RESEARCH

Swiss Finance Institute Research Paper Series N°23-61

Which is Worse: Heavy Tails or Volatility Clusters?

Joshua Traut
University of St. Gallen

Wolfgang Schadner
University of St. Gallen

Which is Worse: Heavy Tails or Volatility Clusters?

Joshua Traut*‡ Wolfgang Schadner†‡

August 11, 2023

Abstract

Heavy tails and volatility clusters are both stylized facts of financial returns that destabilize markets and are often neglected using the simplifying assumptions of normally distributed and iid returns respectively. This work disentangles the two sources and is the first to assess which one does the greater damage to financial stability and whether the threat can be reduced via diversification. As such, it also quantifies the potential shortfalls of the two commonly used simplifying assumptions. The analysis is carried out for index return series representing seven different asset classes and for individual stock portfolios. The stylized facts are isolated using recent developments in surrogate analysis (IAAFT, IAAWT). Our analysis shows that volatility clusters have a greater impact on maximum drawdowns and aggregate losses across all markets and that diversification does not yield any protection from those risks. In fact, diversification amplifies the translation of the two stylized facts into drawdowns, exacerbating their potential negative effects. We further demonstrate the practical relevance of our findings as we can replicate the results of our surrogate analysis using real portfolios. Moreover, we show that regulators should consider the impact of volatility clusters and discard the simplifying assumption of iid returns in order to enhance the accuracy of capital buffers.

Keywords: Financial Stability, Tail Risk, Autocorrelation, Volatility Clustering, Heavy Tails, Risk Management

JEL Classification: G12, G18, G15, G01.

*University of St. Gallen, Swiss Institute of Banking and Finance, Unterer Graben 21, 9000 St. Gallen, Switzerland | joshua.traut@unisg.ch | https://www.joshuatraut.com

†University of St. Gallen, Swiss Institute of Banking and Finance, Unterer Graben 21, 9000 St. Gallen, Switzerland | wolfgang.schadner@unisg.ch | https://www.w-schadner.com

‡We are grateful for insightful comments from Zeno Adams, Andrea Barbon, Daniel Batista Da Silva (discussant), Thea Kolasa (discussant), Philipp Lentner (discussant), Volodymyr Novykov (discussant), Vitaly Orlov, Angelo Ranaldo, Kevin Schneider, Paul Schneider, Paul Söderlind, Sascha Steffen, and Andrea Vedolin, as well as participants of the EFMA Annual Meeting 2023, FMA European Conference 2023, SFI Research Days 2023, World Finance Conference, and Finance Research Seminar as well as the Research Workshop at the University St. Gallen. All errors are our own.

1 Introduction

Asset returns are known to be neither normally distributed nor of perfect random order. In contrast, they appear to exhibit a heavy-tailed distribution and are ordered in a complex, non-random way that causes large (small) fluctuations to be followed by large (small) fluctuations, a phenomenon that is known as volatility clustering. Empirical evidence for heavy tails and volatility clusters in asset returns is extensive, which is why these two statistical properties are called stylized facts by now (see Cont, 2001; Jiang et al., 2019; Lux & Alfarano, 2016, for a profound review). Although mutually agreed on, these facts are often neglected due to simplifying assumptions about returns. These are the assumptions of normally distributed prices or returns, that neglects heavy tails and of independent and identically distributed (iid) returns, that is commonly employed when scaling volatility with $\sqrt{\text{time}}$ across time horizons and neglects the presence of volatility clusters. These assumptions can not only be found in many seminal works that build the basis for many modern financial theories but are also taught in entry level finance classes all over the globe.[1]

Despite making things simpler, the neglection of heavy tails and volatility clusters can have severe consequences as both are known to adversely affect financial stability. Heavy tails are extreme and thus potentially destabilizing events by definition. Volatility clusters can accelerate adverse market developments. Though the presence of both facts is widely accepted, little is known about how much each of them drives adverse market movements. Given that the two destabilizing forces are often ignored, the important question as to which one is worse for financial stability arises.

Answering this question is not only important to assess the potential shortcomings of simplifying assumptions. The answer also directly affects risk management, investment management, policy making, and academic research. For risk and investment managers sound knowledge about the destabilizing effects of heavy tails and volatility clusters can be insightful for proper and cost-effective hedging strategies. Also, exposure to such risks may be rewarded with a risk premium and could be used as a source of returns. For policy makers, the answer can

[1] For example, the option pricing model by Black and Scholes (1973) neglects heavy tails with its assumption of log-normally distributed stock prices and the efficient market hypothesis by Fama (1970) neglects volatility clusters in returns with its assumption of (iid) returns.

be used to evaluate on how to create efficient policies that ensure financial stability. From an academic perspective, the influence of both facts on financial stability may answer questions on how to realistically model asset returns. In a broader sense, determining which stylized fact has greater influence on adverse market movements is of societal interest, given the well-documented impact of financial market developments on the real economy through compound feedback loops (Arestis et al., 2001; Benhabib et al., 2016, 2019; Sockin, 2015; Straub & Ulbricht, 2015).[2] We shortly address the first two implications in this paper to highlight the practical relevance of our analysis.

Within this work, we disentangle the two sources of financial instability via surrogate analysis. The method allows us to transform a return series into surrogates that first, remove heavy tails while preserving volatility clusters, second, keep the initial distribution while removing volatility clusters, and third, remove both stylized facts. The resulting surrogate series can be interpreted as the original return series for which one or both stylized facts were removed while leaving other statistical properties untouched. We prefer the surrogate method over a simulation because we can perform these transformations without imposing any assumptions on the underlying stochastic process. Upon the three disentangled series we compute maximum drawdowns and worst year-over-year losses to assess the adverse impact of each of the two stylized facts. The analysis is carried out for daily index log-return series representing seven different asset classes. Further, we analyze daily stock log-return series of S&P 500 constituents of which we build portfolios with different degrees of diversification. From this analysis, we examine both, how diversification impacts the presence of heavy tails and volatility clusters and how those two stylized facts translate into drawdown measures.

The main contribution of our paper is twofold. To our knowledge, we are the first to evaluate risk metrics on return series from which heavy tails and volatility clusters were isolated while preserving all other statistical properties. Thus, our first contribution is that we quantify which of the two stylized facts more adversely affects financial stability. These results can also be interpreted as a quantification of potential shortcomings that arise when the facts are overlooked due to the simplifying assumptions related to normality and iid returns. Our second contribution is that we expand this analysis to a portfolio context. This allows us

[2]See Bond et al. (2012) for a comprehensive overview of that literature.

to evaluate the benefits of diversification, i.e., whether it reduces the presence and severity of the two risk drivers in portfolios. In general, this paper attempts to build a bridge between the latest developments in surrogate analysis and real-world applications. We observe that especially in the literature on nonlinear dependence—one of our measures for volatility clusters—asset returns are taken as a go-to dataset in empirical analyses, but most works solely focus on the measurement of statistical characteristics. Our goal is to close the gap between this rather technical literature and broader finance applications. In doing so, we try to avoid overly technical jargon and excessive descriptions of model specifications and instead focus more on the practical interpretation of our results.

Our findings show that all financial markets in our analysis are exposed to heavy tails and volatility clusters. Diversification seems to increase the presence of both stylized facts. We observe that removing heavy tails from the original return series reduces maximum drawdowns by about 5% to 19%. In contrast, removing volatility clusters has an even greater potential for market stability with a respective reduction in maximum drawdowns of about 15% to 30%. We further find that the removal of both stylized facts only leads to marginal differences compared to the cases when only volatility clusters are removed. The results for worst year-over-year returns yield a very similar picture. In a portfolio context, we discover that both drivers of drawdowns can not be tamed by using diversification as the weapon of choice. We find that the better diversified the portfolio, the stronger the effect of removing heavy tails and/or volatility clusters on maximum drawdowns and worst year-over-year returns. This result is valid for equity portfolios and portfolios constructed from different asset classes. Overall, we thus conclude that financial markets across asset classes are clearly more destabilized by volatility clusters than by heavy tails.

Our main finding raises numerous follow-up questions of which we address two that we consider as the most pressing: How to hedge against the two facts in practice and whether the use of $\sqrt{\text{time}}$ volatility scaling in the calculation of banks' regulatory capital buffers yields systematic errors. We show that the hedges against the two facts can be achieved with simple portfolio strategies. We further demonstrate that the results of our surrogate series carry over to these real portfolio strategies, strengthening the claim that the surrogates can be interpreted as synthetic returns from which stylized facts were removed. Regarding regulation, we demonstrate that the $\sqrt{\text{time}}$ scaling of volatility introduces an error irrespective of the underlying risk model that can only be fixed by using local scaling exponents.

This paper serves as a first attempt to quantify the effects of heavy tails and volatility clusters on financial stability. While we try to address some of the above stated follow up questions in this paper, more research is necessary to fully understand the impact of heavy tails and volatility clusters. In general, our findings call for a rethinking of risk for financial managers and regulators should consider discarding the iid assumption for the calculation of capital buffers.

The structure of the remaining article is as follows: Section 2 gives an overview of the relevant literature for our work. Section 3 introduces our measures for heavy tails and volatility clusters before going over the methodological concepts for the segregation of the two stylized facts from the return series. Section 4 contains the empirical analysis of the paper. It begins by presenting the data and subsequently examines the impacts of heavy tails and volatility clusters on financial stability across diverse markets and varying levels of diversification. Lastly, it shortly presents the practical relevance of our findings for hedging and financial regulation. Section 5 concludes and lays out the implications of our work. The appendix holds supplementary material such as additional tables and figures, an introduction to nonlinear dependence, and technical details of our model specification.

2 Related Literature

We identify three strains of academic literature that touch our analyses. We expand on studies of higher-order moments and nonlinear dependence in financial returns. The former is related to the heavy-tailedness while the latter is associated with the volatility clustering. Also, our paper is related to the literature on risk-modeling and return forecasting.

There is a vast literature that studies higher moments in financial returns. However, the literature gets thinner the higher the moment of interest. Regarding the second moment, return volatility, Ang et al. (2006), Ang et al. (2009) find that idiosyncratic volatility negatively affects future returns of stocks. Blitz et al. (2013) and Blitz and van Vliet (2007) make similar discoveries for total return volatility. These findings were reinforced by other studies that find similar results in different financial markets (Baker & Haugen, 2012; Cao & Han, 2013; Chung et al., 2019; Detzel et al., 2019; Joshipura & Joshipura, 2016, 2019).

Empirical works on the third moment, skewness, of asset returns include those of Amaya et al. (2015), Bali and Murray (2013), Barberis and Huang (2008), Boyer et al. (2010), Brunnermeier et al. (2007), Conrad et al. (2013), Green and Hwang (2012), Harvey and Siddique (2000), and Mitton and Vorkink (2007). Others focus on the co-skewness of assets with the market (Dittmar, 2002; Harvey & Siddique, 2000; Schneider et al., 2020). Theoretical foundations for why skewness and co-skewness levels influence asset prices are laid out by (Barberis & Huang, 2008; Brunnermeier et al., 2007; Harvey & Siddique, 2000; Kraus & Litzenberger, 1976, 1983; Rubinstein, 1973). It is worth noting that Jin et al. (2022) reports that the investment horizon plays a significant role for the pricing of co-skewness. They find that the co-skewness of a stock scales differently between investment horizons and is only priced for a small range of short horizons.

Literature on the fourth moment, kurtosis, of asset returns appears to be relatively scarce in comparison to that on the second and third moments. Studies that examine kurtosis and co-kurtosis in asset returns include Ang et al. (2006), Chang et al. (2013), Conrad et al. (2013), and Dittmar (2002). Theoretical foundations for the pricing of kurtosis are laid out by Dittmar (2002).

Empirical evidence about the existence of nonlinear dependence in asset returns was first gathered for US stocks (Ding et al., 1993; Lo, 1991). Later studies confirm the findings for the US and expand the analysis to other financial markets. Those include developed stock markets (Alvarez-Ramirez et al., 2008; Al-Yahyaee et al., 2018; Di Matteo, 2007; Di Matteo et al., 2003; Turiel & Pérez-Vicente, 2003, 2005), emerging stock markets (Cajueiro & Tabak, 2004; Caraiani, 2012; Di Matteo, 2007; Di Matteo et al., 2003; Du & Ning, 2008; Kumar & Deo, 2009; Wang & Wang, 2018), including high-frequency analysis of both (Gu & Huang, 2019; Kwapień et al., 2005), fixed-income markets (Aloui et al., 2018; Shahzad et al., 2017; Wang & Wang, 2018), currency markets (Al-Yahyaee et al., 2018; Bogachev et al., 2007; Di Matteo, 2007; Di Matteo et al., 2005; Drozdz et al., 2010; Schmitt et al., 2000), cryptocurrency markets (Al-Yahyaee et al., 2018; Ghazani & Khosravi, 2020; Stosic et al., 2019), and commodity markets (Al-Yahyaee et al., 2018; Bogachev et al., 2007; Ghazani &

Khosravi, 2020; Jiang et al., 2014; Shao, 2020).

