Don't draw the downs apart – How to best simulate asset price drawdowns

Hubert Dichtl[1], Wolfgang Drobetz[1], Tizian Otto[2], and Tatjana Puhan[3,4]

This Version: June 2025

Abstract

This paper evaluates bootstrap simulation techniques for calculating the distribution of the maximum drawdown (MDD), an important risk indicator in stock and cryptocurrency markets. Using stochastic dominance tests, we assess the full distributional properties of MDD under different methods. Our findings reveal that the standard Efron (1979) bootstrap, which assumes independence and identically distributed random variables, systematically underestimates the true MDD. While the moving block bootstrap provides reasonable estimates, it is subject to non-stationarity bias, particularly when large drawdowns occur at the boundaries of a return series. Alternative procedures, such as the block-block bootstrap and the tapered bootstrap, do not lead to better results. Of all the methods studied, the stationary bootstrap of Politis and Romano (1994) produces the most accurate and robust results, particularly with longer block lengths. We recommend this method as the preferred choice for researchers and practitioners modelling drawdown risk.

JEL Classification: G11, G12, G14

Keywords: Bootstrap simulation, maximum drawdown, stock markets, crypto markets, serial dependence.

[1] Faculty of Business Administration, University of Hamburg, 20148 Hamburg, Germany.

[2] Bloomberg L.P., London EC4N 4TQ, United Kingdom.

[3] Swiss Re Management Ltd, 8002 Zurich, Switzerland.

[4] Faculty of Business Administration, University of Mannheim, 68161 Mannheim, Germany.

1. Introduction

In academia and the asset management industry, it is well known that the standard deviation of asset returns (i.e., the return volatility) does not accurately measure investors' true risk sensitivity (Sortino and Van der Meer, 1991). Goetzmann et al.'s (2017) survey results confirm that investors are pronounced fearful of sudden and large market drawdowns. Although this fear is not justified from a longer-term perspective (Goetzman and Kim, 2018), many investors are constrained by solvency and/or regulation considerations such that they cannot endure major losses even for short periods of time. For example, pension funds face large liabilities toward their beneficiaries and the failure of their assets to meet those liabilities carries significant penalties (Ang, Chen, and Sundaresan, 2013). Therefore, such investors have limited loss budgets and face downside risk constraints. The evidence also indicates that mutual fund managers and their shareholders consider downside risk in their investment decisions (Artavanis, Eksi, and Kadlec, 2019; Bodnaruk, Chokaev, and Simonov, 2019). Finally, while standard mean–variance investors would be more focused on volatility than downside risks, central assumptions in this framework are violated. For example, although the mean–variance framework relies on the assumption that asset returns are jointly normally distributed, the literature acknowledges that the distribution of stock returns exhibits skewness (Mandelbrot, 1961; Fama, 1963; Garcia et al., 2011; among others), suggesting downside risk as an additional consideration.[5]

The recent literature has suggested various downside risk measures. For example, Harvey and Siddique (2000) present a conditional skewness measure that captures nonlinear risks in stock prices, Bali and Cakici (2004) describe the value-at-risk as a tail-risk measure, and Conrad et al. (2013) use option market data to directly identify how the market prices tail risks. The simplest downside risk measure is the maximum drawdown (MDD), defined as the largest peak-

[5] An old theoretical literature explains why tail risks in returns matter to investors (Rubinstein, 1973; Kraus and Litzenberger, 1976) and presents asset allocation models with higher moments (Bawa and Lindenberg, 1977; Harlow and Rao, 1989).

to-trough loss associated with a series of returns (Gray and Vogel, 2013; Riley and Yan, 2022). MDD captures the worst possible performance scenario experienced by a buy-and-hold investor dedicated to a specific portfolio strategy. The concept was initially propagated by practitioners (Young, 1991) and has recently gained traction as an intuitive and investor-relevant risk measure. Most investors pay close attention to MDD because it is easy to interpret: How much can they lose when buying at the peak and selling at the bottom? Korn et al. (2022, p. 105) affirm that “the maximum drawdown is undoubtedly the most well-known measure” among drawdown-based risk metrics. Therefore, there is a good reason why Morningstar’s website reports maximum drawdown alongside volatility and beta on the same tab.

Technically, MDD is different from other downside risk measures in that it depends on the order in which returns occur (Van Hemert et al., 2020; Korn et al., 2022). Unlike standard risk measures such as volatility, semivariance, or skewness, MDD is path-dependent—its value is not only determined by return magnitudes but also by their temporal ordering (Van Hemert et al., 2020; Korn et al., 2022). This property makes it particularly sensitive to the clustering of negative returns, a common feature in financial time series.

In order to adequately capture the implications of drawdown risk for investment decisions, it is insufficient to analyse realised maximum drawdown (MDD) alone. Instead, the full distribution of possible MDD outcomes must also be considered. Bootstrap simulation techniques provide a useful framework for estimating such distributions. However, it is important to choose the right bootstrap method, as each approach carries different assumptions and statistical properties. The classical bootstrap method introduced by Efron (1979), which assumes independent and identically distributed (i.i.d.) observations, may not be suited for financial time series data, which often exhibit serial dependence. To address dependencies in time series, a variety of block bootstrap methods have been developed (Hall, 1985). These include non-overlapping block bootstrap methods (Carlstein, 1986) and overlapping or moving block bootstrap methods (Künsch, 1989). A problem is that non-overlapping blocks may result in a small number of usable resamples, especially for short time series. Although overlapping blocks increase the number of resampling units, they introduce another problem: non-stationarity. This means that data points near the ends of a return series are selected with a lower probability than those in the middle. To address this issue, Politis and Romano (1994) propose a stationary bootstrap method that uses randomly sized blocks to ensure an equal probability of selection for all data points, while preserving the stationarity of the resample.

These three bootstrap techniques have been widely used in financial applications.[6] For example, Fama and French (2018a, 2018b) rely on the Efron (1979) bootstrap, justifying their choice with the low autocorrelation of monthly stock returns. Arnott et al. (2019) and Khang et al. (2023) employ fixed-length moving block bootstraps, while Anarkulova et al. (2022) use the stationary bootstrap of Politis and Romano (1994). Among these, Arnott et al. (2019) is the only study that explicitly applies bootstrapping to MDD estimation. They document that MDD estimates of factor portfolios derived from moving block bootstraps more closely align with observed drawdowns than those from standard i.i.d.-based simulations.

Our paper extends the literature in several important ways. First, we systematically compare the three bootstrap methods that have been most widely used in asset management applications—the standard Efron bootstrap (1979), the moving block bootstrap (Künsch, 1989), and the stationary bootstrap (Politis and Romano, 1994)—and assess the effect of different block lengths on the accuracy of simulated MDD distributions. Second, unlike prior studies that focus on point estimates, we evaluate the entire distribution of MDD using stochastic dominance tests. Thirdly, we broaden the scope of the analysis by including both the equity market, represented

[6] For example, Efron's (1989) standard bootstrap is used in Benartzi and Thaler (1995), Hickman et al. (2001), Sinha and Sun (2005), Dierkes et al. (2010), Fama and French (2018a, 2018b). Künsch's (1989) moving block bootstrap is applied in Hansson and Persson (2000), Annaert et al. (2009), Choi and Mukherji (2010), Cogneau and Zakamouline (2013), Fong and Koh (2015), Dichtl et al. (2017), Arnott et al. (2019) and Khang et al. (2023). White (2000), Hansen (2005), Cogneau and Zakamouline (2013), and Anarkulova et al. (2022) are examples for the use of the bootstrap approach proposed by Politis and Romano (1994). Lahiri (1999) provides a systematic comparison of various block bootstrap methods in terms of statistical properties.

by the MSCI World Index and the Nasdaq 100 Index, as well as the cryptocurrency market, represented by Bitcoin. The behaviour of these markets in terms of drawdown is distinctly different. Fourthly, we investigate whether the absence of autocorrelation returns justifies the use of the Efron bootstrap method for MDD estimation. Fifthly, we assess the impact of the non-stationarity problem on the results of the moving block bootstrap method by analysing truncated return series in which extreme drawdowns occur at the start or end of the sample period.

Another important contribution we make is testing alternative procedures for dependent data. In particular, we use alternatives to the block bootstrap and analyse the block-block bootstrap (Andrews, 2004) and the tapered block bootstrap (Paparoditis and Politis, 2001) in robustness tests. To the best of our knowledge, these methods have not yet been employed in the asset management literature. Finally, we complement our MDD analysis by examining average drawdown (ADD), which captures the average severity of all underwater periods rather than a single extreme event.

We summarise our main findings as follows: The time series of the markets under investigation vary considerably. The MSCI World Index has a long data history with relatively few significant drawdowns. By contrast, our data history for Bitcoin is short and includes several large drawdowns. Using Efron's (1979) simple bootstrap method, we observe severe underestimation of the true MDD in most cases. This result remains consistent even when no significant autocorrelation is observed in the time series. Our simulation results for the moving block bootstrap method are generally good, provided that the non-stationarity property is not an issue. Otherwise, we observe a clear underestimation of the true MDD. In particular, the stationarity problem becomes serious when MDD occurs at the left or right boundary of a sample. We obtain the best and most robust simulation results using Politis and Romano's (1994) stationary bootstrap approach. The two alternative procedures — the block-block bootstrap and the tapered block bootstrap — do not outperform the stationary bootstrap, at least in our application.

Therefore, we conclude that the stationary bootstrap method proposed by Politis and Romano (1994) is the most suitable method for estimating drawdown risk and helping investors to make their asset allocation decisions.

The remainder of this paper is organized as follows. Section 2 describes our data. Section 3 outlines our empirical methodology, describing the bootstrap simulation methods and evaluation criteria. Our main results are presented in section 4. Robustness tests based on alternative bootstrap procedures are discussed in section 5. Finally, section 6 concludes.

2. Drawdowns in stock and cryptocurrency markets

Drawdowns are a common feature across asset classes, yet they tend to be particularly pronounced, and often exhibit marked asymmetry, for riskier assets such as equities and, to an even greater extent, cryptocurrencies. Financial markets for stocks and cryptocurrencies therefore provide a natural empirical setting for evaluating bootstrap methods that aim to accurately capture drawdown dynamics. We use monthly stock market data from the MSCI World Total Return Index (MSCI World Index), denominated in USD. This is one of the most widely used global equity benchmark indices.[7] For cryptocurrency markets, we use monthly Bitcoin data. As Harvey et al. (2022) observe, Bitcoin is the most popular cryptocurrency and has the longest available data history. Due to its high correlation with other cryptocurrencies (Hu et al., 2019), Bitcoin appears to be an appropriate representative of this asset class in our study.

Monthly stock prices for the MSCI World Index are available from 31 December 1969 onwards. For Bitcoin, however, data is only available from 31 August 2011 onwards. Exhibit 1, Panel A, shows the five largest drawdowns for the MSCI World Index from December 1969 to December 2023. Panel B shows the five largest drawdowns for Bitcoin from August 2011 to December 2023. For the MSCI World Index, the MDD is -53.65%, coinciding with the global

[7] We follow Van Hemert et al. (2020) and conduct our analysis using monthly rather than daily data. Economic arguments for this decision are provided in Van Hemert et al. (2020, p. 36), footnote 5.

financial crisis (GFC). For Bitcoin, the MDD is significantly higher at -79.77%. The five largest drawdowns of Bitcoin are much larger than those of the global stock market index. In particular, despite Bitcoin's very short data history, we observe three severe drawdowns of over 70%.

[Insert Exhibit 1 here]

Exhibit 2 illustrates the underwater charts for the MSCI World Index in Panel A and for Bitcoin in Panel B. Three of the five largest drawdowns for the MSCI World Index have occurred in the last twenty-four years. The prevailing economic environments during these periods are well known: the bursting of the dotcom bubble, the global financial crisis, and the conflict in Ukraine. The underwater chart of Bitcoin illustrates its even higher potential for drawdown. During its brief recorded history, there have been four drawdowns of over 60% and five of over 40%. Given the pronounced differences between the two data sets, it is instructive to explore the specific effects in our bootstrap simulations.

[Insert Exhibit 2 here]

3. Empirical methodology

3.1 Bootstrap simulation methods

In our bootstrap simulations, we apply the standard Efron (1979) bootstrap, where N is the full sample size of our observed data points $X = \{X_1, \dots, X_N\}$. We randomly select N elements with replacement: $X^* = \{X_1^*, \dots, X_N^*\}$. Repeating this procedure several times, and computing the estimator of interest $\hat{\theta}(X^*)$ for each resample, we obtain an approximate probability distribution of the estimator $\hat{\theta}(X)$. In our application, the observed data points $\{X_1, \dots X_N\}$ are monthly returns, and the estimated parameter $\hat{\theta}(X)$ is MDD. This simple approach assumes that the underlying data are i.i.d., which can be problematic for financial return data because serial dependence is not taken into account (Arnott et al., 2019).

Implementation of the Efron bootstrap is straightforward. Let $X = \{X_1^*, \dots, X_{648}^*\}$, denote the observed monthly returns of the MSCI World Index spanning January 1970 to December 2023, yielding $N = 648$ observations. To generate a bootstrap sample, we draw N observations from X with replacement, where each return is equally likely to be selected. Specifically, we sample N integers uniformly from the set $\{1, \dots, 648\}$, where each integer represents an index used to retrieve the corresponding return from the original dataset. The resulting bootstrap sample $X^* = \{X_1^*, \dots, X_{648}^*\}$ is then used to compute the statistic of interest—in our application, the maximum drawdown, denoted as $\hat{\theta}(X^*) = \widehat{MDD}^*$. Repeating this resampling procedure many times (e.g., 1,000 times) yields an empirical distribution of the estimated maximum drawdown.

In terms of block bootstrap methods, we distinguish between non-overlapping (Carlstein, 1986) and overlapping (Künsch, 1989) approaches.[8] However, due to the relatively small size of our data sample (particularly for Bitcoin), a block bootstrap with non-overlapping blocks would not be feasible.[9] Therefore, we use the bootstrap method with overlapping blocks, also known as the "moving block bootstrap" method, hereafter referred to as MB.

In particular, let l denote the block length, where we have $N - l + 1$ different data blocks $\{(X_1, \dots, X_l), \dots, (X_{N-l+1}, \dots X_N)\}$, labeled as $\{B_1, \dots, B_{N-l+1}\}$. We randomly select b data blocks with replacement, so $l \times b = N$. If this ideal identity does not hold, we follow Hall et al. (1995) and set b as the integer part of N/l. In this way, we have $N' = l \times b$, where $N' < N$. Additionally, we select an extra data block with length $N' - N$. Combining all b data blocks, and possibly the $(N' - N)$ block (if necessary), leaves us with our bootstrap sample $X^* = \{X_1^*, \dots, X_N^*\}$.

To illustrate the implementation of the moving block bootstrap, we adopt a fixed block length of $l = 36$ monthly returns of the MSCI World index. Given the full sample of $N = 648$

[8] Hall (1985) discusses both approaches in the context of spatial data.

