


Article

Large Drawdowns and Long-Term Asset Management

Eric Jondeau ^{1,*}  and Alexandre Pauli ^{2,†}

¹ Faculty of Business and Economics (HEC Lausanne), Swiss Finance Institute and CEPR, University of Lausanne, CH 1015 Lausanne, Switzerland

² Ecole Polytechnique Fédérale de Lausanne, Route Cantonale, CH 1015 Lausanne, Switzerland; alexandre.pauli@epfl.ch

* Correspondence: eric.jondeau@unil.ch

† The authors contributed equally to this work.

Abstract: Long-term investors are often hesitant to invest in assets or strategies prone to significant drawdowns, primarily due to the challenge of predicting these drawdowns. This study presents a multivariate Markov-switching model for small- and large-cap returns in the U.S. equity markets, demonstrating that three distinct regimes are necessary to capture the negative trends in expected returns during financial crises. Our findings indicate that this framework enhances the prediction of conditional drawdowns compared to standard alternative models of financial returns. Furthermore, out-of-sample analysis shows that investment strategies based on these predictions outperform those relying on models with one or two regimes.

Keywords: large drawdowns; stock-market returns; Markov-switching model; portfolio allocation model

1. Introduction

In recent decades, the recurrence of disasters has raised the issue of protecting investors' portfolios from large market drawdowns.¹ In the first quarter of 2020, the U.S. equity market experienced a 36% decline—one of the five largest one-quarter drawdowns in the last century. Such substantial drawdowns are particularly significant for long-term asset managers, including insurers, pension funds, and sovereign wealth funds, as these losses can not only impact annual portfolio performance but may also threaten the survival of the fund. To mitigate these losses, some managers incorporate drawdown objectives into their portfolio optimization processes. However, modeling and predicting the temporal evolution of large drawdowns remains a considerable challenge, particularly as these events often unfold over extended periods (e.g., a quarter or a year), a time scale inconsistent with most conventional financial econometric models.

In this paper, we address both modeling and prediction challenges related to market drawdowns. We introduce a model for daily financial returns based on regime switches, which facilitates the prediction of significant market drawdowns. Our findings demonstrate that three distinct regimes are necessary to capture the long-term dynamics of U.S. stock market returns. Leveraging this model, we design an investment strategy aimed at minimizing the expected value of a large drawdown measure, possibly subject to an expected return target. We then evaluate this strategy out of sample over the last 30 years, providing evidence that investing according to this approach effectively allows investors to reduce their losses due to large market downturns, outperforming standard alternative models such as GARCH-type models.

Most of the literature exploring the prediction of large drawdowns relies on some form of historical simulation. For instance, Chekhlov et al. (2005) simulate scenarios based on historical data to predict next-period drawdowns. This approach is likely to work well in-sample but may not perform well when market conditions vary over time, possibly due to climate change. A key challenge in using parametric models to predict large drawdowns is



Citation: Jondeau, Eric, and Alexandre Pauli. 2024. Large Drawdowns and Long-Term Asset Management. *Journal of Risk and Financial Management* 17: 552. <https://doi.org/10.3390/jrfm17120552>

Academic Editors: W. Brent Lindquist and Svetlozar (Zari) Rachev

Received: 22 October 2024

Revised: 1 December 2024

Accepted: 3 December 2024

Published: 10 December 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

that most models fail to capture large losses accumulating over relatively extended periods. In particular, standard GARCH models, even with nonnormal distributions, are not able to generate the negative trend that we observe during financial market 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. Therefore, when a given return process enters a bear regime, it can accumulate negative values for relatively long periods of time. Papers describing these models for financial returns often overlook this long-term property because their main focus is on other features of the returns' data generating process. In particular, [Ang and Bekaert \(2002\)](#), [Ang and Chen \(2002\)](#), and [Guidolin and Timmermann \(2004, 2007, 2008\)](#) used MS models to capture the nonnormality or the asymmetric correlation of returns. More recently, similar approaches were adopted to analyze the consequences in financial markets of disasters arising from climate change ([Karydas and Xepapadeas 2019](#), and [Barnett et al. 2020](#)) or from the COVID-19 pandemic ([Pagano et al. 2023](#)). For shorter prediction horizons (e.g., 10 days), [Peng et al. \(2022\)](#) find that an MS model with two regimes is sufficient for measuring large drawdowns. For longer horizons, we provide evidence that three regimes are necessary to capture the occurrence of financial crises and subsequent drawdowns effectively.

The literature defines several concepts of large drawdowns. One widely accepted measure is the period maximum drawdown (MDD), which corresponds to the largest loss from peak to trough over 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 \(2008\)](#) in discrete time. [Chekhlov et al. \(2005\)](#) define the concept of conditional drawdown (CDD), which corresponds to an average of the largest drawdowns in a given period, with 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 values 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.

In this paper, we first evaluate the number of regimes necessary to predict large drawdowns. We perform this analysis using a long sample of daily returns for both small- and large-cap stocks, covering nearly 100 years, with the last 30 years used for out-of-sample evaluation.² Given that small caps are more vulnerable to drawdowns than large caps, we assess whether a strategy focused on minimizing an expected large drawdown measure produces optimal weights that differ from those minimizing the portfolio variance. We estimate multivariate MS-GARCH models with one to four regimes and different distributions for the innovation process. We empirically demonstrate that three regimes are statistically necessary to fit the data while four regimes would not provide any additional information. 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 fail to generate such extensive drawdowns.

We proceed by running an out-of-sample investment exercise. Using a rolling window, we estimate the various models and allocate the portfolio for the subsequent period. Although this experiment is time-consuming, it effectively mimics an investor's real-time allocation process, penalizes overparameterized models, and helps mitigate the winner's curse problem ([Hansen 2009](#)). We find that models with three regimes provide the best predictions of large drawdowns in this experiment. When allocating investor wealth by minimizing the predicted large drawdown measure for the upcoming period, the ex post drawdown of the portfolio is consistently lower with the three-regime model compared to the GARCH model or the two-regime model. Investors using the three-regime model tend to allocate long positions in large caps and short positions in small caps—an allocation that proves effective in mitigating large drawdowns.

Finally, we consider an investor aiming to maximize an expected return—large draw-down criterion, which also takes into account the model’s ability to predict expected returns. The out-of-sample results again demonstrate that one- and two-regime models are nearly always dominated by the three-regime models, particularly for extreme drawdowns.

2. Materials and Methods

2.1. Definitions and Measurement

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 notations and the various concepts of large drawdowns that we will analyze in the remainder of the paper.

Let $p_t = (p_{t,1}, \dots, p_{t,H})$ 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).³ The drawdown within this period on day h corresponds to the return loss between the last peak and the current price:

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

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

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

[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 of the drawdown distribution, 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 the average drawdown, $ADD_t = E_t[DD_t]$. When $\theta \rightarrow 1$, $CDD_{\theta,t}$ 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 MDD values 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 the sample average period MDD, i.e., $E[MDD]$.

Similar to the well-known expected shortfall, which measures the tail conditional expectation of the return distribution, CDD and CED correspond to the tail conditional expectation of the drawdown distribution and the tail conditional expectation of the maximum drawdown distribution, respectively. Importantly, as shown by [Chekhlov et al. \(2005\)](#) and [Goldberg and Mahmoud \(2015\)](#), MDD, CDD, and CED satisfy the properties of deviation measures, i.e., nonnegativity, 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 three large drawdown measures: CDD, MDD, and CED. An essential feature of these measures is that drawdowns can develop over different time frames.⁴ 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

We now briefly describe how to measure the ex post drawdown of an asset or a portfolio of assets. 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 wants to control for 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 $p_t = (p_{t,1}, \dots, p_{t,H})$ for $t = 1, \dots, T$. For each subsample, the vector of drawdowns is denoted by $DD_t = (DD_{t,1}, \dots, DD_{t,H})$, with $DD_{t,h} = \max_{1 \leq j \leq h} p_{t,j} - p_{t,h}$, as before.

We obtain the drawdown-based measures for each subsample as follows. The period ADD is simply given by the sample mean of the drawdowns, $ADD_t = \frac{1}{H} \sum_{h=1}^H DD_{t,h}$; the period MDD is given by the maximum drawdown of the subsample, $MDD_t = \max_{1 \leq h \leq H} DD_{t,h}$; and the period 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 $I_x = 1$ if x is true and 0 otherwise, and $Th_{\theta,t} = \inf\{s \mid \frac{1}{H} \sum_{h=1}^H I_{(DD_{t,h} > s)} \leq 1 - \theta\}$.

Finally, CED is based on the distribution of the period MDD measures over the full sample: we collect the MDD values over all the subsamples $MDD = (MDD_1, \dots, MDD_T)$ and the sample CED corresponds to the average of the worst $(1 - \tilde{\theta}) \times 100\%$ MDD values 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 $\alpha_t = (\alpha_{1,t}, \dots, \alpha_{n,t})$ 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: $p_{t+1}(\alpha_t) = (p_{t+1,1}(\alpha_t), \dots, p_{t+1,H}(\alpha_t))$, 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 2.2.⁵ The investment criterion consists of minimizing the expected value of one of the large drawdown measures for the next investment period (period CDD, MDD, or CED), which we denote generically as $XDD_{t+1}(\alpha_t)$. The optimal weight is:

$$\alpha_t^* \in \arg \min_{\{\alpha_t\}} E_t[XDD_{t+1}(\alpha_t)]. \tag{5}$$

We also consider a criterion based on the trade-off between the expected return and the risk of a large drawdown:

$$\alpha_t^* \in \arg \max_{\{\alpha_t\}} E_t[R_{p,t+1}(\alpha_t)] - \frac{\lambda}{2} E_t[XDD_{t+1}(\alpha_t)], \tag{6}$$

where λ denotes the aversion for large drawdowns.

The minimization problem (5) corresponds to the case where the aversion for large drawdowns λ goes to infinity. Although the optimization problem (6) may be more attractive for an investor willing to combine risk and return, problem (5) allows us to investigate more specifically the ability of the various models to predict large drawdowns. In Section 3.4, we will focus our comments on the results based on problem (5) and briefly discuss problem (6).

To obtain predictions of portfolio large drawdown measures, i.e., $E_t[XDD_{t+1}(\alpha_t^*)]$, we proceed as follows: First, we assume a multivariate MS-GARCH model to describe the data generating process (DGP) for daily log-returns and endogenize the path dependence of the return process. Two-regime and three-regime models would capture the drawdown trend if switching probabilities are sufficiently low. Using this DGP, we simulate assets' daily returns for the next investment period. Second, for a given portfolio weight vector, we obtain simulated paths 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.

