$CED_{\tilde{\theta}} = \frac{1}{(1-\tilde{\theta})T}\sum_{t=1}^{T} MDD_t\ I_{(MDD_t>Th_{\tilde{\theta}})}, \tag{4}$$

where $Th_{\tilde{\theta}} = \inf\{s \mid \frac{1}{T}\sum_{t=1}^{T} I_{(MDD_t>s)} \leq 1-\tilde{\theta}\}$.

2.3 Investor's Problem

We now consider the investment strategy of a long-term investor with investment horizon H. At the end of period t, n risky assets are available. We denote by $r_{i,t+1,h}$ the log-return of asset i on day h of the period $t+1$. The vector of cumulated log-returns over h days is denoted by $R_{t+1,h} = \{R_{i,t+1,h}\}_{i=1}^{n}$, where $R_{i,t+1,h} = \sum_{j=1}^{h} r_{i,t+1,j}$. Portfolio weights, determined at the end of period t, are denoted by $\alpha_t = (\alpha_{1,t}, \cdots, \alpha_{n,t})$ with $\sum_{i=1}^{n} \alpha_{i,t} = 1$. The (unknown) value of the portfolio at the end of period $t+1$ is $P_{t+1}(\alpha_t) = \alpha_t' \exp(R_{t+1,H})$ in the absence of rebalancing during period $t+1$. The sequence of daily log-values of the portfolio in period $t+1$ is given by: $p_{t+1}(\alpha_t) = (p_{t+1,1}(\alpha_t), \cdots, p_{t+1,H}(\alpha_t))$, where $p_{t+1,h}(\alpha_t) = \log P_{t+1,h}(\alpha_t)$ and $P_{t+1,h}(\alpha_t) = \alpha_t' \exp(R_{t+1,h})$.

For a given weight vector α_t, we compute daily log-values in period $t+1$ and obtain large drawdown measures as described in Section 2.2. The investment criterion consists in minimizing the expected value of one of the large drawdown measures for the next investment period (period ADD, CDD, MDD, or CED). If we denote these measures generically as $XDD_{t+1}(\alpha_t)$, the objective function is:[6]

$$\min_{\{\alpha_t\}} \quad E_t[XDD_{t+1}(\alpha_t)]. \tag{5}$$

To obtain predictions of portfolio large drawdown measures, i.e., $E_t[XDD_{t+1}(\alpha_t)]$, we proceed as follows: first, we define a data generating process (DGP) for daily log-returns to endogenize the path dependence of the return process. We assume a multivariate MS-GARCH model, thereby capturing the drawdown trend if switching probabilities are sufficiently low. Using this DGP, we simulate assets' daily returns for the next investment period. Then, for a given portfolio weight vector, we obtain a simulated path of the portfolio return from which we deduce the large drawdown measures. This approach provides us with predictions of the large drawdown measures for the next investment period as a function of the weight vector, thereby enabling us to pinpoint the optimal weight that minimizes the objective function (Equation (5)). The details of our approach are the subject of the next section.

3 Methodology

3.1 Multivariate MS-GARCH Model

To capture the possible impact of a large drawdown on the performance of the long-term portfolio, we assume a multivariate MS-GARCH model for the return process. The vector

[6] We also considered the multiperiod mean-large drawdown criterion, for a risk aversion parameter λ, as follows: $\max_{\{\alpha_t\}} \; E_t[R_{p,t+1}(\alpha_t)] - \frac{\lambda}{2} E_t[XDD_{t+1}(\alpha_t)]$. The minimization problem corresponds to the case where λ goes to infinity.

of daily log-returns for the n assets is denoted by $\tilde{r}_{d+1} = (\tilde{r}_{1,d+1}, \cdots, \tilde{r}_{n,d+1})$. The temporal index $d = 1, \cdots, D$, represents days and runs over the full sample.[7]

The model is written as follows:

$$\tilde{r}_{d+1} = \mu_{d+1}(S_{d+1}) + \varepsilon_{d+1},$$

where $\mu_{d+1}(S_{d+1})$ is the vector of expected returns, conditional on state S_{d+1}, and ε_{d+1} is the vector of unexpected returns. It is defined as:

$$\varepsilon_{d+1} = \Omega_{d+1}(S_{d+1})^{1/2} z_{d+1},$$

where $\Omega_{d+1}(S_{d+1})$ denotes the $(n \times n)$ covariance matrix of unexpected returns and z_{d+1} is a sequence of iid innovations with distribution $D(0, I_n)$ with zero mean and identity covariance matrix.

States are defined by the Markov chain $\{S_{d+1}\}$ with K regimes and transition matrix $P = (p_{kk'})_{k,k'=1,\cdots,K}$, where the transition probabilities are $p_{kk'} = \Pr(S_{d+1} = k'|S_d = k)$, $k, k' \in \{1, \cdots, K\}$.

Expected returns are constant within each state: $\mu_{d+1}(S_{d+1}) = \mu^{(k)}$ when $S_{d+1} = k$.[8] The covariance matrix $\Omega_{d+1}(S_{d+1})$ is time- and state-dependent. In a given state k, it is driven by a multivariate GARCH process with state-dependent conditional correlation matrix, as in Pelletier (2006) or Haas and Liu (2018). The conditional variance of asset i in state k is

[7] The model is estimated over a long sample of daily returns, $r = (\tilde{r}_1, \cdots, \tilde{r}_D)$, where D is the number of days in the full sample. In contrast, large drawdown measures are computed over relatively short subsamples (e.g., one quarter or one year) with H days, which we have denoted by $r_t = (r_{t,1}, \cdots, r_{t,H})$, $t = 1, \cdots, T$ in Section 2. Since we use nonoverlapping subsamples, we also have $r = (r_1, \cdots, r_T)$, where $D = H \times T$.

[8] Assuming an autoregressive process would have a very limited effect for a long-term investment objective because persistence of financial returns is low. The first-order autocorrelation of the market return is equal to 0.05 over the 1926–2020 period and equal to -0.06 over the 1990–2020 period.

defined as a standard univariate GARCH(1,1) process:[9]

$$\sigma_{i,d+1}^{(k)2} = \omega_i^{(k)} + \alpha_i^{(k)}(\tilde{r}_{i,d} - \mu_i^{(k)})^2 + \beta_i^{(k)}\sigma_{i,d}^{(k)2},$$

with different parameters for each state, as in Haas et al. (2004).

The $(n \times n)$ correlation matrix is constant in a given state: $\Gamma^{(k)} = \left(\rho_{ij}^{(k)}\right)_{i,j=1,\cdots,n}$, so that the covariance matrix $\Omega_{d+1}(S_{d+1}) = \Omega_{d+1}^{(k)}$ in state k is:

$$\Omega_{d+1}^{(k)} = D_{d+1}^{(k)1/2}\ \Gamma^{(k)}\ D_{d+1}^{(k)1/2},$$

where $D_{d+1}^{(k)}$ is the diagonal matrix with $\sigma_{i,d+1}^{(k)2}$ on the diagonal.

We consider two types of multivariate distributions for innovations z_{d+1}: a Gaussian distribution $N(0, I_n)$ and a standardized Student's t distribution $t(0, I_n, \nu)$, where ν denotes the degree of freedom. The choice of the innovation distribution may matter for two reasons. First, in the model, lower returns can be captured by a higher probability of a bear market, or by lower expected return, or by a lower degree of freedom of the Student's t distribution. Therefore, allowing for Student's t innovations may change the dynamics of regime shifts. Second, for a standard mean-variance investor, we do not expect the distribution of the innovation process to matter. However, as soon as investors care about higher moments (such as for an investment criterion including a concern about large drawdowns), the choice of the innovation distribution may play an important role.

To make inferences about the regimes, we calculate the probability of being in each regime. We denote by $f(\tilde{r}_{d+1} \mid \underline{\tilde{r}}_d, \theta)$ the distribution of the daily log-return process conditional upon past log-returns, with $\underline{\tilde{r}}_d = \{\tilde{r}_d, \tilde{r}_{d-1}, \cdots\}$ and θ denoting the vector of unknown parameters. Parameters include expected returns ($\mu^{(k)}$), volatility parameters ($\omega^{(k)}$, $\alpha^{(k)}$, $\beta^{(k)}$), correlations ($\Gamma^{(k)}$), probabilities ($p_{kk'}$), and the degree of freedom (ν). Using Hamilton (1989)'s filter, we obtain the predicted probabilities $\pi_{k,d+1} = \Pr[S_{d+1} = k \mid \underline{\tilde{r}}_d]$ and the

[9] In this expression, we follow the suggestion of Klaassen (2002) and Haas et al. (2004) and define shocks with respect to a given state using the conditional mean $\mu_i^{(k)}$ instead of the unconditional mean μ_i adopted by Gray (1996).

filtered probabilities $\phi_{k,d+1} = \Pr[S_{d+1} = k \mid \underline{\tilde{r}}_{d+1}]$ as:

$$\pi_{d+1} = P\ \phi_{d+1} \qquad \text{and} \qquad \phi_{d+1} = \frac{\pi_{d+1} \odot l_{d+1}}{e'(\pi_{d+1} \odot l_{d+1})},$$

$$\text{with } l_{d+1} = \begin{bmatrix} f(\tilde{r}_{d+1} \mid \underline{\tilde{r}}_d, S_{d+1} = 1; \theta) \\ \vdots \\ f(\tilde{r}_{d+1} \mid \underline{\tilde{r}}_d, S_{d+1} = K; \theta) \end{bmatrix} \quad \text{and} \quad e = (1, \cdots, 1).$$

The estimation of the model is based on standard likelihood maximization, where the log-likelihood is defined as: $\log L_D(\theta) = \sum_{d=1}^{D-1} f(\tilde{r}_{d+1} \mid \underline{\tilde{r}}_d, \theta) = \sum_{d=1}^{D-1} \log\left(\sum_{k=1}^{K} \pi_{d+1} \odot l_{d+1}\right)$. We impose stationarity conditions as described by Haas et al. (2004) and Abramson and Cohen (2007) in the univariate case and Haas and Liu (2018) in the multivariate case.

3.2 Minimizing the Expected Large Drawdown of a Portfolio

In some specifications of the multivariate MS model, analytical formulas for portfolio characteristics are available. For instance, Guidolin and Timmermann (2003, 2008) provide formulas for the high-order moments in a model with regime-dependent (but time-independent) means and variances. Other nonlinear characteristics, such as the VaR or the expected shortfall of the portfolio, cannot be computed analytically, even in this simple model (Guidolin and Timmermann, 2003). Additionally, in models such as MS-GARCH models, analytical expressions are usually not available because variances are path dependent. For this reason, we use Monte Carlo simulations to compute expected large drawdowns.

For ease of exposition, we again assume a quarterly investment horizon, with H representing the number of days in a quarter. We solve the allocation problem at the end of quarter t through the following steps:

1. We estimate the parameters of the MS-GARCH model using daily log-returns available in quarters $1, \cdots, t$. The last day of the estimation period is denoted by $d = T \times H$. The next quarter, $t+1$, contains days $d+1, \cdots, d+H$.

2. For a given estimated model, we simulate Q samples of length H of daily log-returns for the n assets: $\{r_{t+1,h}^{(q)}\}_{h=1}^{H}$, $q = 1, \cdots, Q$. As the probability of being in state k at the end of period t is given by predicted probabilities $\{\pi_{d+1}^{(k)}\}_{k=1}^{K}$, we simulated a fraction $\pi_{d+1}^{(k)}$ of the draws using $\Omega_{d+1}^{(k)}$ as an initial condition for the covariance matrix in period $t+1$. From the simulated daily log-returns, we compute cumulative log-returns in quarter $t+1$ as: $R_{t+1,h}^{(q)} = \sum_{j=1}^{h} r_{t+1,j}^{(q)}$ for $h = 1, \cdots, H$.

3. For a portfolio weight vector α_t, we transform the simulations to daily log-values of the portfolio: $p_{t+1}^{(q)}(\alpha_t) = (p_{t+1,1}^{(q)}(\alpha_t), \cdots, p_{t+1,H}^{(q)}(\alpha_t))$, where $p_{t+1,h}^{(q)}(\alpha_t) = \log(P_{t+1,h}^{(q)}(\alpha_t))$ and $P_{t+1,h}^{(q)}(\alpha_t) = \alpha_t' \exp(R_{t+1,h}^{(q)})$.

4. We predict the risk measures with simulated daily log-prices of the portfolio. For each simulation q, we compute the drawdown measures using the definitions given in Section 2.2, yielding $XDD_{t+1}^{(q)}(\alpha_t)$. The predictions of the drawdown measures are then given by the average over the Q simulations: $X\hat{D}D_{t+1}(\alpha_t) = \frac{1}{Q}\sum_{q=1}^{Q} XDD_{t+1}^{(q)}(\alpha_t)$, except for CED. To generate CED predictions, we rely on the MDDs obtained over all simulations $MDD_{t+1}(\alpha_t) = (MDD_{t+1}^{(1)}(\alpha_t), \cdots, MDD_{t+1}^{(Q)}(\alpha_t))$ and take the average of the worst $(1-\tilde{\theta}) \times 100\%$ MDDs as in Equation (4).

5. We iterate points 3 and 4 over α_t until the optimal portfolio weight vector α_t^* is found for the investment criterion, $\min_{\{\alpha_t\}} X\hat{D}D_{t+1}(\alpha_t)$.