[9] For example, if we have ten years of monthly return data $(N = 120)$, and we select data blocks with $l = 24$ elements, we will only have $b = 5$ different data blocks.

monthly index returns (from January 1970 to December 2023), we have $N - l + 1 = 613$ overlapping data blocks of the form: $\{(X_1, \dots, X_{36}), (X_2, \dots, X_{37}), \dots, (X_{613}, \dots, X_{648})\}$. To construct $N = 648$, we draw $b = 18$ blocks ($l \times b = 36 \times 18 = 648$) with replacement from the set of 613 available blocks. This is achieved by generating 18 random integers ranging from 1 to 613, each with equal probability, and using them as indices to select the corresponding blocks. Because selection is with replacement, some blocks may be chosen multiple times. The concatenation of the eighteen blocks forms the bootstrap sample, preserving local temporal dependence structures. We then calculate the statistic of interest — the maximum drawdown $\widehat{MDD}^*$. Repeating this procedure multiple times yields an empirical distribution of the estimated maximum drawdown.

A main drawback of the moving block bootstrap is the non-stationarity problem, because the elements $\{X_1, \dots, X_N\}$ have different probabilities for being selected in our subsamples. For example, the first sample element, X_1, is only part of a subsample when the first block is drawn. In contrast, X_2 is part of the subsample when either the first or second block is chosen (and so on). In general, elements at the beginning or end of our sample are less likely to be included in the resamples than those in the middle.[10]

As a third bootstrap method, we also use the stationary bootstrap approach of Politis and Romano (1994), hereafter referred to as PR, to overcome this problem. In contrast to the moving block bootstrap, where block length l is fixed, PR derives the length of block k (l_k) from a

[10] An example can help to illustrate this problem. Imagine we have ten return values labelled from 1 to 10 and select data blocks of length three. In this case, eight different data blocks can be selected: (1;2;3), (2;3;4), (3;4;5), (4;5;6), (5;6;7), (6;7;8), (7;8;9), and (8;9;10). Each block has an equal probability of being selected, i.e. 1/8 = 0.125. However, the probability of selecting each of the ten elements is not equal. For instance, there is only one scenario in which the first and last elements will be selected. The first (last) element is chosen only when the first (eighth) block is chosen. The second and ninth elements can be selected in two scenarios (blocks 1 and 2, and blocks 7 and 8, respectively). All the other elements (3, 4, 5, 6, 7 and 8) have a threefold chance of being selected. Clearly, elements closer to the left and right sample boundaries are less likely to be selected than those in the middle.

geometric distribution with probability p. Accordingly, mean block length $\bar{l}$ equals $1/p$. If we specify average block length $\bar{l}$, the corresponding probability p equals $1/\bar{l}$.

The algorithm of the stationary bootstrap from PR is as follows: In the first step, we generate a random integer number i between 1 and N. This number specifies the first element we draw, X_i^*. In the next step, we generate a uniform random number, $j \in [0;1)$. If $j \geq p$, we select the next element X_{i+1}^*, and add it to the existing data block (with element X_i^*). However, if $j < p$, we generate a new random integer i between 1 and N and start a new data block. The procedure ends when our linked data blocks yield a sample with N elements $\{X_1^*, \ldots, X_N^*\}$.

The main advantage of this bootstrap method is its circular sampling structure. If element X_N is selected as a sample element and the current data block is enhanced, the next element can "wrap around" to element X_1 and the following elements. Therefore, unlike a moving block bootstrap, this method is not affected by non-stationarity bias because all return observations have equal selection probabilities.

The implementation of the PR stationary bootstrap is relatively simple. Unlike the moving block bootstrap, which uses a fixed block size, the PR bootstrap uses random block lengths drawn from a geometric distribution. The length of each block l_k is chosen stochastically based on a geometric distribution, determined by the geometric parameter p. For example, if the expected block length is $\bar{l} = 36$, then $p = 1/36 = 0.027777$. The procedure starts with the random selection of an initial index from the set $\{1, \ldots, 648\}$, which corresponds to a starting point in our monthly return series. Assume that we select the monthly return at position fifty. At each subsequent step, we draw a uniform random number $j \in [0;1)$. If $j \geq p$, the next return in the sequence is added to the current block. In our example, this is the monthly return at position fifty-one. However, if $j < p$, we create a new block by drawing a new integer number uniformly at random from the full set of indices. This algorithm provides block of monthly returns with a mean block size of thirty-six elements.

To ensure continuity in the event of edge cases, for example, if the last selected index is 648 and the algorithm calls for the "next" return, the method wraps around a circle to the start of the original series, i.e., the monthly return at position 1. This circular resampling mechanism ensures that each element has an equal selection probability, thereby preserving the stationarity assumption. The algorithm stops when 648 returns have been selected to form one bootstrap resample of monthly returns, after which the maximum drawdown $\widehat{MDD}^*$ is computed. Repeating this procedure many times produces an empirical distribution of the estimated MDD.

3.2 Evaluation criteria

In contrast to the Efron bootstrap, the moving block bootstrap and the stationary bootstrap require the specification of a block length. We follow the literature and choose values of twelve, twenty-four, and thirty-six months.[11] We can thus infer the effect of block length on our simulation results. All simulations are based on $B = 1{,}000$ runs.[12] We calculate the mean MDD over all B simulation runs and compare the results with the true (population) MDD. The difference between the mean of all B bootstrapped statistics and the "plug-in" estimator of observed values represents the estimated bias of a bootstrap approach (Efron and Tibshirani, 1994):

$$\widehat{bias}_B = 1/B \sum_j^B \hat{\theta}_j^* - \hat{\theta} \qquad \rightarrow \qquad \widehat{bias}_B = 1/B \sum_j^B \widehat{MDD}_j^* - \widehat{MDD} \tag{1}$$

Our first evaluation criterion is the estimated bias based on means, defined in equation (1). However, due to the mean statistic's sensitivity to outliers, we also evaluate the median statistic, which is less sensitive to outliers.

Arnott et al. (2019) measure the percentage of simulation runs that produce a worse MDD than that observed. We also provide this statistic, denoting it as "% < real MDD." Under perfect

[11] There exist specific algorithms for computing optimal block length. However, they strongly depend on specific simulation contexts, which are not comparable to our application (Cogneau and Zakamouline, 2013). Arnott et al. (2019) use a block length of twelve months, Van Hemert et al. (2020) use twenty-four months, and Khang et al. (2023) use thirty-six months.

[12] We are able to attain sufficiently robust results with 1,000 simulation runs.

simulation conditions, 50% of all simulated values are expected to be above the population value and 50% below. If there are deviations, we strictly prefer overestimation of the population MDD to underestimation. When overestimating the true MDD, we potentially forego upside by calibrating the risk of a strategy too conservatively or by applying costly hedging strategies. Conversely, underestimating the true MDD may lead to losses for drawdown-sensitive investors, such as insurance companies and pension funds, that exceed their risk limits and force them to sell risky assets at the worst possible moment.

Efron and Tibshirani (1994) measure the asymmetry of the interval around the point estimate $\hat{\theta}$ ($\widehat{MDD}$). The so-called "shape" measure is defined as: $shape = \frac{\hat{\theta}_{up} - \hat{\theta}}{\hat{\theta} - \hat{\theta}_{lo}}$, where $\hat{\theta}_{up}$ and $\hat{\theta}_{lo}$ represent the upper and lower quantile, respectively. For example, considering a 90% confidence interval, shape is defined as $\frac{\hat{\theta}_{0.95} - \hat{\theta}}{\hat{\theta} - \hat{\theta}_{0.05}}$. Given $B = 1{,}000$ bootstrapped MDD values, $\hat{\theta}_{0.05}$ labels the 5%-quantile and $\hat{\theta}_{0.95}$ the 95%-quantile. A $shape < 1$ indicates greater distance from $\hat{\theta}$ to $\hat{\theta}_{0.05}$ than from $\hat{\theta}_{0.95}$ to $\hat{\theta}$, which is desirable if we prefer an overestimation of the true MDD over an underestimation. In contrast, a $shape > 1$ indicates an underestimation of the true MDD. We also report the mean squared error (MSE): $MSE(\hat{\theta}^*; \hat{\theta}) = E\left([\hat{\theta}^* - \hat{\theta}]^2\right)$. With $B = 1{,}000$ bootstrapped $\widehat{MDD}^*$ values, the MSE is $1/1000 \times \sum_{i=1}^{1000}\left(\widehat{MDD}_i^* - \widehat{MDD}\right)^2$, where $\widehat{MDD}$ denotes the population MDD. This measure is purely informational and provides no evidence of over- or underestimation of the true MDD.[13]

Our bootstrap simulations provide not only a single statistical measure (e.g., the mean or median MDD), but the complete distribution of the MDD as well. We therefore also compare the MDD distributions resulting from the various bootstrap simulations in terms of stochastic dominance (SD), using a suitable hypothesis test. To account for serial dependence in the data,

[13] Efron and Tibshirani (1994) discuss how the mean squared error (MSE) can be split into its two components: squared bias and variance. However, these measures do not provide any additional insight into our analysis.

we apply the SD test proposed by Linton et al. (2005). In particular, we test the null hypothesis that prospect y (with data points $\{X_{y,1}, \ldots, X_{y,N}\}$) statistically dominates prospect z (with data points $\{X_{z,1}, \ldots, X_{z,N}\}$) at degree *s* (i.e., $s = 1,2, \ldots$). In our context, the prospect y and z are two different bootstrap methods (e.g., the Efron and the moving block approaches), which should be compared in terms of their $B = 1{,}000$ bootstrapped $\widehat{MDD}^*$ values.

Since most of the literature formulates the SD assumption in the null hypothesis (Barrett and Donald, 2003; Lee and Whang, 2023), we follow this approach. In particular, the test statistic is defined as (Linton et al., 2005):

$$T_N^{(s)}(y) = sup_{x \in \chi} \sqrt{N} \left[\widehat{D}_y^{(s)}(x) - \widehat{D}_z^{(s)}(x) \right]. \quad (2)$$

Since we only test for first-order stochastic dominance ($s = 1$), $\widehat{D}_y^{(s)}(x)$ represents the cumulative distribution of our prospect y (which is z analogous):

$$\widehat{D}_y^{(1)}(x) = \widehat{F}_{y,N}(x) = \frac{1}{N} \sum_{i=1}^{N} 1\left(X_{y,i} < x\right). \quad (3)$$

In equation (2), χ is the union of the support sets of the two distributions $\widehat{F}_{y,N}$ and $\widehat{F}_{z,N}$. In equation (3), $1(\cdot)$ is the indicator function. To derive the distribution of the test statistic, we generate $N - l + 1$ subsamples *S* with block length *l*: $S_1 = \{(X_{y,1}; X_{z,1}), \ldots, (X_{y,l}; X_{z,l})\}$, $S_2 = \{(X_{y,2}; X_{z,2}), \ldots, (X_{y,l+1}; X_{z,l+1})\}$,…, $S_{N-l+1}\{(X_{y,N-l+1}; X_{z,N-l+1}), \ldots, (X_{y,N}; X_{z,N})\}$. For each $N - l + 1$ subsample, we compute the test statistic $t_l^{(1)}(y)$ defined in Equation (2).[14] With $M = N - l + 1$, the *p*-value of the test, $\hat{p}_M^{(1)}(y)$, can be determined as (Linton et al., 2005):

$$\hat{p}_M^{(1)}(y) = \frac{1}{N-l+1} \sum_{i=1}^{N-l+1} 1\left(t_{l,i}^{(1)}(y) > T_N^{(1)}(y)\right). \quad (4)$$

[14] In this case, we multiply by $\sqrt{l}$ rather than $\sqrt{N}$ (as in equation (2)).

Based on equation (4), large values of the test statistic $T_N^{(1)}(y)$ will lead to low p-values $\hat{p}_M^{(1)}(y)$, suggesting the alternative hypothesis "No stochastic dominance at degree one."

As already explained, one issue with the moving block bootstrap, as applied in Linton et al.'s (2005) SD test, is that observations at the beginning or end of a return series are less likely to be included in the bootstrapped samples. To mitigate this problem, Kläver (2006) proposes the use of the circular block bootstrap method (Politis and Romano, 1992), adding $l-1$ additional blocks to the data blocks from the moving block bootstrap: $S_{N-l+2} = \{(X_{y,N-l+2}; X_{z,N-l+2}), \dots, (X_{y,1}; X_{z,1})\}$, $S_{N-l+3} = \{(X_{y,N-l+3}; X_{z,N-l+3}), \dots, (X_{y,2}; X_{z,2})\}$, …, $S_N = \{(X_{y,N}; X_{z,N}), \dots, (X_{y,l-1}; X_{z,l-1})\}$. Equation (4) can be written as follows based on these N data blocks:

$$\hat{p}_N^{(1)}(y) = \frac{1}{N}\sum_{i=1}^{N} 1\left(t_{l,i}^{(1)}(y) > T_N^{(1)}(y)\right). \quad (5)$$

To parameterize our stochastic dominance tests, we follow Lee and Whang (2023) for the union of the support sets χ and underlay 100 equally spaced grid points. Following Kläver, we determine the block length as $l(N) = 10\sqrt{N}$. With $N = 1{,}000$ data points (simulation runs), the block length is set to 316.

4. Simulation results

4.1 Main simulation results and stochastic dominance tests

Exhibit 3 presents our main simulation results for the MSCI World Index (Panel A) and for Bitcoin (Panel B). Using the Efron bootstrap method, the population MDD is severely underestimated in Panel A: while the MDD on global stock markets is -53.65%, the mean bootstrapped MDD is -39.50% and the median MDD is -38.30%. The statistic "% < real MDD" indicates that the absolute value of the bootstrapped MDD is higher than that of the true MDD in only 7.5% of the 1,000 bootstrap simulation runs. For a 90% confidence interval, we observe a high shape value of 9.88. This implies that most of the observations between the 5% and 95%

quantiles lie within the interval $[-53.65\%;\ -27.61\%]$, with very few lying within the interval $[-56.28\%;\ -53.65\%]$. In other words, the Efron bootstrap generates many more underestimations than overestimations of the true MDD.

Using the moving block bootstrap method, notably higher mean and median MDD values are observed for all three block lengths compared to the simple Efron method. When analysing the percentage overestimation of the population MDD, the "% < real MDD" metric now ranges from 31% (with a block length of twelve months) to 42.2% (with a block length of thirty-six months) across the 1,000 simulation runs. These figures for the MB method are more favourable than those for the Efron bootstrap method, where the MDD (in absolute values) is higher than the true MDD in only 7.5% of the 1,000 bootstrap simulation runs. This result is also supported by the shape statistic, which is much lower for the MB method than for the Efron bootstrap in all three cases (1.70, 1.38, and 1.42 versus 9.88).

[Insert Exhibit 3 here]

Similar results are obtained for the stationary bootstrap. The mean MDD is moderately below the population MDD (-53.65%) for all three block lengths, while the median MDD exactly matches the true MDD for twenty-four- and thirty-six-month block lengths. In terms of overestimating the true MDD, the PR method performs slightly more favourable than the MB method for twelve-month block lengths (33.7% vs. 31.0%), but this result is not replicated for twenty-four- or thirty-six-month block lengths. As expected, comparing mean squared errors, the MSE values are lower for both the MB method and the PR method than for the Efron bootstrap. However, the MSEs are consistently lower for the PR method than for the MB method across all three block lengths. Overall, these findings seem to suggest that a block length of thirty-six months is preferable to twelve months in the context of MDD simulation.

Panel B shows the corresponding simulation results for Bitcoin. Once again, the Efron bootstrap leads to an underestimation of the population MDD. While the true MDD is -79.77%, the mean resampled MDD is -75.90% and the median is -76.11%. Moreover, the absolute value of the simulated MDD exceeds its true value in only 38% of the 1,000 simulations. In all other simulation setups, whether using the MB or PR method, the mean MDD is slightly higher than 80%, thus being close to the population MDD. As one would expect, the percentage overestimation of the true MDD, as indicated by the "% < real MDD" metric, fluctuates around 50%.