A theoretical foundation to model the nonlinear dependence in financial markets was proposed by Peters (1994) based on his criticism of the efficient market hypothesis (Peters, 1991). At its core, his fractal market hypothesis focuses on the heterogeneity of agents concerning their investment horizons and information processing. With equally represented agents, supply and demand clear smoothly as the same information may cause different actions among investors. Periods of crisis occur if one group of investors dominates the market and their actions cannot be cleared by others. Works that build on the fractal market hypothesis include Blackledge et al. (2019), Dar et al. (2017), Kristoufek (2013), Lamphiere et al. (2021), Li et al. (2014), Rachev et al. (1999), and Weron and Weron (2000).[3]

Volatility clusters and heavy tails are usually observed together and research suggests that both of these stylized facts are interrelated. It is found that volatility clusters can either be caused by heavy-tailed distributions or temporal correlation (Kantelhardt et al., 2002). In literature, no consensus is found as to which of the two contribute most to the volatility clusters in financial data. While some find that the distribution has a higher contribution (Barunik et al., 2012; Buonocore et al., 2016; Matia et al., 2003; Zhou, 2009) others have provided evidence of the opposite (Chen & Wu, 2011; Green et al., 2014; Kwapień et al., 2005; Suárez-García & Gómez-Ullate, 2014; Wang et al., 2011) or that both sources are significantly present (Kumar & Deo, 2009; Norouzzadeh & Rahmani, 2006).

In the history of literature on risk modeling, there have been several milestones. Until Mandelbrot (1963) proposed that financial returns exhibit heavy tails and should be modeled following a Pareto-Lévy distribution, prices were thought to follow geometric Brownian motion (Bachelier, 1900; Osborne, 1959). The following works neglected Mandelbrot's theory and found that while returns are indeed heavy-tailed, they are not so heavy as to be described by Lévy laws (Fama, 1976; Hsu et al., 1974; Officer, 1972). Thereafter, the seminal autoregressive conditional heteroskedasticity (ARCH) model by Engle (1982) became widely acknowledged as it was able to capture autocorrelation in return fluctuations and thus volatility clustering of returns. Because of this, volatility clustering is often called 'ARCH effect'

[3] A current more in-depth overview of the fractal market hypothesis and its applications in modeling can be found in Blackledge and Lamphiere (2022).

but one should keep in mind that volatility clustering is a 'non-parametric' property that is not intrinsically linked to an ARCH specification. A generalization of the ARCH model (GARCH) was proposed by Bollerslev (1986) and many more variations of the ARCH model were developed to better capture certain characteristics of financial returns. For example, the exponential (EGARCH) model by Nelson (1991) that accounts for the asymmetric autocorrelation between positive and negative price variations, or the integrated (IGARCH) model by Engle and Bolerslev (1986) and its fractional successor (FIGARCH) (Baillie et al., 1996) that better capture long-term memory of variance. The ARCH model (or some derivation thereof) is one of the most acknowledged models for modeling financial returns to date. However, these models are argued to spuriously handle the fractal nature of financial returns (Calvet & Fisher, 2002; He & Wang, 2017; Lv & Shan, 2013; Schmitt et al., 1999) and various studies have shown that fractal models have a higher goodness of fit than different variations of ARCH models (Chen & Wu, 2011; Chen et al., 2014; Frezza, 2014; Günay, 2016; Lux, 2001; Segnon & Trede, 2018; Wei & Wang, 2008).

3 Methodology

To examine the stylized facts in isolation, we employ surrogate algorithms to generate synthetic return data. Surrogate analysis is a widely used methodology for hypothesis testing in various scientific fields (see e.g., Keylock, 2019). It allows to transform time series to match a specific feature—like the distribution or the dependence structure—while preserving the other characteristics of the original series. The advantage over a comparable simulation approach is that surrogates do not require assumptions about the underlying stochastic process. For our purpose we prefer the surrogate methodology as it is less assumption sensitive and its results are directly derived from the original data. As with simulations, the surrogate methodology allows the creation of multiple new datasets, enabling more robust conclusions with an increasing number of surrogates. In general, our methodology is closely related to that of Barunik et al. (2012), Buonocore et al. (2016), Green et al. (2014), and Zhou (2009) among many others who compare financial data to surrogate time series and draw conclusions on certain characteristics of the original data.

In the following sections, we will first specify our measures of heavy tails and volatility clusters. Then, we introduce the IAAFT and IAAWT algorithms that we use for our surrogate creation. Thereafter, we will provide deeper insights into statistical properties of the resulting surrogates to confirm that we can remove heavy tails and volatility clusters from a return series while leaving other characteristics of the original series untouched.

3.1 Measurement of Stylized facts

To capture the presence of the two stylized facts in our dataset, we employ two measures for each. The first is always related to the simplifying assumptions of normality and iid distributed returns. The second measures the two stylized facts from a practical perspective.

3.1.1 Heavy Tails

Our *assumption related* measures to quantify the presence of heavy tails are excess skewness and kurtosis. With these two measures it can be evaluated by how much the return distribution differs from a normal distribution, allowing for inferences about its heavy tails.

As our *practical measure* we follow Poon et al. (2004) and employ the reciprocal of the Hill estimator for the left tail of the return distribution. We use this measure because tail risk is commonly thought of as the lower tail of returns and the Hill estimator gives a measure with which the strength of tails can be compared between time series. This is not possible with skewness and kurtosis.[4]

We calculate the Hill estimator as follows. Let $X_1, X_2, \ldots X_n$ be independent random variables.[5] The tail of this heavy-tailed variable above a high threshold X_k satisfies

$$Pr(X > x) = x^{-1/\xi} L(x) \tag{1}$$

[4]For example, take two distributions. The first has higher excess kurtosis than the other but positive skewness, while the other has negative skewness. In this scenario it is hard to assess what distribution has the larger tail purely based on kurtosis and skewness as both influence the shape of the tail.

[5]We acknowledge that returns are typically not iid (which is one of the main subjects of this paper). However, this assumption does not bias the estimator but leads to standard errors that are too small. As we are mostly interested in the value of the estimator, we do not filter/de-cluster the data to eliminate the clustering.

for all $X > X_k$ where L is a slowly varying function and $1/\xi$ is the index of regular variation, or the tail index. We focus on the best known estimator of ξ, the Hill (1975) estimator,

$$\hat{\xi}(k) := \frac{1}{k} \sum_{j=1}^{k} \log(X_{j,n}) - \log(X_{k,n}) \tag{2}$$

where $X_{j,n}$ are the observations of variable X that exceed X_k. For our purpose $X_{j,n}$ are losses that exceed a loss threshold level X_k.

The performance of the tail index depends crucially on the choice of the threshold parameter k. In practice, there exist graphical and analytical tools to determine its optimal value for the estimation. We use the most common graphical method, which is to plot k against $\hat{\xi}(k)$ and choose k where the plot first becomes stationary. Whenever we want to compare series of equal length, we use the same k for all series to make the results more comparable. To verify our choice of k we check the robustness of our choice for the smoothing method of Resnick and Stărică (1997) and the bootstrap method of Danielsson et al. (2001) that analytically determines the optimal value of k. In general, a higher tail index indicates a heavier tail.

3.1.2 Volatility Clusters

Our *assumption related* measure of volatility clusters is the degree of nonlinear dependence in the return series. We employ this measure because the presence of volatility clusters in asset returns indicates that they are not independently distributed across time. However, the absence of constant linear dependence (another stylized fact (see Cont, 2001; Jiang et al., 2019; Lux & Alfarano, 2016)) shows that their dependence must be nonlinear (Cont, 2007). Therefore, volatility clusters must be caused by the nonlinear dependence in asset returns.[6]

For the assessment of nonlinear dependence we compute spreads in Hölder exponents, $\Delta\alpha$, using the basic MF-DFA method of Kantelhardt et al. (2002).[7] These spreads are corrected

[6] For readers who are unfamiliar with the matter of nonlinear dependence, we provide a more detailed elaboration on this topic and its connection to volatility clusters in Appendix B.

[7] MF-DFA is one of the standard tools that is most widely used for characterizing nonlinear dependence (see Grech, 2016; Jiang et al., 2019). Its implementation is discussed in thorough detail by Ihlen (2012), thus not repeated here. A recap on the MF-DFA algorithm and the choice of our model specifications are laid out

for spurious inflation by deducting the mean Hölder spread from IAAFT surrogates.[8] The true nonlinear dependence can be retrieved by subtracting the Hölder spreads of the IAAFT surrogates, $\Delta\alpha_{surr}$, from the Hölder spreads of the original time series, $\Delta\alpha_r$:

$$\Delta\alpha_{nl} = \Delta\alpha_r - \Delta\alpha_{surr} \tag{3}$$

The excess spreads, $\Delta\alpha_{nl}$, thus represent the magnitude of pure nonlinear dependence in the original time series. We follow Schreiber and Schmitz (2000) for the assessment of the statistical significance of our results. Also, we create 100 IAAFT surrogates and use their mean Hölder spread as $\Delta\alpha_{surr}$ in the above equation.[9] Among others, this method to retrieve a pure measure of nonlinear dependence was also carried out by Pamuła and Grech (2014), Rak and Grech (2018), Schadner (2022), and Zhou (2012).

As our *practically related* measure of volatility clustering, we compute the cluster index by Tseng and Li (2011). The method quantifies how strong the typical pattern of volatility clusters—large (small) observations tend to follow large (small) observations—is present. In doing so, it uses a rolling-window approach over a window of size w and counts how many times an extreme value is among the largest percentile p of absolute returns. More formally, let $r_1, r_2, \ldots r_T$ be a series of returns. Then the process can be expressed in a vector f with length $T - w + 1$ for which each element is defined as

in Appendix E.

[8] An example of such spurious sources is that false nonlinear dependence is introduced as a result of using data sets with finite length and/or heavy tails (Grech & Pamuła, 2012; Grech & Pamuła, 2013; López & Contreras, 2013; Mukli et al., 2015; Pamuła & Grech, 2014; Rak & Grech, 2018). Both false drivers of nonlinear dependence can be estimated by transforming the original time series into IAAFT surrogate series. As described in subsection 3.2 the transformed IAAFT surrogate series keep the same distribution as the original but do not exhibit nonlinear dependence. Therefore, nonlinear dependence measured on the surrogate time series should only be composed of the two spurious sources.

[9] We find a strong convergence of $\Delta\alpha_{surr}$ for IAAFT surrogate samples larger than 20. From this, we conclude that the use of a larger surrogate sample would not alter the results of our nonlinear dependence estimation.

$$f_t = \sum_{i=0}^{w} d(r_{t+i}) = \begin{cases} 1, & |r_{t+i}| > |r^p| \\ 0, & |r_{t+i}| < |r^p| \end{cases} \quad , \quad \forall t \in \{1, \ldots, T-w+1\} \tag{4}$$

where $|r^p|$ is the quantile of absolute returns with respect to p, i.e., $|r^p| = \{r : Pr(|R| \leq |r|) = (1-p)\}$, and $d(r_{t+i})$ is a dummy variable that indicates whether the respective absolute return $|r_{t+i}|$ is larger than the threshold absolute return $|r^p|$.

When the whole time series was scanned, the standard deviation of f is calculated as σ_r and compared to that of Gaussian noise with the same parameters p and w as $\sigma_G = \sqrt{wp(1-p)}$. The cluster index is then defined as

$$C(r) = \frac{\sigma_r}{\sigma_G}. \tag{5}$$

If the return series exhibits volatility clusters, σ_r will be larger than σ_G. This is the case because when small (large) absolute observations tend to follow small (large) observations, more elements of f will have values close to 0 (w). This means that the distribution of f will be wider than predicted by the random Gaussian noise so $\sigma_r > \sigma_G$.[10] Henceforth, a value of $C(r) = 1$ shows that there are no clusters while $C(r) > 1$ indicates clustering behavior. For the calculation of the parameters we follow Tseng and Li (2011) and set $p = 20\%$. We set $w = 20$ which roughly equals one trading month.[11] Confidence intervals whether $C(r)$ is statistically different from 1 can be calculated using asymptotic theory.

3.2 Keep Heavy Tail, remove Volatility Clusters: IAAFT

The simplest example to create a surrogate series that removes all dependence among observations is a random shuffle. While the distribution itself stays untouched, its order in time is randomized which in turn destroys all linear and nonlinear dependence. However, asset returns do not follow a perfect random walk empirically and typically have a modest

[10]Fig. 7. in Tseng and Li (2011) illustrates this relationship nicely.

[11]We check our results for different choices of p and w and find that they do not influence our results.

degree of linear dependence (see e.g., Li et al., 2016). The level of linear dependence is one of the factors that determine the scaling of volatility and consequently impacts time-aggregated risk metrics. Thus, keeping the same linear dependence structure of the original series while removing all nonlinear dependence is crucial to make a fair evaluation of the latter in terms of risk. A simple random shuffle of returns is thus inappropriate in our context.

To keep the original level of linear dependence, we use the iterated amplitude adjusted Fourier transform (IAAFT) of Schreiber and Schmitz (1996) to create our surrogate data. This algorithm allows us to destroy the nonlinear dependence while preserving the original return distribution as well as their original linear dependence. In a nutshell, the algorithm reorders the original returns in time while preserving their linear dependence structure.

In more detail, the algorithm can be split into three steps:

1. Perform a Fourier transform of the original data. This transformation splits the original series into its power spectrum and phase. The power spectrum describes the periodicity of the original series and reveals repetitive patterns as well as their strength. It is the linear dependence structure of the original data. The phase gives information about how the different patterns of the original series are ordered in time.

2. Perform a random shuffle of the original data

3. Iterate the following two steps until a convergence criterion is fulfilled or any changes are so small that they do not cause a re-ordering of the values of the previous iteration:

    (a) Perform a Fourier transform of the shuffled (newly generated) series and retrieve its power spectrum and phase. Replace the power spectrum with the power spectrum of the original series that was retrieved in step 1. Perform an inverse Fourier transform to generate a new dataset. Given the initial random sort, this means that this new dataset has the same spectrum (linear dependence structure) as the original data but with phases (different order in time) of the shuffled (newly generated) series.