We obtain accurate estimates of the optimal weights by simulating a large number of draws. We set $Q = 50,000$.

4 Data

Our empirical application is based on two size portfolios constructed using the Fama and French (1993) methodology. Small caps include the firms with the lowest market capitalization (bottom 30%), while large caps consist of the firms with the largest market capitalization (top 30%).[10] The portfolios comprise all NYSE, AMEX, and NASDAQ stocks for which market equity data are available. The sample covers the period from July 1926 to December 2020, for a total of 24,896 daily returns.

Size portfolios offer several advantages for analyzing a portfolio based on large drawdowns. First, data are available for a long period of time (almost 100 years). Second, these portfolios have attracted attention for a long time, and hence, relatively well-known properties have been established. Several studies analyze small caps and large caps, in particular regarding their risk and return characteristics. Perez-Quiros and Timmermann (2000) provide evidence that small caps display a high degree of asymmetry between recession and expansion states. In particular, in recession, small caps are more strongly affected than large caps by worsening credit market conditions. Guidolin and Timmermann (2007) also analyze properties of small and large caps in a multivariate MS model with four regimes. Ang and Chen (2002) and Patton (2004) investigate the dependence between small and large caps, particularly the asymmetry between bull and bear times. Empirical research provides evidence that the higher average return of small caps in fact compensates investors for the occurrence of larger drawdowns. In particular, Huang et al. (2012) report that small firms are more exposed to extreme downside risk. The COVID-19 market crash illustrates this phenomenon: small caps experienced a 45% decline, while large caps only decreased by 33% in the first quarter of 2020.

Table 1 reports statistics on small caps and large caps. Panel A corresponds to the full sample (1926–2020, 378 quarters), while Panel B focuses on the out-of-sample period that we used for the investment analysis (1990–2020, 124 quarters). The first part of the table displays descriptive statistics and standard risk measures: the VaR and the expected shortfall (ES) based on one-day returns, and MDD over the full sample. The second part of the table reports the four sample measures of large drawdowns described in Section 2.2. Different horizons of one quarter, two quarters, and four quarters are considered. For ADD,

[10] We have also investigated the cases of the firms with the bottom 20% and top 20% market capitalization and with the bottom 10% and top 10% market capitalization, with limited impact on the main results. The data are available on the website of Kenneth French https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

CDD, and MDD, we report the average and the persistence across the subsamples defined by the investment horizon. A unique CED is measured over the overall sample. We compute CDD for a probability $\theta = 0.8$, which means that we consider the worst 20% drawdowns in a given subsample (e.g., the worst 12 drawdowns in a given quarter) (see Chekhlov et al., 2005). CED has a probability $\tilde{\theta} = 0.9$, which corresponds to the worst 10% of MDDs in the sample (the worst 12 MDDs in the out-of-sample period) (see Goldberg and Mahmoud, 2016).

[Insert Table 1 here]

With regards to descriptive statistics, small caps exhibit higher annual return and higher volatility on average. The VaR and the ES demonstrate that small caps are more prone to large adverse shocks. The 0.1% VaR is equal to 8.5% for small caps and 6.9% for large caps. The overall MDD is equal to 92% for small caps and 86.5% for large caps. These values occurred during the stock market crash of 1929–1932.

By examining the four large drawdown measures, we find that they are all greater for small caps than large caps. On average, the one-quarter MDD on small caps is larger by approximately 2.3% and the two-quarter MDD is larger by 4%. ADD and CDD exhibit similar patterns. CED is also substantially higher for small caps than for large caps (by roughly 8% over one quarter and by 11% over one year). This evidence suggests that, despite higher expected returns for small caps, investors may be reluctant to invest in small caps because they are more exposed to extreme downside risk, particularly over the long term, as suggested by Ang and Chen (2002) and Huang et al. (2012).

The table also reports relatively high first-order autocorrelation coefficients in drawdown measures, which reflects persistence in the measures in the full sample. For MDD in small caps, the AR(1) parameters are equal to 33% for the one-quarter horizon and 48% for the four-quarter horizon. It should be noted that large drawdowns are much less persistent in the out-of-sample period (Panel B: 1990–2020). The AR(1) coefficient is usually low and close to 0 for small caps at all horizons, although it remains higher for large caps. For MDD, the autocorrelations for the four-quarter horizon are equal to 11% and 38% for small caps and large caps, respectively.

Figure 1 presents the temporal evolution of the large drawdown measures. It reveals that, over the last century, four periods have been accompanied by large drawdowns: the Great Depression (1929–1933), the oil crisis (1973–1979), the subprime crisis (2008–2012), and the COVID-19 downturn (2020). The figure also displays the 10% CED for each horizon over the full sample (horizontal lines on the right-hand side plots). We observe that CED for the four-quarter horizon identifies only two exceptional drawdowns (the Great Depression and the subprime crisis episodes), whereas the two-quarter CED would also include the COVID-19 downturn as an exceptional event.

[Insert Figure 1 here]

5 Full-Sample Analysis

In this section, we evaluate the ability of MS-GARCH models to predict large drawdowns over the full sample (1926–2020). This unique estimation helps interpret the parameter estimates and formally test the number of regimes and the choice of the innovation distribution. We then analyze the in-sample optimal allocations obtained by an investor using these models to predict large drawdowns.

5.1 Model Estimation

Tables 2 and 3 report parameter estimates for the models with one, two, and three regimes, with normal and Student's t innovations. Table 4 reports likelihood ratio (LR) test statistics, which we use to identify the model that best reproduces the data properties. We first consider the one-regime model, i.e., the standard multivariate GARCH model with constant conditional correlation. Parameter estimates of expected returns, volatility dynamics, and their correlation are fairly standard and similar for both innovation distributions. The degree of freedom of the Student's t distribution is equal to 5.68, which suggests that innovations have relatively heavy tails. The LR test rejects the null hypothesis that the distribution is normal.

The results for the models with two regimes have very different properties depending on the distribution assumed. With normal innovations, the second regime has large negative expected returns for both assets. Two distinct regimes are clearly identified: the first regime pertains to normal conditions and the second regime corresponds to the bear state, possibly associated with market downturns. Our estimates are consistent with the high degree of asymmetry of small caps highlighted by Perez-Quiros and Timmermann (2000): in the bear market, expected returns of small caps are much lower than expected returns of large caps, probably reflecting tighter credit market conditions. The probability of remaining in the bear state is relatively low ($p_{22} = 68.3\%$), which corresponds to a stationary probability equal to 22%.[11]

In contrast, with Student's t innovations, the expected returns in Regime 2 are close to 0. The probability of remaining in the low regime is the same as the probability of remaining in the high regime ($p_{11} = 96.3\%$ and $p_{22} = 96.7\%$), so that the stationary probability of being in Regime 2 is as high as $\pi_{\infty,2} = 53\%$. As a consequence, Regime 2 cannot be interpreted as a pure bear state. These results suggest that in this model the occurrence of large drawdowns is not captured by large negative expected returns but instead by the heavy-tailedness of the Student's t innovations.[12] For both two-regime models, the gain in likelihood relative to the one-regime models is substantial, and thus, we reject the null hypothesis that the model has one regime. The model with Student's t innovations is also preferred to the model with normal innovations.

[Insert Tables 2 to 4 here]

[11] Stationary (or steady-state) probabilities are defined as: $\pi_\infty = P\ \pi_\infty$. In the two-regime case, this relation boils down to $\pi_{\infty,2} = \Pr[S_t = 2] = (1 - p_{11})/(1 - p_{11} - p_{22})$.

[12] This conclusion appears very robust and not driven by the choice of starting values. We have experimented with several sets of starting parameter values with the same parameter estimates for both models. A similar phenomenon, that the probability of remaining in the bear state is lower with normal innovations than with Student's t innovations, has been reported by Haas and Paolella (2012) and Haas and Liu (2018).

For the three-regime models, expected returns, volatility dynamics, and the correlation are similar for both innovation distributions. The results align with our expectations: Regime 1 captures the long periods during which the stocks are in a bull market (with high expected returns). Regime 2 corresponds to the slow growth or recovery regime, with intermediate expected returns.[13] Regime 3 accounts for bear market conditions. As in the two-regime case, compared to large caps, small caps have much higher expected returns in Regime 1 and much lower expected returns in Regime 3. Regime 3 also exhibits a higher correlation than Regime 1, reflecting the asymmetry in dependence found by Ang and Chen (2002) and Jondeau (2016).

The degree of freedom of the Student's t distribution is equal to 7.8, suggesting relatively thinner tails than in the two-regime model. Similar to the two-regime models, the Student's t innovation partly captures the occurrence of large negative returns, as expected returns in Regime 3 are higher with Student's t innovations. As a result, being in Regime 3 is more likely than in the model with normal innovations: the stationary probability of being in Regime 3 is equal to 19.5% with normal innovations and 30.4% with Student's t innovations. For two-regime models, it appears that heavy-tailed innovations capture some large negative events that otherwise would be captured by the bear state. Finally, we reject the null hypothesis that the model has two regimes only and the null hypothesis that the distribution is normal. In conclusion, all the LR tests point in favor of the three-regime model with a Student's t distribution.

In Figure 2, we represent the filtered probability of being in the bear state for the two-regime and three-regime models. First, we note that the two-regime/Student's t model produces a high filtered probability (above 50%), suggesting that this regime actually does not capture bear markets. Second, the filtered probabilities in the two-regime and three-regime models with normal innovations exhibit similar temporal evolution. Peaks occur at

[13] Guidolin and Timmermann (2007) estimate a four-regime model for financial returns, in which the intermediate regime is decomposed into a slow growth regime and a recovery regime. As we discuss in Section 5.2, we also estimated four-regime models with normal and Student's t innovations with our data. The gain in likelihood was substantial, but the underperformance in our in-sample allocation exercise suggests that these models may be overparameterized. For this reason, we do not continue with these models.

the same times with similar probability. However, these peaks do not always coincide with actual market downturns. The first peak occurs in June 1932, after the Wall Street crash of October 1929. The second peak corresponds to the oil crash in mid-1973. The third peak in May 1984 could not be associated with any market downturn. The fourth peak, which occurred in October 1999, corresponds to the dotcom crash. The last peak in September 2014 again does not correspond to any large market decline.

Third, for the three-regime model with Student's t innovations, we find that most peaks in the filtered probability actually correspond to market events associated with a long-lasting bear market. We can identify three main episodes: the first one corresponds to the bear market at the beginning of the period (with peaks in the second half of 1929 and at the end of 1937, associated with the Wall Street crash and the economic recession, respectively). The second episode corresponds to the inflationary bear market of the seventies (with peaks at the end of 1969 and mid-1973). The third episode is associated with the market crashes at the turn of the new century (Russian crisis in 1998, dotcom crash in 2001 and financial crisis in 2008). In the more recent period, the filtered probability also increased in mid-2015 (associated with the stock market sell-off following the ending of quantitative easing by the Federal Reserve) and at the beginning of 2020 (associated with the COVID-19 market crash).[14]

The analysis of filtered probabilities clearly suggests that the three-regime/Student's t model provides a better description of the market downturns observed in the sample period than do the other competing models.

[Insert Figure 2 here]

5.2 Allocation Based on Full-Sample Model

For the in-sample analysis, we consider a unique optimal allocation for each model. We use the parameters estimated over the full sample (1926–2020), assume that the current state

[14] We note that one of the peaks in the filtered probability (in mid-1984) could not be associated with any particular stock market event.

corresponds to the average sample state, and simulate trajectories of daily returns for small and large caps for horizons from one to four quarters. Then, we determine the optimal weight that minimizes the predicted drawdown measure of the portfolio. Finally, we calculate what would have been the average period drawdown measure, calculated over the full period if the investor had maintained the same portfolio weight. Clearly, this analysis suffers from a forward-looking bias, as such an allocation could not have been implemented in real time. However, with this approach, we can analyze and compare the characteristics of the optimal portfolios for the various drawdown measures and for various investment horizons. We also compare these portfolios with the allocation based on the full-sample empirical distribution of period drawdowns. As the investment strategy is implemented in the same sample, the empirical distribution accurately reflects the true (unknown) distribution of large drawdowns, and the allocation based on this empirical distribution is similar to the first-best allocation. The out-of-sample analysis will be performed in the next section.

Results are reported in Table 5. The first column corresponds to the allocation based on the empirical distribution (model-free allocation). In all cases, the optimal weight for small caps is negative, indicating that the investor would be long large caps and short small caps. Figure 3 illustrates the optimal portfolio weight α_t^* and the corresponding level of large drawdown achieved in sample. The figure confirms that minimizing large drawdowns would have required a rather extreme allocation. For instance, with a two-quarter horizon, the optimal weights are equal to -25%, -34%, -35%, and -49% for ADD, CDD, MDD, and CED, respectively. This finding is consistent with results previously established in the literature, suggesting that the higher expected return of small caps is the compensation for a higher tail risk. In addition, the ex post drawdown measure of the optimal portfolio is always substantially smaller than the corresponding drawdown measures for small and large caps (see Table 1 Panel A). This result suggests that diversification allows the investor to reduce the drawdown of the portfolio relative to the constituent assets.