As explained above, the main advantage of bootstrap simulations is that, in addition to providing estimates of statistical moments (such as mean and median), they offer a full distribution of the parameters of interest. Exhibit 4 shows the quantiles ranging from 10% (Q(10)) to 90% (Q(90)) for the MSCI World Index (Panel A) and Bitcoin (Panel B). For the MB and the PR methods, where we have assumed block lengths of 12, 24, and 36 months, we now select the least biased versions for the comparison. As discussed in Section 3.2, this is the bootstrap method where the mean MDD over all 1,000 bootstrap runs is closest to the population MDD (with $\widehat{bias} = 1/1000 \sum_j^{1000} \widehat{MDD}_j^* - \widehat{MDD}$). As can be inferred from Exhibit 3 above, for both the MSCI World Index and Bitcoin, the optimal block length is 36 months, labelled as MB36 and PR36 methods, respectively.

[Insert Exhibit 4 here]

Panel A confirms that the median MDD (Q50) of the Efron simulation (-38.30%) significantly underestimates the population MBB (-53.65%), whereas the MB36 and PR36 methods precisely match the true MDD. This underestimation holds not only for the median MDD, but for the entire distribution of the MDD. For all quantiles, the values estimated using the Efron bootstrap are consistently lower than those based on the MB36 and PR36 methods.

Next, to verify the statistical significance of our results, we conduct stochastic dominance tests.[15] We test three relationships for first-order stochastic dominance. First, we compare the

[15] We conduct our SD tests using the free Python package, PySDTest, provided by Lee and Whang (2023). We are grateful to both authors for providing this powerful toolbox.

MB36 and PR36 methods with the Efron bootstrap; the null hypotheses for first-order stochastic dominance are $MDD^{MB36} \succcurlyeq_{SD1} MDD^{Efron}$ and $MDD^{PR36} \succcurlyeq_{SD1} MDD^{Efron}$. Second, we test whether the PR36 method stochastically dominates the MB36 method; in this case, the corresponding null hypothesis is $MDD^{PR36} \succcurlyeq_{SD1} MDD^{MB36}$.

The results for the MSCI World Index (Panel A) suggest that the null hypothesis of first-order stochastic dominance of the MB36 and PR36 methods against the Efron bootstrap cannot be rejected at the usual 5% statistical significance level. The *p*-value for not rejecting the null hypothesis of stochastic dominance for the PR36 method is much higher than that for the MB36 method (1.00 vs. 0.12). Nevertheless, when PR36 is tested against MB36, the null hypothesis is rejected, with a *p*-value of 0.00. This indicates that the moving block bootstrap is not first-order stochastically dominated by the stationary bootstrap.

The results for Bitcoin (Panel B) are not as clear. For the higher quantiles (Q(40) to Q(90)), the MB36 and PR36 methods produce higher estimated MDDs than the Efron bootstrap method. The reverse is true for lower quantiles (Q(10) to Q(30)). In such ambiguous cases, a test for stochastic dominance can provide valuable insights. The null hypothesis that the MB36 method dominates the Efron bootstrap must be rejected at the 5% significance level (with a *p*-value of 0.00). However, the opposite holds for the PR36 method; comparing PR36 with the Efron bootstrap, the *p*-value of 0.18 fails to reach standard levels of statistical significance. But most important, directly comparing PR36 with MB36, the null hypothesis of first-order dominance, written as $MDD^{PR36} \succcurlyeq_{SD1} MDD^{MB36}$, cannot be rejected (with a *p*-value of 0.16).

To illustrate these findings for Bitcoin more clearly, we plot the cumulative distribution functions (CDFs) for the two stochastic dominance tests. Exhibit 5, Panel A, shows the CDFs of maximum drawdown for the MB method and the Efron bootstrap. The horizontal axis represents the absolute value of MDD, while the vertical axis indicates the cumulative distribution. The MB method exhibits strong dominance up to a probability of around 80%. From this point onwards, dominance shifts in favour of the Efron bootstrap: when drawdowns become very large, the MB method provides lower MDD estimates than the Efron bootstrap. This explains why the null hypothesis of first-order stochastic dominance ($MDD^{MB36} \succcurlyeq_{SD1} MDD^{Efron}$) must be rejected. In Panel B, the effects of very large drawdowns are less pronounced in the comparison between the PR method and the Efron bootstrap, and thus the null hypothesis of first-order stochastic dominance ($MDD^{PR36} \succcurlyeq_{SD1} MDD^{Efron}$) cannot be rejected in this case.[16]

[Insert Exhibit 5 here]

4.2 The role of autocorrelation

As Fama and French (2018a) emphasise, low autocorrelation in time series data may justify applying the standard bootstrap, which depends on the independence and identically distributed (i.i.d.) assumption. They argue that "[…] independence is a good approximation. Autocorrelations of monthly returns are close to zero" (p. 235). Nevertheless, the absence of autocorrelation does not necessarily imply that neighbouring data points are independent. Autocorrelation only measures the degree of linear cohesion, failing to capture more complex or non-linear temporal structures in the data. This distinction is particularly pertinent in our context, given that the risk metric under consideration — maximum drawdown, which is defined as the maximum loss from a peak to a trough over a given time horizon — is highly sensitive to the temporal clustering of large negative returns. When such extreme returns are interspersed with large positive returns, the MDD declines or even vanishes. This "dilution" effect precisely

[16] As can be seen in Exhibit 5, the cumulative distribution function of the Efron bootstrap is much smoother than those of the MB and PR methods. There are 148 monthly Bitcoin returns (from September 2011 to December 2023). For the Efron bootstrap, 148 monthly returns are randomly selected with replacement. To obtain 148 numbers using the MB method, a total of five data blocks are selected with replacement: four blocks containing thirty-six monthly returns each and one block containing four monthly returns. Due to the occurrence of many severe drawdowns within a short period of time in the data (see Panel B in Exhibit 2), there is a high probability that the MB method will produce a simulated MDD close to the population MDD. This phenomenon is evident in Panel B of Exhibit 5, where a significant increase in the cumulative distribution function for the MB simulations is observed at the MDD with an absolute value of 79.77%.

occurs when returns are resampled independently under the i.i.d. assumption of the Efron bootstrap. As shown above, this method systematically underestimates the true MDD.

In light of this finding, it is crucial to determine whether the presence of statistically significant autocorrelation should be regarded as an indication that the i.i.d.-based bootstrap is not suitable for risk measures such as MDD. We calculate the first-order autocorrelations for the MSCI World Index and Bitcoin. For the MSCI World Index, we measure a first-order autocorrelation of 0.062 for the series of monthly simple returns ranging from January 1970 to December 2023. As this value lies within the 95% confidence interval around zero (CI [-0.115; 0.115]), the null hypothesis of no autocorrelation cannot be rejected. For the Bitcoin return series from September 2011 to December 2023, the autocorrelation coefficient is 0.115, which is again within the 95% confidence interval around zero (CI [-0.161; 0.161]).

To gain further insights into the serial dependence of monthly returns, we measure the rolling thirty-six-month first-order autocorrelations of the MSCI World Index and Bitcoin returns. Exhibit 6, Panel A, illustrates the dynamics of MSCI World Index returns from January 2000 to December 2023. This time period encompasses the two largest drawdowns. With only a few exceptions, first-order autocorrelations are not significantly different from zero. Panel B illustrates the same analysis for Bitcoin. Once again, the rolling autocorrelation coefficients are, in most cases, not significantly different from zero in a statistical sense. In light of these findings, the idea that low (or almost non-existent) autocorrelation justifies the use of the Efron bootstrap appears to be a non-viable conclusion, at least in the context of MDD analyses.

[Insert Exhibit 6 here]

4.3 The non-stationarity problem

As discussed above, a key limitation of the moving block bootstrap is the non-stationarity problem. Elements at the left and right boundaries of the sample are less likely to be selected than those in the middle. In our context, this implies that we should expect the population MDD to be underestimated using moving block bootstrap simulations if the largest drawdown occurs at the beginning or end of the sample. To examine this scenario, we truncate our MSCI World Index sample in February 2009, when drawdowns reach their peak during the global financial crisis, and repeat our simulation analyses.

Exhibit 7 shows the results. Once again, the Efron bootstrap leads to an underestimation of the population MDD (-53.65%), as indicated by both the mean MDD (-39.22%) and the median MDD (-38.14%).[17] As expected, given that the global financial crisis is now at the end of the truncated sample, there is also a significant underestimation of the MDD for the MB method using block lengths of twelve, twenty-four and thirty-six months. The simulated mean MDDs range from -43.70% (twenty-four months) to -44.68% (thirty-six months), while the median values range from -42.21% (twenty-four months) to -44.58% (thirty-six months). All values clearly underestimate the population MDD of -53.65%, indicating that the non-stationarity problem is now occurring severely in the truncated sample. The poor estimation quality of the MB method is confirmed by our two additional statistics: "% < real MDD" and shape. Notably, for all three block lengths, the percentage of simulation runs with an MDD higher (in absolute terms) than the population MDD is below 20% (19.8% for a block length of twelve months and 16.8% for a block length of thirty-six months). The corresponding value for the shape statistic is greater than one in all three cases (2.56, 2.82, and 2.97), which also indicates an underestimation of the population MDD.

[Insert Exhibit 7 here]

Most importantly, when the results of the MB and PR methods are compared, the latter performs better; for all three block lengths, its mean MDD is much closer to the population MDD in the truncated sample. The estimation error results from a moderate degree of

[17] This result is also supported by the percentage of simulation runs, in which MDD is larger than real-world MDD. At 7.30%, this value is much lower than a 50% chance of overestimation.

overestimation, which we view as more preferable than an underestimation (see the discussion above). For all three block lengths, the PR method provides absolute MDD values that are in 57.5% to 61.5% of all simulation runs larger than the population MDD. As expected, this is further reflected in shape statistics below 1.

Panel B shows the quantiles of the simulated MDD values. Based on the bias minimisation criterion (see Section 3.2 above for details), we compare the moving block bootstrap method with a block length of thirty-six months (MB36) with the stationary bootstrap method with an expected block length of twelve months (PR12). Underestimation of MDD using MB36 compared to PR12 is observable not only for the mean and median MDDs, but for the entire distribution of the MDD. This is confirmed by the first-order stochastic dominance test: as expected, the null hypothesis $MDD^{PR12} \succcurlyeq_{SD1} MDD^{MB36}$ cannot be rejected (with a *p*-value of 1.0).

5. Robustness tests

5.1. The non-stationarity problem revisited

To shed further light on the severity of the non-stationarity problem when using the moving block bootstrap method, we implement additional simulations to ensure robustness. As the truncated sample analysis is not applicable to Bitcoin, we repeated our bootstrap simulations using the Nasdaq 100 Index.[18] For this stock market index, monthly price data are available from January 1983 onwards. Unlike the MSCI World Index, the Nasdaq 100 Index experienced its maximum drawdown when the dotcom bubble burst. From March 2000 to September 2002, the index lost 81.07% of its value. It took until February 2015 for the index to fully recover from this massive drawdown, approximately twelve years after the MDD occurred. To illustrate

[18] With only 148 monthly returns, our Bitcoin sample size is small. Therefore, it is not feasible to limit the sample further by shifting the MDD to the left or right borders. In addition to the MDD with a loss of -79.77%, there are two other severe drawdowns with losses of -76.20% and -73.35%. If the returns associated with the maximum drawdown are not chosen because of non-stationarity, selecting from these other severe drawdowns instead may lead to simulated MDDs that are very close to the population drawdown, i.e., the potential negative effects of the non-stationarity problem do not show up.

the negative effects of the non-stationarity problem in combination with the moving block bootstrap once again, we truncate the Nasdaq 100 Index return series so that the MDD lies either at the end (using data from February 1983 to September 2002) or at the beginning (using data from January 2000 to December 2023) of the sample.

The results are shown in Exhibits 8 and 9, respectively. In both cases, the Efron bootstrap dramatically underestimates the population MDD, as indicated by the mean and median of the resampled MDDs (in absolute terms). In approximately 5% of all simulations runs, the absolute value of the bootstrapped MDD is higher than the population value (4.8% in Exhibit 8, Panel A, and 4.4% in Exhibit 9, Panel A).[19] The non-stationarity problem is clearly visible for the moving block bootstrap. In Exhibit 8, Panel A, the mean MDD ranges from -62.88% (block length of twelve months) to -54.64% (block length of thirty-six months). In absolute terms, all these values are below the population MDD of -81.07%. Using a block length of thirty-six months, the simulated MDD underestimates the population MDD in 94.70% of all simulation runs (as indicated by the "% < real MDD" statistic).[20] Most important, compared with the MB method, the PR method yields higher mean and median MDD values. While the simulated mean MDD remains lower than the population MDD when using a block length of thirty-six months, the median MDD is -78.08%, close to the population MDD of -81.07%. Moreover, the simulated MDD is higher than the population MDD in 44.30% of all simulation runs.

[Insert Exhibit 8 and Exhibit 9 here]

Exhibit 8, Panel B, illustrates the distribution of the MDD values for the Efron bootstrap, the MB12 method (with a block length of twelve months), and the PR36 method (with a block

[19] In Exhibit 8, the shape statistic becomes -58.85. This negative value is due to the 5% quantile (-80.35%) being below the real MDD of -81.07%. With these numbers, the denominator in the shape formula becomes negative $(-81.07 - (-80.35) = -0.72)$.

[20] The moving block bootstrap yields a value of 655.68 for the shape statistic. This high value is due to the 5% quantile (-81.15%) being close to the population MDD (-81.07%). Given these numbers, the denominator in the shape statistic is computed as $(-81.07 - (-81.15) = 0.08)$, resulting in a very high shape statistic.

length of thirty-six months). The PR36 method clearly outperforms both the Efron bootstrap and the MB12 method for almost all quantiles. However, at the 90th percentile, the PR method provides a lower MDD estimate than the two competing methods. Therefore, based on the test statistics for tests of first-order stochastic dominance, the two null hypotheses that the stationary bootstrap (PR36) dominates the moving block method (MB12) as well as the Efron bootstrap must be rejected. Panel C shows the cumulative distribution functions underlying the test of the null hypothesis $MDD^{PR36} \succcurlyeq_{SD1} MDD^{MB12}$.

Similar results are obtained in Exhibit 9 under the assumption that the maximum drawdown occurs at the beginning of a truncated sample. Panel A shows that the Efron bootstrap and the MB method underestimate the population MDD. As the block length increases, the mean and median MDDs (in absolute terms) of the MB method decrease. In about 80% of the simulation runs, the MB method delivers a resampled MDD that is below the population MDD. In contrast, the median MDD under the PR method with a thirty-six-month block length is -78.48%, again relatively close to the population MDD of -81.07%. In this case, the simulated MDD is higher than the population MDD (in absolute terms) in 39.9% of all bootstrap runs. As shown in Panel B, the null hypothesis of first-order stochastic dominance of the PR36 method against the Efron bootstrap cannot be rejected. However, the null hypothesis of first-order stochastic dominance of the PR36 method against the MB12 method ($MDD^{PR36} \succcurlyeq_{SD1} MDD^{MB12}$) must be rejected again. A graphical visualization of this comparison is shown in Panel C.

Overall, the simulation results based on truncated samples of the Nasdaq 100 Index suggest that the moving block bootstrap is limited by the non-stationarity problem. Elements at the left or right boundaries of the sample are less likely to be selected than those in the middle, which leads to resampling bias. The stationary bootstrap provides more accurate risk classifications, regardless of when the maximum drawdown occurs during the sample period.