2.4. Methodology

2.4.1. 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 of daily log-returns for the n assets is denoted by $\tilde{r}_{d+1} = (\tilde{r}_{1,d+1}, \dots, \tilde{r}_{n,d+1})$. The temporal index $d = 1, \dots, D$, represents days and runs over the full sample.⁶

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 $P = (p_{kk'})_{k,k'=1,\dots,K}$, where transition probabilities are $p_{kk'} = \Pr(S_{d+1} = k' | S_d = k)$, $k, k' \in \{1, \dots, K\}$.

Expected returns are constant within each state: $\mu_{d+1}(S_{d+1}) = \mu^{(k)}$ when $S_{d+1} = k$.⁷ 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 defined as a standard univariate GARCH(1,1) process:⁸

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

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

The $(n \times n)$ correlation matrix is constant in a given state: $\Gamma^{(k)} = (\rho_{ij}^{(k)})_{i,j=1,\dots,n}$, 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 $(\sigma_{i,d+1}^{(k)2})_{i=1,\dots,n}$ 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, for investors who care about higher moments, the investment criterion might involve metrics, such as large drawdowns, that depend on the properties of the innovation distribution. Second, in our model, large drawdowns can be captured in principle by a higher probability of being in a bear market or by negative expected returns in the bear regime. However, a lower degree of freedom of the Student's t distribution can also affect the dynamics of regime shifts and possibly result in lower returns.⁹

To make inferences about the regimes, we calculate the probability of being in each regime. We denote by $f(\tilde{r}_{d+1} | \tilde{\mathbf{x}}_d, \theta)$ the distribution of the daily log-return process conditional upon past log-returns, with $\tilde{\mathbf{x}}_d = \{\tilde{r}_d, \tilde{r}_{d-1}, \dots\}$ 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 $\pi_{k,d+1} = \Pr[S_{d+1} = k | \tilde{\mathbf{x}}_d]$ and the filtered probabilities $\phi_{k,d+1} = \Pr[S_{d+1} = k | \tilde{\mathbf{x}}_{d+1}]$ as:

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

$$\text{with } l_{d+1} = \begin{bmatrix} f(\tilde{r}_{d+1} | \tilde{\mathbf{x}}_d, S_{d+1} = 1; \theta) \\ \vdots \\ f(\tilde{r}_{d+1} | \tilde{\mathbf{x}}_d, S_{d+1} = K; \theta) \end{bmatrix} \quad \text{and} \quad e = (1, \dots, 1)'$$

The estimation of the model is based on standard likelihood maximization, where the log-likelihood is defined as: $\log L_D(\theta) = \sum_{d=1}^{D-1} f(\tilde{r}_{d+1} | \tilde{\mathbf{x}}_d, \theta) = \sum_{d=1}^{D-1} \log(\sum_{k=1}^K \pi_{d+1} \odot l_{d+1})$. 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.¹⁰

While MS-GARCH models are valuable tools, they have certain limitations. To avoid overfitting, it is essential to carefully determine the number of regimes and ensure the model accurately captures the key patterns in the data. Estimating these models can be computationally demanding, especially because large datasets are needed for reliable results. This makes them less practical for systems with many variables or a high number of regimes. Additionally, the regimes must correspond to clear economic or financial conditions to make their interpretation meaningful. In our analysis, we address these challenges by using a simple model with only two processes, a long dataset, and by rigorously checking how well the model fits the data.

2.4.2. 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 (2004, 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 return distribution, cannot be computed analytically, even in this simple model (Guidolin and Timmermann 2004). Additionally, in models such as MS-GARCH, analytical expressions are usually not available because variances are path

dependent. For this reason, we compute expected large drawdowns with Monte Carlo simulations.

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 (Equation (5)) 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 $1, \dots, t$. The last day of the estimation period is denoted by $d = t \times H$. The next quarter, $t + 1$, contains days $d + 1, \dots, d + H$.
2. For a given estimated model, we simulate Q samples of length H of daily log-returns for the n assets: $\{r_{t+1,h}^{(q)}\}_{h=1}^H, q = 1, \dots, Q$. 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 simulate 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 $h = 1, \dots, H$.
3. For a portfolio weight vector α_t , we obtain daily log-values of the portfolio: $p_{t+1}^{(q)}(\alpha_t) = (p_{t+1,1}^{(q)}(\alpha_t), \dots, p_{t+1,H}^{(q)}(\alpha_t))$, where $p_{t+1,h}^{(q)}(\alpha_t) = \log(\alpha'_t \exp(R_{t+1,h}^{(q)}))$.
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 MDD values obtained over all simulations $MDD_{t+1}(\alpha_t) = (MDD_{t+1}^{(1)}(\alpha_t), \dots, MDD_{t+1}^{(Q)}(\alpha_t))$ and take the average of the worst $(1 - \tilde{\theta}) \times 100\%$ MDD values 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 problem (5). To solve investment problem (6), we also predict the expected return of the portfolio for the next quarter $t + 1$ as $E_t[R_{p,t+1}(\alpha_t)] = \frac{1}{Q} \sum_{q=1}^Q [\exp(p_{t+1,H}^{(q)}(\alpha_t) - p_{t,H}^{(q)}(\alpha_t)) - 1]$.

To obtain accurate estimates of the optimal weights, we simulate a large number of draws ($Q = 50,000$).

3. Results

3.1. 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%).¹¹ The portfolios comprise all NYSE, AMEX, and NASDAQ stocks for which market equity data are available. The sample spans from July 1926 to December 2020, encompassing a total of 24,896 daily returns.¹²

Size portfolios offer several advantages for analyzing the construction of a portfolio in the context of large drawdowns. First, data on these portfolios is available over an extensive period (nearly 100 years). Second, size portfolios have been the focus of large attention for decades, making their properties relatively well understood, particularly in terms of their risk and return characteristics. Third, numerous studies have examined small-cap and large-cap stocks due to their distinct behaviors during crises and varying market conditions, such as bull and bear markets. [Perez-Quiros and Timmermann \(2000\)](#) provide evidence that small caps exhibit a high degree of asymmetry between recessionary and expansionary states. During recessions, small caps are more strongly impacted than large caps by deteriorating credit market conditions. [Ang and Chen \(2002\)](#) and [Patton \(2004\)](#) investigate the dependence between small and large caps, focusing on their asymmetric behavior during bull and bear markets. [Huang et al. \(2012\)](#) report that small firms are more exposed

to extreme downside risks, and that the higher average returns of small caps actually compensate investors for the occurrence of larger drawdowns. The COVID-19 market crash illustrates this phenomenon: in the first quarter of 2020, small caps experienced a 45% decline, whereas large caps decreased by only 33%. From this perspective, we compare the optimal weights allocated to size portfolios based on the minimization of large drawdowns and the minimization of standard portfolio variance.

Table 1 reports statistics on small 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. Small caps exhibit higher annual return and higher volatility on average. The Value-at-Risk (VaR) and expected shortfall (ES) measures demonstrate that small caps are more prone to large adverse shocks. The overall MDD is equal to 92% for small caps and 86.5% for large caps, corresponding in both cases to the stock market crash of 1929–1932.

The second part of the table reports the four sample measures of large drawdowns described in Section 2.2, for horizons of one quarter, two quarters, and four quarters. We compute CDD for a probability $\theta = 0.8$, i.e., we consider the average of the worst 20% drawdowns in a given subsample (e.g., the worst 12 drawdowns in a given quarter) (see Chekhlov et al. 2005). We compute CED with a probability $\tilde{\theta} = 0.9$, which corresponds to the worst 10% of MDD values in the sample (the worst 12 MDD values in the out-of-sample period) (see Goldberg and Mahmoud 2016). By examining the four period drawdown measures, we find that they also 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 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 ADD, CDD, and MDD measures in the full sample. The four-quarter MDD has AR(1) parameters equal to 48% for small caps and 45% for large caps. In the out-of-sample period (Panel B: 1990–2020), large drawdowns are much less persistent. The AR(1) parameter is usually low and close to 0 for small caps at all horizons, although it remains slightly higher for large caps. For the four-quarter MDD, the AR(1) parameters are equal to 11% for small caps and 38% for large caps.

Figure 1 presents the temporal evolution of the large drawdown measures. As expected, 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). With the four-quarter CED as a threshold, we identify only two exceptional drawdowns (the Great Depression and the subprime crisis episodes), whereas with the two-quarter CED we would also include the COVID-19 downturn as an exceptional event.

3.2. Full-Sample Model Estimation

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. 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 dynam-

ics, 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 properties of the models with two regimes are very different 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 are much lower for small caps than for large caps, probably reflecting tighter credit market conditions. The probability of remaining in the bear state is relatively low ($p_{22} = 68.3\%$), with a stationary probability equal to 22%.¹³

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
20% CDD - 1Q	7.75	0.25	5.76	0.28	7.28	0.12	5.27	0.23
20% CDD - 2Q	11.80	0.41	8.20	0.46	10.81	0.05	7.14	0.30
20% CDD - 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
10% CED - 1Q	31.22		22.84		25.97		20.60	
10% CED - 2Q	41.97		31.35		35.40		25.41	
10% CED - 4Q	51.37		40.25		38.36		31.19	

Note: This table reports statistics (in percentage) 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. VaR(α) and ES(α) denote the Value-at-Risk and Expected Shortfall computed for probability α . “XDD - n Q” means that the XDD measure is computed over n quarters. “20% CDD” means that the CDD measure is computed for probability $\theta = 20\%$. “10% CED” means that the CED measure is computed for probability $\tilde{\theta} = 10\%$.