The table also reveals that the one-regime (GARCH) models and the two-regime/Student's t model fail in predicting a negative portfolio weight. The reason is that these models do not generate bear market conditions and therefore fail to penalize small caps. Allocations based on these models produce the largest ex post values of the portfolio drawdown measures, often substantially higher than the ex post values obtained with the model-free approach. We note that the two-regime/Student's t model performs much worse than the two-regime/normal model for all investment strategies. This result contrasts with the LR test reported in Section 5.1, which strongly rejects the null hypothesis of normal innovations against the alternative of Student's t innovations. Clearly, properties of stock returns captured by large drawdowns are of a different nature than properties reflected in the model likelihood. Our interpretation is that the two-regime/Student's t model provides a good description of the large negative returns through the heavy-tailed distribution of the innovation processes. However, it fails in reproducing the large negative trends in the bear markets.

For a one-quarter horizon, the two-regime/normal model and the three-regime/normal model generate the lowest drawdown measures. When the investment criterion is ADD, CDD, or MDD, the optimal weight is negative and often close to the model-free allocation (Columns 4 and 6 compared to Column 1). It follows that the ex post drawdown measures are very close to each other. For instance, the optimal weight found with the two-regime model for the MDD objective is equal to -0.26 (with an ex post period MDD equal to 7.43%), compared to a weight of -0.25 (with the same ex post period MDD) with the model-free approach. The three-regime/Student's t model is usually close to the best models. For CED, all three models perform similarly, with the same optimal weight (equal to -0.25), while it is equal to -0.56 for the empirical distribution.

For the two-quarter and four-quarter horizons, the best model is the three-regime/Student's t model for all drawdown measures. The gain is substantial in most cases relative to the other parametric models. For instance, the optimal MDD is equal to 10.56% for the three-regime/Student's t model and 10.94% for the second-best model (two-regime/normal). For the four-quarter horizon, the ex post MDD measures are at least 0.4% below the ex post MDD obtained with the other models. The table also reveals that the optimal weights are usually negative, although higher than those obtained with the empirical distribution. The

CED gain of using the three-regime/Student's t model is close to 0.5% compared with the second-best model (two-regime/normal).

It is worth noting that we also analyzed the in-sample allocation based on the four-regime models. For all investment criteria, the four-regime models perform as well as the three-regime models for the one-quarter investment horizon but are dominated by the best three-regime model for the two-quarter and four-quarter horizons. As four-regime models require the estimation of a large number of parameters (53 parameters for the model with Student's t innovations), the risk of an overparameterization is high and we do not continue the investigation with these models.

These results support the finding that three regimes and Student's t innovations are necessary to correctly capture and control drawdown-based risk measures. Since this conclusion is obtained from an in-sample analysis and therefore may suffer from a forward-looking bias, we now consider a fully out-of-sample allocation.

[Insert Table 5 and Figure 3 here]

6 Out-of-Sample Analysis

The out-of-sample strategy that we consider now is implemented as follows. We use a rolling window to estimate the parameters of the MS-GARCH models. For each model, the first set of parameters is estimated over the sample of daily returns from January 1927 to December 1989. These parameter estimates are used to simulate paths of daily log-returns of length H, which we use to predict next-period drawdown measures. We determine the optimal portfolio weight that minimizes the expected drawdown measures. Next, we roll the window by one period (H days) and proceed in the same way until we reach the last subsample (from October 1958 to September 2020). With this approach, we evaluate the ex post performance out of sample over 123 quarters.

6.1 Model Estimation and Adequacy Tests

The temporal evolution of parameter estimates for the competing models is displayed in Appendix A. Comparison with the full sample estimates in Tables 2 and 3 reveals a remarkable match between the two sets of estimates on average. The figures demonstrate that parameter estimates are usually rather stable over time, although a few parameters exhibit trends. In particular, the correlation between small and large caps tends to decrease in Regimes 1 and 3 in the three-regime model. The degree of freedom of the Student's t distribution tends to increase in the two-regime and three-regime models.[15]

We evaluate the adequacy of the out-of-sample estimation of the models with respect to the returns for small and large caps by backtesting the predicted $A\hat{D}D_t$, $C\hat{D}D_{\theta,t}$, and $M\hat{D}D_t$ obtained via simulations.[16] We adopt the approach proposed by Acerbi and Szekely (2014) for backtesting expected shortfall estimates. We adjust their testing framework for the unconditional coverage of the three large drawdown measures. The logic of the test is the following: for each risk measure, we define a Z statistic that is equal to zero under the null hypothesis that the model correctly predicts this risk measure with simulations. Then, we compute the realized Z statistic for the out-of-sample period. Finally, we test the significance of the test by comparing the realized Z statistic to the distribution of the Z statistics under H_0, which is obtained by computing Z from simulations of the model over the out-of-sample period.

Adequacy Test Statistics

For ease of exposition, we assume that the distribution is continuous and strictly increasing, as in Section 2.1. For CDD, we build the Z statistic from the fact that $CDD_{\theta,t} = E_t[DD_t \ I_{(DD_t > Th_{\theta,t})}]$, where $Th_{\theta,t} = \inf\{s \mid \frac{1}{H}\sum_{h=1}^{H} I_{(DD_{t,h}>s)} \leq 1-\theta\}$, with $E_t[Th_{\theta,t}] =$

[15] We also investigated subsamples of 40 years with similar but more erratic evolutions.

[16] We developed a similar test for CED, but the number of observations is insufficient in this case.

$H(1-\theta)$. The test statistic is defined as:

$$Z_{CDD}(DD) = \frac{1}{T}\sum_{t=1}^{T}\sum_{h=1}^{H}\frac{DD_{t,h}\ I_{(DD_{t,h}>\hat{Th}_{\theta,t})}}{H\ (1-\theta)\ C\hat{D}D_{\theta,t}} - 1,$$

where $C\hat{D}D_{\theta,t}$ and $\hat{Th}_{\theta,t}$ are predictions of $CDD_{\theta,t}$ and $Th_{\theta,t}$ based on simulations of the model estimated using data until quarter $t-1$. T corresponds to the number of quarters in the out-of-sample period.

The hypotheses for this test are as follows:

$$H_0 : CDD_{\theta,t} - C\hat{D}D_{\theta,t} = 0 \text{ for all } t$$

$$H_1 : |CDD_{\theta,t} - C\hat{D}D_{\theta,t}| > 0 \text{ for some } t \text{ and } |Th_{\theta,t} - \hat{Th}_{\theta,t}| \geq 0 \text{ for all } t$$

Under the null hypothesis, we have $E_{H_0}[Z_{CDD}(DD)] = 0$. Under the alternative hypothesis, we have that $E_{H_1}[Z_{CDD}(DD)] > 0$. As in the second test of Acerbi and Szekely (2014), this test does not require the predicted threshold to be correct. The test actually jointly evaluates the frequency and the magnitude of the tail events because the test statistics do not impose that $\sum_{h=1}^{H} I_{(DD_t>\hat{Th}_{\theta,t})} = H(1-\theta)$, i.e., the frequency is correct. Therefore, the test statistic will be close to 0 if both the predicted threshold is close to the realized threshold and the predicted CDD is close to the realized CDD.

For ADD, we use the fact that $ADD_t = E_t[DD_t]$ and define the Z statistic as follows:

$$Z_{ADD}(DD) = \frac{1}{T}\sum_{t=1}^{T}\sum_{h=1}^{H}\frac{DD_{t,h}}{H\ A\hat{D}D_t} - 1,$$

where $A\hat{D}D_t = \frac{1}{Q}\sum_{q=1}^{Q} ADD_t^{(q)}$ is the prediction of ADD_t based on simulations of the model.

The hypotheses of this test are:

$$H_0 : ADD_t - A\hat{D}D_t = 0 \text{ for all } t$$

$$H_1 : |ADD_t - A\hat{D}D_t| > 0 \text{ for some } t.$$

It follows that $E_{H_0}[Z_{ADD}(DD)] = 0$ under the null hypothesis that the model correctly describes the realized ADD. Under the alternative hypothesis, we have that $E_{H_1}[Z_{ADD}(DD)] > 0$ when $ADD_t > A\hat{D}D_t$ (underestimation) and $E_{H_1}[Z_{ADD}(DD)] < 0$ when $ADD_t < A\hat{D}D_t$ (overestimation).[17]

We use the same logic for MDD and define the Z statistic as:

$$Z_{MDD}(DD) = \frac{1}{T}\sum_{t=1}^{T}\frac{MDD_t}{M\hat{D}D_t} - 1,$$

with the following hypotheses:

$$H_0 : MDD_t - M\hat{D}D_t = 0 \text{ for all } t$$

$$H_1 : |MDD_t - M\hat{D}D_t| > 0 \text{ for some } t.$$

Details on the implementation of the test and the power of these tests are provided in Appendix B.

Backtesting Results

Results of the adequacy tests are reported in Table 6. The table indicates that most models fail to capture the time-series properties of large drawdown statistics. The only model that performs relatively well for small caps and large caps is the three-regime/Student's t model. For CDD, it produces a number of exceedances and an average drawdown above the threshold, which is close to the numbers observed in the data. For the one-quarter horizon, the numbers of exceedances equal 1,800 and 1,412 for the small caps and large caps, respectively, while the expected number is 1,562 (i.e., 20% of the total number of daily observations in the out-of-sample period). The p-value of the test statistic is equal to 2.8% for small caps and to 66.7% for large caps, suggesting that the 20% threshold on drawdowns

[17] Note that Acerbi and Szekely (2014) assume a one-sided test in line with Basel VaR tests, which are designed to detect excesses of VaR exceptions. In our case, we assume a two-sided test, since we test whether a given model correctly predicts large drawdown measures.

is slightly too small for small firms. Similarly, this model performs relatively well for the two-quarter horizon, with a number of exceedances and an average drawdown above the threshold, which is in line with the data.

The two-regime/normal model and the three-regime/normal model provide a good description of the ADD, CDD, and MDD of small caps, with p-values above 5% for all three horizons. However, both models fail at correctly predicting the drawdown measures of large caps. They systematically overestimate the realized ADD, CDD, and MDD for large caps.

The other models fail in capturing the properties of realized drawdown measures. The one-regime models and the two-regime/Student's t model systematically underestimate the drawdown measures for small and large caps. The reason for this failure is that the models do not allow for a negative trend in expected returns, and therefore, they cannot reproduce long-lasting market declines.

[Insert Table 6 here]

6.2 Out-of-Sample Allocation

We now consider an investor who allocates her wealth in real time period by period, with an investment horizon from one quarter to one year between 1990 and 2020. This out-of-sample analysis corresponds to 124 nonoverlapping quarterly allocations and 31 nonoverlapping annual allocations. We evaluate the performances of the various models by comparing the ex post drawdown measures obtained from the final value of the portfolio.

Table 7 reports results for the out-of-sample allocation over 1990–2020. We compare the performance of the allocation based on parametric models with the performance of the model-free allocation. For the model-free approach, we use the empirical distribution based on subsamples for the last 25 and 63 years (columns 1 and 2, respectively).[18] We start again with the allocation based on ADD, CDD, and MDD (Panels A to C). The three-regime/normal

[18] We also consider shorter subsamples, such as 10 years, but such a short subsample would not provide us with relevant estimates of the four-quarter drawdown measures, as the predictions would be based only on 10 observations. The window of 63 years corresponds to the window used for the estimation of the parametric models. As the table reveals, the length of the window has a limited impact for ex post drawdown measures.

model performs well for the one-quarter horizon, while the three-regime/Student's t model performs the best for the two- and four-quarter horizons. In all cases, the best models generate an optimal weight that is negative or close to 0. In comparison with the model-free allocation, we find that the ex post levels of ADD, CDD, and MDD are in general below all three horizons. For instance, the average CDD is equal to 9.22% over four quarters with the three-regime/Student's model and to 9.56% and 9.83% for the empirical distributions based on 25 and 63 years, respectively. Similarly, the average MDD is equal to 9.92% over two quarters with the three-regime/Student's model and to 10.01% and 10.14% for the empirical distributions based on 25 and 63 years, respectively.

For the CED allocation (Panel D), the results suggest that the three-regime/Student's t model again produces better out-of-sample performances than other parametric models. For the one-quarter and two-quarter horizons, the parametric model also clearly outperforms model-free allocation. For the two-quarter horizon, the ex post CED is equal to 24.84% with the three-regime/Student's model, more than 2.5% below the ex post CED produced by the model-free approach based on 25 years.[19]

Figure 4 displays the evolution of the optimal weight of the strategies based on model-free predictions and on two-regime and three-regime model predictions for the two-quarter horizon. In general, the optimal weights are ordered in the same way for the various investment criteria: the weight of small caps is higher for ADD, then for CDD, then for MDD, and finally for CED. This result is expected because an investor reluctant to take more extreme risks (such as being exposed to MDD) would want to short more small caps.

As the figure illustrates, the evolution of the optimal weights exhibits very different patterns among approaches for prediction. Model-free approaches usually produce negative weights, in particular for the CED criterion. For the two-regime and three-regime models with normal innovations, patterns are similar, with weights that are close to 0. This finding

[19]For four quarters, the model-free allocations dominate parametric approaches, probably because the number of observations is very low. Indeed, CED is based in this case on the worst 10% of MDDs among the 31 estimates of four-quarter MDDs; i.e., it is based on only three observations.

suggests that the innovation process matters when it comes to generating sufficiently large drawdowns for small caps and therefore to producing a large and negative weight.