5.2 Alternative bootstrap methods for dependent data

In our main analysis, we employ three bootstrap simulation techniques, which have been applied in various contexts in the asset management literature (Arnott et al., 2019; Anarkulova et al., 2022; Khang et al., 2023). These methods can be regarded as standard tools in quantitative investment research. In this section, we test alternative bootstrap procedures suitable for dependent data, which could potentially improve MDD estimations further. Specifically, we employ the block-block bootstrap approach proposed by Andrews (2004) and the tapered block bootstrap method introduced by Paparoditis and Politis (2001).

The block-block bootstrap: The block-block bootstrap, introduced by Andrews (2004), addresses asymptotic shortcomings of the standard block bootstrap. The non-parametric i.i.d. Efron (1979) bootstrap as well as parametric bootstrapping typically have superior asymptotic properties compared to the block bootstrap approach. This is due to the join-point problem, which arises when the independence between the blocks in the block bootstrap does not reflect the dependence structure of the original sample. To solve this problem, the block-block bootstrap does not modify the block bootstrap algorithm itself. Instead, it alters the sample statistic to which the block bootstrap is applied.

Andrews (2004) introduces block statistics (for the original sample) that have join-point features that resemble those of the block bootstrap versions of these statistics. Join-points occur at positions $l+1, 2 \times l + 1, \ldots, (b-1) \times l + 1$, where l is the block length, and b denotes the number of blocks such that $l \times b = N$ (the total sample size). Before each join-point, $[\pi \times l]$ observations are removed. $[\pi \times l]$ denotes the smallest integer greater than or equal to $\pi \times l$, where π represents the proportion of observations that are deleted from the bootstrapped samples and the observed sequence (Iglesias, 2013), and $[\cdot]$ is the ceiling function. This adjustment ensures that the last non-zero element of one block is separated from the first element of the subsequent block, thereby addressing the join-point problem and improving asymptotic behaviour. Implementation of the block-block bootstrap requires to specify the block length l and the deletion fraction π. Following Andrews (2004), we set $\pi = 0.125$.[21] For example, with a block length of thirty-six months ($l = 36$), $\pi l = 0.125 \times 36 = 4.5$ and $[\pi \times l] = 5$. As before, we implement the overlapping (moving block) bootstrap proposed by Künsch (1989)[22] and apply the deletion procedure to the observed sequence and the bootstrapped samples.

The simulation setup for the MSCI World Index is identical to that described above. Panel A in Exhibit 10 shows the simulation results using the full sample period from January 1970 to December 2023, and Panel B those using the truncated sample that ranges from January 1970 to February 2009, when the drawdowns reached their peak during the global financial crisis. The block-block bootstrap (BBB) is implemented with block lengths of 12, 24, and 36 months. To properly assess the BBB method, we select the least biased methods from the corresponding simulations above as benchmarks; the MB36 method for the full sample period (see Panel A in Exhibit 3) and the PR12 method for the truncated sample period (see Panel A in Exhibit 7).

[Insert Exhibit 10 here]

The first column in Exhibit 10, Panel A, contains the population MDD values, which – because of the way the BBB method is constructed – are also modified. Comparing mean and median MDD values of the three BBB specifications with the corresponding values of MB36 as well as the "true" population MDD, there is a tendency to underestimate. Underestimation is only moderate for the BBB36 approach. Starting from the population MDD of -53.65%, the mean MDD of MB36 is -51.50%, and the mean MDD of BBB36 is -49.98%. While the median MDD of MB36 matches exactly the true parameter (-53.65%), the median MDD of BBB36 is marginally lower (-53.25%) in absolute terms. Nevertheless, the null hypothesis of first-order stochastic dominance of MB36 over BBB36 cannot be rejected (with a *p*-value of 0.22).

---

[21] Andrews (2004) suggests a data-dependent nested bootstrap procedure for selecting (l, π), which can be time-consuming to implement. We leave the application of such methods in asset management for future research.

[22] The block-block bootstrap can be applied with both the non-overlapping block bootstrap (Carlstein, 1986) and the overlapping moving block bootstrap (Künsch, 1989). See Andrews (2004) and Iglesias (2013) for details.

Next, the results for the truncated sample in Panel B indicate all three BBB specifications lead to an underestimation of the population MDD. All mean and median MDD values are around -40.0%, which, in absolute terms, is well below the true value of -53.65%. In other words, the block-block bootstrap method does not perform any better than the standard moving block bootstrap (see Exhibit 7 for the results of the MB method for the truncated sample). The opposite is true for the PR12 method. The mean and median MDD values are -54.31% and -56.27%, respectively, and both are close to the population MDD value of -53.65%. This result is further supported by the number of simulated MDD values that are below the population MDD (see column "% < real MDD"). For all three BBB specifications, the resulting MDD underestimates the population MDD in roughly 90% of all simulation runs. For the PR12 method, underestimation occurs in only 43.4% of all simulation runs, a percentage that is close to the expected value of 50%. Therefore, it seems consistent that the null hypothesis of first-order stochastic dominance of PR12 against BBB36 cannot be rejected (with a *p*-value of 1.00).

In the context of MDD estimation for portfolio management decisions, our findings suggest that, overall, Andrew's (2004) block-block bootstrap method produces inferior results compared to the moving block and stationary bootstrap methods, both of which are well-established in the asset management literature.

The tapered block bootstrap: Another alternative to the block bootstrap for dependent data is the tapered block bootstrap of Paparoditis and Politis (2001). They demonstrate that the tapered block bootstrap outperforms the traditional block bootstrap of Künsch (1989), particularly with regard to reducing the asymptotic bias in time series data. The main idea of the tapered bootstrap is to assign diminishing weights to observations near the endpoints of each data block, i.e., the first and the last elements of a data block are down-weighted. Let h be the index of all the elements within a block, e.g., with a block length of five $(l = 5)$, $h = 1, \dots, 5$. The data tapering window is a single function w, with $w: \mathbb{R} \to [0,1]$, so that $w_l(h) = w\left(\frac{h-0.5}{l}\right)$. For the weighting function, Paparoditis and Politis (2001) suggest a trapezoidal function, w_c^{trap}, which is defined as:

$$w_c^{trap}(t) = \begin{cases} t/c, & \text{if } t \in [0, c] \\ 1, & \text{if } t \in [c, 1-c] \\ (1-t)/c, & \text{if } t \in [1-c, 1] \end{cases} \tag{6}$$

where c is some fixed constant in $(0; \frac{1}{2}]$ and $t = w_l(h)$. Given this interval, Paparoditis and Politis (2001) show that it is optimal to choose $c = 0.43$.

An example illustrates the computation steps. Assuming a block length of five ($l = 5$), the data tapering function $w_l(h)$ is applied on the following values: $\left(\frac{1-0.5}{5}\right) = 0.1$, $\left(\frac{2-0.5}{5}\right) = 0.3$, $\left(\frac{3-0.5}{5}\right) = 0.5$, $\left(\frac{4-0.5}{5}\right) = 0.7$ and $\left(\frac{5-0.5}{5}\right) = 0.9$. For $t = 0.1$ and $t = 0.3$, the resulting weights according to equation (6) are $0.1/0.43 = 0.2326$ and $0.3/0.43 = 0.6977$, respectively. For $t = 0.5$, the weight is 1.0. For $t = 0.7$ and $t = 0.9$, the weighting function yields $(1 - 0.7)/0.43 = 0.6977$ and $(1 - 0.9)/0.43 = 0.2326$, respectively.

Paparoditis and Politis (2001) note that the variance of the bootstrapped values decreases due to the shrinking caused by the weighting function in equation (6). To correct for this effect, they multiply the weights by the 'inflation factor' $l^{1/2}/\|w_l\|_2$, with $\|w_l\|_2 = \left\{\sum_{h=1}^{l} w_l^2(h)\right\}^{1/2}$. In our example, $\|w_l\|_2 = \{0.2326^2 + 0.6977^2 + 1^2 + 0.6977^2 + 0.2326^2\}^{1/2} = 1.4428$ and $l^{1/2}/\|w_l\|_2 = 5^{0.5}/1.4428 = 1.5498$. All weights are multiplied by this factor so that the five "inflation-adjusted" factors are 0.3604, 1.0813, 1.5498, 1.0813, and 0.3604.

To ensure that the block endpoints shrink to zero, Paparoditis and Politis (2001) centralize the original sample datapoints by subtracting the mean value of the population data: $Y_t = X_t - \bar{X}_N$, where $t = 1, 2, \dots, N$.[23] The tapered block bootstrap is implemented in the following way:

[23] As Paparoditis and Politis (2001) state: "The tapered block bootstrap shrinkage idea could also be implemented using the uncentered $\{X_t\}$ data, but in this case the block endpoints should be shrunk towards $\bar{X}_N$, instead to zero, and the resulting procedure becomes less transparent" (p. 1110).

As in the moving block bootstrap of Künsch (1989), we select b blocks from the centralized population data $\{Y_t\}$, each with block length l $(l \times b = N)$. For $m = 0, 1, \dots, b-1$, let:

$$Y^*_{ml+j} \coloneqq w_l(j) \frac{l^{1/2}}{\|w_l\|_2} Y_{i_m+j-1} \qquad \text{with } j = 1, 2, \dots, l, \tag{7}$$

where i_m denotes the starting index of the m[th] picked data block. To obtain the original data, we add the mean $\bar{X}_N$ to the centralized data: $X^*_{ml+j} = Y^*_{ml+j} + \bar{X}_N$. This procedure is repeated for all 1,000 simulation runs.

Again, the simulation setup for the MSCI World Index is identical to that in our baseline analysis. Panel A in Exhibit 11 shows the simulation results using the full sample period from January 1970 to December 2023, and Panel B those using the truncated sample that ranges from January 1970 to February 2009, when the drawdowns reached their peak during the global financial crisis. The tapered block bootstrap (TBB) is implemented with block lengths of 12, 24, and 36 months. To evaluate the TBB method, we again choose the appropriate benchmarks: the MB36 method for the full sample period (see Panel A in Exhibit 3) and the PR12 method for the truncated sample period (see Panel A in Exhibit 7).

[Insert Exhibit 11 here]

In Panel A, while the mean MDD of the TBB36 method (-53.01) is closer to the population MDD (-53.65%) than that of the MB36 method (-51.50), the opposite is true for the median values (-52.29% for TBB36 and -53.56% for MB36). As a result, comparing TBB36 and MB36 in terms of first-order stochastic dominance, the null hypothesis must be rejected in both directions, i.e., TBB36 does not dominate MB36, but MB36 also does not dominate TBB36.

Based on the truncated sample in Panel B, all three tapered block bootstrap specifications (TBB12, TBB24, and TBB36) tend to underestimate the population MDD. The mean and median MDD values of the TBB models are never below -44%, which, in absolute terms, is clearly below the population MDD (-53.65%). Conversely, the mean (-54.31%) and median (-56.27%)

MDDs of the PR12 benchmark model are much closer to the population value. Most important, the null hypothesis of first-order stochastic dominance of PR12 against TBB 36 cannot be rejected (with a p-value of 1.00).

Taken together, these results allow us to conclude that the tapered block bootstrap is inferior compared to the stationary bootstrap in the truncated sample, in which the non-stationarity problem seems to be highly pronounced. When modelling drawdown risk in investor portfolios, stationary bootstrap simulations appear to be the preferred method for researchers and practitioners.

5.3 Average drawdown

So far, the analysis has focused solely on maximum drawdown, which is a single risk number resulting from an extreme event. To gain further insights into drawdown behaviour, we next investigate the average drawdown, denotes as ADD, and defined as $ADD := \frac{1}{N}\sum_{i=1}^{N} D_i$, where $D_i := \frac{M_i - S_i}{M_i}$ and $M_i := \max_{t=0,\dots,i} S_t$ (Korn et al., 2022; Chekhlov et al., 2005). The population ADD is -7.96% for the MSCI World Index and -38.18% for Bitcoin. The corresponding simulation results are presented in Exhibit 12.

For the MSCI World Index in Panel A, median values fit better than mean values; the latter overestimate the population ADD in all cases. The metric "% < real ADD" varies from 48.30% to 54.40%, indicating that approximately 50% of all simulated ADD values are above and 50% are below the population ADD. In Panel B, Bitcoin exhibits a much larger ADD of -38.18%. As previously mentioned, Bitcoin has a high concentration of large drawdowns within a fairly small data sample (see Panel B in Exhibit 2). The Efron bootstrap clearly underestimates the population ADD. The ADD resulting from the MB method is higher for all three block lengths; however, even in these cases, the population ADD is underestimated. Finally, the PR method with a block length of thirty-six months yields the best results. With this parameterisation, the mean ADD is -36.20% and the corresponding median is -35.86%. These values are closest to the population ADD.

[Insert Exhibit 12 here]

6. Concluding remarks

This study examines the effectiveness of various bootstrap methods for simulating the distribution of maximum drawdown (MDD), which is a path-dependent risk measure particularly relevant in traditional and digital asset markets. Using monthly data from the MSCI World Index and Bitcoin, we conduct a comprehensive comparison of three bootstrap procedures that have been used in the asset management literature, including Efron's (1979) independent and identically distributed (i.i.d.) standard bootstrap, the moving block bootstrap (Künsch, 1989). and the stationary bootstrap (Politis and Romano, 1994). In addition, we assess two less known bootstrap procedures for dependent data for MDD estimation: the block-block bootstrap (Andrews, 2004) and the tapered block bootstrap (Paparoditis and Politis, 2001).

Our results highlight several important findings. First, the standard i.i.d. bootstrap significantly underestimates the true MDD, particularly in equity market simulations, even when the return series exhibits no statistically significant autocorrelation. This finding calls into question the common justification for using Efron's bootstrap in weakly autocorrelated financial data. We demonstrate that the absence of linear autocorrelation does not imply statistical independence, particularly for risk measures such as MDD that are sensitive to the temporal clustering of extreme negative returns. Second, the moving block bootstrap method works reasonably well, provided that the non-stationarity problem is not too severe. However, when the maximum drawdown occurs near the boundaries of the sample—which cannot be ruled out in financial time series—its performance deteriorates markedly. Our simulations confirm that this non-stationarity issue can lead to significant MDD underestimation. Third, the stationary bootstrap of Politis and Romano (1994) consistently delivers the most accurate and robust MDD estimates, especially with longer expected block lengths. Unlike the moving block approach, it is designed to preserve stationarity and captures the dependence structure of return series more effectively. Stochastic dominance tests further confirm the superiority of the stationary bootstrap for estimating not just central moments, but the entire distribution of drawdowns. Finally, extensions of the block bootstrap, including the block-block and tapered block methods, do not provide any significant improvements to the stationary bootstrap when it comes to MDD estimation. These findings remain consistent across different market environments and asset classes, including both equity and cryptocurrency markets.

We therefore recommend the stationary bootstrap method proposed by Politis and Romano (1994)—specifically with a longer average block length—as the preferred approach for simulating drawdown risk. This applies to academic researchers conducting empirical risk analyses, as well as to investment practitioners involved in asset allocation, portfolio stress testing, or risk-based strategy design. Despite its popularity and simplicity, the Efron (1979) bootstrap may yield misleading results in applications where path-dependent risk metrics such as maximum drawdown are central.