    (b) Replace the values of the series created above with those of the original series in a rank-order matching process. This means that the largest observation of the new series is replaced with the largest observation of the original series and so forth until all observations of the new series are replaced. This preserves the original distribution of the series. However, the replacement process influences the power spectrum of the resulting series and thus deteriorates the quality of the matching of the linear dependence structure. Because of this, steps (a) and (b) are repeated until the difference between the linear dependence structure of the original series and the surrogate series is minimized.

The resulting surrogate series of this algorithm precisely matches the distribution function of the original series as it simply reorders the original observations in time. The matching of the linear dependence structure is fulfilled until a given error tolerance. Because the algorithm performs a random shuffle before entering the iterative matching process, multiple different surrogates can be created from the same original time series.[12]

3.3 Remove Heavy Tail, keep Volatility Clusters: IAAWT

To preserve the nonlinear dependence structure while transforming the original returns into a surrogate series that has equal statistical properties we use Keylock's (2017) innovation of the iterated amplitude adjusted wavelet transform (IAAWT). This algorithm resembles the nonlinear dependence structure and not solely the overall degree of nonlinearity (like Paluš, 2008). A natural consequence of keeping the nonlinear dependence structure is that the surrogate's volatility clusters around the same time when crises occurred. In short, the algorithm works relatively similarly to the IAAFT—it reorders the observations of the original time series while preserving its linear and nonlinear dependence structure.

In more detail the algorithm can be split into the following steps:

1. Perform a wavelet transform of the original data. Same to the Fourier transform, the transformation allows splitting the original series into its power spectrum and phase. However, the wavelet transform provides more detailed insights into the dependence structure of the original series. While the Fourier transform only retrieves the overall

[12] For readers who seek further explanatory material on this algorithm: A good visualization of the algorithm can be found in Venema et al. (2006). An explanation of why the power spectrum of the Fourier transform equals the linear dependence structure of the original series can be found in Vaseghi (2008).

linear dependence structure, the wavelet transform also documents where and how that structure changed in time. In other words, the wavelet transform captures the nonlinear dependence of the series. Same as with the Fourier transform, that information is captured in the power spectrum component of the transformation.

2. Perform a random shuffle to the original data.

3. Iterate the following two steps until a convergence criterion is fulfilled or any changes are so small that they do not cause a re-ordering of the values of the previous iteration:

   (a) Perform a wavelet transform to the randomly shuffled (newly generated) data series and obtain the new phases, combine these with the original power spectrum that was retrieved in step 1. Perform an inverse wavelet transform to generate a new dataset. Again, the resulting series has the same linear and nonlinear dependence structure as the original series but that structure is differently ordered in time.

   (b) Replace the values of the new series with those of the original series in the same rank-order matching process as in the IAAFT algorithm. The highest observation of the newly created series is replaced with the highest observation of the original series and so forth until all observations of the new series are replaced. Again, the replacement process influences the power spectrum of the resulting series so the above two steps have to be repeated until the difference in nonlinear dependence structure between the original series and the surrogate series is minimized.

Again, the resulting surrogate series of the algorithm precisely matches the distribution function of the original series as it simply reorders the original observations in time. Matching the nonlinear dependence structure is undertaken until a given error tolerance. Same as with the IAAFT, because of the initial random shuffle, multiple different surrogates can be created from the same original time series.[13] The attentive reader may have noticed that unlike promised by the header of this section, the IAAWT algorithm preserves both, the heavy tails and volatility clusters of the original series. How this algorithm can be modified to remove the heavy tails while keeping the volatility clusters will be laid out in the next section.

[13]For readers who seek further explanatory material on this algorithm: A more technical layout of the algorithm can be found in Keylock (2017).

3.4 Illustrative Example of the Surrogate Algorithms

As explained above, the IAAFT and IAAWT algorithms can be used to create surrogates that have the same linear/nonlinear dependence structure as the original time series while preserving the original distribution. For parts of our application, however, we want to change the distribution of the surrogate data while keeping all other statistical properties of the original data. We achieve this simply by using a different dataset in the rank-order matching process within step 3(b) of the above two algorithms. For our purposes, we fit a normal distribution of the same mean and variance to the original data and thus create a normalized version of the original dataset with the same number of observations. Now, instead of replacing the observations in the surrogate time series with the observations of the original data in step 3(b), we take the observations of our normalized series. This change does not affect the matching of the linear/nonlinear dependence structure because it is explicitly controlled for as the convergence criterion in both algorithms after the match-making process. Within this paper, whenever we perform a normalization of the distribution function we will mark the respective algorithm with the subscript n. Hence, if we perform a normalization within the IAAFT (IAAWT) algorithm, we will show it in the notation as IAAFT_n (IAAWT_n).[14]

An illustration of how we use the IAAFT and IAAWT algorithms to change the statistical properties of return series can be found in Figure 1. The upper panel holds the original log-return series of the S&P 500 index between January 1926 and December 2021. It can be observed that the volatility persists over longer periods, especially during periods of financial distress. Further, it can be seen that while most returns fall closely around 0% there are a couple of very extreme observations, that make the return distribution heavy-tailed.

Below the original return series, three exemplary surrogate time series are illustrated. The series underneath the original returns was created using the IAAWT_n algorithm. It can be seen that the algorithm removed all heavy tails, making the whole distribution much more compact. This claim is supported by the gray probability density plot on the right end of the figure. The volatility clusters of the original return series remain untouched as the bursts in times of financial distress can still easily be identified. The series in the third row of the

[14] Note, that this change of distribution can also be performed using any other distribution function without affecting the desired dependence structure matching of both algorithms.

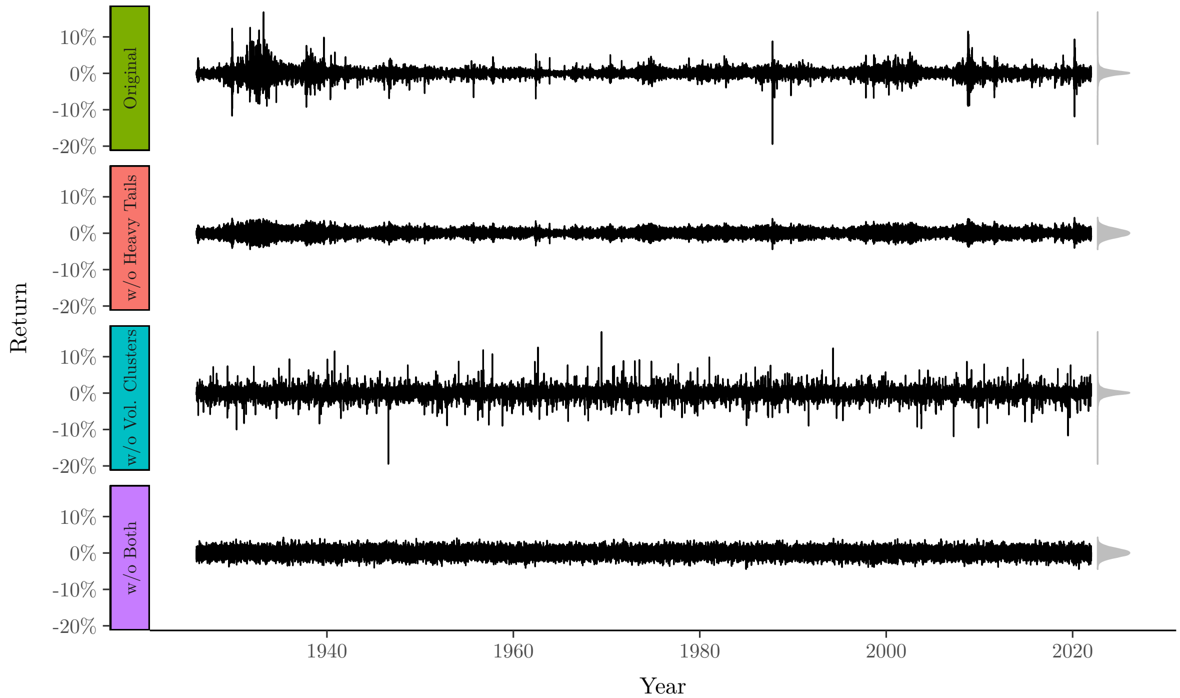

Figure 1: Exemplary comparison of original and surrogate return series for the S&P 500 index

The figure shows the returns for the original S&P 500 index and for three surrogate series that do not exhibit heavy tails, volatility clusters, or both. The time series range from the beginning of 1926 until the end of 2021. The surrogate series without heavy tails was created using the IAAWT_n method. The surrogate series without volatility clusters was created using the IAAFT method. The surrogate series without both stylized facts was created using the IAAFT_n method. The gray areas on the right side of the time axis are probability density plots of the return distributions.

figure was created using the IAAFT algorithm. Visually, it can be estimated that it exhibits the same distribution and thus number and amplitude of extreme observations as the original return series. This comes as no surprise because the individual observations of this surrogate series and original series are the same. When looking at the course of this series it appears much more random than the original and does not exhibit volatility clusters. The bottom series was created using the IAAFT_n algorithm. Similar to the data series in the second panel, because of the normalization, its extreme observations are not as severe as that of the original return series. Additionally, and very similar to the series in the third panel, its course appears to be much more random and no volatility clusters are apparent.

To bolster the takeaways of the above visual analysis quantitatively, we provide descriptive statistics of the four series in Table 1. We can see that all series have the same mean and variance. However, the Jarque Bera and Kolmogorov-Smirnov statistics indicate that only the original and surrogate series without volatility clusters differ significantly from a normal distribution. As for the left tail, we observe that the original tail is preserved by the surrogate without volatility clusters while it is strongly reduced by the two series without heavy tails. With respect to volatility clustering, we observe that the two surrogates without volatility clusters exhibit no more nonlinear dependence while the series without heavy tails shows the same level of nonlinear dependence as the original. This observation is confirmed by the cluster index which only shows significant clustering for the original series and the surrogate without heavy tails. Lastly, to get insights into the significance of autocorrelations we perform Ljung Box tests with lags of 20. We find that all series have a modest degree of linear dependence in returns. This supports our choice to employ the IAAFT algorithm instead of a random shuffle. We also apply this test to absolute returns and find that there is no statistically significant autocorrelation for the surrogates without volatility clusters. For the original series and the surrogate series without heavy tails, however, the autocorrelation in absolute returns is highly significant.

To further investigate this last observation, we plot the autocorrelation of absolute returns up to 20 lags for the four series in Figure 2. Studies that focus on the measurement of volatility clusters find that the slow decay behavior of the function of absolute returns is strongly related to the degree of volatility clustering (Tseng & Li, 2011, 2012). Also, this decay is regarded as a typical phenomenon for financial return series (Cont, 2001, 2007). More generally, examining the autocorrelations of return series is a suggested measure to quantify volatility clustering (Ding et al., 1993).[15] We can see that for the original S&P 500 index, there is significant correlation between absolute returns that decays slowly. The same pattern is present in the surrogate series without heavy tails. We further see that for the two surrogates without volatility clusters, there is no statistically significant autocorrelation in absolute returns, indicating that indeed all volatility clusters were removed.

To further see what properties of the return series are preserved and how they relate to other commonly known properties of financial return series, such as reversals, we provide an

[15] It is worth noting that the autocorrelation of absolute returns is also an example of nonlinear dependence between returns, which again shows the tight link between nonlinear dependence and volatility clusters.

	Original	w/o Heavy Tails	w/o Vol. Clusters	w/o Both
Mean	11.01%	11.01%	11.01%	11.01%
Variance	18.15%	18.15%	18.15%	18.15%
Skewness	-0.07	0.01	-0.07	0.01
Kurtosis	16.92	-0.01	16.92	-0.01
Tail Index	0.29***	0.11***	0.29***	0.11***
Jarque-Bera	3.02×10^{5}***	0.74	3.02×10^{5}***	0.74
Kol.-Smir.	0.09***	0.01	0.09***	0.01
Nonlin. Dep.	0.17***	0.17***	0.02	0.02*
Cluster Index	2.11***	2.71***	1.00	0.99
Ljung-Box	106.56***	83.51***	100.81***	106.22***
| Ljung-Box |	3.85×10^{4}***	8.12×10^{4}***	23.52	28.38

Table 1: Exemplary comparison of descriptive statistics of original and surrogate return series for the S&P 500 index

The table compares descriptive statistics for the returns for the original S&P 500 index and for three surrogate series that do not exhibit heavy tails, volatility clusters, or both. The time series range from the beginning of 1926 until the end of 2021. The surrogate series without heavy tails was created using the IAAWT_n method. The surrogate series without volatility clusters was created using the IAAFT method. The surrogate series without both stylized facts was created using the IAAFT_n method. The tail index is calculated for the lower tail of the return distribution using the reciprocal of the Hill (1975) estimator with a threshold of $k = 100$. The Jarque-Bera and Kolmogrov Smirnov tests both assess whether the series is different from a normal distribution. Nonlinear dependence is measured by the excess spreads in Hölder exponents using the MF-DFA method of Kantelhardt et al. (2002). The cluster index is calculated following Tseng and Li (2011). The two Ljung-Box tests assess whether there is significant autocorrelation in the series or absolute sereies from top to bottom. Statistical significance at confidence levels of 10%, 5%, and 1% are indicated with *, **, and *** respectively.

additional regression based analysis in Appendix C.