Last, two-regime and three-regime models with Student's t distribution display more temporal variation. Optimal weights are usually higher at the beginning of the sample and tend to decrease over time. Although patterns are similar, optimal weights for the model with normal innovations are almost always positive, while those associated with Student's t innovations are usually negative. This finding can be explained by the failure of the two-regime/Student's t model to generate negative expected returns in Regime 2, which prevents large drawdowns in small caps from developing. In contrast, the three-regime/Student's t model is able to predict large negative trends in stock markets and therefore to identify that holding small caps implies higher tail risk. As a consequence, this model produces large negative weights for most allocation criteria.

This out-of-sample analysis demonstrates that two main components are necessary in the multivariate MS-GARCH model to produce large and negative weights for small caps. First, three regimes are necessary to generate large negative trends in stock returns. Second, Student's t innovations are necessary to generate a longer-lasting bear state and therefore to produce more drawdowns for small caps.

7 Conclusion

In this paper, we consider the modeling and prediction of large drawdowns in financial markets. Since the standard GARCH model cannot capture long-lasting market declines, we investigate multivariate MS-GARCH models, which are able to generate regimes with positive or negative market trends. Provided that the probability of remaining in the bear regime is sufficiently high, these models are able to reproduce some of the large drawdown properties. More specifically, we find that in three-regime models, one of the regimes is characterized by large negative expected returns (bear market) and the generation of large windfalls.

In sample, predictions based on the empirical distribution of large drawdowns (model-free approach) are relatively accurate so that minimizing the expected portfolio drawdown based on these predictions provides an approximately first-best allocation. Among parametric models, best performances are obtained by three-regime models for all investment criteria and horizons. In addition, for one-quarter and two-quarter horizons, the models often perform as well as the model-free approach does, although they impose parametric restrictions upon the properties of financial returns.

Out of sample, when the distribution of drawdowns varies through time, parametric models based on three regimes serve as better tools to predict expected drawdowns because they impose restrictions that are consistent with the data. Therefore, they always dominate other parametric models and often produce lower ex post drawdown measures than does the model-free approach.

As our objective is to demonstrate the superiority of three-regime MS-GARCH models in predicting large drawdowns, we keep our specification as simple and straightforward as possible. In particular, we do not attempt to identify factors driving the dynamics of state probabilities, which may be affected by variables related to government or monetary policy, geopolitical issues, or climate change. Such an analysis is an important avenue for future research.

References

Abramson, A., Cohen, I., 2007. On the stationarity of Markov-switching GARCH processes. Econometric Theory 23, 485–500.

Acerbi, C., Szekely, B., 2014. Backtesting expected shortfall. MSCI Working Paper.

Alexander, G.J., Baptista, A.M., 2004. A comparison of VaR and CVaR constraints on portfolio selection with the mean-variance model. Management Science 50(9), 1261–1273.

Ang, A., Bekaert, G., 2002. International asset allocation with regime shifts. Review of Financial Studies 15, 1137–1187.

Ang, A., Chen, J., 2002. Asymmetric correlations of equity portfolios. Journal of Financial Economics 63, 443–494.

Ang, A., Chen, J., Xing, Y., 2006. Downside risk. Review of Financial Studies 19, 1191–1239.

Artzner, P., Delbaen, F., Eber, J.M., Heath, D., 1999. Coherent measures of risk. Mathematical Finance 9, 203–228.

Barnett, M., Brock, W., Hansen, L.P., 2020. Pricing uncertainty induced by climate change. Review of Financial Studies 33, 1024–1066.

Barro, R., 2006. Rare disasters and asset markets in the twentieth century. Quarterly Journal of Economics 121, 823–866.

Barro, R., 2009. Rare disasters, asset prices, and welfare costs. American Economic Review 99, 243–264.

Barro, R., Ursua, J., 2008. Macroeconomic crises since 1870. Brookings Papers on Economic Activity, Spring, 255–335.

Bollerslev, T., Todorov, V., 2011. Tails, fears, and risk premia. Journal of Finance 66, 2165–2211.

Chekhlov, A., Uryasev, S., Zabarankin, M., 2003. Portfolio optimization with drawdown constraints, asset and liability management tools, in: Scherer, B. (Ed.), Risk Books, pp. 263–278. London.

Chekhlov, A., Uryasev, S., Zabarankin, M., 2005. Drawdown measure in portfolio optimization. Journal of Theoretical and Applied Finance 8, 13–58.

Engle, R.F., Giglio, S., Kelly, B., Lee, H., Stroebel, J., 2020. Hedging climate change news. Review of Financial Studies 33, 1184–1216.

Fama, E.F., French, K.R., 1993. Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 33, 3–56.

Gabaix, X., 2012. Variable rare disasters: An exactly solved framework for ten puzzles in macro-finance. Quarterly Journal of Economics 127, 645–700.

Goldberg, L.R., Mahmoud, O., 2015. A convex measure of drawdown risk. Working Paper.

Goldberg, L.R., Mahmoud, O., 2016. Drawdown: from practice to theory and back again. Mathematics and Financial Economics 11, 275–297.

Gourio, F., 2012. Disaster risk and business cycles. American Economic Review 102, 2734–2766.

Gourio, F., 2013. Credit risk and disaster risk. American Economic Journal: Macroeconomics 5, 1–34.

Gray, S., 1996. Modeling the conditional distribution of interest rates as regime-switching process. Journal of Financial Economics 42, 27–62.

Grossman, S., Zhou, Z., 1993. Optimal investment strategies for controlling draw-downs. Mathematical Finance 3, 241–276.

Guidolin, M., Timmermann, A., 2003. Value at risk and expected shortfall under regime switching. Working Paper.

Guidolin, M., Timmermann, A., 2007. Asset allocation under multivariate regime switching. Journal of Economic Dynamics and Control 31, 3503–3544.

Guidolin, M., Timmermann, A., 2008. International asset allocation under regime switching, skew, and kurtosis preferences. Review of Financial Studies 21, 889–935.

Haas, M., Liu, J.C., 2018. A multivariate regime-switching GARCH model with an application to global stock market and real estate equity returns. Studies in Nonlinear Dynamics and Econometrics 22, 1–27.

Haas, M., Mittnik, S., Paolella, M.S., 2004. A new approach to Markov-switching GARCH models. Journal of Financial Econometrics 2, 493–530.

Haas, M., Paolella, M.S., 2012. Mixture and regime-switching GARCH models, in: Bauwens, L., Hafner, C.M., Laurent, S. (Eds.), Handbook of Volatility Models and their Applications. John Wiley & Sons.

Hamilton, J., 1989. A new approach to the economic analysis of nonstationary time series and the business cycle. Econometrica 57, 357–384.

Hansen, P.R., 2009. In-sample fit and out-of-sample fit: Their joint distribution and its implications for model selection. Working Paper.

Hasler, M., Marfe, R., 2016. Disaster recovery and the term structure of dividend strips. Journal of Financial Economics 122, 116–134.

Huang, W., Liu, Q., Rhee, S.G., Wu, F., 2012. Extreme downside risk and expected stock returns. Journal of Banking and Finance 36, 1492–1502.

Jondeau, E., 2016. Asymmetry in tail dependence in equity portfolios. Computational Statistics and Data Analysis 100, 351–368.

Karydas, C., Xepapadeas, A., 2019. Climate change financial risks: Pricing and portfolio allocation. Economics Working Paper Series, No. 19/327.

Kelly, B., Jiang, H., 2014. Tail risk and asset prices. Review of Financial Studies 27, 2841–2871.

Klaassen, F., 2002. Improving GARCH volatility forecasts with regime-switching GARCH. Empirical Economics 27, 363–394.

Pagano, M., Wagner, C., Zechner, J., 2020. Disaster resilience and asset prices. CEPR Working Paper No. 14773.

Patton, A., 2004. On the out-of-sample importance of skewness and asymmetric dependence for asset allocation. Journal of Financial Econometrics 2, 130–168.

Pelletier, D., 2006. Regime switching for dynamic correlations. Journal of Econometrics 131, 445–473.

Peng, C., Kim, Y.S., 2020. Portfolio optimization on multivariate regime switching GARCH model with normal tempered stable innovation. Working Paper, available at arXiv:2009.11367v2.

Perez-Quiros, G., Timmermann, A., 2000. Firm size and cyclical variations in stock returns. Journal of Finance 55, 1229–1262.

Reveiz, A., Leon, C., 2008. Efficient portfolio optimization in the wealth creation and maximum drawdown space. Banco de la Republica de Colombia. Technical report.

Rietz, T.A., 1988. The equity risk premium: A solution. Journal of Monetary Economics 22, 117–131.

Tsai, J., Wachter, J.A., 2015. Disaster risk and its implications for asset pricing. Annual Review of Financial Economics 7, 219–252.

Wachter, J.A., 2013. Can time-varying risk of rare disasters explain aggregate stock market volatility? Journal of Finance 68, 987–1035.

Table 1: Summary statistics on daily returns and period drawdowns for small caps and large caps

	Panel A: 1926–2020				**Panel B: 1990–2020**			
	Small caps		Large caps		Small caps		Large caps	
Daily returns	Stat.		Stat.		Stat.		Stat.	
Annualized Mean	10.63		8.99		10.38		9.88	
Annualized Std dev.	19.50		17.15		20.36		18.06	
Skewness	-0.39		-0.48		-0.81		-0.40	
Kurtosis	23.61		21.83		13.40		14.03	
Maximum	20.42		14.15		8.02		11.16	
Minimum	-16.75		-20.94		-14.26		-12.57	
VaR (0.1%)	8.53		6.86		9.01		7.56	
VaR (1%)	3.70		3.11		3.72		3.26	
VaR (5%)	1.82		1.56		1.97		1.74	
ES (0.1%)	10.76		9.16		11.24		9.32	
ES (1%)	5.55		4.64		5.50		4.76	
ES (5%)	3.09		2.62		3.17		2.77	
Overall MDD	92.02		86.54		67.13		58.29	
Period drawdowns	Stat.	AR(1)	Stat.	AR(1)	Stat.	AR(1)	Stat.	AR(1)
ADD - 1Q	3.84	0.15	2.78	0.20	3.65	0.12	2.47	0.25
ADD - 2Q	5.62	0.29	3.85	0.42	5.19	0.00	3.30	0.35
ADD - 4Q	7.52	0.46	4.82	0.42	6.16	-0.10	3.83	0.35
CDD (20%) - 1Q	7.75	0.25	5.76	0.28	7.28	0.12	5.27	0.23
CDD (20%) - 2Q	11.80	0.41	8.20	0.46	10.81	0.05	7.14	0.30
CDD (20%) - 4Q	16.65	0.45	10.96	0.43	13.99	0.01	8.87	0.33
MDD - 1Q	9.87	0.33	7.55	0.34	9.34	0.17	7.10	0.26
MDD - 2Q	15.05	0.44	10.93	0.47	13.99	0.06	9.93	0.28
MDD - 4Q	21.52	0.48	15.23	0.45	18.67	0.11	13.08	0.38
CED (10%) - 1Q	31.22		22.84		25.97		20.60	
CED (10%) - 2Q	41.97		31.35		35.40		25.41	
CED (10%) - 4Q	51.37		40.25		38.36		31.19	

Note: This table reports statistics on daily returns for small caps and large caps. Panel A covers the period from 1926 to 2020. Panel B covers the period from 1990 to 2020.