References

Anarkulova, A., Cederburg, S., and O'Doherty, M.S. (2022): Stocks for the Long Run? Evidence from a Broad Sample of Developed Markets, in: *Journal of Financial Economics*, Vol. 143, 409-433.

Andrews, D.W.K. (2004): The Block-Block Bootstrap: Improved Asymptotic Refinements, in: *Econometrica*, Vol. 72, 673-700.

Ang, A., Chen, B., and Sundaresan, S. (2013): Liability-Driven Investment with Downside Risk, in: *Journal of Portfolio Management*, Vol. 40, 71-87.

Annaert, J., Van Osselaer, S., and Verstraete, B. (2009): Performance Evaluation of Portfolio Insurance Strategies using Stochastic Dominance Criteria, in: *Journal of Banking and Finance*, Vol. 33, 272-280.

Arnott, R., Harvey, C.R., Kalesnik, V., and Linnainmaa, J. (2019): Alice's Adventures in Factorland: Three Blunders That Plague Factor Investing, in: *Journal of Portfolio Management*, Vol. 45, 18-36.

Artavanis, N., Eksi, A., and Kadlec, G. (2019): Downside Risk and Mutual Fund Flows. Working Paper, available at SSRN: https://ssrn.com/abstract=3302876.

Bali, T.G., and Cakici, N. (2004): Value at Risk and Expected Stock Returns, in: *Financial Analysts Journal*, Vol. 60(2), 57-73.

Barrett, G.F., and Donald, S.G. (2003): Consistent Tests for Stochastic Dominance, in: *Econometrica*, Vol. 71, 71-104.

Bawa, V.S., and Lindenberg, E.B. (1977): Capital Market Equilibrium in a Mean-Lower Partial Moment Framework, in: *Journal of Financial Economics*, Vol. 5, 189-200.

Benartzi, S., and Thaler, R.H. (1995): Myopic Loss Aversion and the Equity Premium Puzzle, in: *Quarterly Journal of Economics*, Vol. 110, 73-92.

Bodnaruk, A., Chokaev, B., and Simonov, A. (2019): Downside Risk Timing by Mutual Funds, in: *Review of Asset Pricing Studies*, Vol. 9, 171-196.

Carlstein, E. (1986): The Use of Subseries Values for Estimating the Variance of a General Statistic from a Stationary Sequence, in: *The Annals of Statistics*, Vol. 14, 1171-1179.

Chekhlov, A., Uryasev, S., and Zabarankin, M. (2005): Alice's Adventures in Factorland: Three Blunders That Plague Factor Investing, in: *International Journal of Theoretical and Applied Finance*, Vol. 8, 13-58.

Choi, B.P., and Mukherji, S. (2010): Optimal Portfolios for Different Holding Periods, in: *Journal of Business and Economics Research*, Vol. 8, 1-6.

Cogneau, P., and Zakamouline, V. (2013): Block Bootstrap Methods and the Choice of Stocks for the Long Run, in: *Quantitative Finance*, Vol. 13, 1443-1457.

Conrad, J., Dittmar, R., and Ghysels, E. (2013): Ex Ante Skewness and Expected Stock Returns, in: *Journal of Finance*, Vol. 68, 85-124.

Dichtl, H., Drobetz, W., and Wambach, M. (2017): A Bootstrap-Based Comparison of Portfolio Insurance Strategies, in: *European Journal of Finance*, Vol. 23, 31-59.

Dierkes, M., Erner, C., and Zeisberger, S. (2010): Investment Horizon and the Attractiveness of Investment Strategies: A Behavioural Approach, in: *Journal of Banking and Finance*, Vol. 34, 1032-1046.

Efron, B. (1979): Bootstrap Methods: Another Look at the Jackknife, in: *The Annals of Statistics*, Vol. 7, 1-26.

Efron, B., and Tibshirani, R.J. (1994): *An Introduction to the Bootstrap*, Chapman & Hall/CRC, Boca Raton (FL).

Fama, E. F. (1963): Mandelbrot and the Stable Paretian Hypothesis, in: *Journal of Business*, Vol. 36, 420-429.

Fama, E.F., and French, K.R. (2018a): Long-Horizon Returns, in: *Review of Asset Pricing Studies*, Vol. 8, 232-252.

Fama, E.F., and French, K.R. (2018b): Volatility Lessons, in: *Financial Analysts Journal*, Vol. 74, 42-53.

Fong, W.M., and Koh, T. (2015): Strategic Asset Allocation with Low-Risk Stocks: A Bootstrap Analysis, in: *Journal of Investment Management*, Vol. 13, 39-58.

Garcia, R., Renault, E., and Veredas, D. (2011): Estimation of Stable Distributions by Indirect Inference, in: *Journal of Econometrics*, Vol. 161, 325-337.

Goetzmann, W.N., Kim, D., and Shiller, R.J. (2017): Crash Beliefs from Investor Surveys, NBER Working Paper No. 22143.

Goetzmann, W.N., and Kim, D. (2018): Negative bubbles: What Happens After a Crash, in: *European Financial Management*, Vol. 24, 171-191.

Gray, W.R., and Vogel, J.R. (2013): Using Maximum Drawdown to Capture Tail Risk, Working Paper.

Hall, P. (1985): Resampling a Coverage Pattern, in: *Stochastic Processes and their Applications*, Vol. 20, 231-246.

Hall, P., Horowitz, J.L., and Jing, B.Y. (1995): On Blocking Rules for the Bootstrap with Dependent Data, in: *Biometrica*, Vol. 82, 561-574.

Hansen, P.R. (2005): A Test for Superior Predictive Ability, in: *Journal of Business & Economic Statistics*, Vol. 23, 365-380.

Hansson, B., and Persson, M. (2000): Time Diversification and Estimation Risk, in: *Financial Analysts Journal*, Vol. 56(5), 55-62.

Harlow, W.V., and Rao, R. (1989): Asset Pricing in a Generalized Mean-Lower Partial Moment Framework: Theory and Evidence, in: *Journal of Financial and Quantitative Analysis*, Vol. 24, 285-311.

Harvey, C., and Siddique, A. (2000): Conditional Skewness in Asset Pricing Tests, in: *Journal of Finance*, Vol. 55, 1263-1295.

Harvey, C.R., Zeid, T.A., Draaisma, T., Luk, M., Neville, H., Rzym, A., and Van Hemert, O. (2022): An Investor's Guide to Crypto, in: *Journal of Portfolio Management*, Vol. 49, 146-171.

Hickman, K., Hunter, H., Byrd, J., Beck, J., and Terpening, W. (2001): Life Cycle Investing, Holding Periods, and Risk, in: *Journal of Portfolio Management*, Vol. 27, 101-111.

Hu, A.S., Parlour, C.A., and Rajan, U. (2019): Cryptocurrencies: Stylized facts on a new investible instrument, in: *Financial Management*, Vol. 48, 1049-1068.

Iglesias, E.M. (2013): The Block-Block Bootstrap for Time Series, in: *Communications in Statistics – Theory and Methods*, Vol. 42, 2584-2600.

Khang, K., Picca, A., Zhang, S., and Zhu, M. (2023): How Inefficient is the 1/N Strategy for a Factor Investor? in: *Journal of Investment Management*, Vol. 21, 103-119.

Kläver, H. (2006): Tests of Stochastic Dominance for Time Series Data: Theory and Empirical Application, Shaker Verlag, Düren.

Korn, O., Möller, P.M., and Schwehm, C. (2022): Drawdown Measures: Are They All the Same? in: *Journal of Portfolio Management*, Vol. 48, 104-120.

Kraus, A., and Litzenberger, R.H. (1976): Skewness Preference and the Valuation of Risk Assets, in: *Journal of Finance*, Vol. 31, 1085-1100.

Künsch, H.R. (1989): The Jackknife and the Bootstrap for General Stationary Observations, in: *The Annals of Statistics*, Vol. 17, 1217-1241.

Lahiri, S.N. (1999): Theoretical Comparisons of Block Bootstrap Methods, in: *The Annals of Statistics*, Vol. 27, 386-404.

Lee, K., and Whang, Y.-J. (2023): PySDTest: A Python Package for Stochastic Dominance Tests, Working Paper.

Linton, O., Maasoumi, E., and Whang, Y.-J. (2005): Consistent Testing for Stochastic Dominance under General Sampling Schemes, in: *Review of Economic Studies*, Vol. 72, 735-765.

Mandelbrot, B. (1961): Stable Paretian Random Functions and the Multiplicative Variation of Income, in: *Econometrica*, Vol. 29, 517-543.

Paparoditis, E., and Politis, D.N. (2001): Tapered Block Bootstrap, in: *Biometrica*, Vol. 88, 1105-1119.

Politis, D.N., and Romano, J.P. (1992): A Circular Block-Resampling Procedure for Stationary Data. In Lepage, R. and Billard, L., eds., *Exploring the Limits of Bootstrap*. New York: Wiley, 263-270.

Politis, D.N., and Romano, J.P. (1994): The Stationary Bootstrap, in: *Journal of the American Statistical Association*, Vol. 89, 1303-1313.

Riley, T., and Yan, Q. (2022): Maximum Drawdown as Predictor of Mutual Fund Performance and Flows, in: *Financial Analysts Journal*, Vol. 78, 59-76.

Rubinstein, M.E. (1973): The Fundamental Theorem of Parameter-Preference Security Valuation, in: *Journal of Financial and Quantitative Analysis*, Vol. 8, 61-69.

Sinha, A.K., and Sun, M.Y. (2005): Asset Choice and Time Diversification Benefits, in: *Journal of Business & Economics Research*, Vol. 3, 23-34.

Sortino, F.A., and van der Meer, R. (1991): Downside Risk, in: *Journal of Portfolio Management*, Vol. 17, 27-31.

Van Hemert, O., Ganz, M., Harvey, C.R., Rattray, S., Sanchez Martin, E., and Yawitch, D. (2020): Drawdowns, in: *Journal of Portfolio Management*, Vol. 46, 34-50.

White, H. (2000): A Reality Check for Data Snooping, in: *Econometrica*, Vol. 68, 1097-1126.

Young, T.W. (1991): Calmar Ratio: A Smoother Tool, in: *Futures Magazine*, Vol. 20, 40.

Exhibit 1: Five largest declines in the MSCI World Index and in bitcoin

Panel A: MSCI World Index

Crisis/Crash	Start date	End date	MDD	MDD date	# of periods
Financial crisis	2007-10-31	2013-04-30	-53.65%	2009-02-27	65
Dotcom bubble	2000-03-31	2006-01-31	-46.31%	2002-09-30	69
Market decline 1973-1974	1973-03-30	1977-12-30	-40.12%	1974-09-30	56
Ukraine crash	2021-12-31	2023-12-29	-25.13%	2022-09-30	23
Market decline 1990	1989-12-29	1993-03-31	-24.01%	1990-09-28	38

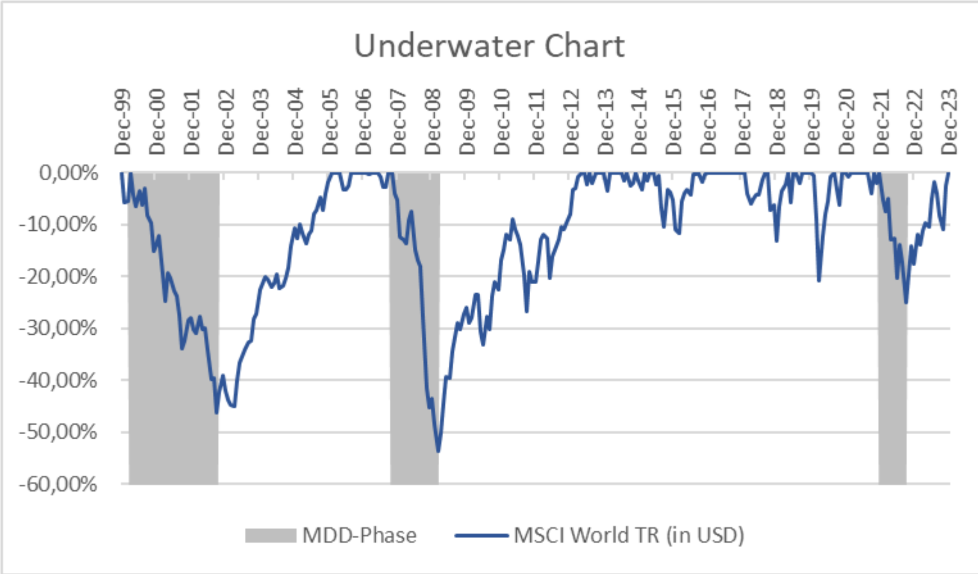
Panel B: Bitcoin

Crisis/Crash	Start date	End date	MDD	MDD date	# of periods
First crash	2013-11-29	2017-02-28	-79.77%	2015-01-30	38
Second crash	2017-12-29	2020-11-30	-76.20%	2019-01-31	34
Third crash	2021-10-29	2023-12-29	-73.35%	2022-12-30	26
Fourth crash	2011-08-31	2012-07-31	-60.13%	2011-11-30	10
Fifth crash	2021-03-31	2021-10-29	-40.40%	2021-06-30	6

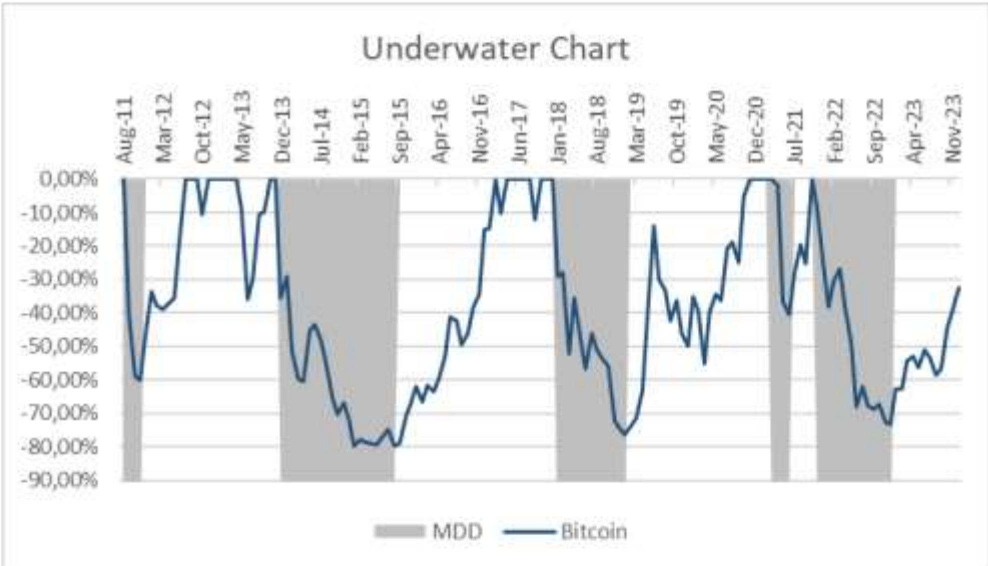
Panel A shows the five largest drawdowns of the MSCI World Index between December 1969 and December 2023, based on 649 monthly index values. 'Start date' marks the beginning of the drawdown and 'End date' marks the date at which the drawdown caught up. 'MDD' is the maximum drawdown during a crisis, and 'MDD date' is the date of occurrence. "# of periods" quantifies the number of underwater months (not counting the 'Start date' or 'End date' months). Panel B lists the five largest drawdowns for Bitcoin from August 2011 to December 2023 (149 monthly prices). All measures are the same as described in the notes to Panel A.