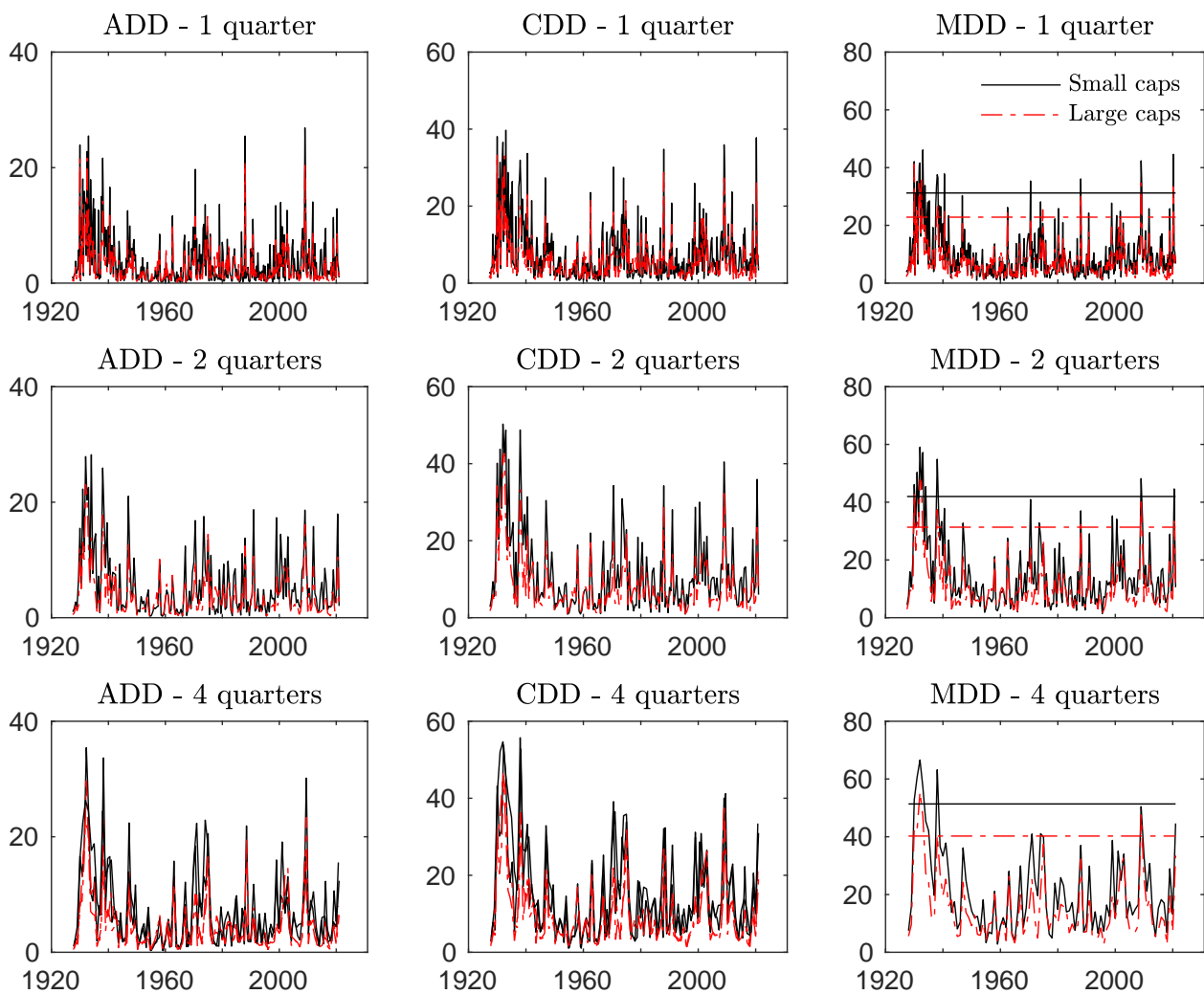


Figure 1. Evolution of ADD, CDD, and MDD over non-overlapping subsamples. The figure displays the evolution in percentage of ADD, 20% CDD, and MDD over various non-overlapping subsamples (from one to four quarters) between 1926 and 2020. The straight line on right plots corresponds to 10% CED. The black 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.

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-tailed nature of the Student’s *t* innovations.¹⁴

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. Regime 3 accounts for bear market conditions. As in the two-regime case, small caps have much higher expected returns than large caps in Regime 1 and much lower expected returns than large caps in Regime 3. Regime 3 also exhibits a

higher correlation than Regime 1, reflecting the asymmetry in dependence found by [Ang and Chen \(2002\)](#).

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			

Note: See Table 3 for details.

The degree of freedom of the Student’s t distribution is equal to 7.8, suggesting relatively thin tails compared to 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, the stationary probability of being in Regime 3 is higher than in the model with normal innovations (30.4% versus 19.5%).

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 (0.015)	0.0169 (0.003)	0.0453 (0.018)		0.9402 (0.009)	0.0170 (0.004)	0.0266 (0.007)	
$P_{2,:}$	0.0350 (0.016)	0.8954 (0.008)	0.2663 (0.026)		0.0165 (0.007)	0.9116 (0.011)	0.1139 (0.013)	
$P_{3,:}$	0.0445 (0.008)	0.0877 (0.006)	0.6884 (0.021)		0.0433 (0.008)	0.0715 (0.009)	0.8595 (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 report parameter estimates for the models 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 spanning from 1926 to 2020. BIC denotes the Bayesian Information Criterion.

A formal test of the number of regimes can be performed using the Likelihood-Ratio (LR) test but the usual asymptotic distribution of the test statistic does not hold. The reason is that, in the test of the null hypothesis that $n - 1$ regimes are sufficient against the alternative of n regimes, parameters associated with the n -th regime are not identified under the null hypothesis and the regularity conditions justifying the χ^2 approximation to the LR test do not hold. Hansen (1992, 1996) proposed an LR test procedure that addresses this problem (see also Garcia 1998). Specifically, we adopt the strategy proposed by Ang and Bekaert (2002), which is based on Monte-Carlo simulations to obtain the finite-sample distribution of the LR test statistic.¹⁵ We implement this approach for all tests of the number of regimes from 1 to 3 (see Table 4). For 1 and 2 regimes, we reject the null hypothesis, with p -values all below 0.5%, whatever the distribution of the innovation process. These tests indicate that the one-regime model should be rejected against the two-regime model and the two-regime model should be rejected against the three-regime model.

We also estimate four-regime models (with normal and Student’s t innovations) with our data and test whether 3 regimes are sufficient to match the data. Compared to three regimes, the gain in likelihood with four regimes is relatively large, but using the simulation-

based LR distribution, we do not reject the null hypothesis of three regimes against four regimes, with a *p*-value equal to 20.2% with the normal distribution and to 19% with the Student’s *t* distribution.¹⁶

Table 4. Likelihood ratio tests.

Null Hypothesis	Alternative Hypothesis	dof	LR Stat.	<i>p</i> -Value
H0(N1): 1 regime—normal	Ha(N1): 1 regime—Student’s <i>t</i>	1	4216.3	<0.5%
H0(NR1): 1 regime—normal	Ha(NR1): 2 regimes—normal	11	4637.0	<0.5%
H0(TR1): 1 regime—Student’s <i>t</i>	Ha(TR1): 2 regimes—Student’s <i>t</i>	11	1528.0	<0.5%
H0(N2): 2 regimes—normal	Ha(N2): 2 regimes—Student’s <i>t</i>	1	1107.3	<0.5%
H0(NR2): 2 regimes—normal	Ha(NR2): 3 regimes—normal	13	1385.3	<0.5%
H0(NT2): 2 regimes—Student’s <i>t</i>	Ha(NT2): 3 regimes—Student’s <i>t</i>	13	1137.2	<0.5%
H0(N3): 3 regimes—normal	Ha(N3): 3 regimes—Student’s <i>t</i>	1	859.2	<0.5%
H0(NR3): 3 regimes—normal	Ha(NR3): 4 regimes—normal	15	924.0	20.2%
H0(NT3): 3 regimes—Student’s <i>t</i>	Ha(NT3): 4 regimes—Student’s <i>t</i>	15	772.0	19.0%

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 (dof) of the test (number of restrictions under the null hypothesis). The fourth column reports the LR test statistics. As explained in the main text, the *p*-values in the fifth column are based on the asymptotic χ^2 distribution for the test of the null hypothesis of the normal distribution (N1, N2, and N3) and on simulations of the finite-sample distribution for the test of the null of $n - 1$ regimes against n regimes.

In Figure 2, we represent the filtered probability of being in the low regime for the two-regime and three-regime models. First, we note that the two-regime/Student’s *t* model produces a high filtered probability (on average 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 the same times with similar probabilities. 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. Consequently, models with normally distributed innovations do not accurately capture observed market downturns.

In the three-regime model with Student’s *t* innovations, 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).¹⁷ 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).¹⁸

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 the other competing models.

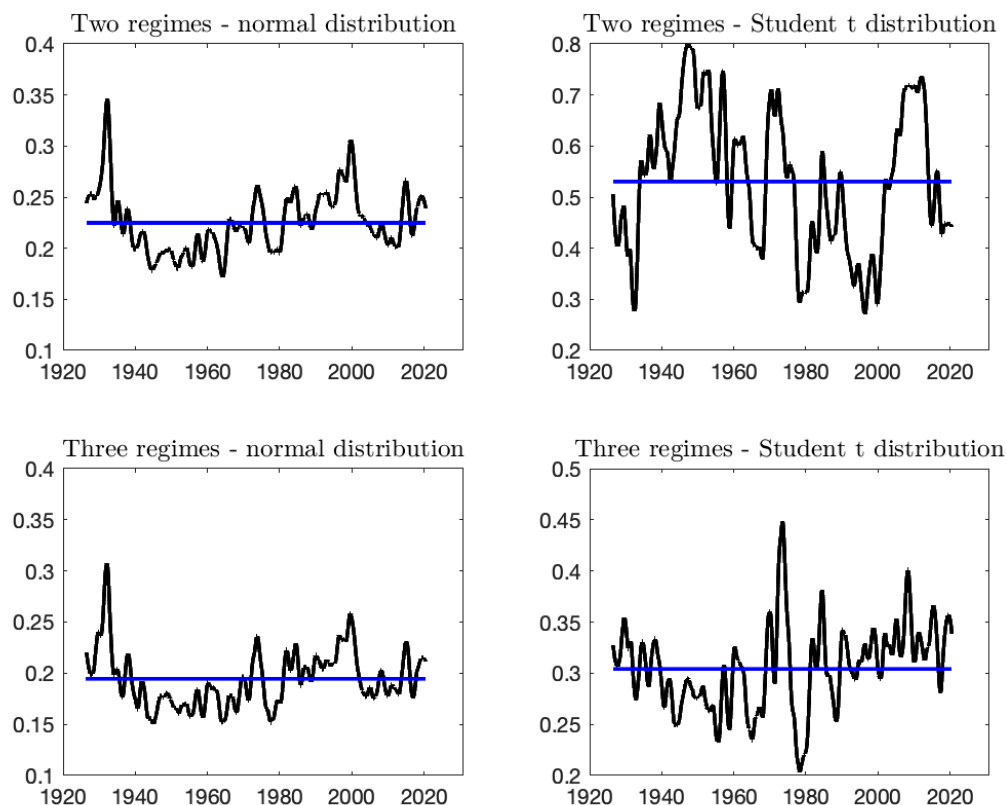


Figure 2. Filtered Probability of Being in the Bear State. The figure displays the filtered probability ϕ_{d+1} of being in the low expected return regime (bear state), for the two-regime and three-regime models. The horizontal blue line corresponds to the stationary probability of being in the bear state $\pi_{b,\infty} = \Pr[S_{d+1} = k_b]$, where k_b denotes the bear state.