Table 2: Parameter estimates for one-regime and two-regime models

	One regime - Normal distribution				One regime - Student's t distribution			
	Small caps		Large caps		Small caps		Large caps	
	param.	std err.	param.	std err.	param.	std err.	param.	std err.
Expected returns								
μ ($\times 100$)	0.0679	(0.006)	0.0588	(0.005)	0.0865	(0.004)	0.0687	(0.004)
Volatility dynamics								
ω ($\times 100$)	1.4840	(0.201)	0.9667	(0.122)	0.9902	(0.093)	0.7532	(0.071)
α	0.1298	(0.008)	0.0984	(0.006)	0.1220	(0.007)	0.0970	(0.004)
β	0.8654	(0.008)	0.8962	(0.006)	0.8764	(0.006)	0.9010	(0.004)
Correlation								
ρ	0.8291	(0.003)			0.8212	(0.002)		
Degree of freedom								
ν	–				5.6810	(0.154)		
Log-lik.	-48,291.3				-46,183.1			
BIC	96.6736				92.4674			

	Two regimes - Normal distribution				Two regimes - Student's t distribution			
	Small caps		Large caps		Small caps		Large caps	
	param.	std err.	param.	std err.	param.	std err.	param.	std err.
Expected returns								
$\mu^{(1)}$ ($\times 100$)	0.1265	(0.007)	0.0965	(0.006)	0.2093	(0.035)	0.1155	(0.012)
$\mu^{(2)}$ ($\times 100$)	-0.2706	(0.043)	-0.1771	(0.030)	-0.0276	(0.016)	0.0207	(0.010)
Volatility dynamics								
$\omega^{(1)}$ ($\times 100$)	0.2974	(0.052)	0.3132	(0.049)	1.2510	(0.293)	0.8278	(0.140)
$\alpha^{(1)}$	0.0418	(0.005)	0.0417	(0.004)	0.2055	(0.021)	0.1309	(0.017)
$\beta^{(1)}$	0.9209	(0.008)	0.9289	(0.006)	0.7968	(0.020)	0.8663	(0.019)
$\omega^{(2)}$ ($\times 100$)	0.1550	(0.380)	0.5501	(0.279)	0.1576	(0.169)	0.2482	(0.195)
$\alpha^{(2)}$	0.1868	(0.024)	0.1463	(0.019)	0.0411	(0.023)	0.0422	(0.021)
$\beta^{(2)}$	0.9173	(0.011)	0.9291	(0.008)	0.9563	(0.023)	0.9565	(0.022)
Correlation								
$\rho^{(1)}$	0.8179	(0.004)			0.7450	(0.014)		
$\rho^{(2)}$	0.8334	(0.006)			0.8884	(0.010)		
Transition probabilities								
p_{11}	0.9081	(0.009)			0.9626	(0.024)		
p_{22}	0.6828	(0.027)			0.9669	(0.019)		
Degree of freedom								
ν	–				6.2260	(0.252)		
Log-lik.	-45,974.7				-45,420.1			
BIC	92.1478				91.0507			

Table 3: Parameter estimates for three-regime model

	Three regimes - Normal distribution				Three regimes - Student's t distribution			
	Small caps		Large caps		Small caps		Large caps	
	param.	std err.	param.	std err.	param.	std err.	param.	std err.
Expected returns								
$\mu^{(1)}$ ($\times$100)	0.3304	(0.022)	0.1174	(0.011)	0.3194	(0.017)	0.1251	(0.011)
$\mu^{(2)}$ ($\times$100)	0.0754	(0.008)	0.0966	(0.007)	0.1193	(0.008)	0.1133	(0.008)
$\mu^{(3)}$ ($\times$100)	-0.3451	(0.024)	-0.1897	(0.024)	-0.2316	(0.017)	-0.1049	(0.020)
Volatility dynamics								
$\omega^{(1)}$ ($\times$100)	0.0001	(0.191)	0.1797	(0.108)	0.5529	(0.205)	0.4921	(0.141)
$\alpha^{(1)}$	0.0403	(0.019)	0.0325	(0.009)	0.1303	(0.022)	0.0927	(0.013)
$\beta^{(1)}$	0.9480	(0.022)	0.9534	(0.012)	0.8585	(0.022)	0.8945	(0.013)
$\omega^{(2)}$ ($\times$100)	0.2705	(0.046)	0.3461	(0.060)	0.0201	(0.200)	0.0858	(0.040)
$\alpha^{(2)}$	0.0401	(0.005)	0.0439	(0.004)	0.0122	(0.003)	0.0177	(0.003)
$\beta^{(2)}$	0.9180	(0.008)	0.9233	(0.006)	0.9739	(0.006)	0.9691	(0.006)
$\omega^{(3)}$ ($\times$100)	0.0001	(0.640)	0.8492	(0.405)	0.1357	(0.310)	0.6340	(0.358)
$\alpha^{(3)}$	0.2262	(0.033)	0.1655	(0.022)	0.1328	(0.031)	0.1050	(0.022)
$\beta^{(3)}$	0.9005	(0.018)	0.9246	(0.011)	0.9145	(0.022)	0.9306	(0.017)
Correlation								
$\rho^{(1)}$	0.7814	(0.025)			0.7300	(0.016)		
$\rho^{(2)}$	0.8487	(0.006)			0.8465	(0.009)		
$\rho^{(3)}$	0.8462	(0.008)			0.8764	(0.007)		
Transition matrix								
$P_{1,:}$	0.9206	0.0169	0.0453		0.9402	0.0170	0.0266	
	(0.015)	(0.003)	(0.018)		(0.009)	(0.004)	(0.007)	
$P_{2,:}$	0.0350	0.8954	0.2663		0.0165	0.9116	0.1139	
	(0.016)	(0.008)	(0.026)		(0.007)	(0.011)	(0.013)	
$P_{3,:}$	0.0445	0.0877	0.6884		0.0433	0.0715	0.8595	
	(0.008)	(0.006)	(0.021)		(0.008)	(0.009)	(0.014)	
Degree of freedom								
ν	–				7.7750	(0.399)		
Log-lik.	-45,282.0				-44,852.6			
BIC	90.9244				90.0753			

Note: Tables 2 and 3 reports parameter estimates for the model with daily returns on small caps and large caps. Table 2 reports estimates of the one-regime and two-regime models. Table 3 reports estimates of the three-regime models. The estimation is based on the period from 1926 to 2020.

Table 4: Likelihood ratio tests

Null hypothesis	Alternative hypothesis	restr.	LR stat.
H0: One regime - normal	Ha: One regime - Student's t	1	4,216.3
H0: One regime - normal	Ha: Two regimes - normal	11	4,637.0
H0: One regime - Student's t	Ha: Two regimes - Student's t	11	1,528.0
H0: Two regimes - normal	Ha: Two regimes - Student's t	1	1,107.3
H0: Two regimes - normal	Ha: Three regimes - normal	16	1,385.3
H0: Two regimes - Student's t	Ha: Three regimes - Student's t	16	1,137.2
H0: Three regimes - normal	Ha: Three regimes - Student's t	1	859.2

Note: The table reports the likelihood ratio test statistics for various tests of interest. The first two columns indicate the null and alternative hypotheses. The third column reports the degree of freedom (restr.) of the test (number of restrictions under the null hypothesis). The last column reports the LR test statistics, which is distributed as a χ^2(restr.) under the null hypothesis. The p-values are all below 0.01%.

Table 5: In-Sample Allocation Based on Large Drawdowns (1926–2020)

Horizon	Stat.	Model -free	1-regime normal	1-regime Student t	2-regime normal	2-regime Student t	3-regime normal	3-regime Student t
				Panel A: Minimization of ADD				
1 quarter	Weight	-0.22	0.31	0.47	-0.15	0.33	-0.16	0.06
	ADD	2.74	2.97	3.13	**2.74**	3.00	**2.74**	2.81
2 quarters	Weight	-0.25	0.32	0.55	0.05	0.35	0.09	-0.08
	ADD	3.76	4.21	4.59	3.89	4.24	3.92	**3.80**
4 quarters	Weight	-0.18	0.31	0.59	0.07	0.40	0.09	0.00
	ADD	4.75	5.29	6.06	4.89	5.50	4.91	**4.83**
				Panel B: Minimization of CDD				
1 quarter	Weight	-0.26	0.30	0.45	-0.25	0.35	-0.25	-0.02
	CDD (20%)	5.64	6.14	6.41	**5.64**	6.23	**5.64**	5.74
2 quarters	Weight	-0.34	0.29	0.50	0.03	0.34	0.05	-0.24
	CDD (20%)	7.87	8.94	9.65	8.26	9.09	8.29	**7.90**
4 quarters	Weight	-0.38	0.27	0.53	0.10	0.36	0.04	-0.06
	CDD (20%)	10.35	12.03	13.41	11.31	12.47	11.10	**10.80**
				Panel C: Minimization of MDD				
1 quarter	Weight	-0.25	0.28	0.44	-0.26	0.30	-0.25	-0.07
	MDD	7.43	7.97	8.28	**7.43**	7.99	**7.43**	7.50
2 quarters	Weight	-0.35	0.28	0.48	0.01	0.30	0.02	-0.26
	MDD	10.53	11.75	12.49	10.94	11.83	10.98	**10.56**
4 quarters	Weight	-0.40	0.25	0.48	0.08	0.36	0.06	-0.09
	MDD	14.43	16.28	17.60	15.52	16.89	15.45	**14.95**
				Panel D: Minimization of CED				
1 quarter	Weight	-0.56	0.25	0.36	-0.25	0.18	-0.25	-0.25
	CED (10%)	21.61	24.34	25.16	21.97	23.84	**21.95**	21.97
2 quarters	Weight	-0.49	0.26	0.39	0.02	0.25	0.05	-0.29
	CED (10%)	29.67	33.51	34.72	31.51	33.36	31.66	**29.92**
4 quarters	Weight	-0.46	0.22	0.40	0.10	0.32	0.08	-0.03
	CED (10%)	39.78	41.40	42.74	40.62	42.14	40.52	**40.14**

Note: This table reports the optimal weight of small caps (α^*) and the value of the objective function at the optimum, when the investor minimizes ADD, 20% CDD, MDD, and 10% CED (Panels A to D, respectively). We consider three investment horizons: one quarter, two quarters, and four quarters. These results are based on the simulation of the model estimated over the full sample 1926–2020.

Table 6: Adequacy Tests for Out-of-sample Predictions (1990–2020)

	1-regime normal		1-regime Student t		2-regime normal		2-regime Student t		3-regime normal		3-regime Student t	
	Small	Large	Small	Large	Small	Large	Small	Large	Small	Large	Small	Large
Panel A: ADD												
1 quarter												
Stat.	0.793	0.225	1.232	0.382	0.057	-0.211	1.012	0.373	0.071	-0.191	0.263	0.009
p-val.	(0.000)	(0.000)	(0.000)	(0.000)	(0.411)	(0.001)	(0.000)	(0.000)	(0.393)	(0.007)	(0.007)	(0.904)
2 quarters												
Stat.	1.039	0.242	1.747	0.450	-0.047	-0.345	1.263	0.417	0.018	-0.296	0.106	-0.165
p-val.	(0.000)	(0.000)	(0.000)	(0.000)	(0.590)	(0.000)	(0.000)	(0.000)	(0.856)	(0.001)	(0.289)	(0.055)
4 quarters												
Stat.	1.017	0.139	1.930	0.380	-0.191	-0.470	1.308	0.317	-0.089	-0.400	-0.093	-0.372
p-val.	(0.000)	(0.088)	(0.000)	(0.002)	(0.084)	(0.000)	(0.000)	(0.004)	(0.462)	(0.001)	(0.446)	(0.002)
Panel B: 20% **CDD**												
1 quarter												
Nb exc.	2597	1781	2961	1964	1490	1015	2666	1961	1512	1082	1800	1412
Stat.	1.693	0.653	2.310	0.876	0.052	-0.307	1.693	0.829	0.039	-0.282	0.308	0.055
p-val.	(0.000)	(0.000)	(0.000)	(0.000)	(0.666)	(0.007)	(0.000)	(0.000)	(0.762)	(0.018)	(0.028)	(0.667)
2 quarters												
Nb exc.	2930	1938	3443	2221	1168	697	2967	2173	1278	784	1442	1124
Stat.	2.165	0.687	3.134	0.987	-0.134	-0.545	2.167	0.891	-0.063	-0.484	0.037	-0.266
p-val.	(0.000)	(0.000)	(0.000)	(0.000)	(0.210)	(0.000)	(0.000)	(0.000)	(0.367)	(0.001)	(0.396)	(0.033)
4 quarters												
Nb exc.	3163	1728	3863	2117	1037	510	3243	2037	1244	613	1193	683
Stat.	2.061	0.449	3.285	0.794	-0.328	-0.649	2.237	0.662	-0.186	-0.567	-0.238	-0.545
p-val.	(0.000)	(0.015)	(0.000)	(0.001)	(0.066)	(0.000)	(0.000)	(0.003)	(0.218)	(0.002)	(0.138)	(0.002)
Panel C: MDD												
1 quarter												
Stat.	0.622	0.266	0.885	0.377	-0.047	-0.202	0.739	0.363	-0.033	-0.192	0.078	-0.009
p-val.	(0.000)	(0.000)	(0.000)	(0.000)	(0.327)	(0.000)	(0.000)	(0.000)	(0.538)	(0.000)	(0.176)	(0.852)
2 quarters												
Stat.	0.802	0.284	1.168	0.423	-0.057	-0.285	0.919	0.383	-0.010	-0.237	-0.033	-0.188
p-val.	(0.000)	(0.000)	(0.000)	(0.000)	(0.305)	(0.000)	(0.000)	(0.000)	(0.862)	(0.000)	(0.595)	(0.002)
4 quarters												
Stat.	0.873	0.290	1.286	0.435	-0.045	-0.299	1.017	0.363	0.021	-0.232	-0.059	-0.269
p-val.	(0.000)	(0.000)	(0.000)	(0.000)	(0.503)	(0.000)	(0.000)	(0.000)	(0.773)	(0.001)	(0.404)	(0.000)