Exhibit 2: Underwater charts of the MSCI World Index and Bitcoin

Panel A: MSCI World Index



Panel B: Bitcoin



Panel A shows the underwater chart of the MSCI World Index from December 1999 to December 2023. Of the five largest drawdowns between December 1969 and December 2023, three occurred during this period: the dot-com bubble burst, the global financial crisis, and the Ukraine crash. All three of these events are marked with a grey bar. Panel B shows the underwater chart of Bitcoin from August 2011 to December 2023. The five largest drawdowns are also marked with a grey bar.

Exhibit 3: Simulation of the maximum drawdown of the MSCI World Index and Bitcoin

Panel A: MSCI World Index

Real MDD: -53.65%							
	Mean	**Q(50)**	**% < real MDD**	**Q(5)**	**Q(95)**	**Shape**	**MSE**
Efron	-39.50%	-38.30%	7.50%	-56.28%	-27.61%	9.88	0.0279
MB 12M	-48.74%	-49.82%	31.00%	-67.35%	-30.42%	1.70	0.0152
MB 24M	-51.07%	-52.61%	38.30%	-69.15%	-32.20%	1.38	0.0117
MB 36M	-51.50%	-53.65%	42.20%	-67.86%	-33.47%	1.42	0.0102
PR 12M	-49.65%	-51.08%	33.70%	-66.96%	-31.48%	1.67	0.0128
PR 24M	-50.42%	-53.65%	36.00%	-64.18%	-32.18%	2.04	0.0095
PR 36M	-50.86%	-53.65%	34.20%	-64.50%	-38.68%	1.38	0.0079

Panel B: Bitcoin

Real MDD: -79.77%							
	Mean	**Q(50)**	**% < real MDD**	**Q(5)**	**Q(95)**	**Shape**	**MSE**
Efron	-75.90%	-76.11%	38.00%	-92.34%	-58.58%	1.69	0.0127
MB 12M	-81.39%	-82.23%	57.80%	-96.18%	-63.85%	0.97	0.0100
MB 24M	-82.50%	-82.18%	60.50%	-93.66%	-70.14%	0.69	0.0066
MB 36M	-80.96%	-79.77%	45.90%	-89.85%	-75.77%	0.40	0.0024
PR 12M	-81.72%	-80.75%	55.30%	-94.37%	-68.31%	0.79	0.0066
PR 24M	-81.42%	-79.77%	47.20%	-92.88%	-73.35%	0.49	0.0041
PR 36M	-80.99%	-79.77%	43.90%	-92.13%	-73.35%	0.52	0.0030

Panel A shows the results of the bootstrap simulations of the maximum drawdown for the MSCI World Index. The population consists of monthly returns from January 1970 to December 2023. During this period, the MSCI World Index experienced a maximum drawdown (MDD) of -53.65%. The analysed bootstrap simulation algorithms are Efron's (1979) bootstrap, the moving block bootstrap (MB), and the Politis and Romano (1994) (PR) bootstrap. The MB and PR bootstrap algorithms are implemented using block lengths of twelve, twenty-four and thirty-six months. Each bootstrap simulation comprises 1,000 runs. 'Mean' and 'Q(50)' provide the mean and median MDDs. The column '% < real MDD' lists the percentage of all simulation runs where the simulated MDD is higher than the true population MDD (i.e., overestimation). 'Q(5)' and 'Q(95)' provide the 5^{th} and 95^{th} quantiles of the 1,000 simulated MDD values. Both statistical measures are necessary to compute the 'Shape' measure, as defined in Section 3.2. The final column, 'MSE', contains the mean squared errors, as defined in Section 3.2. Panel B provides the bootstrap simulation results for Bitcoin. The population consists of monthly Bitcoin returns from September 2011 to December 2023. During this period, Bitcoin exhibits an MDD of -79.77%. The bootstrap simulation methods applied, their parameterisations and all evaluated statistical measures are exactly the same as for the MSCI World Index.

Exhibit 4: Distribution of maximum drawdowns for the MSCI World Index and Bitcoin

Panel A: MSCI World Index

Real MDD: -53.65%									
	Q(10)	**Q(20)**	**Q(30)**	**Q(40)**	**Q(50)**	**Q(60)**	**Q(70)**	**Q(80)**	**Q(90)**
Efron	-51.30%	-45.83%	-42.92%	-40.73%	-38.30%	-36.23%	-34.08%	-31.75%	-29.13%
MB 36M	-62.83%	-57.88%	-54.58%	-53.65%	-53.65%	-51.65%	-48.40%	-44.25%	-40.09%
PR 36M	-59.38%	-54.33%	-53.65%	-53.65%	-53.65%	-51.49%	-46.31%	-45.25%	-40.12%
H_0	$MDD^{MB36} \succcurlyeq_{SD1} MDD^{Efron}$			$MDD^{PR36} \succcurlyeq_{SD1} MDD^{Efron}$			$MDD^{PR36} \succcurlyeq_{SD1} MDD^{MB36}$		
p-value	0.12			1.00			0.00		

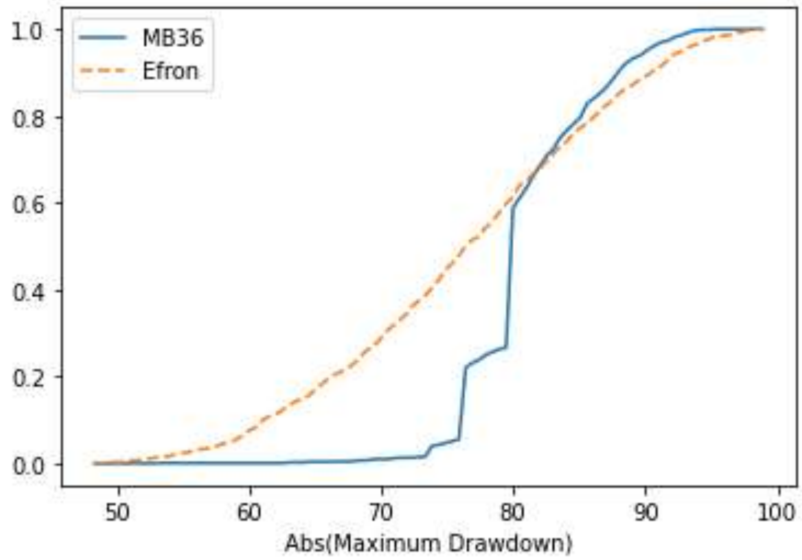
Panel B: Bitcoin

Real MDD: -79.77%									
	Q(10)	**Q(20)**	**Q(30)**	**Q(40)**	**Q(50)**	**Q(60)**	**Q(70)**	**Q(80)**	**Q(90)**
Efron	-90.12%	-86.02%	-82.25%	-79.12%	-76.11%	-73.26%	-70.09%	-66.02%	-60.96%
MB 36M	-87.90%	-85.18%	-82.47%	-80.19%	-79.77%	-79.77%	-79.77%	-76.20%	-76.20%
PR 36M	-88.78%	-85.35%	-82.15%	-79.77%	-79.77%	-79.77%	-79.77%	-76.20%	-76.20%
H_0	$MDD^{MB36} \succcurlyeq_{SD1} MDD^{Efron}$			$MDD^{PR36} \succcurlyeq_{SD1} MDD^{Efron}$			$MDD^{PR36} \succcurlyeq_{SD1} MDD^{MB36}$		
p-value	0.00			0.18			0.16		

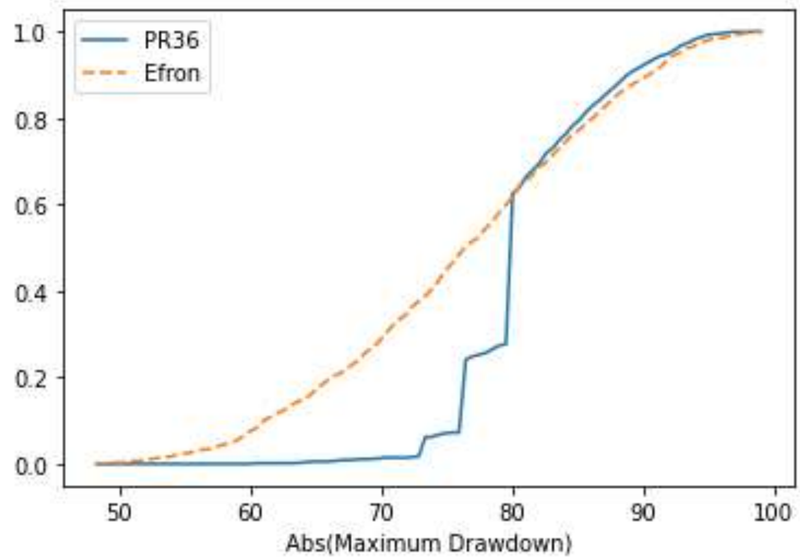
Panel A shows the distribution of maximum drawdowns based on the bootstrap simulations for the MSCI World Index. The population comprises monthly returns from January 1970 to December 2023. During this period, the MSCI World Index experienced a maximum drawdown (MDD) of -53.65%. The analysed bootstrap simulation algorithms are Efron's (1979) bootstrap, the moving block bootstrap (MB), and the Politis and Romano (1994) (PR) bootstrap. The MB and PR bootstrap algorithms are implemented using block lengths of thirty-six months. These are the least biased of all three implemented block sizes (12, 24, and 36 monthly returns) (see equation (1) in section 3.2). Each bootstrap simulation comprises 1,000 runs. 'Q(10)' provides the 10% quantile of the 1,000 simulated MDD values, 'Q(20)' provides the 20% quantile, and so on. The table also contains the results of stochastic dominance tests of order one. We test three hypotheses: The first hypothesis is that the MDD distribution resulting from the MB36 method stochastically dominates the distribution resulting from the Efron approach. The second hypothesis is that the MDD distribution resulting from the PR36 method stochastically dominates the distribution resulting from the Efron approach. The third hypothesis is that the MDD distribution resulting from the PR36 method stochastically dominates the distribution resulting from the MB36 approach. The applied hypothesis test is described in Linton et al. (2005) and incorporates improvements by Kläver (2006) (see section 3.2 for details). The *p*-value provides the probability of rejecting the null hypothesis of stochastic dominance despite the null hypothesis being correct. Panel B shows the bootstrap simulation results for Bitcoin. The population consists of monthly returns from September 2011 to December 2023. During this period, Bitcoin exhibited a maximum drawdown (MDD) of -79.77%. The bootstrap simulation methods applied, their parameterisations and all evaluated statistical measures are exactly the same as for the MSCI World Index.

Exhibit 5: Cumulative distributions of the maximum drawdown for Bitcoin

Panel A: SD1 test $MDD^{MB36} \succcurlyeq_{S1} MDD^{Efron}$



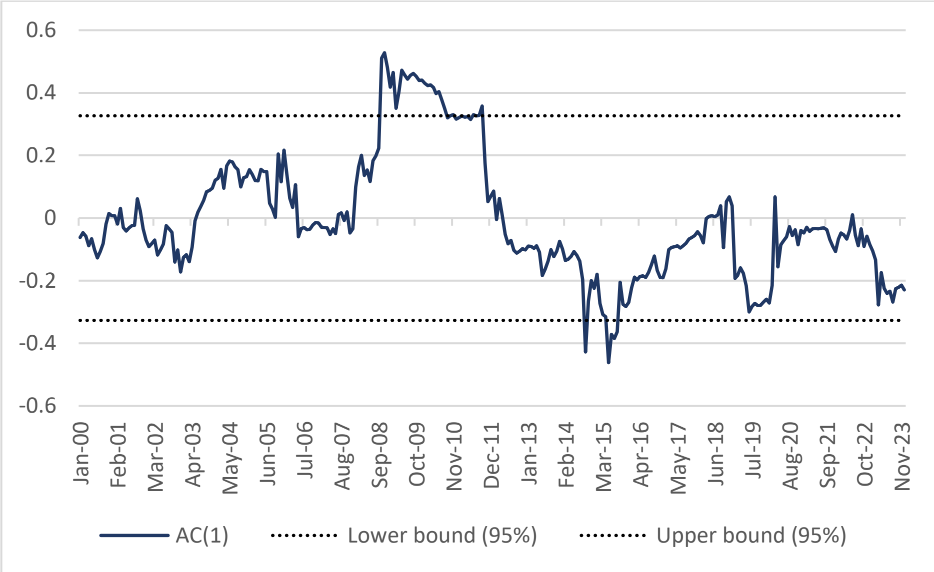
Panel B: SD1 test $MDD^{PR36} \succcurlyeq_{S1} MDD^{Efron}$



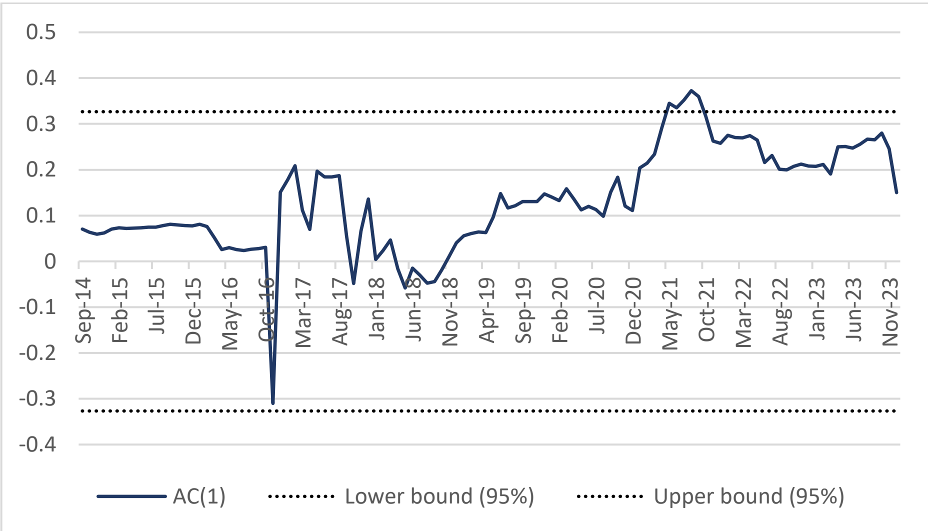
Panel A shows the cumulative distribution of 1,000 simulated maximum drawdown (MDD) values for Bitcoin, resulting from: the moving block bootstrap approach with a block length of thirty-six months; and Efron's (1979) bootstrap method. The x-axis shows the absolute values of MDDs, and the y-axis shows the corresponding cumulative probabilities. The population underlying the simulations is based on monthly Bitcoin returns from September 2011 to December 2023. Similarly, Panel B illustrates the cumulative distribution of the 1,000 simulated MDD values for Bitcoin resulting from: the Politis and Romano (1994) bootstrap approach, applied with a block length of thirty-six months; and Efron's bootstrap method.

Exhibit 6: Thirty-six-month first-order rolling autocorrelations

Panel A: MSCI World Index



Panel B: Bitcoin



Panel A shows the rolling first-order autocorrelations (AC1) of the MSCI World Index, calculated using a rolling window of thirty-six monthly returns. These autocorrelations cover the period from January 2000 to December 2023. The first autocorrelation, for January 2000, is computed using the thirty-six monthly MSCI World Index returns ranging from January 1997 to January 2000. The second autocorrelation, for February 2000, is computed using the thirty-six monthly MSCI World Index returns ranging from February 1997 to February 2000. This process continues for each subsequent autocorrelation. The figure also shows the 5% and 95% confidence bounds. Panel B illustrates the rolling first-order autocorrelations for Bitcoin, computed over a rolling window of thirty-six monthly discrete returns. These autocorrelations cover the period from September 2014 to December 2023. The computation of the rolling first-order autocorrelation is identical to that used for the MSCI World Index.