4 Empirical Analysis

The empirical analysis of our work can be separated into four parts. First, we provide an overview of the data. Second, we disentangle the two stylized facts for seven index return series to assess how the stability of different financial markets is affected by heavy tails and volatility clusters. Third, we analyze whether diversification can protect investors from the destabilizing effects of the two stylized facts. Lastly, we provide two practical implications of our work. First, we demonstrate how the two stylized facts can be hedged in practice. Second, we demonstrate how the neglection of volatility clusters can cause errors in an exemplary risk

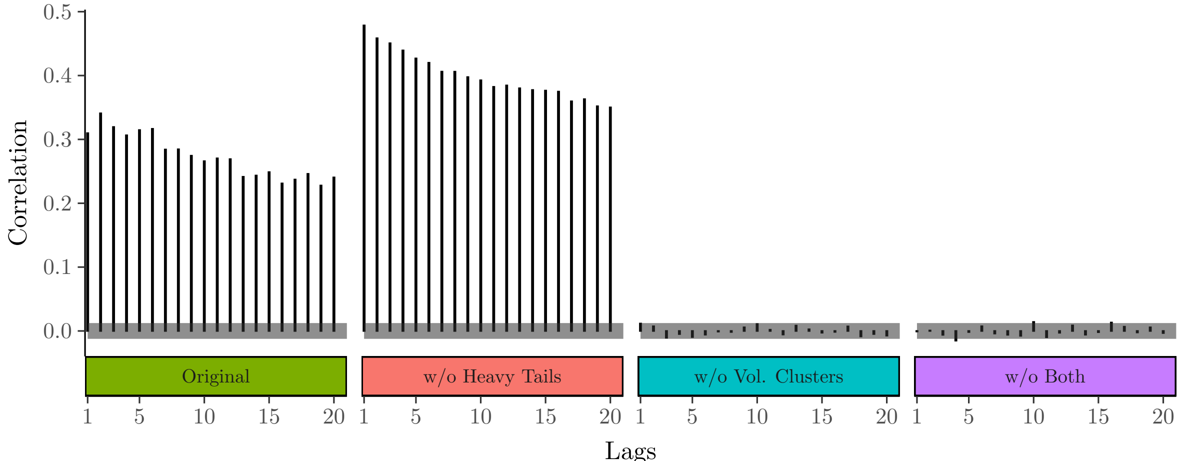

Figure 2: Exemplary comparison of original and surrogate autocorrelation function of absolute returns for the S&P 500 index

The figure shows the autocorrelation of absolute returns between 1 to 20 lags for the original S&P 500 index and for three surrogate series that do not exhibit heavy tails, volatility clusters, or both. The time series range from the beginning of 1926 until the end of 2021. The surrogate series without heavy tails was created using the IAAWT_n method. The surrogate series without volatility clusters was created using the IAAFT method. The surrogate series without both stylized facts was created using the IAAFT_n method. The gray area illustrates the 95% confidence interval for the null hypothesis that there is no autocorrelation.

modeling routine that is inspired by current regulatory framework.

4.1 Data

Our analysis focuses on two datasets. First, we analyze seven index time series of daily log-returns, each representing a different asset class that we source from CRSP, Bloomberg, Refinitiv and MSCI. The series have a common ending date of 12/31/2021, the starting date was chosen individually with respect to data availability. Details can be found in Table 2. Second, to gain insights into the performance of portfolios, we analyze daily log-returns of constituents of the S&P 500 index that we pull from Datastream. We solely use constituents for which uninterrupted return data is available between 01/01/1992 and 31/12/2021. Also, we exclude stocks for which returns did not change for a trading week to exclude highly illiquid stocks from the analysis. The total number of stocks in this dataset equals 144.[16] The choice for the time horizon of this dataset was made based on two counteracting objectives.

[16] A list of the respective company names can be found in Appendix D. We acknowledge that our sample of stocks is prone to survivorship and size bias. However, since our focus is on understanding the impact of diversification, these biases are expected to attenuate our findings. Consequently, they can be interpreted as a conservative baseline, with the expectation that potential effects of fully diversified portfolios is even larger.

First, the sample size was to be maximized to strengthen statistical inferences. Second, the time horizon of the series was to be kept as long as possible because analyses for nonlinear dependence yield more robust results for longer time series (Shao et al., 2012). The sample of 144 stocks over a period of 30 years finds a compromise between the two.

From these 144 stocks, we form portfolios at different levels of diversification. We chose the number of portfolio holdings following literature. Studies examining the effect of diversification report that its benefits decrease in the number of holdings and plateau with roughly 40 holdings (Evans & Archer, 1968; Mao, 1970; Statman, 1987; Upson et al., 1975). Thus, the number of portfolio holdings we use are 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, and 45. For each, we create 40 equal-weighted portfolios of which the constituents are randomly chosen from the 144 available stocks.[17] We further create the minimum variance (MV) and maximum Sharpe ratio (SR) portfolios to also include optimized portfolios in our analysis. Overall, our total number of portfolios is 802.

We assess the presence of the two stylized facts in Table 2 for the index data and in Figure 3 for the stock portfolio data. The excess kurtosis and negative skewness as well as the tail index for both data sets indicates that all series are characterized by a heavy-tailed distribution. From the excess Hölder spreads and the cluster index we observe that all series experience volatility clustering. Concerning the stock portfolios two more patterns can be observed. First, heavy-tailedness and volatility clustering seems to increase with the number of portfolio holdings. The latter only being observed for the cluster index whereas nonlinear dependence is constant across holdings. Second, the distribution of the measures decreases with the number of holdings. This is likely attributable to the fact that the more holdings a portfolio has, the more idiosyncratic risk is diversified away, making portfolios more similar. Also, at higher degrees of diversification it is more likely that similar stocks are held in the portfolios. Regarding the optimal portfolios, we find that they have smaller tails than well-diversified portfolios. Regarding volatility clustering no conclusive comparison can be made.

[17] The choice to create 40 portfolios is based on the assumption that the central limit theorem holds at that sample size and that results are not strongly biased due to outliers.

Assets *Instrument*	Start	Skew.	Kurt.	Tail	Nonl.	Clust.
Foreign Exchange *GBP/USD rate*	08/1971	0.25	6.73	0.23***	0.82***	1.69***
Commodity *Bloomberg Commodity Index*	01/1991	-0.38	4.37	0.28***	0.14***	1.74***
Stocks *S&P 500 Index*	01/1926	-0.07	16.92	0.24***	0.17***	2.11***
Private Equity *LPX Composite Listed P.E. Index*	04/2004	-0.74	23.82	0.37***	0.06	2.19***
Real Estate *MSCI U.S. Liquid Real Estate Index*	06/2001	-1.05	18.89	0.35***	0.26***	2.20***
Bonds *S&P 500 Bond Index*	01/1995	-0.59	6.52	0.29***	0.43***	1.50**
Cryptocurrency *FTSE Bitcoin Index*	10/2015	-0.13	3.83	0.33***	0.16*	1.43**

Table 2: Description and summary statistics of the index data

This table holds data description and summary statistics for daily log-returns of the seven indices that were used in the analysis. All series end on 12/31/2021. The tail index is calculated for the lower tail of the return distribution using the reciprocal of the Hill (1975) estimator with a threshold of $k = 60$. Nonlinear dependence is measured by the excess spreads in Hölder exponents using the MF-DFA method of Kantelhardt et al. (2002). The cluster index is calculated following Tseng and Li (2011). Statistical significance at confidence levels of 10%, 5%, and 1% are indicated with *, **, and *** respectively.

4.2 Disentangling the Sources in Financial Markets

For each of the seven return series, we create 200 IAAFT, IAAWT$_n$, and IAAFT$_n$ surrogates. Recall, that the IAAFT surrogates can be interpreted as if the original returns had no volatility clusters, the IAAWT$_n$ surrogates as if returns would be normally distributed but with the same volatility clusters as the original returns, and the IAAFT$_n$ as a benchmark where both stylized facts are removed. Next, upon the surrogate series, we compute maximum drawdowns and worst year-over-year return and compare these measures between surrogates and the original returns. We focus on these two risk measures because they are path dependent and can thus be influenced by volatility clusters and heavy tails in the distribution. Other risk measures such as the empirical VaR or the semi-standard deviation solely focus on the

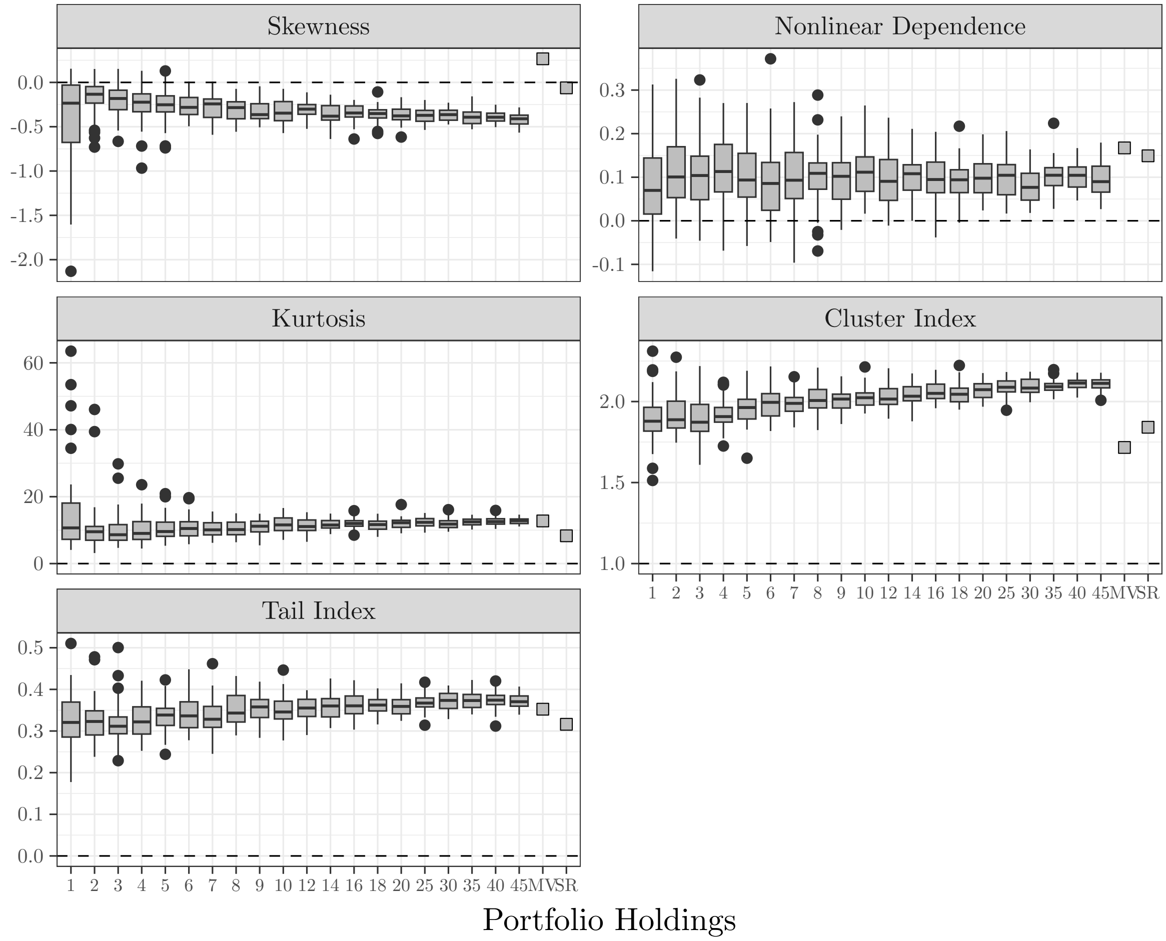

Figure 3: Summary statistics of the stock portfolio data

The figure holds skewness, excess kurtosis, and nonlinear dependence measured by the excess spreads in Hölder exponents using the MF-DFA method of Kantelhardt et al. (2002) in the subfigures from top to bottom respectively. The portfolios are composed using daily log-returns of randomly chosen constituents of the S&P 500 index for which return data is available between 01/01/1992 and 31/12/2021, resulting in an investible pool of 144 stocks. For each number of holdings 40 equal-weighted portfolios were formed. MV and SR are the minimum variance and maximum Sharpe ratio portfolios respectively. Measures for heavy-tailedness and volatility clustering are displayed on the left and right respectively. The tail index is calculated for the lower tail of the return distribution using the reciprocal of the Hill (1975) estimator with a threshold of $k = 60$. Nonlinear dependence is measured by the excess spreads in Hölder exponents using the MF-DFA method of Kantelhardt et al. (2002). The cluster index is calculated following Tseng and Li (2011).

distribution of returns and are unaffected by volatility clusters. In our case, because the distribution of the surrogates is either equal to that of the original dataset or set to be normal, risk measures that solely focus on the distribution will yield inconclusive results.[18] Also, we

[18]Results for the original series will be identical to those of the IAAFT series and results for the IAAWT_n series will be identical to those of the IAAFT_n series because either of the two pairs share the same distribution. Note, that this holds only if non-path-dependent risk measures are calculated with the same time scale as the

prefer drawdown measures because they show how prone a return series is to shocks and can thus be interpreted as proxies for financial stability. The results for maximum drawdowns and worst year-over-year returns are reported in Table 3. Figure 4 visualizes the maximum drawdowns of the disentangled return series. A visualization for worst year-over-year returns is shifted to the Appendix in Figure A.1, because it looks very similar.