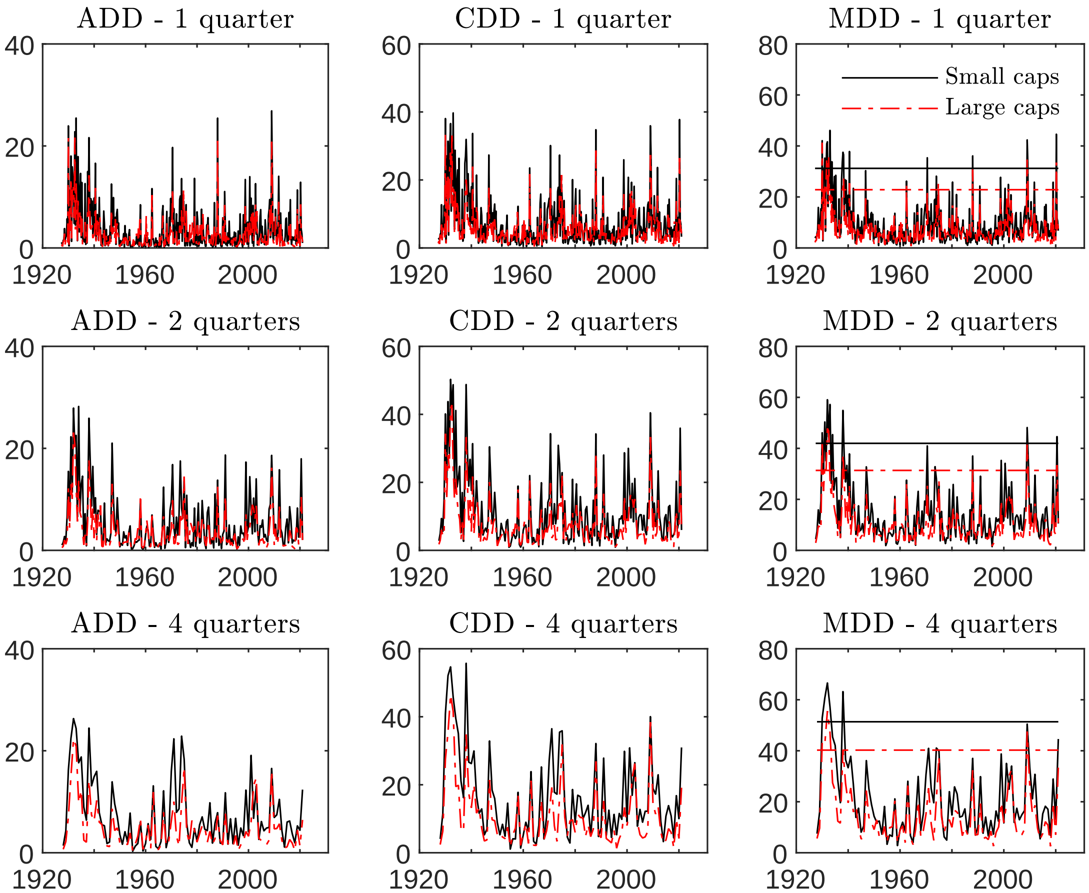
Note: This table reports the test statistics and the p-value for ADD, 20% CDD, and MDD. For CDD, the table also reports the number of exceedances. We consider three investment horizons: one quarter, two quarters, and four quarters. These results are based on the predictions of the models between 1990 and 2020.

Table 7: Out-of-Sample Allocation Based on Large Drawdowns (1990–2020)

Horizon	Statistics	Model-free 25 years	Model-free 63 years	1-regime normal	1-regime Student t	2-regime normal	2-regime Student t	3-regime normal	3-regime Student t
				Panel A: Minimization of ADD					
1 quarter	Weight	0.05	-0.08	0.44	0.55	0.04	0.40	-0.01	0.17
	ADD	2.49	2.51	2.75	2.84	2.48	2.69	**2.47**	2.52
2 quarters	Weight	-0.05	-0.14	0.49	0.66	0.00	0.41	0.06	-0.03
	ADD	3.31	3.34	4.00	4.35	3.36	3.85	3.35	**3.32**
4 quarters	Weight	-0.02	-0.12	0.56	0.77	0.20	0.49	0.31	-0.01
	ADD	4.03	4.08	4.83	5.56	4.21	4.72	4.37	**3.93**
				Panel B: Minimization of CDD					
1 quarter	Weight	0.05	-0.06	0.44	0.54	-0.11	0.41	-0.17	0.07
	20% CDD	5.25	5.32	5.75	5.90	5.24	5.66	**5.23**	5.27
2 quarters	Weight	-0.12	-0.25	0.46	0.63	-0.06	0.39	0.02	-0.18
	20% CDD	7.18	7.32	8.47	9.16	7.23	8.30	7.26	**7.15**
4 quarters	Weight	-0.28	-0.45	0.51	0.67	0.42	0.47	0.87	-0.07
	20% CDD	9.56	9.83	11.36	12.66	11.62	11.34	14.28	**9.22**
				Panel C: Minimization of MDD					
1 quarter	Weight	0.08	-0.13	0.43	0.53	-0.12	0.39	-0.17	0.05
	MDD	7.16	7.20	7.56	7.71	7.04	7.40	7.10	**7.02**
2 quarters	Weight	-0.09	-0.23	0.44	0.59	-0.08	0.39	0.00	-0.21
	MDD	10.01	10.14	11.31	11.95	10.10	11.14	10.06	**9.92**
4 quarters	Weight	-0.18	-0.31	0.48	0.64	0.09	0.49	0.18	-0.10
	MDD	13.87	13.87	16.08	17.15	14.42	16.16	14.80	**13.82**
				Panel D: Minimization of CED					
1 quarter	Weight	-0.21	-0.31	0.37	0.44	-0.16	0.26	-0.21	-0.14
	10% CED	21.62	22.32	22.25	22.53	20.69	21.43	**20.30**	20.53
2 quarters	Weight	-0.21	-0.35	0.37	0.47	-0.06	0.30	0.03	-0.25
	10% CED	27.39	27.51	29.24	30.03	25.26	28.80	25.56	**24.84**
4 quarters	Weight	0.04	-0.24	0.41	0.52	0.02	0.43	0.00	-0.06
	10% CED	33.52	32.47	35.22	36.40	33.96	36.51	33.96	**33.56**

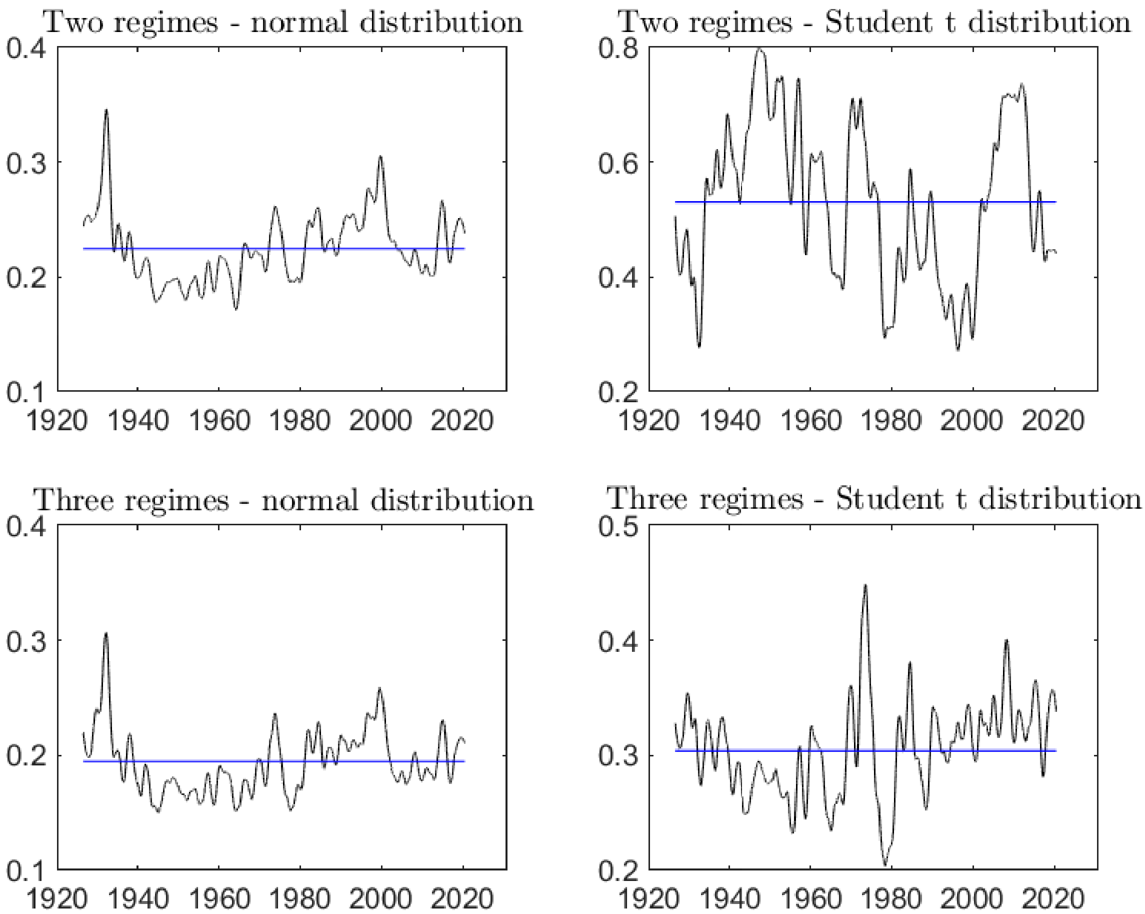
Note: This table reports the optimal weight of small caps (α^*) and the value of the objective function at the optimum, when the investor minimizes ADD, 20% CDD, the MDD, and 10% CED (Panels A to D, respectively). We consider three investment horizons: one quarter, two quarters, and four quarters. These results are based on the simulation of the model estimated over subsamples between 1990 and 2020.

Figure 1: Evolution of ADD, CDD, and MDD over non-overlapping subsamples



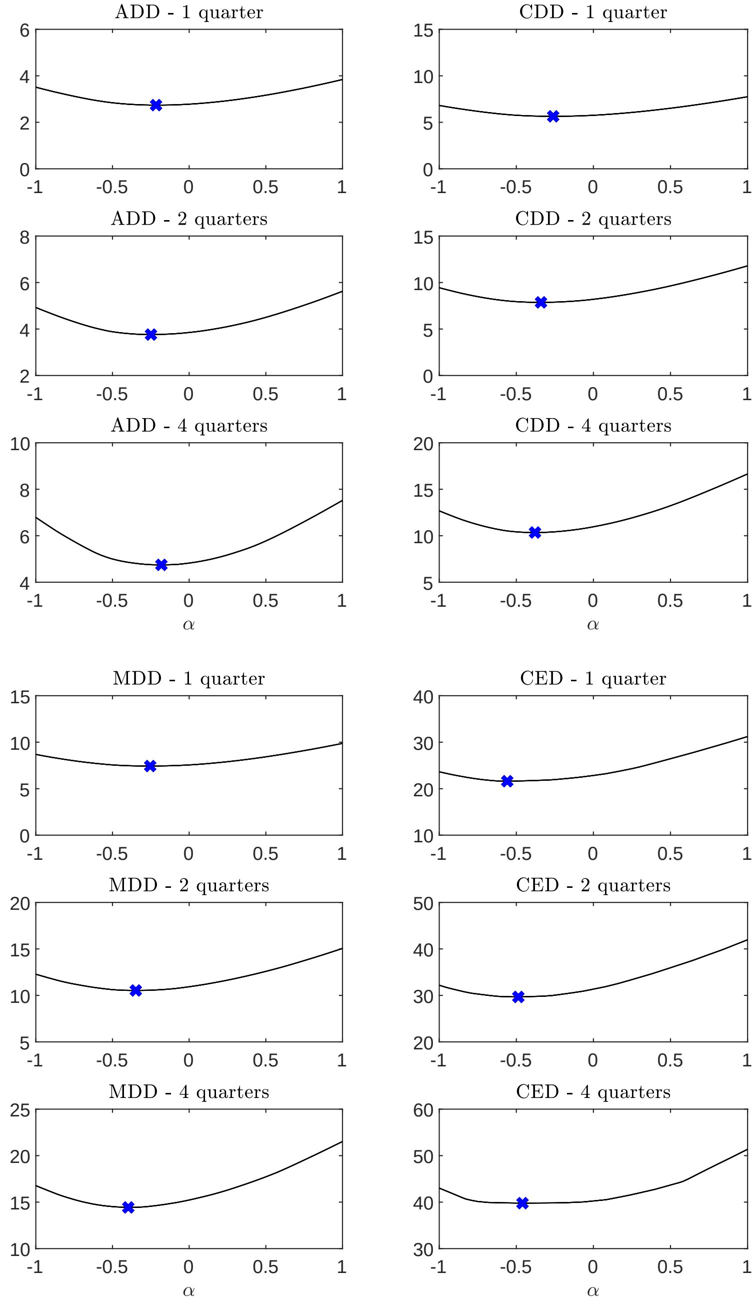
Note: The figure displays the evolution of ADD, CDD, and MDD over various non-overlapping subsamples (from one to four quarters) between 1926 and 2020. The straight line on right plots corresponds to 10% CED. The black lines correspond to the small caps, the red dashed lines to the large caps. CED is computed with 376, 188, and 94 observations for the one-quarter, two-quarter, and four-quarter horizons, respectively.