Exhibit 7: Simulation of the maximum drawdown of the MSCI World Index (based on a truncated sample where the MDD occurs at the sample end)

Panel A: Statistics

Real MDD: -53.65%							
	Mean	**Q(50)**	**% < real MDD**	**Q(5)**	**Q(95)**	**Shape**	**MSE**
Efron	-39.22%	-38.14%	7.30%	-56.16%	-25.67%	11.14	0.0300
MB 12M	-44.21%	-43.90%	19.80%	-64.54%	-25.77%	2.56	0.0229
MB 24M	-43.70%	-42.21%	19.50%	-63.59%	-25.65%	2.82	0.0218
MB 36M	-44.68%	-44.58%	16.80%	-62.62%	-26.97%	2.97	0.0180
PR 12M	-54.90%	-56.37%	57.50%	-77.81%	-32.42%	0.88	0.0195
PR 24M	-56.02%	-59.52%	61.30%	-77.27%	-31.68%	0.93	0.0191
PR 36M	-55.92%	-60.70%	61.50%	-77.16%	-32.40%	0.90	0.0184

Panel B: Distributions

Real MDD: -53.65%									
	Q(10)	**Q(20)**	**Q(30)**	**Q(40)**	**Q(50)**	**Q(60)**	**Q(70)**	**Q(80)**	**Q(90)**
Efron	-51.98%	-47.44%	-43.51%	-40.78%	-38.14%	-35.82%	-33.36%	-30.78%	-27.65%
MB 36M	-58.14%	-52.34%	-48.64%	-46.31%	-44.58%	-41.06%	-40.12%	-38.80%	-32.63%
PR 12M	-72.28%	-66.23%	-62.28%	-59.99%	-56.37%	-52.47%	-46.31%	-40.73%	-35.50%
H_0	$MDD^{MB36} \succcurlyeq_{SD1} MDD^{Efron}$			$MDD^{PR12} \succcurlyeq_{SD1} MDD^{Efron}$			$MDD^{PR12} \succcurlyeq_{SD1} MDD^{MB36}$		
p-value	0.87			1.00			1.00		

Panel A shows the results of the bootstrap simulations of the maximum drawdown for the MSCI World Index. The population comprises monthly returns from January 1970 to February 2009, when the drawdown reached its peak during the global financial crisis. During this period, the MSCI World Index exhibited a maximum drawdown (MDD) of -53.65%. The analysed bootstrap simulation algorithms are Efron's (1979) bootstrap, the moving block bootstrap (MB), and the Politis and Romano (1994) (PR) bootstrap. The MB and PR bootstrap algorithms are implemented using block lengths of twelve, twenty-four and thirty-six months. Each bootstrap simulation comprises 1,000 runs. 'Mean' and 'Q(50)' provide the mean and median MDDs. The column '% < real MDD' lists the percentage of simulation runs where the simulated MDD is higher than the true population MDD (i.e., overestimation). 'Q(5)' and 'Q(95)' provide the 5th and 95th quantiles of the 1,000 simulated MDD values. Both statistical measures are necessary to compute the 'Shape' measure, as defined in section 3.2. The final column, 'MSE', contains the mean squared errors, as defined in section 3.2. Panel B shows the quantiles of the simulated MDD values resulting from the Efron bootstrap, the moving block bootstrap with a block length of thirty-six months (MB36), and the PR bootstrap with a block length of twelve months (PR12). MB36 and PR12 are the least biased of the three implemented block sizes (12, 24 and 36 monthly returns) (see equation (1) in Section 3.2). 'Q(10)' provides the 10% quantile of the 1,000 simulated MDD values, 'Q(20)' provides the 20% quantile, and so on. The table also contains the results of stochastic dominance tests of order one. We test three hypotheses: The first hypothesis is that the MDD distribution resulting from the MB36 method stochastically dominates the distribution resulting from the Efron approach. The second hypothesis is that the MDD distribution resulting from the PR12 method stochastically dominates the distribution resulting from the Efron approach. The third hypothesis is that the MDD distribution resulting from the PR12 method stochastically dominates the distribution resulting from the MB36 approach. The applied hypothesis test is described in Linton et al. (2005) and incorporates improvements by Kläver (2006) (see section 3.2 for details). The *p*-value provides the probability of rejecting the null hypothesis of stochastic dominance despite the null hypothesis being correct.

Exhibit 8: Simulation of the maximum drawdown of the Nasdaq 100 Index (based on a truncated sample where the MDD occurs at the sample end)

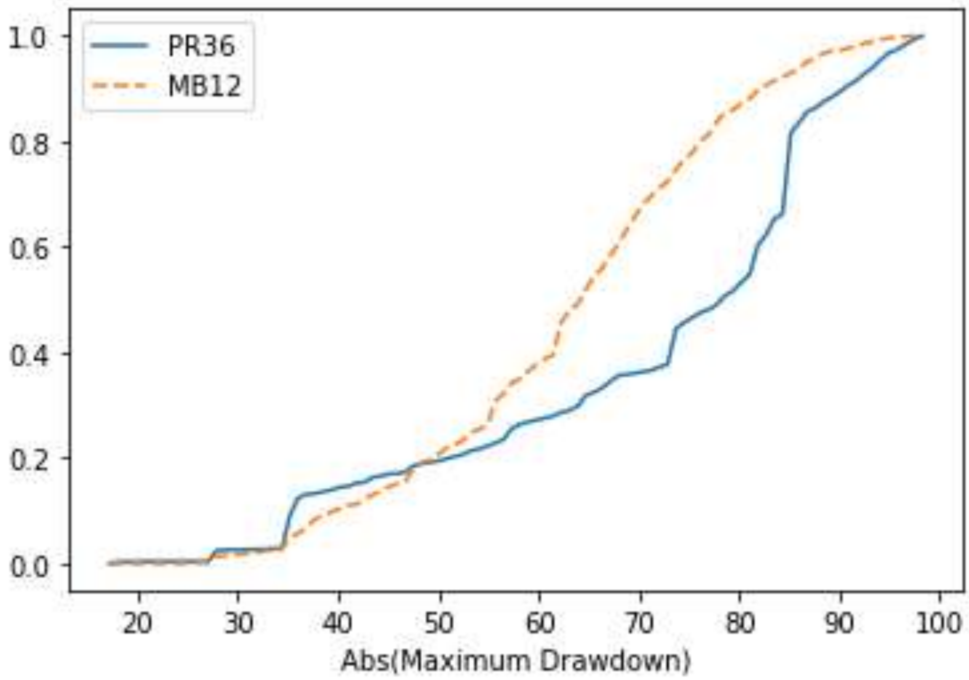
Panel A: Statistics

Real MDD: -81.07%							
	Mean	**Q(50)**	**% < real MDD**	**Q(5)**	**Q(95)**	**Shape**	**MSE**
Efron	-57.49%	-55.84%	4.80%	-80.35%	-38.83%	-58.85	0.0717
MB 12M	-62.88%	-63.98%	12.40%	-87.13%	-35.19%	7.57	0.0568
MB 24M	-60.48%	-65.37%	13.50%	-90.00%	-31.67%	5.53	0.0828
MB 36M	-54.64%	-56.61%	5.30%	-81.15%	-27.64%	655.68	0.1076
PR 12M	-67.08%	-69.32%	25.10%	-92.09%	-35.42%	4.14	0.0503
PR 24M	-69.42%	-73.58%	36.90%	-92.34%	-34.57%	4.13	0.0482
PR 36M	-70.40%	-78.08%	44.30%	-93.81%	-34.57%	3.65	0.0498

Panel B: Distributions

Real MDD: -81.07%									
	Q(10)	**Q(20)**	**Q(30)**	**Q(40)**	**Q(50)**	**Q(60)**	**Q(70)**	**Q(80)**	**Q(90)**
Efron	-75.44%	-68.77%	-63.69%	-59.57%	-55.84%	-52.81%	-49.67%	-46.46%	-42.18%
MB 12M	-82,64%	-76,28%	-71,56%	-67,50%	-63,98%	-61,42%	-55,15%	-49,15%	-39,94%
PR 36M	-90.49%	-84.64%	-84.42%	-81.53%	-78.08%	-73.43%	-64.23%	-50.91%	-35.42%
H_0	$MDD^{MB12} \ngeq_{SD1} MDD^{Efron}$			$MDD^{PR36} \ngeq_{SD1} MDD^{Efron}$			$MDD^{PR36} \ngeq_{SD1} MDD^{MB12}$		
p-value	0.00			0.00			0.00		

Panel C: SD1 test $MDD^{PR36} \ngeq_{SD1} MDD^{MB12}$



Panel A shows the results of the bootstrap simulations of the maximum drawdown for the Nasdaq100 Index. The population comprises monthly returns from February 1983 to September 2002, when the drawdown reached its peak at the bursting of the dotcom bubble. During this period, the Nasdaq 100 Index exhibited a maximum drawdown (MDD) of -81.07%. The analysed bootstrap simulation algorithms are Efron's (1979) bootstrap, the moving block bootstrap (MB), and the Politis and Romano (1994) (PR) bootstrap. The MB and PR bootstrap algorithms are implemented using block lengths of twelve, twenty-four and thirty-six months. Each bootstrap simulation comprises 1,000 runs. ‘Mean’ and ‘Q(50)’ provide the mean and median MDDs. The column ‘% < real MDD’ lists the percentage of simulation runs where the simulated MDD is higher than the true population MDD (i.e., overestimation). ‘Q(5)’ and ‘Q(95)’ provide the 5th and 95th quantiles of the 1,000 simulated MDD values. Both statistical measures are necessary to compute the ‘Shape’ measure, as defined in Section 3.2. The final column, ‘MSE’, contains the mean squared errors, as defined in Section 3.2. Panel B shows the quantiles of the simulated MDD values resulting from the Efron bootstrap, the moving block bootstrap with a block length of twelve months (MB12), and the PR bootstrap with a block length of thirty-six months (PR36). MB12 and PR36 are the least biased of the three implemented block sizes (12, 24 and 36 monthly returns) (see equation (1) in section 3.2). ‘Q(10)’ provides the 10% quantile of the 1,000 simulated MDD values, ‘Q(20)’ provides the 20% quantile, and so on. The table also contains the results of stochastic dominance tests of order one. We test three hypotheses: The first hypothesis is that the MDD distribution resulting from the MB12 method stochastically dominates the distribution resulting from the Efron approach. The second hypothesis is that the MDD distribution resulting from the PR36 method stochastically dominates the distribution resulting from the Efron approach. The third hypothesis is that the MDD distribution resulting from the PR36 method stochastically dominates the distribution resulting from the MB12 approach. The applied hypothesis test is described in Linton et al. (2005) and incorporates improvements by Kläver (2006) (see section 3.2 for details). The *p*-value provides the probability of rejecting the null hypothesis of stochastic dominance despite the null hypothesis being correct. Panel C shows the cumulative distribution of 1,000 simulated maximum drawdown (MDD) values for the Nasdaq 100 Index, resulting from: the Politis and Romano (1994) bootstrap approach, applied with a block length of thirty-six months; and the moving block bootstrap method, applied with a block length of twelve months. The x-axis shows the absolute values of MDDs, and the y-axis shows the corresponding cumulative probabilities. The population underlying the simulations is based on monthly Nasdaq 100 Index returns from February 1983 to September 2002.

Exhibit 9: Simulation of the maximum drawdown of the Nasdaq 100 Index (based on a truncated sample where the MDD occurs at the sample start)

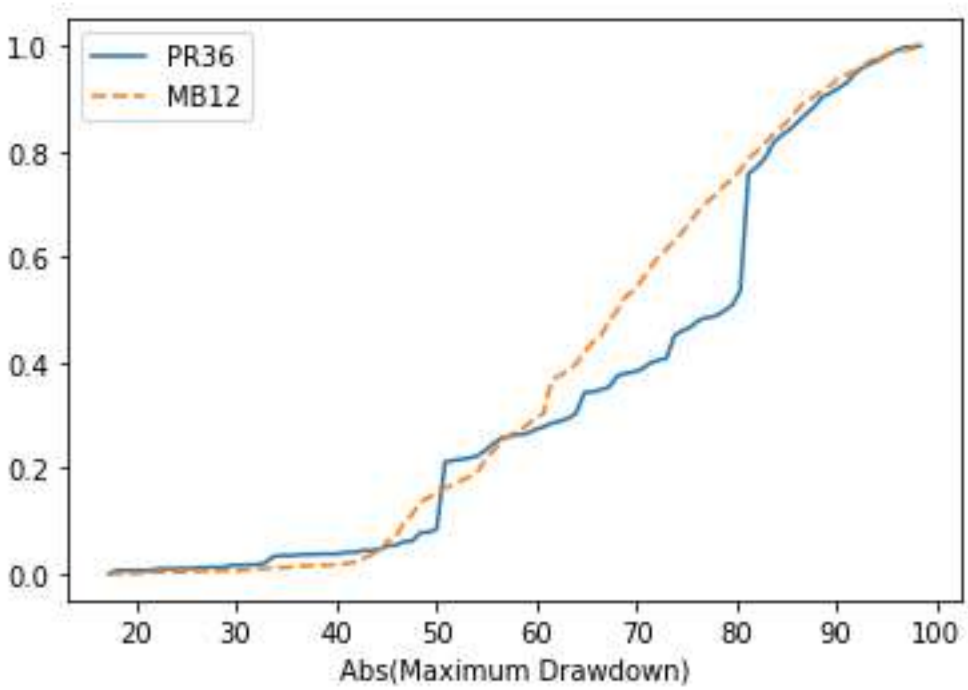
Panel A: Statistics

Real MDD: -81.07%							
	Mean	**Q(50)**	**% < real MDD**	**Q(5)**	**Q(95)**	**Shape**	**MSE**
Efron	-57.99%	-57.12%	4.40%	-80.13%	-37.84%	-45.84	0.0705
MB 12M	-67.71%	-67.99%	21.30%	-91.43%	-45.04%	3.48	0.0399
MB 24M	-64.60%	-67.89%	18.20%	-90.57%	-34.80%	4.87	0.0569
MB 36M	-60.83%	-55.87%	18.10%	-88.37%	-32.97%	6.59	0.0742
PR 12M	-68.01%	-69.63%	24.50%	-91.33%	-42.70%	3.74	0.0417
PR 24M	-69.66%	-73.43%	34.30%	-92.53%	-44.07%	3.23	0.0404
PR 36M	-70.91%	-78.48%	39.90%	-92.30%	-45.05%	3.21	0.0379

Panel B: Distributions

Real MDD: -81.07%									
	Q(10)	**Q(20)**	**Q(30)**	**Q(40)**	**Q(50)**	**Q(60)**	**Q(70)**	**Q(80)**	**Q(90)**
Efron	-76.31%	-69.77%	-65.11%	-60.92%	-57.12%	-53.52%	-50.04%	-45.75%	-41.55%
MB 12M	-87,04%	-81,77%	-76,62%	-71,93%	-67,99%	-63,88%	-60,62%	-54,79%	-46,93%
PR 36M	-88.45%	-83.28%	-81.07%	-81.07%	-78.48%	-71.35%	-63.66%	-50.11%	-50.11%
H_0	$MDD^{MB12} \succcurlyeq_{SD1} MDD^{Efron}$			$MDD^{PR36} \succcurlyeq_{SD1} MDD^{Efron}$			$MDD^{PR36} \succcurlyeq_{SD1} MDD^{MB12}$		
p-value	0.00			0.11			0.00		

Panel C: SD1 test $MDD^{PR36} \succcurlyeq_{SD1} MDD^{MB12}$



Panel A shows the results of the bootstrap simulations of the maximum drawdown for the Nasdaq100 Index. The population comprises monthly returns from January 2000 to December 2023. The drawdown reached its peak at the beginning of the truncated sample, when the dotcom bubble burst. During this period, the Nasdaq 100 Index exhibited a maximum drawdown (MDD) of -81.07%. The analysed bootstrap simulation algorithms are Efron's (1979) bootstrap, the moving block bootstrap (MB), and the Politis and Romano (1994) (PR) bootstrap. The MB and PR bootstrap algorithms are implemented using block lengths of twelve, twenty-four and thirty-six months.