Another related interesting feature of MS models is that they can generate asymmetry in the distribution of returns even if the volatility dynamics and the innovation distribution are symmetric. The reason for this property is the possible shift from one regime to another, which generates relatively large (positive or negative) events. To assess whether the various models are able to generate the asymmetry that we observe in the data (see Table 1), we simulate long trajectories of the small caps and large caps daily returns and compute the skewness of the simulated series. As reported in Table 5, the one-regime models do not generate any asymmetry because this property is absent from the model. Two-regime models are able to produce some asymmetry but it is clearly insufficient to match the data. For instance, in the normal distribution case, the skewness measures of the small caps and large caps returns are equal to -0.19 and -0.16 , respectively. This is well above the sample skewness that we obtained over the full sample (-0.39 for small caps and -0.48 for large caps, respectively). In contrast, three-regime models produce skewness measures that are in the ball park of the sample measures. In the normal model, the skewness estimates are equal to -0.8 and -0.5 for small caps and large caps, respectively. In the Student’s t model, the values are equal to -0.5 and -0.3 . Again, this analysis suggests that three regimes are necessary to match the extreme behavior of actual returns.

The table also reports predictions of the large drawdown measures for the three horizons. Again, one-regime models and the two-regime/Student’s t model fail at capturing the magnitude of large drawdown measures obtained with the data, by substantially underestimating drawdowns. The two-regime/normal model is able to capture the magnitude of drawdowns on average but fail at generating the asymmetry observed between small caps and large caps. The drawdowns are usually similar for both size portfolios. In contrast, three-regime models are able to generate this asymmetry, although it is often not as large as with the observed data.

Table 5. Predictions of Moments and Drawdown Measures.

	Sample Data (1926–2020)		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	Small	Large
Skewness	−0.39	−0.48	0.01	0.00	0.00	0.00	−0.19	−0.17	0.17	0.01	−0.84	−0.47	−0.50	−0.26
Kurtosis	23.92	21.98	16.65	10.73	42.01	36.66	13.43	13.32	72.30	28.86	14.70	14.07	14.04	13.88
ADD														
1 quarter	3.84	2.78	2.55	2.32	2.16	2.13	3.76	3.43	2.40	2.21	3.66	3.29	3.42	2.91
2 quarters	5.62	3.85	3.42	3.11	2.68	2.75	5.62	5.36	3.14	2.91	5.34	4.92	5.33	4.74
4 quarters	7.52	4.82	4.53	4.08	3.22	3.46	7.88	7.68	3.86	3.66	7.31	6.85	7.69	7.18
20% CDD														
1 quarter	7.75	5.76	5.41	4.95	4.72	4.62	7.93	7.40	5.20	4.79	7.74	7.11	7.58	6.47
2 quarters	11.80	8.20	7.35	6.70	6.02	6.06	11.65	11.25	6.95	6.39	11.25	10.41	11.78	10.57
4 quarters	16.65	10.96	10.05	9.01	7.71	7.94	16.14	15.80	8.84	8.27	15.41	14.40	16.83	15.68
MDD														
1 quarter	9.87	7.55	7.05	6.44	6.22	6.05	10.20	9.62	6.80	6.25	9.94	9.25	10.04	8.67
2 quarters	15.05	10.93	9.84	8.92	8.27	8.18	14.86	14.41	9.25	8.51	14.38	13.46	15.49	14.13
4 quarters	21.52	15.23	13.96	12.41	11.26	11.23	20.54	20.15	12.17	11.43	19.78	18.63	21.98	20.68
10% CED														
1 quarter	31.22	22.84	12.02	10.63	11.25	10.37	20.59	19.29	12.14	10.64	20.74	19.10	20.85	17.48
2 quarters	41.97	31.35	17.22	15.06	15.87	14.74	27.90	26.86	16.81	14.83	27.82	25.75	29.74	26.97
4 quarters	51.37	40.25	20.59	17.89	17.73	17.16	31.19	30.59	18.40	16.89	30.64	28.55	33.57	31.51

Note: This table reports predictions of some statistics on interest based on simulations of the various models. The statistics the skewness, the kurtosis, the ADD, the 20% CDD, the MDD, and the 10% CED. For large drawdown measures, we consider three investment horizons: one quarter, two quarters, and four quarters. The simulations are based on the parameter estimates obtained over the full sample (1926–2020). Statistics are computed using 1000 draws.

3.3. Rolling-Window Model Estimation and Adequacy Tests

To implement the out-of-sample allocation strategy, we use a rolling window to estimate the parameters of the MS-GARCH models over subsample periods. For each model, the first set of parameters is estimated over the sample of daily returns from January 1927 to December 1989, while the last window covers the sample from January 1958 to December 2020. These parameter estimates will be used to simulate paths of daily log-returns of length H and predict next-period drawdown measures.

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.

We assess the adequacy of out-of-sample estimation of our models with respect to returns for small and large caps by backtesting predicted $CDD_{\theta,t}$ and MDD_t obtained through simulations.¹⁹ To perform these tests, we adopt the approach proposed by Acerbi and Szekely (2014) for testing expected shortfall estimates. We adjust their testing framework for the unconditional coverage of the CDD measure.

The methodology and the main results are reported in Appendix B. In a nutshell, our adequacy tests reveal that one- and two-regime models inaccurately predict drawdown measures for both small and large caps. Specifically, the one-regime and two-regime/Student’s t models underestimate drawdown measures, while the two-regime/normal and three-regime/normal models overestimate realized CDD and MDD for large caps. The one-regime and two-regime/Student’s t models systematically underestimate drawdown measures for both small and large caps. The only model that perform relatively well for both small and large caps is the three-regime/Student’s t model, with the expected number

of exceedances and average drawdown above the threshold being close to the numbers observed in the data.

3.4. Out-of-Sample Analysis

We now consider investors who allocate their wealth in real time period by period, with an investment horizon from one quarter to one year between 1990 and 2020. The out-of-sample strategy is implemented as follows. We use the rolling-window estimation of the models to predict next-period drawdown measures. With these predictions, 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 (ending in September 2020). This out-of-sample analysis corresponds to 124 nonoverlapping quarterly allocations and 31 nonoverlapping annual allocations.

Figure 3 displays the evolution of the optimal small-cap weights obtained by minimizing large drawdown measures based on the various models for a two-quarter horizon. As illustrated, the evolution of optimal weights exhibits notable differences across the approaches used for prediction. For the one-regime models and the two-regime/Student's t model, the optimal small-cap weights are all positive, regardless of the targeted large drawdown measure. This finding implies that a diversified portfolio, with weights close to 50%, could provide effective diversification against large drawdowns. This result indicates that these models do not effectively generate the large drawdowns observed in the small-cap portfolio. Conversely, the two-regime/normal and three-regime/normal models display similar patterns, with optimal small-cap weights close to 0. This suggests that the innovation process plays a critical role in generating sufficiently large drawdowns for small caps, potentially resulting in negative weights.

The three-regime model/Student's t model displays optimal weights that are usually negative. This finding can be explained by the ability of the three-regime/Student's t model to generate large and negative expected returns in a relatively long-lasting Regime 3, which allows large drawdowns in small caps to develop 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 and investment horizons.²⁰

In general, the optimal weights are ordered in the same way for the various investment criteria: the weight of small caps is higher for CDD, then for MDD, and finally for CED. We compare these weights with those resulting from the minimization of the portfolio variance. Minimum variance (MV) portfolio weights are obtained using the same simulation approach. Results indicate that, for all investment horizons, the optimal weight of the MV portfolio is always negative, in the range $[-40\%; -20\%]$ for the two-regime and three-regime models. Figure 3 also demonstrates that the weight of small caps is systematically lower for the MV criterion than for criteria based on large drawdowns.²¹

These patterns suggest that investors targeting large drawdowns tend to be even more cautious in their allocation than MV investors, as they are reluctant to take substantial short positions in small caps.

Table 6 reports results for the out-of-sample allocation when the investor minimizes the expected value of large drawdowns. We compare the performance of the allocation based on the various parametric models. The standard (one-regime) multivariate GARCH models (with normal and Student's t distributions) and the two-regime/Student's t models fail at capturing that the risk of large drawdowns is higher for small caps. Therefore, these models generate a large weight on average for small caps, for all investment horizons. As small caps experienced larger drawdowns than large caps in our sample, strategies based on these models tend to underperform and suffer from much higher ex post drawdowns on average. For instance, for the two-quarter CDD, strategies based on these models have average small caps weights equal to 0.46, 0.63, and 0.39, respectively. Their ex post two-quarter CDD values are equal to 8.47%, 9.16%, and 8.30% on average. In contrast, the two-regime/normal model and the three-regime models (with normal and Student's t

distributions) tend to have negative small cap weights on average, with substantially lower ex post two-quarter CDD values, equal to 7.23%, 7.26%, and 7.15%, respectively.