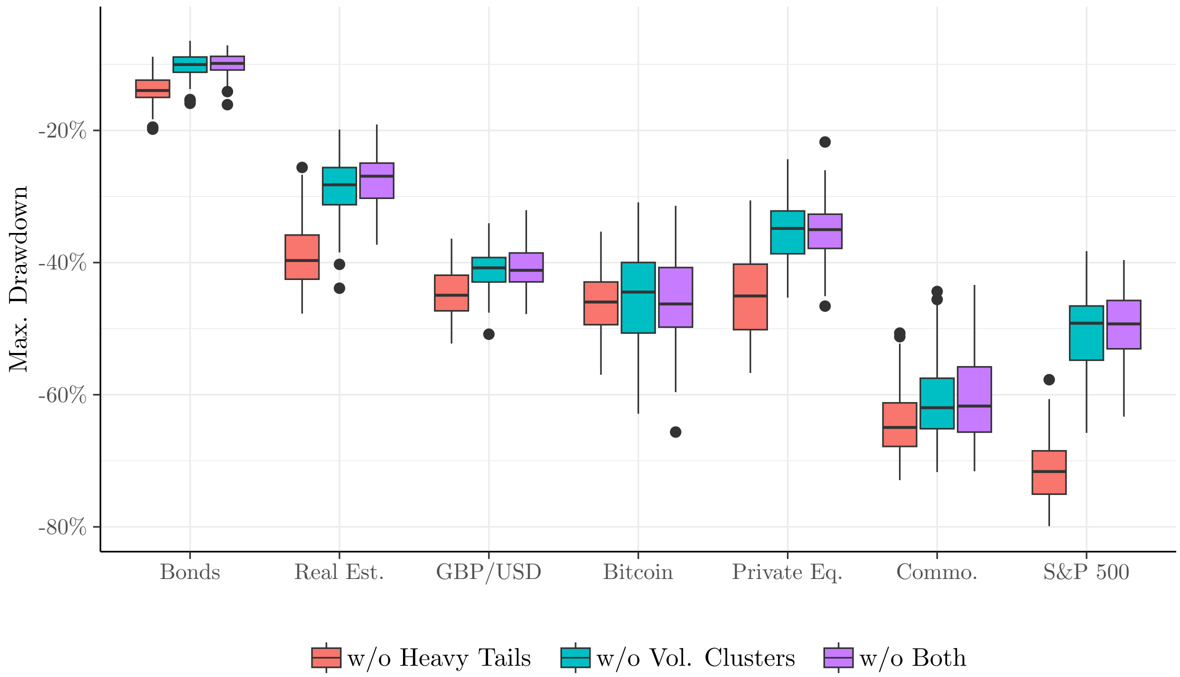

Figure 4: Maximum drawdowns of the different asset classes after disentangling stylized facts
The figure holds maximum drawdowns for three surrogates series of seven indices. Each surrogate series comprises 200 surrogates. The surrogates without heavy tails were created using the IAAWT_n method. The surrogates without volatility clusters were created using the IAAFT method. The surrogates without both stylized facts were created using the IAAFT_n method.

The results show that each of the seven indices has a lower downside risk when volatility clusters rather than heavy tails are removed. We can see that in absolute terms the reduction is greater for asset classes that are riskier, i.e. stocks vs. bonds. However, when looking at the relative change, we find that those are relatively similar across asset classes, at least for maximum drawdowns. In general, we observe that the removal of heavy tails reduces maximum drawdowns by roughly 15%. The removal of volatility clusters reduces

time series under investigation. If those measures were calculated under a different time scale, for example, yearly (rolling) VaR calculations for daily data, the results would be different for all time series as a result of varying degrees of scaling properties among the time series.

	Bonds	Real Est.	F.X.	Crypto	P.E.	Comm.	Stocks
	Max. Drawdown						
Original	-14.8%	-45.8%	-55.0%	-56.3%	-54.4%	-77.2%	-84.8%
w/o Heavy Tails	-14.0%	-39.0%	-44.8%	-46.1%	-45.0%	-64.2%	-72.3%
	(-5%)	(-15%)	(-19%)	(-18%)	(-17%)	(-17%)	(-15%)
w/o Vol. Cluster	-10.2%	-28.7%	-41.0%	-45.3%	-35.1%	-61.0%	-50.7%
	(-31%)	(-37%)	(-25%)	(-20%)	(-36%)	(-21%)	(-40%)
w/o Both	-10.1%	-27.6%	-40.9%	-45.6%	-35.1%	-60.6%	-49.8%
	(-32%)	(-40%)	(-26%)	(-19%)	(-36%)	(-22%)	(-41%)
	Worst YoY Loss						
Original	-12.4%	-35.0%	-28.1%	-50.1%	-43.0%	-52.1%	-64.5%
w/o Heavy Tails	-11.2%	-33.2%	-28.6%	-38.3%	-33.8%	-44.0%	-53.2%
	(-10%)	(-5%)	(2%)	(-24%)	(-22%)	(-16%)	(-18%)
w/o Vol. Clusters	-8.2%	-23.1%	-25.8%	-35.0%	-26.1%	-37.8%	-39.2%
	(-34%)	(-34%)	(-8%)	(-30%)	(-39%)	(-28%)	(-39%)
w/o Both	-8.2%	-21.9%	-26.1%	-36.2%	-25.7%	-38.1%	-38.1%
	(-34%)	(-37%)	(-7%)	(-28%)	(-40%)	(-27%)	(-41%)

Table 3: Maximum drawdowns and worst year-over-year returns of original series compared to mean values of surrogates

The table compares maximum drawdowns and worst YoY returns for surrogates series without heavy tails, without volatility clusters, and without both stylized facts of seven indices. Each surrogate series comprises 200 surrogates. The surrogates without heavy tails were created using the IAAWT$_n$ method. The surrogates without volatility clusters were created using the IAAFT method. The surrogates without both stylized facts were created using the IAAFT$_n$ method. The numbers in parentheses indicate relative reduction compared to the original series.

maximum drawdowns by around 30%. Table A.1 in the Appendix proves that the reduction in drawdowns for the series without volatility clusters is significantly larger than that of the series without heavy tails. For worst year-over-year returns the range of relative reductions is wider. Furthermore, we find that when both stylized facts are removed, the improvement in downside risk compared to a return series without volatility clusters is marginal. Hence, reducing volatility clustering of returns has greater potential for financial stability.

4.3 Evaluating the Effects of Diversification

In practice, the go-to weapon against risk is diversification. Hence, it makes sense to evaluate whether volatility clusters or heavy tails can be influenced by diversification and how these changes translate into risk measures. In Figure 3 we already showed that higher degrees of diversification increase the presence of heavy tails and volatility clusters. Hence, we can already see that diversification does not reduce the presence of the two stylized facts. To determine whether it yields protection regarding how the two stylized facts translate into downside measures, we repeat the analysis that we carried out on an index level below.

We create five IAAFT, IAAWT_n, and IAAFT_n surrogate return series for each portfolio, leading to a total of 200 surrogates for each algorithm-holding pair. Again, upon the surrogate series, we compute maximum drawdowns and worst year-over-year returns. Figure 5 holds the results maximum drawdowns. It also includes the drawdowns of the original portfolios for a better comparison. However, it has to be noted that the data for the original portfolios comprises 40 observations per holding whereas the surrogate time series have 200 observations per holding. In the case of the optimized portfolios, the original data only consists of one observation. We only show results for maximum drawdowns here because those for worst year-over-year returns look very similar and can be found in the Appendix, Figure A.2.

Five important insights can be gained from the figure, some of which confirm the findings for the index analysis. First, diversification works. The average drawdown of better-diversified portfolios is smaller than that of concentrated portfolios. Second, removing the stylized facts from the return series leads to a reduction in maximum drawdowns across all levels of diversification.[19] Third, volatility clusters seem to be a bigger driver for drawdown risk as drawdowns of the surrogate series without volatility clusters are always below those of the surrogates without heavy tails. Fourth, removing both stylized facts from the return series only leads to a marginal increase in drawdowns compared to a series from which only volatility clusters were removed. Fifth, the removal of the two stylized facts has a larger influence on the

[19]The only exception from this pattern are the surrogate portfolios without heavy tails for the MV portfolio. This is because the MV portfolio is one of the only portfolios in our sample that is positively skewed. Therefore, the normalization of returns (to remove the heavy tails) shrinks positive returns more strongly than it shrinks negative ones, leading to higher maximum drawdowns.

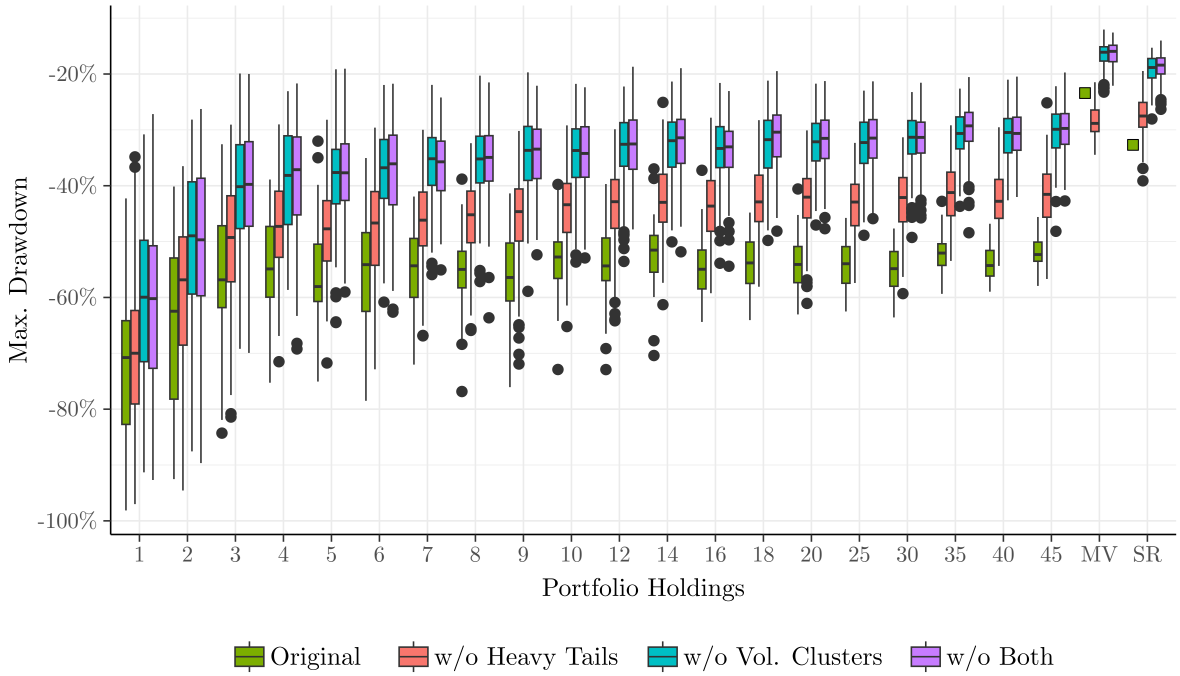

Figure 5: Maximum drawdowns of stock portfolios with different holdings after disentangling stylized facts

The figure holds maximum drawdowns for the original return and three surrogate series of stock portfolios with different holdings. The portfolios are composed using daily log-returns of randomly chosen constituents of the S&P 500 index for which return data is available between 01/01/1992 and 31/12/2021, resulting in an investible pool of 144 stocks. For each number of holdings, 40 equal-weighted portfolios as well as the minimum variance (MV) and maximum Sharpe ratio (SR) portfolios were formed. For each portfolio, we create five surrogates leading to a total number of 200 time series per portfolio holding for the equal-weighted portfolios. For the optimal MV and SR portfolios we create 200 surrogates. The surrogates without heavy tails were created using the IAAWT_n method. The surrogates without volatility clusters were created using the IAAFT method. The surrogates without both stylized facts were created using the IAAFT_n method.

reduction of drawdowns at higher degrees of diversification.[20] These results are consistent with the increase of heavy tails and volatility clusters in portfolio holdings as reported in Figure 3. If we believe that both stylized facts drive drawdowns, it is intuitive that drawdown reduction of the surrogate series without either of the two increases in diversification.

We further analyze the last finding finding—that the removal of the two stylized reduces drawdowns more at higher degrees of diversification—for indices. Figure 6 holds the average relative reduction in drawdowns from the surrogate portfolios consisting of the commodity,

[20]This observation is made clearer by Figure A.4 that compares the relative reduction in maximum drawdowns between the surrogate series without heavy tails, surrogate series without volatility clusters, and the original portfolio data.

real estate, stock, and bond index compared to the original portfolio. The matrices have to be read as follows. Each of the four indices is placed in one corner. Portfolio weights are assigned according to proximity to the corners. For example, the commodity index is located in the upper left of every matrix and the upper left portfolio is 100% invested into the commodity index. The portfolio in the middle of every matrix, however, is an equal-weighted portfolio of all four indices. To see whether the two facts have a larger influence on the reduction of drawdowns at higher degrees of diversification we would expect to see higher relative reductions whenever we move further away from the corners. We can see this pattern for the series where volatility clusters and both stylized facts were removed. For the series without heavy tails, the picture is less clear. To provide better intuition for the matrix plot and give further insights into the underlying pattern we performed the same analysis using only the stock and the bond index in Figure A.3 of the Appendix. This plot thus represents the bottom row of every matrix in Figure 6.

4.4 Practical Implications

The practical implications of our findings are widespread. This paper is certainly not enough to fully address the follow up questions that result from our main finding that volatility clusters are worse for financial stability than heavy tails. Nevertheless, the following paragraphs contain two analyses that address two important implications of our findings.