Figure 2: Filtered Probability of Being in the Bear State



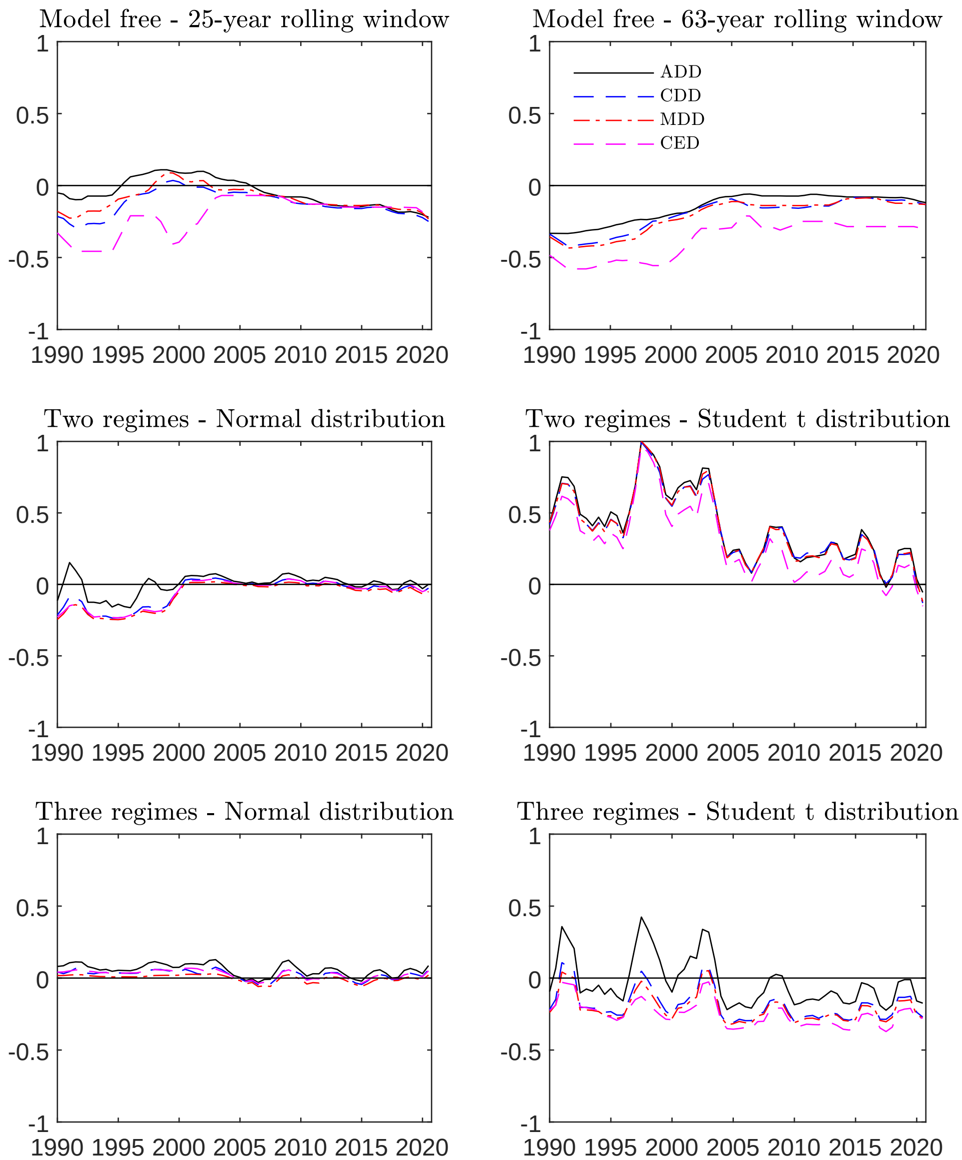
Note: The figure displays the filtered probability ϕ_{d+1} of being in the low expected return regime (bear state), for the two-regime and three-regime models. The horizontal blue line corresponds to the stationary probability of being in the bear state $\pi_{b,\infty} = \Pr[S_{d+1} = k_b]$, where k_b denotes the bear state.

Figure 3: Optimal Weights Based on the Full Sample (1926–2020)



Note: The figure displays the minimum values for ADD, 20% CDD, MDD, and 10% CED that an investor minimizing this criterion would have obtained for a portfolio weight ranging from -1 to 1 and for a one-quarter, two-quarter, and for-quarter, respectively. The optimal allocation is obtained by using the historical distribution over the full sample. The blue cross indicates the optimal portfolio weights.

Figure 4: Out-of-Sample Optimal Weights – Two-quarter Horizon (1990–2020)



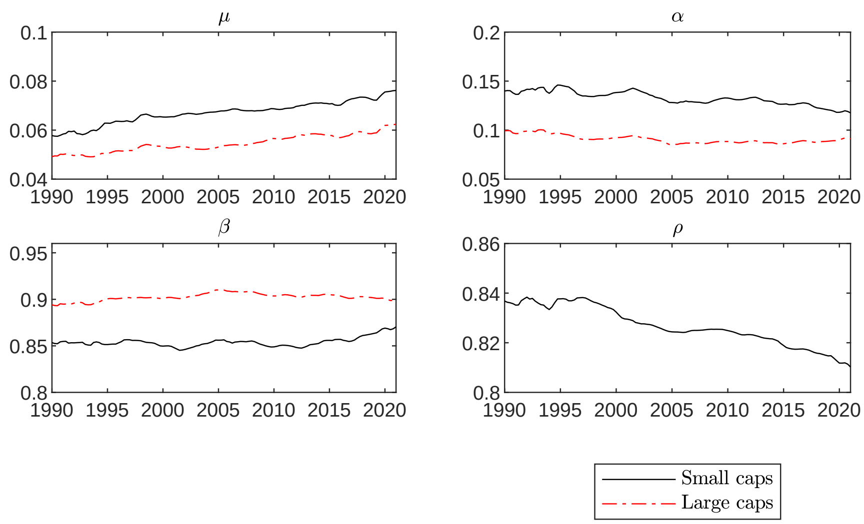
Note: The figure displays the temporal evolution of the optimal weight for the two-quarter horizon when predictions are based on the historical distribution (model free, based on 25-year and 63-year rolling window) and on the two-regime and three-regime models.

Appendices

A Evolution of Model Parameters

Figure A1: Model Parameters: One-regime Models

Panel A: Normal distribution



Panel B: Student's t distribution

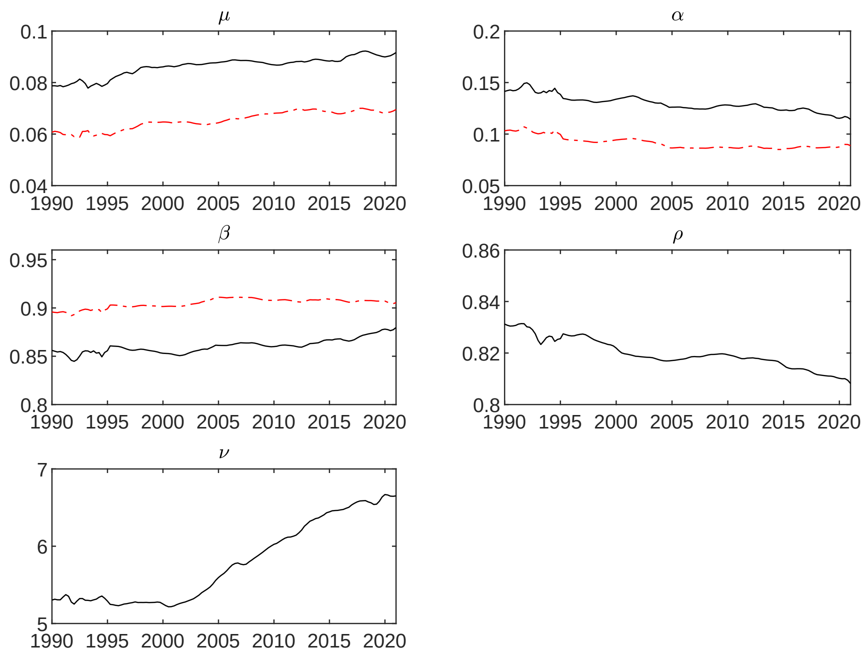
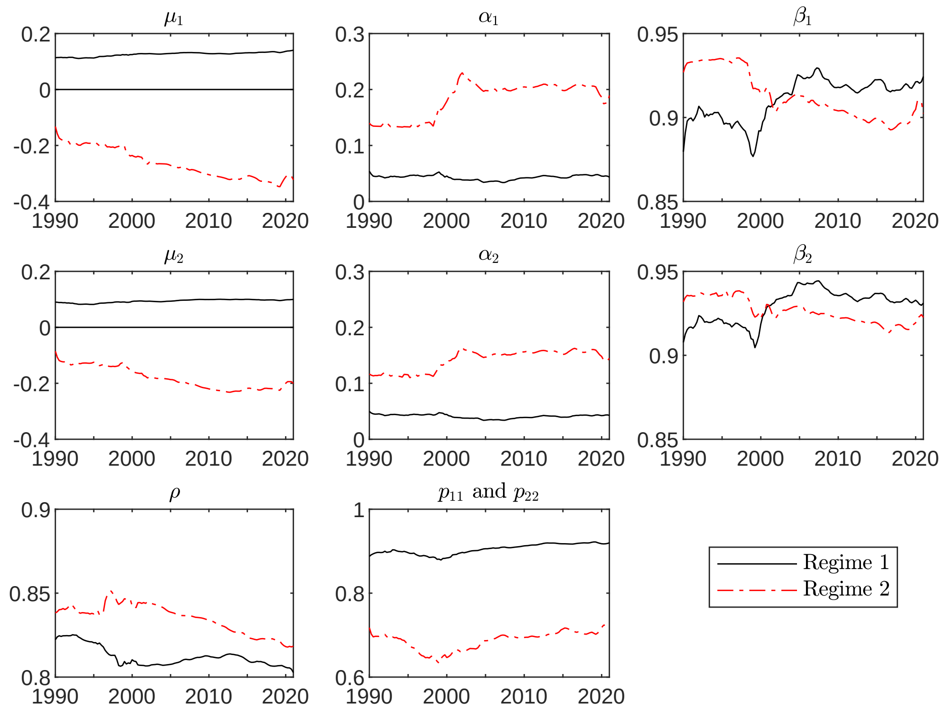


Figure A2: Model Parameters: Two-regime Models

Panel A: Normal distribution



Panel B: Student's t distribution

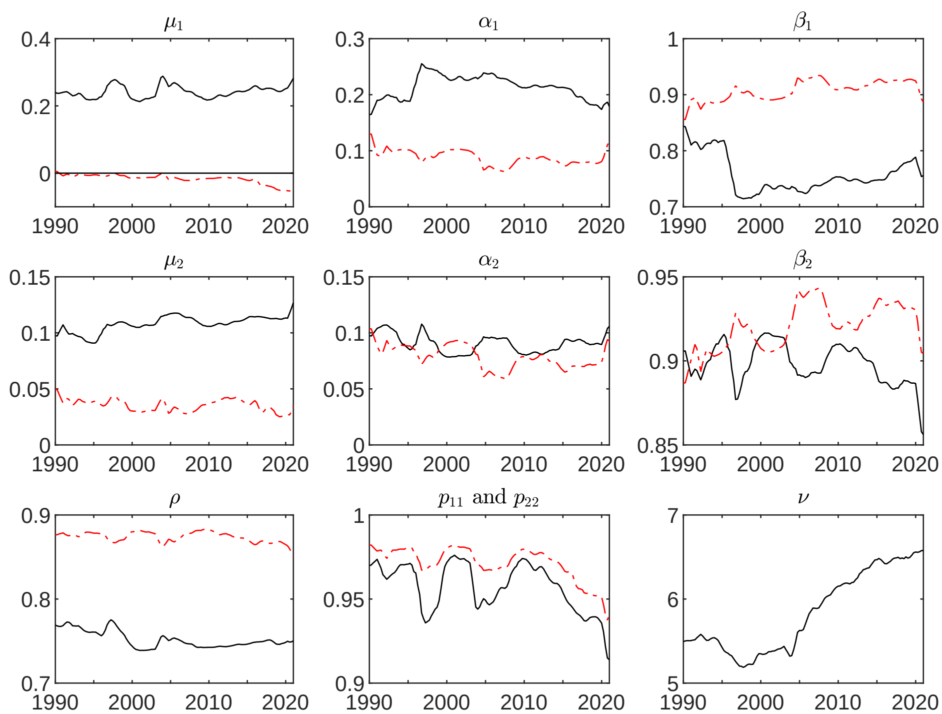
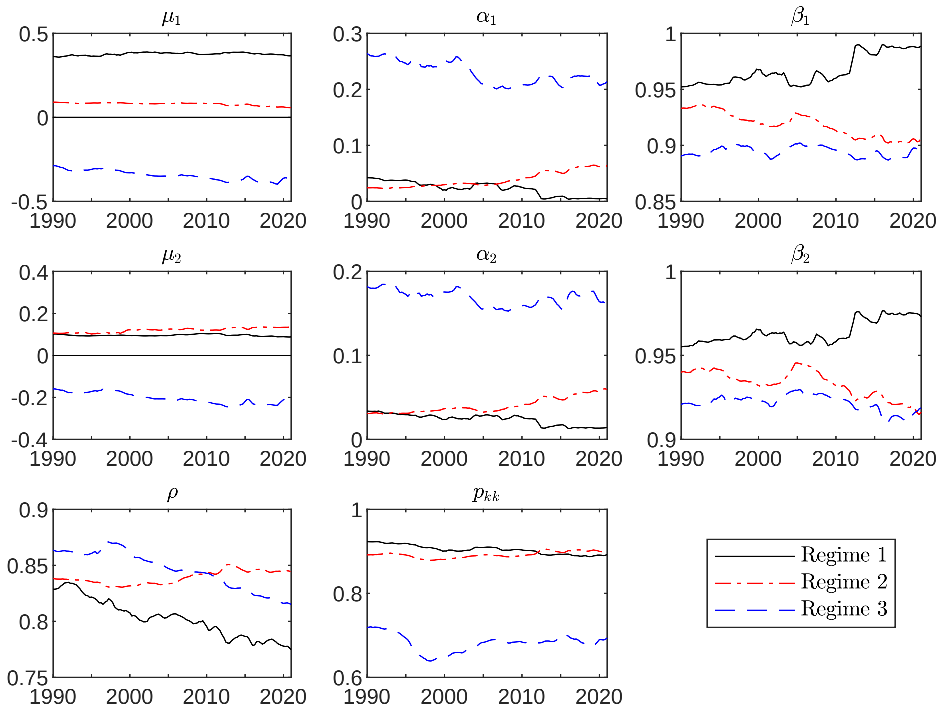
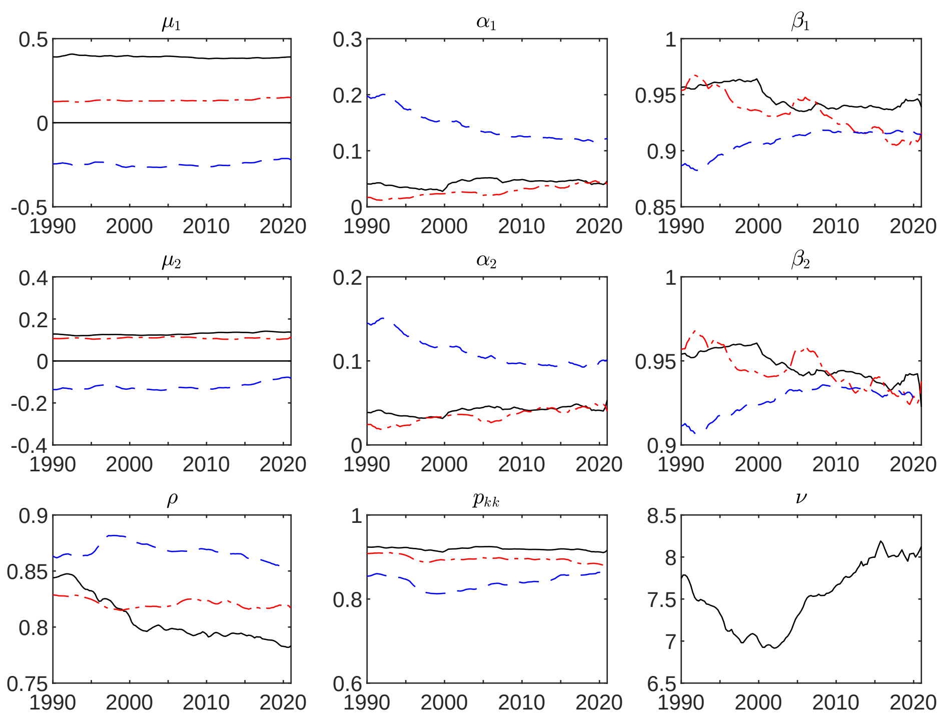