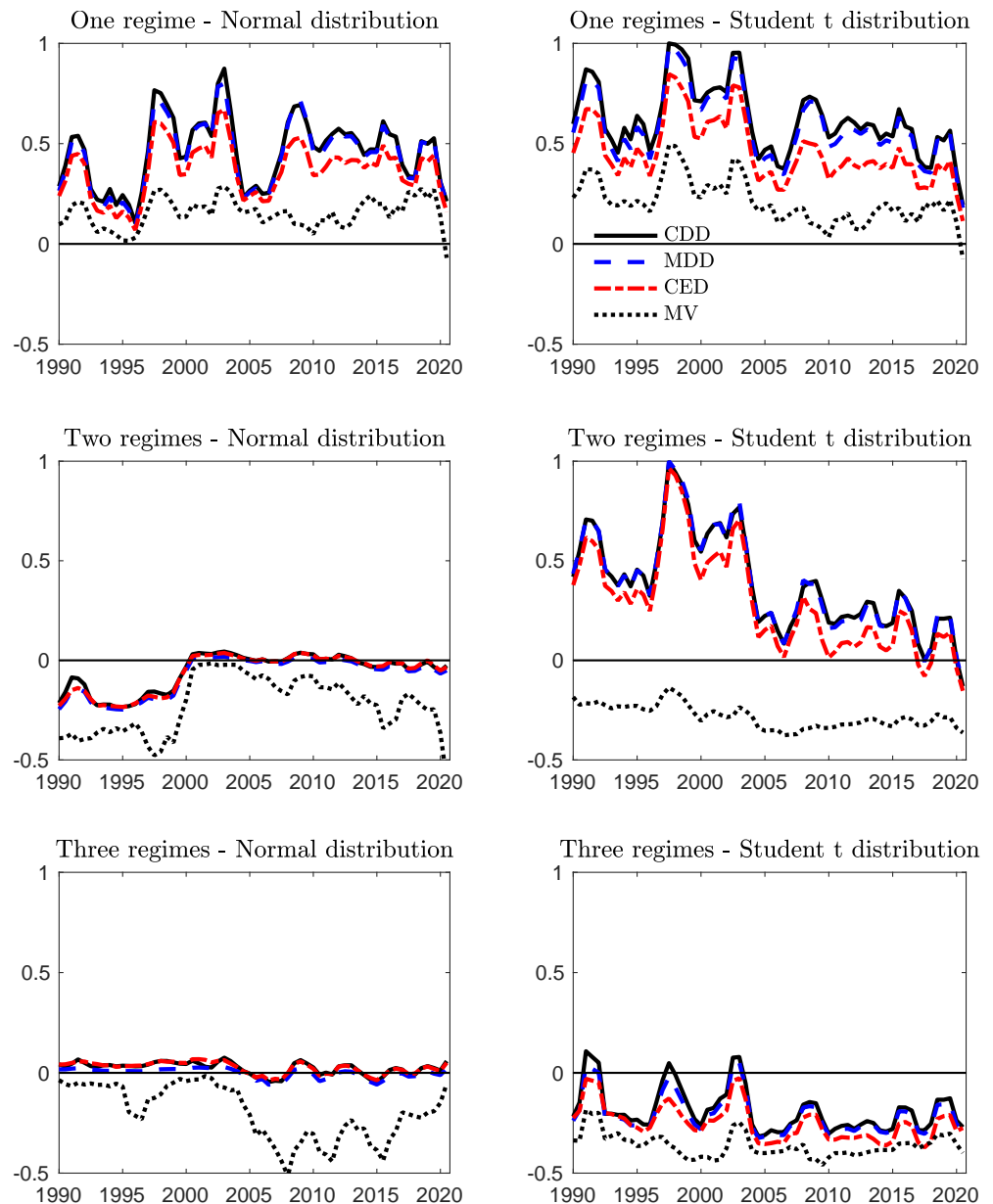


Figure 3. Out-of-Sample Optimal Weights—Two-quarter Horizon (1990–2020). The figure displays the temporal evolution of the optimal weight of small caps for the 20% CDD, MDD, 10% CED, and MV portfolios over the two-quarter horizon, when predictions are based on the one-regime, two-regime, and three-regime models.

It is worth emphasizing that, in all cases, the strategies based on the standard multivariate GARCH models result in much larger drawdown measures than strategies based on three-regime models. The gap is economically substantial for the three measures CDD, MDD, and CED. The ex post four-quarter CDD of the one-regime/normal model is 2.1 points higher than the same measure of the three-regime/Student’s t model. The difference is equal to 2.3 points for the four-quarter MDD and 1.7 points for the four-quarter CED. The gaps are less severe for the two-regime/normal model but they remain large for long horizons. They are equal to 2.4 points for the four-quarter CDD, 0.6 point for the four-quarter MDD, 0.4 point for the four-quarter CED.

Overall, the three-regime/normal 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.

Table 6. Out-of-Sample Allocation: Minimization of Large Drawdowns.

Horizon	Statistics	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 CDD							
1 quarter	Weight	0.44	0.54	−0.11	0.41	−0.17	0.07
	20% CDD	5.75	5.90	5.24	5.66	5.23	5.27
2 quarters	Weight	0.46	0.63	−0.06	0.39	0.02	−0.18
	20% CDD	8.47	9.16	7.23	8.30	7.26	7.15
4 quarters	Weight	0.51	0.67	0.42	0.47	0.87	−0.07
	20% CDD	11.36	12.66	11.62	11.34	14.28	9.22
Panel B: Minimization of MDD							
1 quarter	Weight	0.43	0.53	−0.12	0.39	−0.17	0.05
	MDD	7.56	7.71	7.04	7.40	7.10	7.02
2 quarters	Weight	0.44	0.59	−0.08	0.39	0.00	−0.21
	MDD	11.31	11.95	10.10	11.14	10.06	9.92
4 quarters	Weight	0.48	0.64	0.09	0.49	0.18	−0.10
	MDD	16.08	17.15	14.42	16.16	14.80	13.82
Panel C: Minimization of CED							
1 quarter	Weight	0.37	0.44	−0.16	0.26	−0.21	−0.14
	10% CED	22.25	22.53	20.69	21.43	20.30	20.53
2 quarters	Weight	0.37	0.47	−0.06	0.30	0.03	−0.25
	10% CED	29.24	30.03	25.26	28.80	25.56	24.84
4 quarters	Weight	0.41	0.52	0.02	0.43	0.00	−0.06
	10% CED	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 20% CDD, the MDD, and 10% CED (Panels A to C, 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. Numbers in bold correspond to the smallest values of the large drawdown measures.

We also consider an investor maximizing the expected return—large drawdown criterion (Equation (6)), which also accounts for the ability of the models to predict expected returns. Results are reported in Table 7 with a degree of aversion for large drawdowns equal to $\lambda = 5$. Optimal weights for small caps are usually higher, reflecting the higher expected return of small caps. For the one-regime models and the two-regime/Student’s t model, small caps weights are even often larger than one, indicating the large caps are shorted in these allocations. Turning to the two-regime/normal model and three-regime models, we find that weights are also higher than in the case of the minimization of large drawdowns but to a lesser extent. For the MDD and CED targets, small caps weights are negative for a short horizon but close to 50% for two-quarter and four-quarter horizons. These results indicate that investors also targeting the expected return would invest more in small caps because of their higher return on average.

Table 7. Out-of-Sample Allocation: Maximization of Expected Return—Large Drawdown Criterion.

Horizon	Statistics	1-Regime Normal	1-Regime Student t	2-Regime Normal	2-Regime Student t	3-Regime Normal	3-Regime Student t
Panel A: Expected Return—CDD criterion							
1 quarter	Weight	0.47	0.59	0.12	1.10	0.17	0.62
	Criterion	−0.020	−0.059	0.015	−0.037	0.029	−0.004
	Opp. cost (%)	4.87	8.76	1.43	6.62	0.00	3.32
2 quarters	Weight	0.48	0.64	0.56	1.40	0.62	0.29
	Criterion	−0.070	−0.161	−0.026	−0.119	−0.031	−0.004
	Opp. cost (%)	6.63	15.70	2.23	11.53	2.73	0.00
4 quarters	Weight	0.54	0.73	0.48	1.83	0.50	0.65
	Criterion	−0.175	−0.323	−0.055	−0.293	−0.058	−0.075
	Opp. cost (%)	11.66	26.43	−0.30	23.50	0.00	1.68
Panel B: Expected Return—MDD criterion							
1 quarter	Weight	0.62	0.85	−0.02	0.85	−0.03	0.31
	Criterion	−0.091	−0.111	−0.076	−0.113	−0.067	−0.074
	Opp. cost (%)	2.36	4.40	0.94	4.57	0.00	0.67
2 quarters	Weight	0.68	1.01	0.43	1.01	0.60	−0.01
	Criterion	−0.203	−0.258	−0.181	−0.254	−0.203	−0.147
	Opp. cost (%)	5.57	11.07	3.42	10.71	5.64	0.00
4 quarters	Weight	0.78	1.28	0.49	1.42	0.51	0.50
	Criterion	−0.319	−0.442	−0.281	−0.478	−0.284	−0.277
	Opp. cost (%)	4.23	16.49	0.45	20.14	0.70	0.00
Panel C: Expected Return—CED criterion							
1 quarter	Weight	0.89	1.33	−0.11	0.50	−0.15	−0.03
	Criterion	−0.463	−0.481	−0.424	−0.460	−0.406	−0.409
	Opp. cost (%)	5.66	7.48	1.78	5.41	0.00	0.28
2 quarters	Weight	1.05	1.73	0.45	0.57	0.69	−0.16
	Criterion	−0.648	−0.684	−0.616	−0.665	−0.681	−0.526
	Opp. cost (%)	12.21	15.87	9.03	13.95	15.49	0.00
4 quarters	Weight	1.28	1.97	0.50	0.78	0.51	0.48
	Criterion	−0.788	−0.850	−0.762	−0.889	−0.762	−0.789
	Opp. cost (%)	2.59	8.86	0.03	12.73	0.00	2.78

Note: This table reports the optimal weight of small caps (α^*), the value of the objective function at the optimum, and the opportunity cost ($\zeta^{(sub)}$) relative to the optimal model, when the investor maximizes the expected return—large drawdown criterion, with 20% CDD, the MDD, and 10% CED (Panels A to C, respectively). The degree of aversion for large drawdowns is equal to $\lambda = 5$. 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.

Regarding the performance of the strategies, we obtain essentially the same conclusions as before. One-regime models are always dominated by models with at least two regimes. In addition, three-regime models usually outperform other models, in particular

for extreme drawdowns (MDD and CED). To illustrate the gain of using a three-regime model to predict large drawdowns, we compute the opportunity cost of using a one-regime or two-regime model, using the relation:

$$E_t[R_{p,t+1}(\alpha_t^{(sub)}) - \zeta^{(sub)}] - \frac{\lambda}{2} E_t[XDD_{t+1}(\alpha_t^{(sub)})] = E_t[R_{p,t+1}(\alpha_t^{(opt)})] - \frac{\lambda}{2} E_t[XDD_{t+1}(\alpha_t^{(opt)})],$$

where $\alpha_t^{(sub)}$ and $\alpha_t^{(opt)}$ denote the vector of weights obtained using the suboptimal model and the optimal model, respectively. The opportunity cost $\zeta^{(sub)}$ is defined as the fraction of the expected return that an investor using the suboptimal strategy is ready to pay to get access to the optimal strategy.

The estimates of the opportunity cost are also reported in the table. For the one-quarter investment horizon, the best strategy is the one based on the three-regime/normal model. In this case, an investor using the two-regime/normal model would be ready to pay 1% if the objective function is targeting the MDD and 1.8% if the objective function is targeting the CED. For the two-quarter investment horizon, the best strategy is the one based on the three-regime/Student's t model. In this case, an investor using the two-regime/normal model would be ready to pay 3.4% if the objective function is targeting the MDD and up to 9% if the objective function is targeting the CED. As these estimates illustrate, the gain of using a three-regime model is economically substantial for investors targeting large drawdown measures.