First, we provide an example of how hedging against the two stylized facts can be done in practice and whether the findings of our synthetic return series also carry over to real portfolios. Also, the results can be used to evaluate whether our synthetic time series can be interpreted as perfectly hedged portfolios against the two stylized facts. Second, we provide a short illustrative example of how the $\sqrt{\text{time}}$ scaling of volatility that is employed in the capital requirements regulation for banks yields errors of risk models. Given that the Basel regulation is one of the standard regulatory frameworks worldwide, it is important to evaluate whether a systematic flaw is made and how it can be fixed.

w/o Heavy Tails

-14%	-10%	-5%	-4%	-14%
-10%	-8%	-11%	-7%	-4%
-17%	-15%	-15%	-10%	-5%
-14%	-17%	-15%	-12%	-13%
-12%	-15%	-14%	-11%	-8%

w/o Vol. Clusters

-12%	-12%	-15%	-24%	-31%
-14%	-22%	-28%	-31%	-35%
-30%	-34%	-38%	-39%	-36%
-34%	-39%	-42%	-42%	-37%
-34%	-38%	-41%	-40%	-31%

w/o Both

-11%	-12%	-15%	-24%	-35%
-13%	-21%	-29%	-33%	-37%
-31%	-35%	-39%	-41%	-37%
-34%	-40%	-43%	-43%	-39%
-35%	-39%	-41%	-42%	-33%

Assets in Corners

Comm.	Real Est.
Stocks	Bonds

Avg. DD Reduction

-40% -30% -20% -10%

Figure 6: Relative maximum drawdown reduction of index portfolios after disentangling stylized facts

The figure holds the average relative maximum drawdown reductions for three surrogate series of portfolios that consist of a stock, bond, real estate, and commodity index at different portfolio weights. The matrix can be read as follows. In each corner of the matrix, one of the assets is positioned. Portfolio weights are assigned according to proximity to the corners. The corner location of each index is indicated in the table on the bottom right. For example, the upper left corner represents a portfolio that is 100% invested in the commodity index. The portfolio in the middle of the matrix is an equal-weighted portfolio of all four indices. Drawdown reductions were calculated compared to the respective original portfolio of the four indices. The portfolios were composed using daily log-returns between 06/01/2001 and 31/12/2021. Each surrogate series comprises 200 surrogates. The surrogates without heavy tails were created using the IAAWT_n method. The surrogates without volatility clusters were created using the IAAFT method. The surrogates without both stylized facts were created using the IAAFT_n method.

4.4.1 Hedging

The answer on how to hedge against heavy tails and volatility clusters in practice already exists for decades. Tail risk is commonly hedged by using derivatives, for example, the "protective put" strategy where the investor buys the underlying and put options to protect against extreme losses. Including this downside protection shifts probability mass from the lower tail upwards, inducing positive skewness, and thereby reduces the heavy tails overall. Such strategies are discussed in the literature like Bertrand and Prigent (2011), Figlewski et al. (1993), and McKeon and Svetina (2017). As for hedging against volatility clusters, the industry also provides a comprehensive palette of investment products. The volatility cluster hedge can be achieved more easily, without the requirement of derivatives. The hedge attempts to keep portfolio volatility constant. These strategies are also known under the terms "volatility targeting" or "risk control" and are further discussed in works like Fleming et al. (2001), Harvey et al. (2018), Moreira and Muir (2017), and Mylnikov (2021).

We discuss the heavy tail/volatility cluster hedge upon two practical and easy to replicate examples. For the former we use data of the *Cboe S&P 500 10% Put Protection Index* as a standardized hypothetical portfolio with tail risk protection. This index tracks the value of a hypothetical portfolio composed of a long position in the S&P 500 index and a long position in a one-month 10% out-of-the-money put option on the index that is re-balanced monthly. For the volatility cluster hedge we formulate a strategy that is almost identical to the *S&P 500 Daily Risk Control 15% Index*. This index makes an allocation between the S&P 500 index and the one-month Treasury bill rate that is re-balanced daily, to achieve a portfolio with a constant volatility of 15%. While the index uses historical volatility for the targeting purpose, we use the VIX instead.[21] To keep the setting simple we assume the T-bill to be the risk-free rate with zero expected volatility. The data frequency for both hedging strategies is daily and the observation horizon is 1990-01-02 (inception of the CBOE index) to 2021-12-31. The comparison is presented in Table 4, *Panel A*. For better comparability of results we scale the returns of the two hedging portfolios recursively such that the mean return equals that of the unhedged portfolio. The scaling was done in a way that an investor can take leveraged positions in the two hedged portfolios that are financed at the Treasury bill rate. The results for these scaled portfolios are presented in *Panel B* of Table 4.[22]

We find that the hedged portfolios have lower mean returns but also lower volatility than the original. In case of our cluster hedged portfolio, the volatility targeting strategy, this leads to higher risk-adjusted returns, as expressed by the higher Sharpe Ratio. With regards to heavy tails, we observe that the skewness was indeed largely eliminated for the tail-hedged

[21] With the usage of the VIX as a forward-looking measure of expected volatility (instead of historical volatility as in the *S&P 500 Daily Risk Control 15% Index*) we achieve a more accurate elimination of volatility clusters as changes in volatility are recognized faster. The volatility target of 15% was chosen due to two reasons: (i) it is close to the average standard deviation of the S&P 500 returns in our sample and (ii) it is also a common target level used at indices and investment products.

[22] For robustness, we also present results for a setting where we use the *S&P 500 Daily Risk Control 15% Index* in Table A.2 in the Appendix. The findings are very similar to the ones presented here.

strategy, while it remained similar to the original skewness for the cluster hedge. Both hedged portfolios have a lower kurtosis than the original but the tail index shows that the tail hedged portfolio has the smallest left tail among the three portfolios. In terms of volatility clustering the picture is exactly the other way around. While the tail-hedged portfolio preserves the clusters relatively well, as indicated by the significant nonlinear dependence and a cluster index far above 1, they are thoroughly removed within the volatility targeting strategy, which has no excess nonlinear dependence and a cluster index of 1.26. Therefore, both strategies in isolation perform well in their hedging purpose. With a look on drawdowns we now recognize that both hedges reduce the aggregated downside potential, but in line with our predictions from the before surrogate analysis, it is the volatility clustering that is the greater driver of downside risk. This finding still holds for the analysis where the hedged portfolios were scaled to have the same mean as the original. Also, the result that in this simple context both facts could be isolated relatively well without affecting other properties of the original return series much further indicates that our synthetic surrogate series can indeed be interpreted as the returns of a portfolio that is perfectly hedged against heavy tails and/or volatility clusters.

4.4.2 Policy Making

For the policy implications of our finding, we focus on the standard for the calculation of capital requirements for risks arising from movements in market prices—the MAR framework by the Basel Committee on Banking Supervision (BCBS). This framework lays out a calculation methodology for capital buffers that shall absorb losses from adverse movements of market prices. These capital requirements apply to the 45 Basel members which comprise central banks and bank supervisors from 28 jurisdictions.

When examining the current Basel 3 Framework it is apparent that it focuses on heavy tails but neglects volatility clusters. In the framework banks can choose between a standardized and an internal risk model approach (BCBS, 2020, 11.7). Within the latter, risk is assessed using the 97.5^{th} percentile expected shortfall (ES) (BCBS, 2020, 33.3). Banks are required to calculate the ES at a base horizon of ten days that is not to be scaled from shorter horizons. They further have to specify a set of risk factors for which the regulation sets a respective time horizon. Depending on the risk factor exposure, the bank has to scale a modified version

	Original	Tail Hedge	Cluster Hedge
		Panel A	
Mean	10.33%	7.44%	7.92%
Volatility	18.09%	13.36%	11.58%
Sharpe Ratio	0.57	0.56	0.68
Max. Drawdown	-55.25%	-41.96%	-37.53%
Worst YoY Loss	-47.50%	-31.55%	-27.25%
Skewness	-0.41	0.03	-0.48
Kurtosis	11.42	4.31	1.56
Tail Index	0.35***	0.20***	0.22***
Nonlin. Dep.	0.14***	0.13***	0.01
Cluster Index	2.04***	1.71***	1.26
		Panel B	
Mean	10.33%	10.33%	10.33%
Volatility	18.09%	18.55%	15.11%
Sharpe Ratio	0.57	0.56	0.68
Max. Drawdown	-55.25%	-53.02%	-45.89%
Worst YoY Loss	-42.34%	-29.46%	-26.36%
Skewness	-0.41	0.03	-0.48
Kurtosis	11.42	4.31	1.56
Tail Index	0.35***	0.20***	0.22***
Nonlin. Dep.	0.14***	0.12***	0.01
Cluster Index	2.04***	1.71***	1.26

Table 4: Summary statistics of practical portfolios

The table compares descriptive statistics for the original S&P 500 index returns as well as for a tail hedged and a volatility cluster hedged portfolio of the S&P 500 index. The time series range from the beginning of 1990 until the end of 2021. The tail hedged portfolio is represented by the CBOE S&P 500 Put Protection Index at a level of 10%. The volatility hedged portfolio was built with a volatility targeting strategy similar to the S&P 500 Daily Risk Control Index that uses the VIX to target a volatility of 15%. *Panel A* holds the results for the unscaled portfolios. *Panel B* holds the results for the hedged portfolios that are scaled to have the same mean return as the original portfolio. Nonlinear dependence is measured by the excess spreads in Hölder exponents using the MF-DFA method of Kantelhardt et al. (2002). The tail index is calculated for the lower tail of the return distribution using the reciprocal of the Hill (1975) estimator with a threshold of $k = 60$. The cluster index is calculated following Tseng and Li (2011). Statistical significance at confidence levels of 10%, 5%, and 1% are indicated with *, **, and *** respectively.

of the 10-day ES with the $\sqrt{\text{time}}$ delta that is determined by the prescribed time horizon. The capital requirements are then determined by the sum of the base and scaled ES (BCBS, 2020, 33.4). Heavy tails are captured by the ES measure by construction. Volatility clusters are not rightfully accounted for as the base ES measure is scaled with $\sqrt{\text{time}}$ for periods beyond ten days. This claim is supported by the findings that banks' exceedances of capital requirements cluster in time (Berkowitz & O'Brien, 2002; O'Brien & Szerszeń, 2017).

To evaluate this regulatory focus we provide a simple risk model backtesting example below. We are guided by the regulation regarding our setup. Within Basel 3 banks are required to backtest their risk models based on value-at-risk (VaR) over the most recent 12 months (BCBS, 2020, 32.4). The framework is a straightforward procedure of comparing predictions with actual trading outcomes. To assess the quality of their risk model, banks have to calculate the number of times that the trading outcomes were not covered by the VaR. This means that a 99% VaR should cover 99% of the trading outcomes (BCBS, 2020, 32.5).

Following the regulation, we use VaR as the risk metric of interest for our model backtest. We employ our daily S&P 500 index return series for the most recent 40 years of data. For simplicity, our expected VaR equals the VaR calculated from the observed mean and volatility of the most recent 12 months. Then we compare how many times that estimate was exceeded thereafter. We update our estimate on ten-day non-overlapping rolling windows leading to a total sample of 1,059 observations. To test the overall quality of the VaR model we repeat the above steps for confidence levels between 90% and 99.9% at steps of 0.5%.

We test three VaR models. First, we calculate a daily VaR model that is evaluated against the following one-day return. This model serves as a baseline benchmark as the risk model and the evaluation period have the same time scale and no scaling are required. Second, we use the results of the baseline model and scale them with $\sqrt{\text{time}}$ ($10^{0.5}$) to retrieve a 10-day VaR estimate.[23] This model is then compared to the following 10-day return. A comparison between the quality of this second model and our baseline model should help answer the question of whether ignoring the scaling properties of returns alters VaR estimates. Third, similar to our second model, we scale the results of our base model with the local scaling exponent of the most recent estimate within our in-sample period to retrieve 10-day VaR estimates. The scale exponent estimates are based on the algorithm of (Ihlen & Vereijken,

[23]The choice to use ten days is inspired by the base horizon for the capital requirement regulation.

2014)[24] using the most recent five years of available data. The choice of a longer return history for the scaling exponent estimation stems from the potential for short time series to produce misleading results (Shao et al., 2012). Thus we follow the recommendation to set the sample size greater than 1,000 observations (Ihlen, 2012). Also, we estimate the local 10-day scaling exponent using an interpolated value for local scale estimations of 9, 10, and 11 days because including more scales is recommended to increase robustness (Ihlen & Vereijken, 2014). We arrive at our 10-day VaR estimate by taking our base 1-day VaR and multiplying it with 10^{H_l} where H_l is our local scaling exponent estimation.[25]

The results of our analysis are summarized in Figure 7. It shows the percentiles of the VaR models on the x-axis and the respective fractions of returns that fall out of the confidence interval on the y-axis. The black line refers to the null hypothesis, i.e. that 1% of returns fall below the 99% VaR estimate. The gray area indicates 95% level confidence bands around the null. The results for our base 1-day VaR estimate are represented with black X's. We find that though simple, the VaR estimate performs relatively well as all of its estimates fall within the confidence bands. When comparing the results of our estimate using $\sqrt{10}$ as the uniform scaling coefficient (red dots) we find that the performance of the VaR model becomes worse. At lower confidence levels the model is too conservative whereas at higher levels it underestimates the frequency of extreme events. Lastly, when comparing the 10-day VaR model that was constructed using a local scaling exponent we find that the performance of the model is much more in line with the base model. The flaws that were introduced with the uniform scaling exponent appear to be corrected for when using local time scaling. This observation is supported by the data as the average distance from the null hypothesis is 0.88% for the uniformly scaled VaR model and 0.47% for the locally scaled model.