Each bootstrap simulation comprises 1,000 runs. ‘Mean’ and ‘Q(50)’ provide the mean and median MDDs. The column ‘% < real MDD’ lists the percentage of simulation runs where the simulated MDD is higher than the true population MDD (i.e., overestimation). ‘Q(5)’ and ‘Q(95)’ provide the 5th and 95th quantiles of the 1,000 simulated MDD values. Both statistical measures are necessary to compute the ‘Shape’ measure, as defined in Section 3.2. The final column, ‘MSE’, contains the mean squared errors, as defined in Section 3.2. Panel B shows the quantiles of the simulated MDD values resulting from the Efron bootstrap, the moving block bootstrap with a block length of twelve months (MB12), and the PR bootstrap with a block length of thirty-six months (PR36). MB12 and PR36 are the least biased of the three implemented block sizes (12, 24 and 36 monthly returns) (see equation (1) in section 3.2). ‘Q(10)’ provides the 10% quantile of the 1,000 simulated MDD values, ‘Q(20)’ provides the 20% quantile, and so on. The table also contains the results of stochastic dominance tests of order one. We test three hypotheses: The first hypothesis is that the MDD distribution resulting from the MB12 method stochastically dominates the distribution resulting from the Efron approach. The second hypothesis is that the MDD distribution resulting from the PR36 method stochastically dominates the distribution resulting from the Efron approach. The third hypothesis is that the MDD distribution resulting from the PR36 method stochastically dominates the distribution resulting from the MB12 approach. The applied hypothesis test is described in Linton et al. (2005) and incorporates improvements by Kläver (2006) (see section 3.2 for details). The *p*-value provides the probability of rejecting the null hypothesis of stochastic dominance despite the null hypothesis being correct. Panel C shows the cumulative distribution of 1,000 simulated maximum drawdown (MDD) values for the Nasdaq 100 Index, resulting from: the Politis and Romano (1994) bootstrap approach, applied with a block length of thirty-six months; and the moving block bootstrap method, applied with a block length of twelve months. The x-axis shows the absolute values of MDDs, and the y-axis shows the corresponding cumulative probabilities. The population underlying the simulations is based on monthly Nasdaq 100 Index returns from January 2000 to December 2023.

Exhibit 10: Simulation results using the block-block bootstrap

Panel A: MSCI World Index, 1970-01 to 2023-12

Real MDD: -53.65%								
	Popula-tion MDD	**Mean**	**Q(50)**	**% < real MDD**	**Q(5)**	**Q(95)**	**Shape**	**MSE**
BBB 12M	-49.38%	-45.97%	-45.75%	24.10%	-65.60%	-26.89%	1.39	0.0199
BBB 24M	-51.88%	-49.41%	-50.68%	34.80%	-67.65%	-29.06%	1.33	0.0142
BBB 36M	-42.79%	-49.98%	-53.25%	36.40%	-68.06%	-29.51%	0.53	0.0129
MB 36M	-53.65%	-51.50%	-53.65%	42.20%	-67.86%	-33.47%	1.42	0.0102
H_0	$MDD^{MB36} \succcurlyeq_{SD1} MDD^{BBB36}$				$MDD^{BBB36} \succcurlyeq_{SD1} MDD^{MB36}$			
p-value	0.22				0.00			

Panel B: MSCI World Index, 1970-01 to 2009-02

Real MDD: -53.65%								
	Popula-tion MDD	**Mean**	**Q(50)**	**% < real MDD**	**Q(5)**	**Q(95)**	**Shape**	**MSE**
BBB 12M	-41.58%	-39.26%	-38.43%	9.30%	-57.49%	-24.64%	1.07	0.0317
BBB 24M	-45.25%	-40.45%	-39.70%	10.70%	-59.82%	-24.87%	1.40	0.0283
BBB 36M	-46.45%	-41.49%	-40.48%	11.50%	-59.97%	-24.01%	1.66	0.0252
PR 12M	-53.56%	-54.31%	-56.27%	56.60%	-75.99%	-30.73%	1.03	0.0198
H_0	$MDD^{PR12} \succcurlyeq_{SD1} MDD^{BBB36}$				$MDD^{BBB36} \succcurlyeq_{SD1} MDD^{PR12}$			
p-value	1.00				0.00			

Panel A shows the results of the block-block bootstrap simulations of the maximum drawdown for the MSCI World Index. The population comprises monthly returns from January 1970 to December 2023. During this full sample period, the MSCI World Index experienced a maximum drawdown (MDD) of -53.65%. The analysed bootstrap simulation algorithms are the block-block bootstrap (BBB) and the moving block bootstrap with a block size of thirty-six months (MB36). MB36 is the least biased of all the analysed bootstrap methods within this simulation setup (see Exhibit 3). The BBB method is implemented using block lengths of twelve, twenty-four and thirty-six months. Each bootstrap simulation comprises 1,000 runs. 'Mean' and 'Q(50)' provide the mean and median MDDs. The column '% < real MDD' lists the percentage of simulation runs where the simulated MDD is higher than the true population MDD (i.e., overestimation). 'Q(5)' and 'Q(95)' provide the 5th and 95th quantiles of the 1,000 simulated MDD values. Both statistical measures are necessary to compute 'Shape' measure, as defined in Section 3.2. The final column, 'MSE', contains the mean squared errors, as defined in Section 3.2. Panel A also contains the results of stochastic dominance tests of order one. We test two hypotheses: The first hypothesis is that the MDD distribution resulting from the MB36 method stochastically dominates the distribution resulting from block-block bootstrap with a block size of 36 months. The second hypothesis is the reverse. The applied hypothesis test is described in Linton et al. (2005) and incorporates improvements by Kläver (2006) (see section 3.2 for details). The *p*-value provides the probability of rejecting the null hypothesis of stochastic dominance despite the null hypothesis being correct. Panel B shows the results of the bootstrap simulations for the MSCI World Index using a truncated sample from January 1970 to February 2009, when the drawdowns reached the peak during the global financial crisis. The bootstrap simulation methods applied, their parameterisations, and all evaluated statistical measures are the same as those explained in Panel A. The benchmark bootstrap method is Politis and Romano with a block size of twelve months (PR12), the least biased of the implemented block sizes within this simulation setup (see Exhibit 7).

Exhibit 11: Simulation results using the tapered block bootstrap

Panel A: MSCI World Index, 1970-01 to 2023-12

Real MDD: -53.65%							
	Mean	**Q(50)**	**% < real MDD**	**Q(5)**	**Q(95)**	**Shape**	**MSE**
TBB 12M	-47.79%	-47.73%	32.90%	-69.12%	-29.91%	1.53	0.0182
TBB 24M	-51.22%	-51.65%	46.90%	-71.58%	-30.93%	1.27	0.0168
TBB 36M	-53.01%	-52.29%	48.10%	-71.55%	-31.71%	1.23	0.0153
MB 36M	-51.50%	-53.65%	42.20%	-67.86%	-33.47%	1.42	0.0102
H_0	$MDD^{MB36} \succcurlyeq_{SD1} MDD^{TBB36}$			$MDD^{TBB36} \succcurlyeq_{SD1} MDD^{MB36}$			
p-value	0.00			0.00			

Panel B: MSCI World Index, 1970-01 to 2009-02

Real MDD: -53.65%							
	Mean	**Q(50)**	**% < real MDD**	**Q(5)**	**Q(95)**	**Shape**	**MSE**
TBB 12M	-41.63%	-39.73%	15.10%	-62.19%	-25.10%	3.34	0.0276
TBB 24M	-41.87%	-41.18%	12.60%	-61.22%	-27.94%	3.40	0.0247
TBB 36M	-43.81%	-43.40%	14.30%	-63.13%	-25.05%	3.02	0.0214
PR 12M	-54.31%	-56.27%	56.60%	-75.99%	-30.73%	1.03	0.0198
H_0	$MDD^{PR12} \succcurlyeq_{SD1} MDD^{TBB36}$			$MDD^{TBB36} \succcurlyeq_{SD1} MDD^{PR12}$			
p-value	1.00			0.00			

Panel A shows the results of the tapered block bootstrap simulations of the maximum drawdown for the MSCI World Index. The population comprises monthly returns from January 1970 to December 2023. During this full sample period, the MSCI World Index experienced a maximum drawdown (MDD) of -53.65%. The analysed bootstrap simulation algorithms are the tapered block bootstrap (TBB) and the moving block bootstrap with a block size of thirty-six months (MB36). MB36 is the least biased of all the analysed bootstrap methods within this simulation setup (see Exhibit 3). The TBB method is implemented using block lengths of twelve, twenty-four and thirty-six months. Each bootstrap simulation comprises 1,000 runs. 'Mean' and 'Q(50)' provide the mean and median MDDs. The column '% < real MDD' lists the percentage of simulation runs where the simulated MDD is higher than the true population MDD (i.e., overestimation). 'Q(5)' and 'Q(95)' provide the 5th and 95th quantiles of the 1,000 simulated MDD values. Both statistical measures are necessary to compute 'Shape' measure, as defined in Section 3.2. The final column, 'MSE', contains the mean squared errors, as defined in Section 3.2. Panel A also contains the results of stochastic dominance tests of order one. We test two hypotheses: The first hypothesis is that the MDD distribution resulting from the MB36 method stochastically dominates the distribution resulting from tapered block bootstrap with a block size of 36 months. The second hypothesis is the reverse. The applied hypothesis test is described in Linton et al. (2005) and incorporates improvements by Kläver (2006) (see section 3.2 for details). The *p*-value provides the probability of rejecting the null hypothesis of stochastic dominance despite the null hypothesis being correct. Panel B shows the results of the bootstrap simulations for the MSCI World Index using a truncated sample from January 1970 to February 2009, when the drawdowns reached the peak during the global financial crisis. The bootstrap simulation methods applied, their parameterisations, and all evaluated statistical measures are the same as those explained in Panel A. The benchmark bootstrap method is Politis and Romano with a block size of twelve months (PR12), the least biased of the implemented block sizes within this simulation setup (see Exhibit 7).

Exhibit 12: Simulation of the average drawdown

Panel A: MSCI World Index

Real ADD: -7.96%							
	Mean	**Q(50)**	**% < real ADD**	**Q(5)**	**Q(95)**	**Shape**	**MSE**
Efron	-8.71%	-8.02%	50.50%	-14.39%	-4.98%	0.46	0.0011
MB 12M	-9.26%	-8.10%	51.80%	-17.91%	-4.15%	0.38	0.0021
MB 24M	-8.87%	-7.99%	50.30%	-16.68%	-4.13%	0.44	0.0018
MB 36M	-8.62%	-7.92%	49.70%	-15.57%	-3.92%	0.53	0.0015
PR 12M	-9.16%	-8.38%	54.40%	-16.72%	-4.28%	0.42	0.0021
PR 24M	-8.48%	-8.05%	51.50%	-14.41%	-4.22%	0.58	0.0011
PR 36M	-8.37%	-7.83%	48.30%	-14.21%	-4.36%	0.58	0.0010
H_0	$MDD^{MB36} \succcurlyeq_{SD1} MDD^{Efron}$			$MDD^{PR36} \succcurlyeq_{SD1} MDD^{Efron}$		$MDD^{PR36} \succcurlyeq_{SD1} MDD^{MB36}$	
p-value	0.00			0.00		0.17	

Panel B: Bitcoin

Real ADD: -38.18%							
	Mean	**Q(50)**	**% < real ADD**	**Q(5)**	**Q(95)**	**Shape**	**MSE**
Efron	-25.28%	-23.55%	11.00%	-44.31%	-13.15%	4.09	0.0260
MB 12M	-32.32%	-31.01%	25.40%	-56.27%	-15.89%	1.23	0.0173
MB 24M	-35.87%	-34.97%	38.10%	-55.22%	-18.90%	1.13	0.0125
MB 36M	-35.49%	-35.28%	33.50%	-47.05%	-24.52%	1.54	0.0057
PR 12M	-34.22%	-33.16%	31.80%	-52.99%	-19.37%	1.27	0.0125
PR 24M	-35.84%	-35.13%	37.80%	-51.14%	-23.09%	1.16	0.0078
PR 36M	-36.20%	-35.86%	37.20%	-49.08%	-23.93%	1.31	0.0061
H_0	$MDD^{MB24} \succcurlyeq_{SD1} MDD^{Efron}$			$MDD^{PR36} \succcurlyeq_{SD1} MDD^{Efron}$		$MDD^{PR36} \succcurlyeq_{SD1} MDD^{MB24}$	
p-value	0.36			0.28		0.00	

Panel A shows the results of the bootstrap simulations of the average drawdown for the MSCI World Index. The population consists of monthly returns from January 1970 to December 2023. During this period, the MSCI World Index exhibits an average drawdown (ADD) of -7.96%. The analysed bootstrap simulation algorithms are Efron's (1979) bootstrap, the moving block bootstrap (MB), and the Politis and Romano (1994) (PR) bootstrap. The MB and PR bootstrap algorithms were implemented using block lengths of twelve, twenty-four and thirty-six months. Each simulation comprises 1,000 runs. 'Mean' and 'Q(50)' provide the mean and median ADD. The column '% < real ADD' lists the percentage of all simulation runs where the simulated ADD is higher than the true population ADD (i.e., overestimation). 'Q(5)' and 'Q(95)' provide the 5^{th} and 95^{th} quantiles of the 1,000 simulated ADD values. Both statistical measures are necessary to compute 'Shape' measure, as defined in Section 3.2. The final column, 'MSE', contains the mean squared errors, as defined in Section 3.2. The table also contains the results of stochastic dominance tests of order one. MB36 and PR36 are the least biased versions of all three implemented block sizes (12, 24 and 36 monthly returns) (see equation (1) in Section 3.2). We test three hypotheses: The first hypothesis is that the MDD distribution resulting from the MB36 method stochastically dominates the distribution resulting from the Efron approach. The second hypothesis is that the MDD distribution resulting from the PR36 method stochastically dominates the distribution resulting from the Efron approach. The third hypothesis is that the MDD distribution resulting from the PR36 method stochastically dominates the distribution resulting from the MB36 approach. The applied hypothesis test is described in Linton et al. (2005) and incorporates improvements provided by Kläver (2006). The *p*-value provides the probability of rejecting the null hypothesis of stochastic dominance despite the null hypothesis being correct. Panel B shows the results of the bootstrap simulations for Bitcoin. The population consists of monthly returns from September 2011 to December 2023. During this period, Bitcoin exhibited an average drawdown (ADD) of -38.18%. The bootstrap simulation methods applied, their parameterisations, and all evaluated statistical measures are the same as those explained in Panel A.