This out-of-sample analysis demonstrates that three regimes are necessary in the multivariate MS-GARCH model to produce large and negative weights for small caps. Student's t innovations also help generate a longer-lasting bear state and therefore to produce more drawdowns for small caps. Importantly, as the allocation exercise is performed in a fully out-of-sample fashion, we account for the risk of over-parameterization of the three-regime models.

4. Conclusions

This paper addresses the modeling and prediction of large drawdowns in financial markets. Given that the standard GARCH model struggles to capture prolonged market declines, we explore multivariate MS-GARCH models, which can generate regimes characterized by both positive and negative market trends. Provided that the probability of remaining in a bear regime is sufficiently high, these models are capable of reproducing significant drawdown properties. Specifically, our findings show that in three-regime models, one regime is marked by substantial negative expected returns (representing a bear market) and the capacity to generate large windfalls.

In the out-of-sample investment analysis, as changes in the distribution of drawdowns are updated, three-regime models prove to be superior tools for predicting expected drawdowns by imposing restrictions that align well with the observed data. These models consistently outperform other parametric models featuring one regime (such as standard multivariate GARCH models) or two regimes.

To assess the predictive power of MS-GARCH models for large drawdowns, we intentionally keep our model specification straightforward within a well-established framework, ensuring robust results. Notably, we do not attempt to identify factors influencing the dynamics of state probabilities, which may be affected by variables related to government or monetary policies, geopolitical issues, or climate change, among others.²² Exploring these factors represents an important avenue for future research.

Author Contributions: Conceptualization, E.J. and A.P.; methodology, E.J. and A.P.; validation, E.J. and A.P.; writing—original draft preparation, E.J. and A.P.; writing—review and editing, E.J. and A.P.; supervision, E.J. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: All data used in this paper are from Kenneth French’s website: https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html, accessed on 2 December 2024.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Evolution of Model Parameters

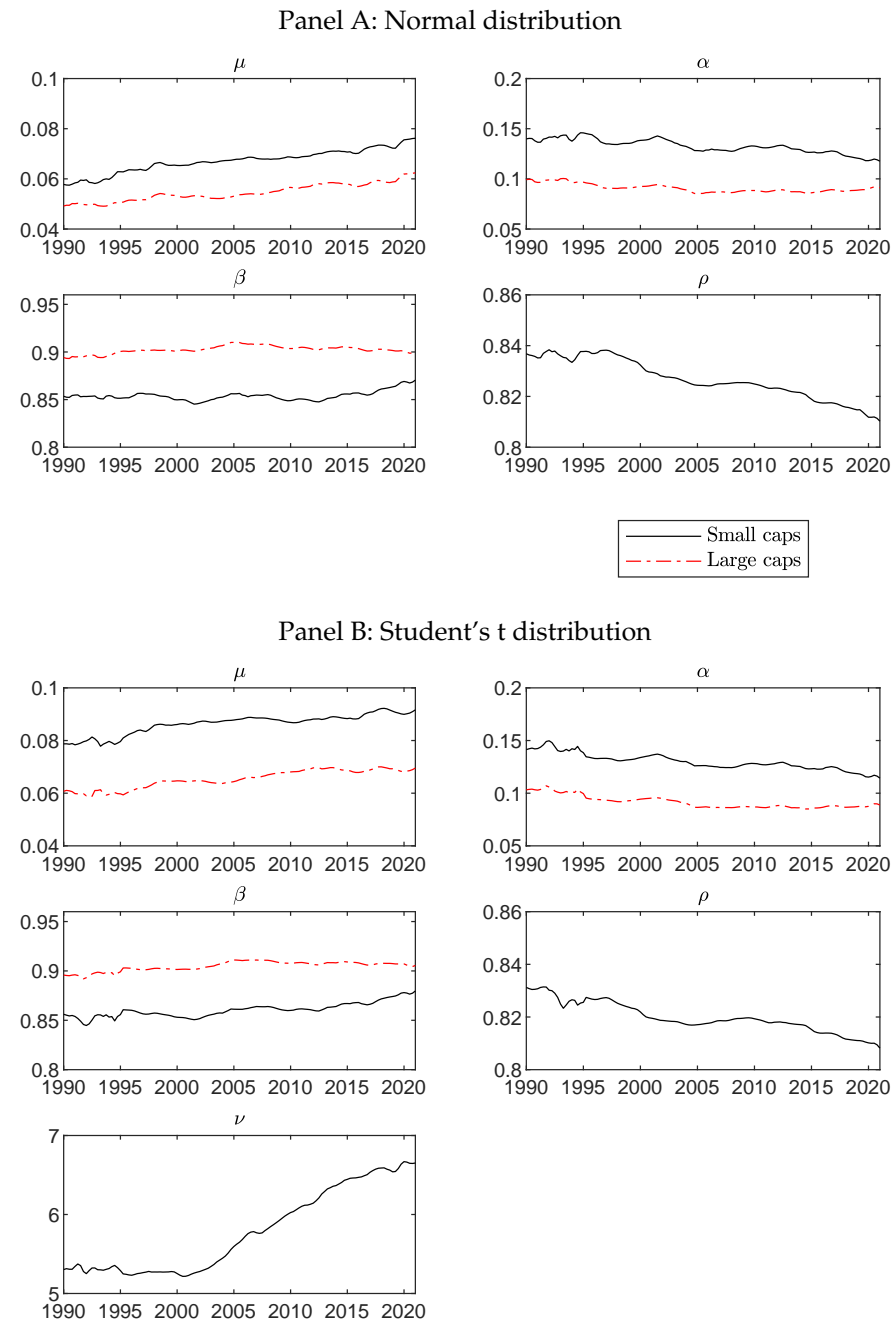


Figure A1. Model Parameters: One-regime Models.

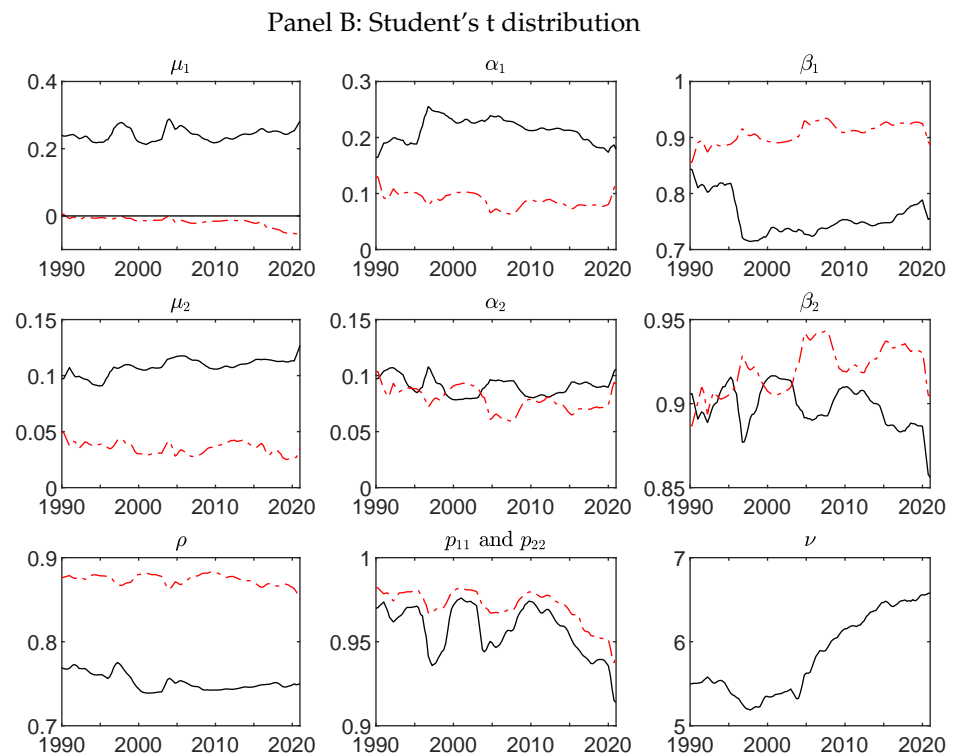
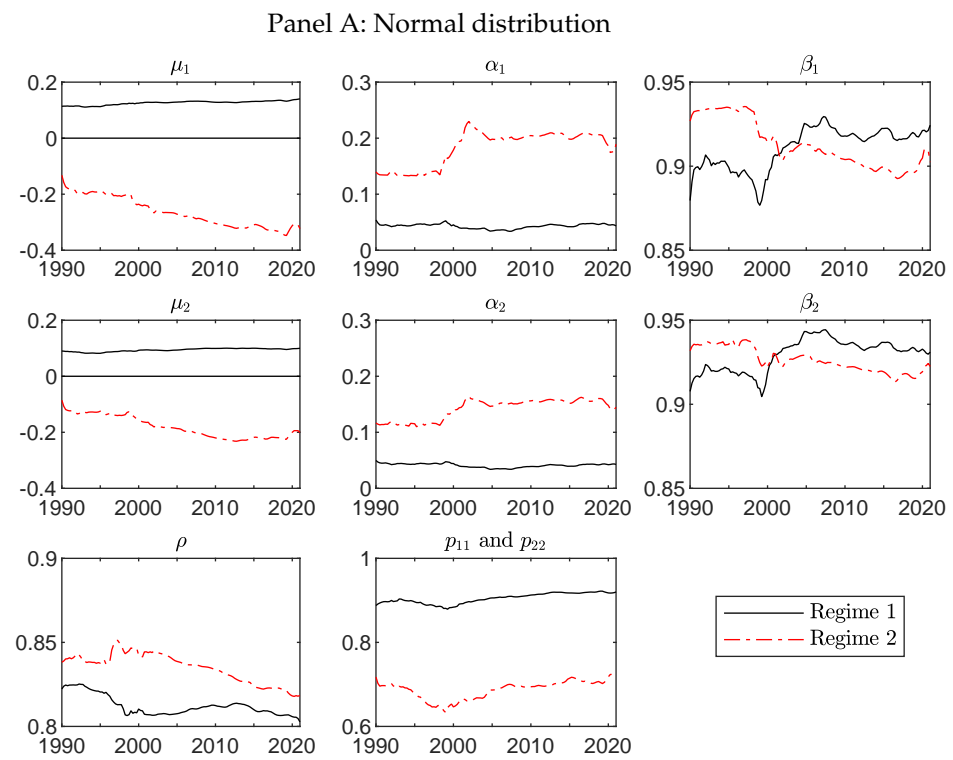


Figure A2. Model Parameters: Two-regime Models.