To demonstrate that the results originate from the $\sqrt{\text{time}}$ scaling of volatility, we repeat the example with a more sophisticated model setup. Again, we estimate three VaR models. First, a base 1-day VaR model and then two scaled 10-day VaR models of which one is scaled with $\sqrt{10}$ and the other with 10^{H_l}. Now we estimate the base model using an AR(1) GARCH(1,1) model. All other specifications are the same as before. The results are reported in Figure 8.

[24]More technical details about the algorithm can be found in Ihlen (2012) and Ihlen and Vereijken (2013).

[25]We acknowledge that multiplying a VaR forecast with a scaling factor comes with the assumption that the mean of the return series is zero. We keep this assumption as it is implicitly embedded in the regulation.

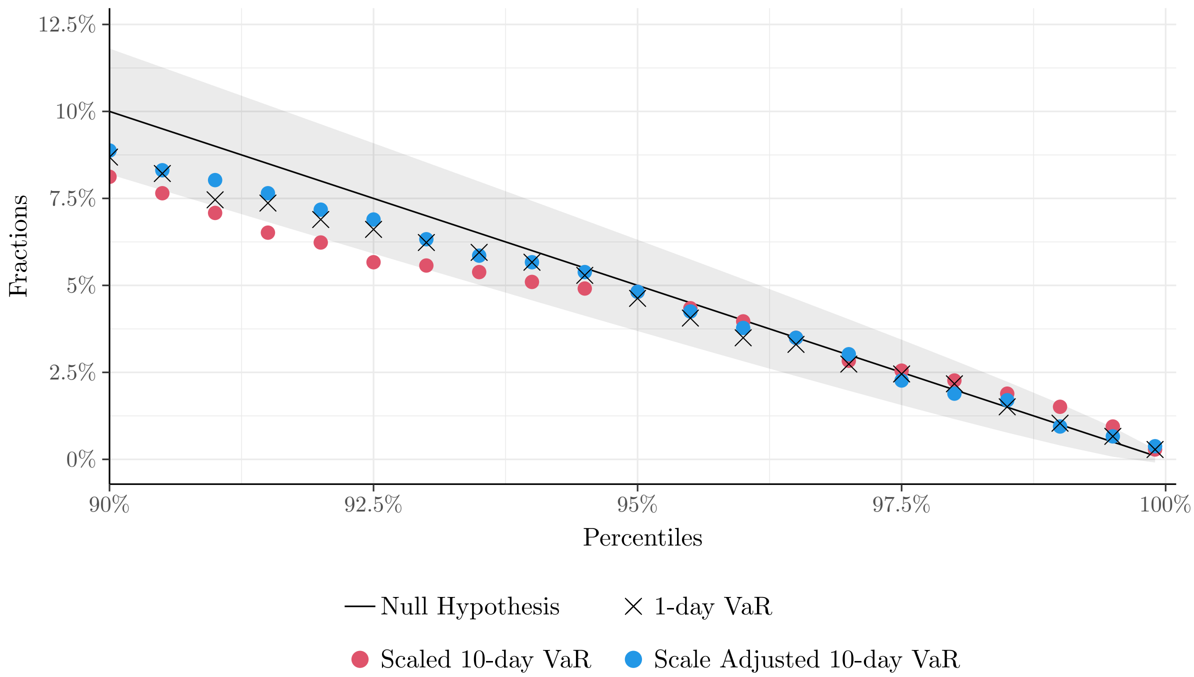

Figure 7: Comparison between scale-adjusted and unadjusted 10-day historical VaR forecast
The figure compares the accuracy of three VaR risk models. The black line refers to the null hypothesis, i.e. that 1% of returns fall below the 99% VaR estimate. The grey area around the hypothesis is the 95% level confidence interval that was calculated assuming normally distributed shocks. The black crosses refer to a 1-day VaR model that was estimated using the historical distribution of the most recent 12 months of data and was evaluated against the following 1-day return. The red dots refer to a scaled version of the 1-day model. This model was scaled with $10^{0.5}$ to retrieve a 10-day VaR estimate and was evaluated against the following 10-day return. The blue dots also refer to a scaled version of the 1-day model. This model was scaled with 10^{H_l} where H_l is the local scaling exponent to arrive at a 10-day VaR estimate and was evaluated against the following 10-day return. The local scaling exponent H_l was estimated based on the algorithm of (Ihlen & Vereijken, 2014) using the past five years of data with interpolation for local scale estimations of 9,10, and 11 days.

The picture is very similar to the one before. Again, the base 1-day VaR forecast is relatively accurate with all forecasts falling into the 95% confidence interval. Scaling this base forecast with $\sqrt{\text{time}}$ worsens its accuracy. Respecting the local scaling behavior corrects the error such that the performance of the forecast is very similar to that of the base. Again, looking at the average distance from the null hypothesis supports this conclusion. The uniformly scaled VaR model misses the null by 1.06% on average whereas the locally scaled model has an average error of 0.62%. Overall, we observe that scaling volatility with $\sqrt{\text{time}}$ worsens the accuracy of VaR forecasts and incorporating the respective nonlinear scaling behavior fixes the problem.

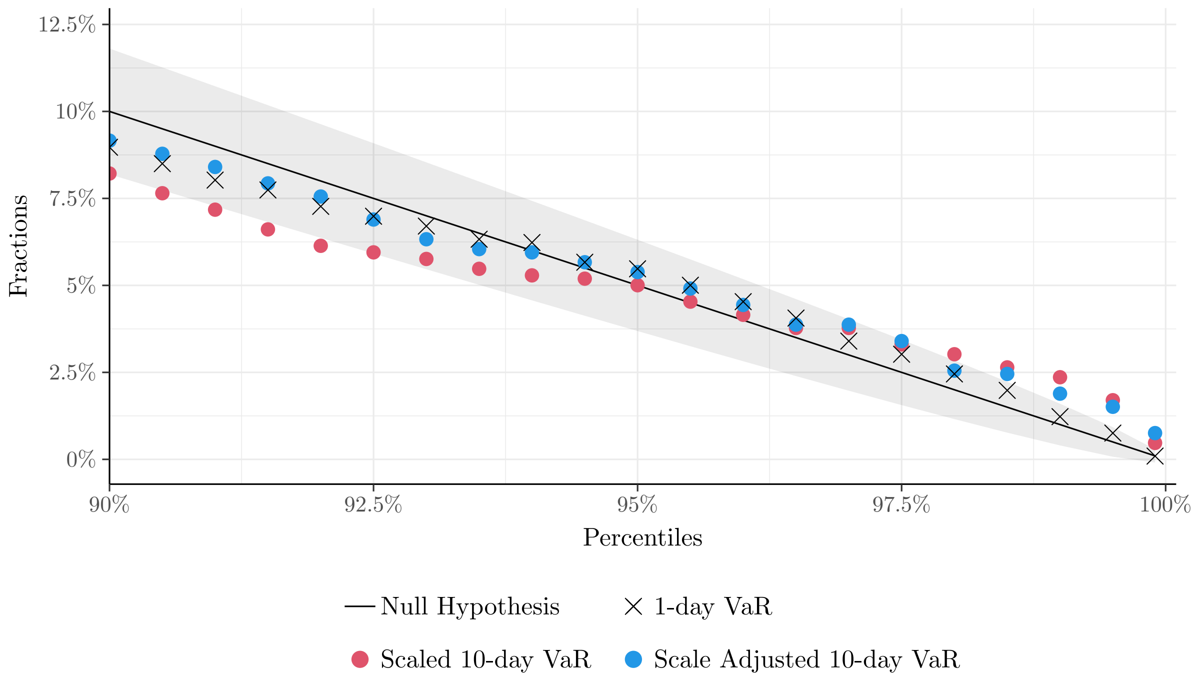

Figure 8: Comparison between scale-adjusted and unadjusted 10-day AR(1) GARCH(1,1) VaR forecast

The figure compares the accuracy of three VaR risk models. The black line refers to the null hypothesis, i.e. that 1% of returns fall below the 99% VaR estimate. The grey area around the hypothesis is the 95% level confidence interval that was calculated assuming normally distributed shocks. The black crosses refer to a 1-day VaR model that was estimated using an AR(1) model for the mean and a GARCH(1,1) model for the variance. We fit both models to the most recent 12 months of data and make a VaR forecast assuming normally distributed shocks that was evaluated against the following 1-day return. The red dots refer to a scaled version of the 1-day model. This model was scaled with $10^{0.5}$ to retrieve a 10-day VaR estimate and was evaluated against the following 10-day return. The blue dots also refer to a scaled version of the 1-day model. This model was scaled with 10^{H_l} where H_l is the local scaling exponent to arrive at a 10-day VaR estimate and was evaluated against the following 10-day return. The local scaling exponent H_l was estimated based on the algorithm of (Ihlen & Vereijken, 2014) using the past five years of data with interpolation for local scale estimations of 9,10, and 11 days.

5 Conclusion

We analyze the effect of the two stylized facts, heavy tails and volatility clusters, on financial stability. We find that financial markets across various asset classes are clearly more destabilized from volatility clusters than from heavy-tailed distributions per se. We also observe that the effect gets more pronounced with an increasing degree of portfolio diversification. Our result has various practical implications of which we explore what we think are the two most pressing: How to hedge against these facts in practice and whether the neglection of volatility clusters cause errors in the regulatory calculation of banks' capital buffers. We find, that the two stylized facts can be hedged with rather simple portfolio strategies for which summary statistics are in line with that of our synthetic portfolios. This, further bolsters our claim that volatility clusters destabilize markets more. With regards to the regulation, we show not only that the neglection of volatility clusters cause errors that can not be corrected by increasing the accuracy of the underlying risk model, but also that the integration of local scaling exponents can fix this issue.

Overall, our findings call for a rethinking of risk. We demonstrate that prioritizing the tails of a distribution for risk management may offer less downside protection compared to a focus on the dependence structure of returns. Additionally, our study is the first to provide estimates of potential shortfalls for two widely used simplifying assumptions: normally distributed and iid returns of which we find that the latter may cause larger damage. Given this results, financial market regulators should consider incorporating measures of volatility clustering into capital requirement regulations. Alternatively, taking proactive steps to reduce volatility clusters in asset returns could also prove beneficial.

The implications of this paper reach beyond the areas of practical hedging and regulation. Therefore, more research is needed to address all of the follow-up questions that evolve from our main finding. For example, for investment managers, the finding that volatility clusters appear to cause more harm to investors raises the question whether a risk premium can be earned with clustering exposure. From a more theoretical perspective, it has to be questioned whether the vast literature of linear asset pricing models is valuable if the nonlinear dependence in asset returns appear to be a large driver of downside risk. We leave the answers to these questions to future analysis.

References

Aloui, C., Shahzad, S. J. H., & Jammazi, R. (2018). Dynamic efficiency of European credit sectors: A rolling-window multifractal detrended fluctuation analysis. *Physica A: Statistical Mechanics and its Applications*, *506*, 337–348.

Alvarez-Ramirez, J., Alvarez, J., Rodriguez, E., & Fernandez-Anaya, G. (2008). Time-varying Hurst exponent for US stock markets. *Physica A: Statistical Mechanics and its Applications*, *387*(24), 6159–6169.

Al-Yahyaee, K. H., Mensi, W., & Yoon, S. M. (2018). Efficiency, multifractality, and the long-memory property of the Bitcoin market: A comparative analysis with stock, currency, and gold markets. *Finance Research Letters*, *27*, 228–234.

Amaya, D., Christoffersen, P., Jacobs, K., & Vasquez, A. (2015). Does realized skewness predict the cross-section of equity returns? *Journal of Financial Economics*, *118*(1), 135–167.

Ang, A., Chen, J., & Xing, Y. (2006). Downside risk. *The Review of Financial Studies*, *19*(4), 1191–1239.

Ang, A., Hodrick, R. J., Xing, Y., & Zhang, X. (2009). High idiosyncratic volatility and low returns: International and further U.S. evidence. *Journal of Financial Economics*, *91*(1), 1–23.

Arestis, P., Demetriades, P. O., & Luintel, K. B. (2001). Financial Development and Economic Growth: The Role of Stock Markets. *Journal of Money, Credit and Banking*, *33*(1), 16–41.

Bachelier, L. (1900). Théorie de la spéculation Annales. *Annales scientifiques de l'É.N.S.*, *17*, 21–86.

Baillie, R. T., Bollerslev, T., & Mikkelsen, H. O. (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *74*(1), 3–30.

Baker, N. L., & Haugen, R. A. (2012). Low Risk Stocks Outperform within All Observable Markets of the World. *SSRN Electronic Journal*.

Bali, T. G., & Murray, S. (2013). Does risk-neutral skewness predict the cross section of equity option portfolio returns? *Journal of Financial and Quantitative Analysis*, *48*(4), 1145–1171.

Barberis, N., & Huang, M. (2008). Stocks as lotteries: The implications of probability weighting for security prices. *American Economic Review*, *98*(5), 2066–2100.

Barunik, J., Aste, T., Di Matteo, T., & Liu, R. (2012). Understanding the source of multifractality in financial markets. *Physica A: Statistical Mechanics and its Applications*, *391*(17), 4234–4251.

BCBS. (2020). MAR: Calculation of RWA for credit risk.