Figure A3: Model Parameters: Three-regime Models

Panel A: Normal distribution



Panel B: Student's t distribution



B Adequacy Tests

To backtest our predicted risk measures, we rely on the test by Acerbi and Szekely (2014). Our predictions of the large drawdown risk measures, generically defined as $X\hat{D}D_t$, are obtained by simulations from the models. Therefore, by backtesting $X\hat{D}D_t$, we test whether the model underlying a particular prediction is adequate. The tests and their statistics are presented in the paper. We describe below the methodology to compute the significance and run different experiments to assess the power of the tests.

B.1 Test Significance

To find the test significance, we generate Z statistics under the null hypothesis by simulating the model used to predict the risk measure under consideration. Then, by comparing the realized Z statistic to the simulated Z statistics under H_0, we evaluate the p-value, i.e. the probability to generate the realized Z statistic from the model underlying the null hypothesis.

We proceed as follows:

1. Using the model under H_0, we simulate Q samples of log-returns for all quarters of the backtesting period: $\{r_t^{(q)}\}_{h=1}^{H}$, $t = 1, \cdots, T$ and $q = 1, \cdots, Q$. The simulations are generated as in Section 3.2.

2. We calculate the risk measure of interest for each simulation, $XDD_t^{(q)}$, $t = 1, \cdots, T$. Then, we compute Z statistic for each simulation, $Z_{XDD}^{(q)}$, $q = 1, \cdots, Q$. They represent the distribution of the Z statistic under the null hypothesis.

3. We compute the realized Z statistic as we did for simulations, but using the observed log-returns (as in Section 6.1).

4. We estimate the significance by comparing the realized Z-statistics to the distribution of the simulated Z-statistics and compute the p-value of the bilateral test as: $p\text{-}val = 1 - \frac{1}{Q}\sum_{q=1}^{Q}(|Z_{XDD}^{(q)}| < |Z_{XDD}|)$.

B.2 Test Power

To gauge the power of the test, we run three different experiments. For each experiment, we specify two distinct models for the null hypothesis and the alternative hypothesis. By running tests on simulations of the alternative model (the true model), we evaluate the frequency at which the tests correctly reject the model used to predict risk measures under the null hypothesis.

For all experiments, we simulate paths of 124 quarters of 60 days, which approximately corresponds to the length of the backtesting period. We assume normal innovations ϵ_t for all models. The model parameters are calibrated on small caps returns.

Experiment 1: The true model (H_1) is a MS model with two regimes, and we assume a standard GARCH model under H_0. The unconditional expected return is the same for both models.

$\mathbf{H}_0$

$$r_t = 0.0508 + \epsilon_t$$
$$\epsilon_t = \sigma_t z_t$$
$$\sigma_t^2 = 0.0149 + 0.8654\sigma_{t-1}^2 + 0.1289\epsilon_{t-1}^2$$

$\mathbf{H}_1$

$$\text{Regime 1} : r_t = 0.1265 + 0.2836\epsilon_t$$
$$\text{Regime 2} : r_t = -0.2106 + 0.3464\epsilon_t$$
$$P = \begin{bmatrix} 0.9081 & 0.0919 \\ 0.3172 & 0.6828 \end{bmatrix}$$

Experiment 2: Both the true model (H_1) and the assumed model under H_0 are MS models with two states. They are identical, except for the probability of staying in Regime 1, which is lower under H_1:

$\mathbf{H}_0$

$$\text{Regime 1} : r_t = 0.1265 + 0.2836\epsilon_t$$
$$\text{Regime 2} : r_t = -0.2106 + 0.3464\epsilon_t$$
$$P = \begin{bmatrix} 0.9081 & 0.0919 \\ 0.4 & 0.6 \end{bmatrix}$$

$\mathbf{H}_1$

$$\text{Regime 1} : r_t = 0.1265 + 0.2836\epsilon_t$$
$$\text{Regime 2} : r_t = -0.2106 + 0.3464\epsilon_t$$
$$P = \begin{bmatrix} 0.9081 & 0.0919 \\ 0.3172 & 0.6828 \end{bmatrix}$$

Experiment 3: The true model (H_1) is a MS model with three regimes, whereas the model under H_0 has two regimes. Their unconditional expected returns are equal (0.0508).

$\mathbf{H}_0$

$$\text{Regime 1} : r_t = 0.1265 + 0.2836\epsilon_t$$
$$\text{Regime 2} : r_t = -0.2106 + 0.3464\epsilon_t$$
$$P = \begin{bmatrix} 0.9081 & 0.0919 \\ 0.3172 & 0.6828 \end{bmatrix}$$

$\mathbf{H}_1$

$$\text{Regime 1} : r_t = 0.3304 + 0.1849\epsilon_t$$
$$\text{Regime 2} : r_t = 0.0754 + 0.2538\epsilon_t$$
$$\text{Regime 3} : r_t = -0.265 + 0.3873\epsilon_t$$
$$P = \begin{bmatrix} 0.9206 & 0.0169 & 0.0453 \\ 0.0350 & 0.8954 & 0.2663 \\ 0.0444 & 0.0877 & 0.6884 \end{bmatrix}$$

Table A1 reports the power of the test and Figure A4 displays the cumulative distribution function of the Z-statistic under the null hypothesis and the alternative hypothesis.

In the first experiment, all the tests reject with a power of 100% the null hypothesis of a GARCH model when the true model is a MS model with two regimes. The second regime is characterized by a negative return that generates larger drawdowns than the assumed GARCH model does. The tests capture unambiguously that the drawdowns emerge from different models.

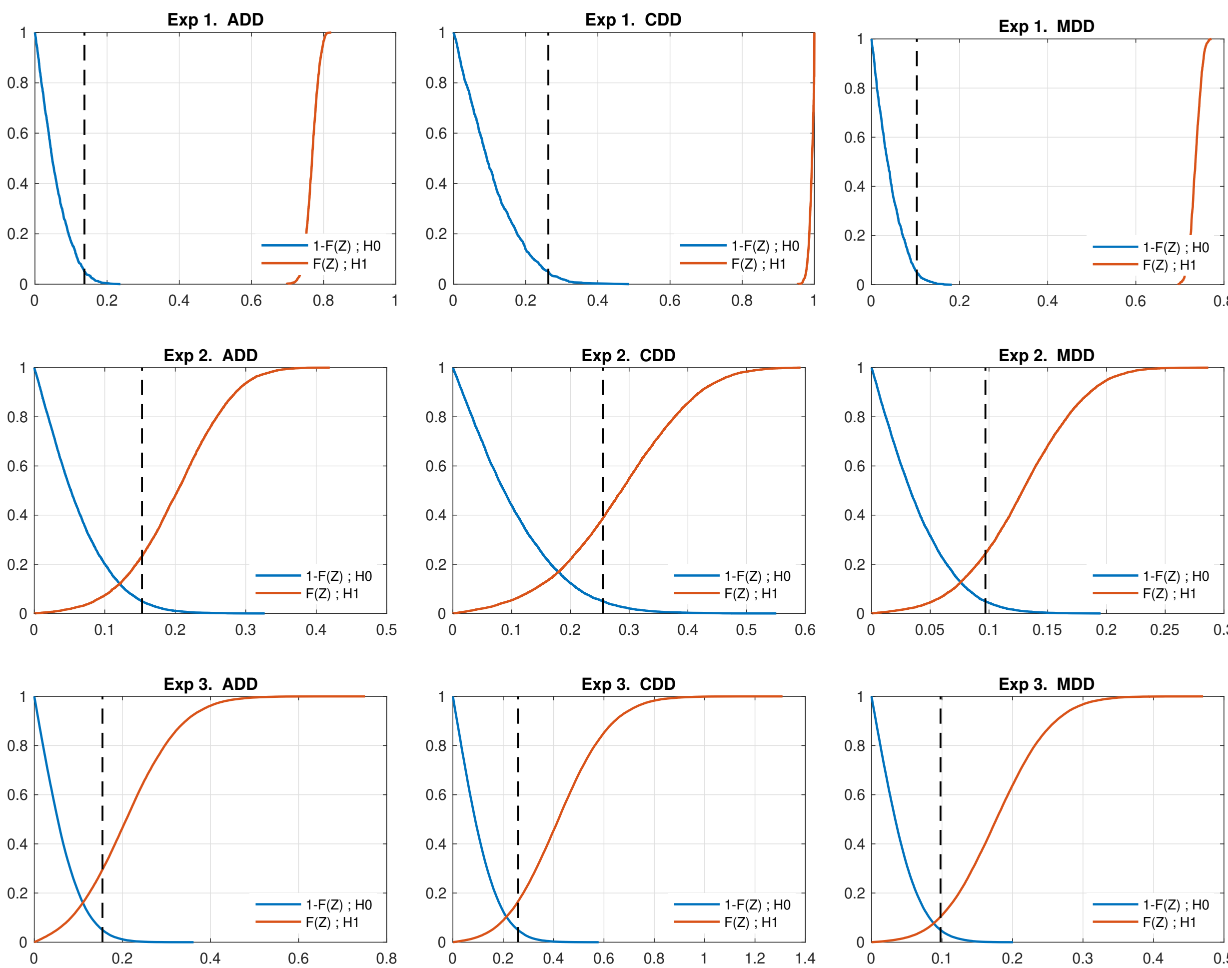
In the second experiment, the powers of the tests remain high. The tests capture that the second regime under the null hypothesis has a lower expected duration, which results in smaller drawdowns.

In the third experiment, we compare two MS models, one with two regimes and the other with three regimes. They have the same unconditional expected return. The null hypothesis, which ignores the existence of a third regime, is rejected with a power ranging between 67% and 90%. This experiment sheds light on the fact that the choice between two and three regimes is consequential for measuring large drawdown risk measures.

Table A1: Test power

Test power ($\alpha = 5\%$)	ADD	CDD	MDD
Experiment 1	100%	100%	100%
Experiment 2	83.05%	80.80%	88.90%
Experiment 3	67.10%	77.20%	81.95%

Figure A4: Distribution of Z statistics



Swiss Finance Institute

Swiss Finance Institute (SFI) is the national center for fundamental research, doctoral training, knowledge exchange, and continuing education in the fields of banking and finance. SFI's mission is to grow knowledge capital for the Swiss financial marketplace. Created in 2006 as a public–private partnership, SFI is a common initiative of the Swiss finance industry, leading Swiss universities, and the Swiss Confederation.

swiss:finance:institute

c/o University of Geneva, Bd. Du Pont d'Arve 42, CH-1211 Geneva 4
T +41 22 379 84 71, rps@sfi.ch, www.sfi.ch