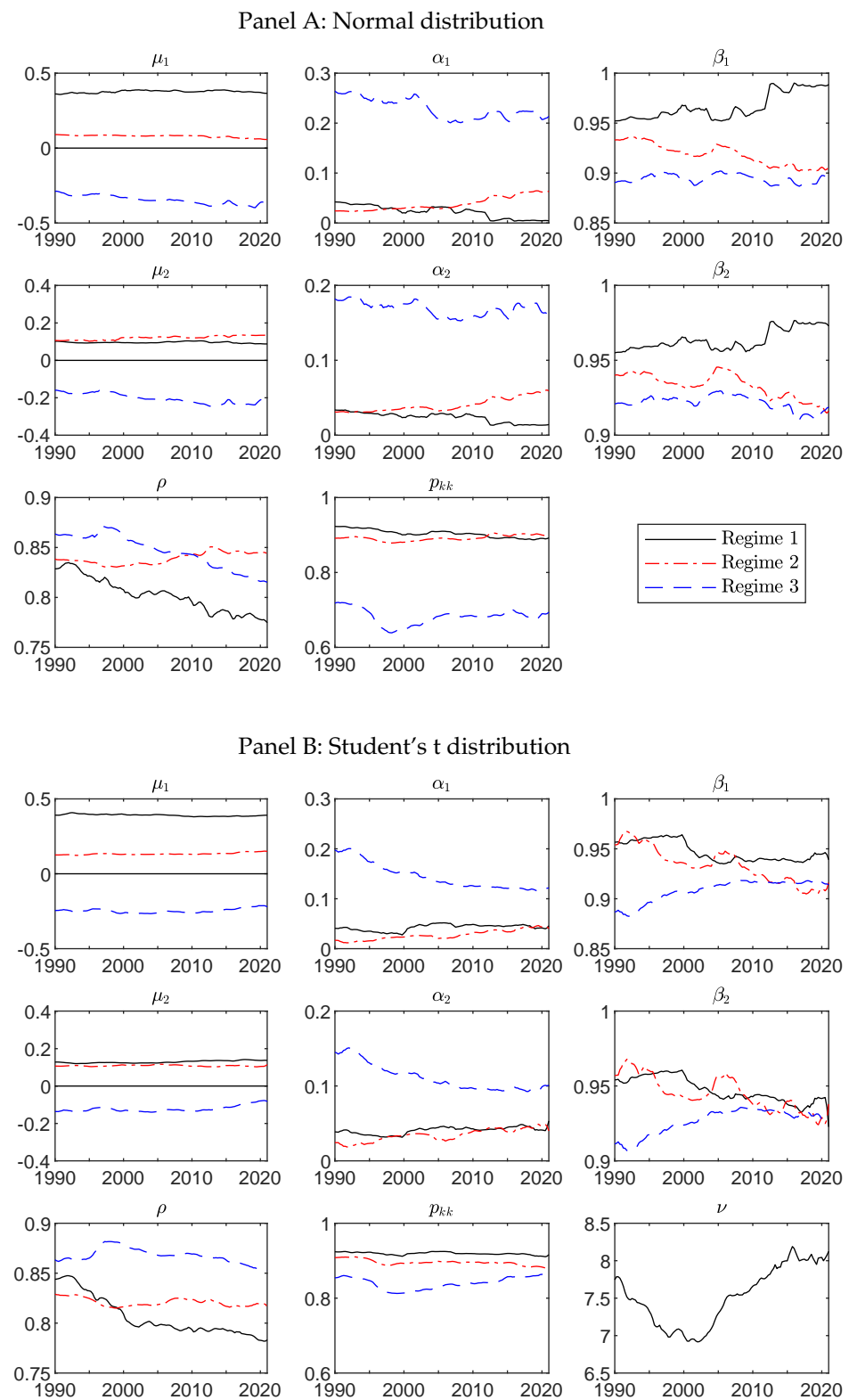


Figure A3. Model Parameters: Three-regime Models.

Appendix B. Adequacy Tests

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 assess whether the model underlying a particular prediction is adequate. To backtest our predicted risk measures, we rely on the second test by [Acerbi and Szekely \(2014\)](#). Note that [Acerbi and Szekely \(2017\)](#) propose a more robust approach for backtesting a risk measure that depends on another statistic, such as ES or CDD, which depend on a threshold. The proposed ridge backtest accounts for the sensitivity of the statistic to the threshold estimation by penalizing errors in estimation. However, this test is only relevant for validating a model from a prudential perspective, that is, testing that the model does not underestimate the risk measure. Therefore, this procedure is not suitable for testing the CDD. As we want to assess if a model tends to overestimate or underestimate the CDD, we do not pursue with this test. The backtest for $M\hat{D}D_t$ does not suffer from the sensitivity to the threshold.

Appendix B.1. Logic of the Adequacy Test

The logic of the test is the following: For a given risk measure, we define a Z statistic that is equal to zero under the null hypothesis that the model correctly predicts this risk measure:

$$E_{H_0}[Z_{XDD}(X\hat{D}D, DD)] = 0,$$

where DD is a random variable representing drawdowns. Then, from the sequence of risk measure predictions and drawdowns realizations, we compute the realized Z statistic over the out-of-sample period:

$$z_{XDD} = \frac{1}{T} \sum_{t=1}^T Z_{XDD}(X\hat{D}D_t, DD_t),$$

where DD_t is the vector of observed drawdowns defined in Section 2.2.

Finally, we test the adequacy of a given model by comparing the realized Z statistic obtained from this model to the distribution of the Z statistics under the null hypothesis. This distribution is obtained by computing Z statistics from simulations of the model over the out-of-sample period.²³

Appendix B.2. 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 $E_t[Th_{\theta,t}] = (1 - \theta)H$. The test statistic is defined as (see Equation (3)):

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

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

The hypotheses for this test are as follows:

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

$$H_1 : |CDD_{\theta,t} - C\hat{D}D_{\theta,t}| > 0 \text{ for some } t.$$

Under the null hypothesis, we have $E_{H_0}[Z_{CDD}(C\hat{D}D, DD)] = 0$. Under the alternative hypothesis, if the model underestimates the CDD, we have $E_{H_1}[Z_{CDD}(C\hat{D}D, DD)] > 0$, and if the model overestimates the CDD, we have $E_{H_1}[Z_{CDD}(C\hat{D}D, DD)] < 0$. The test

actually jointly evaluates the frequency and the magnitude of the tail events because the test statistics do not impose that $\sum_{h=1}^H I_{(DD_t > \hat{T}h_{\theta,t})} = (1 - \theta)H$, 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 MDD, we simply define the Z statistic as:

$$z_{MDD} = \frac{1}{T} \sum_{t=1}^T \frac{MDD_t}{M\hat{D}D_t} - 1,$$

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

The hypotheses of this test are:

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

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

It follows that $E_{H_0}[Z_{MDD}(M\hat{D}D, DD)] = 0$ under the null hypothesis that the model correctly describes the realized MDD. When the model underestimates (overestimates) the risk measure, the Z statistic takes a positive (negative) value, i.e., $E_{H_1}[Z_{MDD}(M\hat{D}D, DD)] \gtrless 0$.

Appendix B.3. Adequacy Test Significance

To find the test significance, we generate Z statistics under the null hypothesis by simulating the model for which we test the adequacy. Then, by comparing the realized Z statistic to the simulated Z statistic under H_0 , we evaluate the p-value, i.e., the probability to obtain 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: $\{r_t^{(q)}\}_{h=1}^H, t = 1, \dots, T$ and $q = 1, \dots, Q$. The simulations are generated as in Section 2.4.2.
2. We calculate the vectors of drawdowns for each simulation, $DD_t^{(q)}, t = 1, \dots, T$. Then, we compute Z statistic for each simulation, $z_{XDD}^{(q)} = z_{XDD}(X\hat{D}D, DD^{(q)}), q = 1, \dots, Q$. They represent the distribution of the Z statistic under the null hypothesis.
3. We compute the realized Z statistic on the drawdowns $z_{XDD}^{(q)} = z_{XDD}(X\hat{D}D, DD)$.
4. We estimate the significance by comparing the realized Z statistic to the distribution of the simulated Z statistics and compute the p-value of the bilateral test as: $p\text{-val} = 1 - \frac{1}{Q} \sum_{q=1}^Q (|z_{XDD}^{(q)}| < |z_{XDD}|)$.

Appendix B.4. Adequacy Test Results

Results of the adequacy tests are reported in Table A1. The statistics are based on 50,000 simulations of the estimated models. The table indicates that most models fail to capture the time-series properties of large drawdown statistics. The one-regime models and the two-regime/Student's t model systematically underestimate the drawdown measures for both small and large caps. For CDD, the expected number of exceedances is equal to 1562 (i.e., 20% of the total number of daily observations in the out-of-sample period). The one-quarter number of exceedances for the two-regimes/Student's t model is equal to 2666 and 1961 for small caps and large caps, respectively, meaning that the estimated threshold is too low. As a consequence, the estimated CDD is too small, resulting in high values of the test statistic $Z_{CDD}(DD)$. For all three models and all three investment horizons, p-values for the CDD and MDD tests are below 1.5%. The reason for this failure is that these models do not allow for a negative trend in expected returns, and therefore, they cannot reproduce long-lasting market declines.

The two-regime/normal model and the three-regime/normal model provide a good description of the 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 CDD and MDD for large caps. For large caps, the one-quarter number of exceedances is close to 1000 for both models, while the expected number is equal to 1562.

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 1800 and 1412 for the small caps and large caps, respectively, while the expected number is 1562. 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 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.

Table A1. 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: 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
<i>p</i> -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
<i>p</i> -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
<i>p</i> -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 B: 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
<i>p</i> -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
<i>p</i> -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
<i>p</i> -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 20% CDD and MDD. For CDD, the table also reports the number of exceedances, with an expected number equal to 1562. 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.

Notes

- 1 Disasters may include severe macroeconomic and financial crises, pandemics, wars, or extreme weather and climate conditions.
- 2 We use a large sample to estimate model’s parameters accurately and perform the out-of-sample analysis over a long period of time. Such a long sample would not be necessary in practice to estimate MS models and to predict subsequent large drawdowns.
- 3 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.
- 4 A drawdown may span over a short period (as for the COVID-19 crisis, with a 36% drawdown in 24 days) or over a window of more than a year (as for the subprime crisis, with a 60% drawdown in 355 days).

- 5 We note that the knowledge of the cumulated log-return at the end of the period is not sufficient to infer the large drawdown measures, as peaks and troughs are likely to occur on random days within the period.
- 6 The model is estimated over a long sample of daily returns, $(\tilde{r}_1, \dots, \tilde{r}_D)$, 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 $r_t = (r_{t,1}, \dots, r_{t,H})$, $t = 1, \dots, T$ in Section 2. Since we use nonoverlapping subsamples, both notations define the same sample: $(\tilde{r}_1, \dots, \tilde{r}_D) = (r_1, \dots, r_T)$, where $D = H \times T$.
- 7 Assuming an autoregressive process would have a very limited effect for a long-term investment objective because the autocorrelation of daily 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.
- 8 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).
- 9 In Section 3.2, we use simulations to demonstrate that a three-regime model with a (symmetric) Student's t innovations can generate some asymmetry in large drawdowns, as observed in the data.
- 10 We note that stationarity conditions apply to the complete distribution and not regime by regime. As a consequence, usual stationarity conditions in a GARCH model might not be satisfied for some regimes. In particular, global stationarity can be obtained even when $\alpha_i^{(k)} + \beta_i^{(k)} \geq 1$ for some asset i and regime k .
- 11 We have also analyzed the cases of firms in the bottom and top 20% and 10% market capitalization, with limited impact on the main results. The data are available on the website of Kenneth French at (https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html, accessed on 2 December 2024). Other long-sample portfolios, such as value versus growth or losers versus winners portfolios, are also available.
- 12 Our sample ends in December 2020, slightly after the end of the COVID-19 market crash. Therefore, the impact of this episode on the financial performance of our investment strategies is reflected in the out-of-sample analysis discussed in Section 3.4.
- 13 Stationary (or steady-state) probabilities are defined as: $\pi_\infty = P \pi_\infty$. In the two-regime case, this relation boils down to $\pi_{\infty,2} = \Pr[S_t = 2] = (1 - p_{11}) / (1 - p_{11} - p_{22})$.
- 14 This conclusion appears very robust and not driven by the choice of starting values. We experimented with several sets of starting parameter values and obtained the same parameter estimates for both models. A similar phenomenon—where the probability of remaining in the bear state is lower with normal innovations than with Student's t innovations—was reported by Haas and Paolella (2012) and Haas and Liu (2018).
- 15 The finite-sample distribution of the LR statistic is obtained by simulating many samples of returns, using the estimated parameters of the $(n - 1)$ -regime model and estimating, for each simulated samples, the $(n - 1)$ -regime model and the n -regime model, from which we compute the LR test statistics. The finite-sample distribution of the LR test statistic is computed from the empirical distribution of the LR statistics based on the simulated samples.
- 16 Guidolin and Timmermann (2007) estimate MS model for U.S. small caps, large caps, and long-term bonds. In a specification with within-regime constant expected returns, volatilities, and correlations, they find that four regimes are necessary to match the data. In their model, the intermediate regime is further decomposed into a slow growth regime and a recovery regime.
- 17 The largest spike in the filtered probability of being in the bear state is associated with the 1973 oil crisis. The drawdown actually started in January 1973, accelerated in October with the surge in oil price and lasted until December 1974, with a drawdown of 48% over this period. A major feature of this 1973–1974 drawdown is its duration. From peak to trough, the downturn lasted for almost 2 years, while the subprime crisis was associated with a 60% drawdown in slightly more than 1 year.
- 18 As in the models with normal innovations, one of the peaks in the filtered probability, in mid-1984, could not be associated with any particular stock market event.
- 19 Although we developed a similar test for CED, the number of observations was insufficient for robust conclusions.
- 20 We note that, even with our preferred model, i.e., the three-regime model with Student's t distribution, the relationship between the probability of being in the bear regime (Figure 2) and the optimal small-cap weight (Figure 3) is far from perfect. The reason is that the probability of being in the bear regime predicts the state for the next day, whereas the portfolio is allocated for a long horizon (from 1 quarter to 1 year) and the regime is likely to change over this period.
- 21 This result of a lower small-cap weights for the MV criterion does not seem to be driven by a theoretical relation between the variance and large drawdown measures but more likely by the properties of our data.
- 22 For instance, Gray (1996) allows state probabilities to depend on lagged interest rates in an MS-GARCH model for short-term interest rates.
- 23 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, as we test whether a given model correctly predicts large drawdown measures.

References

- Abramson, Ari, and Israel Cohen. 2007. On the stationarity of Markov-switching GARCH processes. *Econometric Theory* 23: 485–500. [CrossRef]
- Acerbi, Carlo, and Balazs Szekely. 2014. Backtesting Expected Shortfall. *Risk* 27: 76–81.
- Acerbi, Carlo, and Balazs Szekely. 2017. *General Properties of Backtestable Statistics*. Working Paper. New York: MSCI Inc. Available online: <https://ssrn.com/abstract=2905109> (accessed on 25 February 2023).
- Ang, Andrew, and Geert Bekaert. 2002. International asset allocation with regime shifts. *Review of Financial Studies* 15: 1137–87. [CrossRef]
- Ang, Andrew, and Joseph Chen. 2002. Asymmetric correlations of equity portfolios. *Journal of Financial Economics* 63: 443–94. [CrossRef]
- Barnett, Michael, William Brock, and Lars Peter Hansen. 2020. Pricing uncertainty induced by climate change. *Review of Financial Studies* 33: 1024–66. [CrossRef]
- Chekhlov, Alexei, Stanislav Uryasev, and Michael Zabaranki. 2003. Portfolio optimization with drawdown constraints. In *Asset and Liability Management Tools*. Edited by Berndt Scherer. London: Risk Books, pp. 263–78.
- Chekhlov, Alexei, Stanislav Uryasev, and Michael Zabaranki. 2005. Drawdown measure in portfolio optimization. *International Journal of Theoretical and Applied Finance* 8: 13–58. [CrossRef]
- Fama, Eugene F., and Kenneth R. French. 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 33: 3–56. [CrossRef]
- Garcia, Rene. 1998. Asymptotic null distribution of the likelihood ratio test in Markov switching models. *International Economic Review* 39: 763–88. [CrossRef]
- Goldberg, Lisa R., and Ola Mahmoud. 2015. *On a Convex Measure of Drawdown Risk*. Working Paper. Berkeley: Center for Risk Management Research. Available online: <http://arxiv.org/abs/1404.7493> (accessed on 25 March 2021).
- Goldberg, Lisa R., and Ola Mahmoud. 2016. Drawdown: From practice to theory and back again. *Mathematics and Financial Economics* 11: 275–97. [CrossRef]
- Gray, Stephen F. 1996. Modeling the conditional distribution of interest rates as a regime-switching process. *Journal of Financial Economics* 42: 27–62. [CrossRef]
- Grossman, Sanford, and Zhongquan Zhou. 1993. Optimal investment strategies for controlling draw-downs. *Mathematical Finance* 3: 241–276. [CrossRef]
- Guidolin, Massimo, and Allan Timmermann. 2004. *Value at Risk and Expected Shortfall Under Regime Switching*. Working Paper. San Diego: University of California. Available online: <https://ssrn.com/abstract=557091> (accessed on 1 May 2021).
- Guidolin, Massimo, and Allan Timmermann. 2007. Asset allocation under multivariate regime switching. *Journal of Economic Dynamics and Control* 31: 3503–44. [CrossRef]
- Guidolin, Massimo, and Allan Timmermann. 2008. International asset allocation under regime switching, skew, and kurtosis preferences. *Review of Financial Studies* 21: 889–935. [CrossRef]
- Haas, Markus, and Ji-Chun Liu. 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. [CrossRef]
- Haas, Markus, Stefan Mittnik, and Marc S. Paolella. 2004. A new approach to Markov-switching GARCH models. *Journal of Financial Econometrics* 2: 493–530. [CrossRef]
- Haas, Markus, and Marc S. Paolella. 2012. Mixture and regime-switching GARCH models. In *Handbook of Volatility Models and Their Applications*; Edited by Luc Bauwens, Christian M. Hafner, and Sebastien Laurent. Hoboken: John Wiley & Sons, pp. 71–102.
- Hamilton, James D. 1989. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica* 57: 357–84. [CrossRef]
- Hansen, Bruce E. 1992. The likelihood ratio test under non-standard conditions: Testing the Markov switching model of GNP. *Journal of Applied Econometrics* 7: S61–S82. [CrossRef]
- Hansen, Bruce E. 1996. Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica* 64: 413–30. [CrossRef]
- Hansen, Peter Reinhard. 2009. *In-Sample Fit and Out-of-Sample Fit: Their Joint Distribution and Its Implications for Model Selection*. Working Paper. Stanford: Department of Economics, Stanford University.
- Huang, Wei, Qianqiu Liu, S. Ghon Rhee, and Feng Wu. 2012. Extreme downside risk and expected stock returns. *Journal of Banking and Finance* 36: 1492–502. [CrossRef]
- Karydas, Christos, and Anastasios Xepapadeas. 2019. *Climate Change Financial Risks: Pricing and Portfolio Allocation*. Economics Working Paper Series, No. 19/327, 327. Zurich: Center of Economic Research (CER-ETH). [CrossRef]
- Klaassen, Franc. 2002. Improving GARCH volatility forecasts with regime-switching GARCH. *Empirical Economics* 27: 363–94. [CrossRef]
- Pagano, Marco, Christian Wagner, and Josef Zechner. 2023. Disaster resilience and asset prices. *Journal of Financial Economics* 150: 103712. [CrossRef]
- Patton, Andrew. 2004. On the out-of-sample importance of skewness and asymmetric dependence for asset allocation. *Journal of Financial Econometrics* 2: 130–68. [CrossRef]
- Pelletier, Denis. 2006. Regime switching for dynamic correlations. *Journal of Econometrics* 131: 445–73. [CrossRef]

- Peng, Cheng, Young Shin Kim, and Stefan Mittnik. 2022. Portfolio optimization on multivariate regime switching GARCH model with normal tempered stable innovation. *Journal of Risk and Financial Management* 15: 230. [[CrossRef](#)]
- Perez-Quiros, Gabriel, and Allan Timmermann. 2000. Firm size and cyclical variations in stock returns. *Journal of Finance* 55: 1229–62. [[CrossRef](#)]
- Reveiz, Alejandro, and Carlos Leon. 2008. Efficient portfolio optimization in the wealth creation and maximum drawdown space. In *Interest Rate Models, Asset Allocation and Quantitative Techniques for Central Banks and Sovereign Wealth Funds*. Edited by Arjan B. Berkelaar, Joachim Coche, and Ken Nyholm. London: Palgrave Macmillan, pp. 134–157.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.