Benhabib, J., Liu, X., & Wang, P. (2016). Sentiments, financial markets, and macroeconomic fluctuations. *Journal of Financial Economics*, *120*(2), 420–443.

Benhabib, J., Liu, X., & Wang, P. (2019). Financial Markets, the Real Economy, and Self-Fulfilling Uncertainties. *The Journal of Finance*, *74*(3), 1503–1557.

Berkowitz, J., & O'Brien, J. (2002). How Accurate Are Value-at-Risk Models at Commercial Banks? *The Journal of Finance*, *57*(3), 1093–1111.

Bertrand, P., & Prigent, J.-l. (2011). Omega performance measure and portfolio insurance. *Journal of Banking and Finance*, *35*(7), 1811–1823.

Black, F., & Scholes, M. (1973). The Pricing of Options and Corporate Liabilities. *Journal of Political Economy*, *81*(3), 637–654.

Blackledge, J., Kearney, D., Lamphiere, M., Rani, R., & Walsh, P. (2019). Econophysics and fractional calculus: Einstein's evolution equation, the fractal market hypothesis, trend analysis and future price prediction. *Mathematics*, *7*(11).

Blackledge, J., & Lamphiere, M. (2022). A Review of the Fractal Market Hypothesis for Trading and Market Price Prediction. *Mathematics*, *10*(1), 1–46.

Blitz, D., Pang, J., & van Vliet, P. (2013). The volatility effect in emerging markets. *Emerging Markets Review*, *16*(1), 31–45.

Blitz, D., & van Vliet, P. (2007). The volatility effect. *The Journal of Portfolio Management*, *34*(1), 102–113.

Bogachev, M. I., Eichner, J. F., & Bunde, A. (2007). Effect of nonlinear correlations on the statistics of return intervals in multifractal data sets. *Physical Review Letters*, *99*(24), 1–4.

Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, *31*(3), 307–327.

Bond, P., Edmans, A., & Goldstein, I. (2012). The Real Effects of Financial Markets. *Annual Review of Financial Economics*, *4*, 339–360.

Boyer, B., Mitton, T., & Vorkink, K. (2010). Expected idiosyncratic skewness. *The Review of Financial Studies*, *23*(1), 170–202.

Brunnermeier, M. K., Gollier, C., & Parker, J. A. (2007). Optimal beliefs, asset prices, and the preference for skewed returns. *American Economic Review*, *97*(2), 159–165.

Buonocore, R. J., Aste, T., & Di Matteo, T. (2016). Measuring multiscaling in financial time-series. *Chaos, Solitons and Fractals*, *88*, 38–47.

Cajueiro, D. O., & Tabak, B. M. (2004). The Hurst exponent over time: Testing the assertion that emerging markets are becoming more efficient. *Physica A: Statistical Mechanics and its Applications*, *336*, 521–537.

Calvet, L., & Fisher, A. (2002). Multifractality in Asset Returns: Theory and Evidence. *The Review of Economics and Statistics*, *84*(3), 381–406.

Cao, J., & Han, B. (2013). Cross section of option returns and idiosyncratic stock volatility. *Journal of Financial Economics*, *108*(1), 231–249.

Caraiani, P. (2012). Evidence of multifractality from emerging European stock markets. *PLoS ONE*, *7*(7), 1–9.

Chang, B. Y., Christoffersen, P., & Jacobs, K. (2013). Market skewness risk and the cross section of stock returns. *Journal of Financial Economics*, *107*(1), 46–68.

Chen, H., & Wu, C. (2011). Forecasting volatility in Shanghai and Shenzhen markets based on multifractal analysis. *Physica A: Statistical Mechanics and its Applications*, *390*(16), 2926–2935.

Chen, W., Wei, Y., Lang, Q., Lin, Y., & Liu, M. (2014). Financial market volatility and contagion effect: A copula-multifractal volatility approach. *Physica A: Statistical Mechanics and its Applications*, *398*, 289–300.

Chung, K. H., Wang, J., & Wu, C. (2019). Volatility and the cross-section of corporate bond returns. *Journal of Financial Economics*, *133*(2), 397–417.

Conrad, J., Dittmar, R. F., & Ghysels, E. (2013). Ex Ante Skewness and Expected Stock Returns. *The Journal of Finance*, *68*(1), 85–124.

Cont, R. (2001). Empirical properties of asset returns: Stylized facts and statistical issues. *Quantitative Finance*, *1*(2), 223–236.

Cont, R. (2007). Volatility clustering in financial markets: Empirical facts and agent-based models. In G. Teyssière & A. P. Kirman (Eds.), *Long memory in economics* (pp. 289–309). Springer.

Danielsson, J., De Haan, L., Peng, L., & De Vries, C. G. (2001). Using a Bootstrap Method to Choose the Sample Fraction in Tail Index Estimation. *Journal of Multivariate Analysis*, *76*(2), 226–248.

Dar, A. B., Bhanja, N., & Tiwari, A. K. (2017). Do global financial crises validate assertions of fractal market hypothesis? *International Economics and Economic Policy*, *14*(1), 153–165.

Detzel, A., Duarte, J., Kamara, A., Siegel, S., & Sun, C. (2019). The Cross-Section of Volatility and Expected Returns: Then Then and Now. *SSRN Electronic Journal*.

Di Matteo, T. (2007). Multi-scaling in finance. *Quantitative Finance*, *7*(1), 21–36.

Di Matteo, T., Aste, T., & Dacorogna, M. M. (2003). Scaling behaviors in differently developed markets. *Physica A: Statistical Mechanics and its Applications*, *324*(1-2), 183–188.

Di Matteo, T., Aste, T., & Dacorogna, M. M. (2005). Long-term memories of developed and emerging markets: Using the scaling analysis to characterize their stage of development. *Journal of Banking and Finance*, *29*(4), 827–851.

Ding, Z., Granger, C. W., & Engle, R. F. (1993). A long memory property of stock market returns and a new model. *Journal of Empirical Finance*, *1*(1), 83–106.

Dittmar, R. F. (2002). Nonlinear pricing kernels, kurtosis preference, and evidence from the cross section of equity returns. *The Journal of Finance*, *57*(1), 369–403.

Drozdz, S., Kwapień, J., Oświęcimka, P., & Rak, R. (2010). The foreign exchange market: Return distributions, multifractality, anomalous multifractality and the Epps effect. *New Journal of Physics*, *12*.

Du, G., & Ning, X. (2008). Multifractal properties of Chinese stock market in Shanghai. *Physica A: Statistical Mechanics and its Applications*, *387*(1), 261–269.

Engle, R. F. (1982). Autoregressive Conditional Heteroscedacity with Estimates of variance of United Kingdom Inflation. *Econometrica*, *50*(4), 987–1008.

Engle, R. F., & Bolerslev, T. (1986). Modeling the Persistence of Conditional Variances. *Econometric Reviews*, *5*(1), 1–50.

Evans, J. L., & Archer, S. H. (1968). Diversification and the Reduction of Dispersion: An Empirical Analysis. *The Journal of Finance*, *23*(5), 761–767.

Fama, E. F. (1970). Efficient Capital Markets : A Review of Theory and Empirical Work. *The Journal of Finance*, *25*(2), 383–417.

Fama, E. F. (1976). *Foundations of Finance*. Basic Books.

Figlewski, S., Chidambaran, N., & Kaplan, S. (1993). Evaluating the Performance of the Protective Put Strategy. *Financial Analysts Journal*, *49*(4), 46–56.

Fleming, J., Kirby, C., & Ostdiek, B. (2001). The economic value of volatility timing with realized jumps. *The Journal of Finance*, *56*(1), 329–352.

Frezza, M. (2014). Goodness of fit assessment for a fractal model of stock markets. *Chaos, Solitons and Fractals*, *66*, 41–50.

Ghazani, M. M., & Khosravi, R. (2020). Multifractal detrended cross-correlation analysis on benchmark cryptocurrencies and crude oil prices. *Physica A: Statistical Mechanics and its Applications*, *560*, 125172.

Grech, D., & Pamuła, G. (2012). Multifractal background noise of monofractal signals. *Acta Physica Polonica A*, *121*(2 B), 34–39.

Grech, D. (2016). Alternative measure of multifractal content and its application in finance. *Chaos, Solitons and Fractals*, *88*, 183–195.

Grech, D., & Pamuła, G. (2013). On the multifractal effects generated by monofractal signals. *Physica A: Statistical Mechanics and its Applications*, *392*(23), 5845–5864.

Green, E., Hanan, W., & Heffernan, D. (2014). The origins of multifractality in financial time series and the effect of extreme events. *European Physical Journal B*, *87*(6), 1–9.

Green, T. C., & Hwang, B. H. (2012). Initial public offerings as lotteries: Skewness preference and first-day returns. *Management Science*, *58*(2), 432–444.

Gu, D., & Huang, J. (2019). Multifractal detrended fluctuation analysis on high-frequency SZSE in Chinese stock market. *Physica A: Statistical Mechanics and its Applications*, *521*, 225–235.

Günay, S. (2016). Performance of the multifractal model of asset returns (MMAR): Evidence from emerging stock markets. *International Journal of Financial Studies*, *4*(2).

Harvey, C. R., Hoyle, E., Korgaonkar, R., Rattray, S., Sargaison, M., & Van Hemert, O. (2018). The impact of volatility targeting. *The Journal of Portfolio Management*, *45*(1), 14–33.

Harvey, C. R., & Siddique, A. (2000). Conditional skewness in asset pricing tests. *The Journal of Finance*, *55*(3), 1263–1295.

He, S., & Wang, Y. (2017). Revisiting the multifractality in stock returns and its modeling implications. *Physica A: Statistical Mechanics and its Applications*, *467*, 11–20.

Hill, B. M. (1975). A Simple General Approach to Inference About the Tail of a Distribution. *The Annals of Statistics*, *3*(5), 1163–1174.

Hsu, D. A., Miller, R. B., & Wichern, D. W. (1974). On the stable paretian behavior of stock-market prices. *Journal of the American Statistical Association*, *69*(345), 108–113.

Ihlen, E. A. (2012). Introduction to multifractal detrended fluctuation analysis in Matlab. *Frontiers in Physiology*, *3*(141).

Ihlen, E. A., & Vereijken, B. (2013). Identifying multiplicative interactions between temporal scales of human movement variability. *Annals of Biomedical Engineering*, *41*(8), 1635–1645.

Ihlen, E. A., & Vereijken, B. (2014). Detection of co-regulation of local structure and magnitude of stride time variability using a new local detrended fluctuation analysis. *Gait and Posture*, *39*(1), 466–471.

Jiang, Z. Q., Xie, W. J., & Zhou, W. X. (2014). Testing the weak-form efficiency of the WTI crude oil futures market. *Physica A: Statistical Mechanics and its Applications*, *405*, 235–244.

Jiang, Z. Q., Xie, W. J., Zhou, W. X., & Sornette, D. (2019). Multifractal analysis of financial markets: A review. *Reports on Progress in Physics*, *82*(12).

Jin, C., Conlon, T., & Cotter, J. (2022). Co-Skewness across Return Horizons. *Journal of Financial Econometrics*, *22-14*.

Joshipura, N., & Joshipura, M. (2016). The Volatility Effect: Evidence from Indian Markets. *Applied Finance Letters*, *5*(1), 12–17.

Joshipura, N., & Joshipura, M. (2019). The Volatility Effect: Recent Evidence from Indian Markets. *Theoretical Economics Letters*, *09*(06), 2152–2164.

Kantelhardt, J. W., Zschiegner, S. A., Koscielny-Bunde, E., Havlin, S., Bunde, A., & Stanley, H. E. (2002). Multifractal detrended fluctuation analysis of nonstationary time series. *Physica A: Statistical Mechanics and its Applications*, *316*, 87–114.

Keylock, C. J. (2019). Hypothesis Testing for Nonlinear Phenomena in the Geosciences Using Synthetic, Surrogate Data. *Earth and Space Science*, *6*(1), 41–58.

Keylock, C. J. (2017). A multifractal surrogate data generation algorithm that preserves pointwise Holder regularity structure , with initial applications to turbulence. *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, *95*(032123).

Kraus, A., & Litzenberger, R. H. . (1976). Skewness Preference and the Valuation of Risk Assets. *The Journal of Finance*, *31*(4), 1085–1100.

Kraus, A., & Litzenberger, R. H. . (1983). On the Distributional Conditions for a Consumption-oriented Three Moment CAPM. *The Journal of Finance*, *38*(5), 1381–1391.

Kristoufek, L. (2013). Fractal markets hypothesis and the global financial crisis: Wavelet power evidence. *Scientific Reports*, *3*.

Kumar, S., & Deo, N. (2009). Multifractal properties of the Indian financial market. *Physica A: Statistical Mechanics and its Applications*, *388*(8), 1593–1602.

Kwapień, J., Oświęcimka, P., & Drozdz, S. (2005). Components of multifractality in high-frequency stock returns. *Physica A: Statistical Mechanics and its Applications*, *350*(2-4), 466–474.

Lamphiere, M., Blackledge, J., & Kearney, D. (2021). Carbon futures trading and short-term price prediction: An analysis using the fractal market hypothesis and evolutionary computing. *Mathematics*, *9*(9).

Li, D., Nishimura, Y., & Men, M. (2014). Fractal markets : Liquidity and investors on different time horizons. *Physica A: Statistical Mechanics and its Applications*, *407*, 144–151.

Li, D., Nishimura, Y., & Men, M. (2016). Why the long-term auto-correlation has not been eliminated by arbitragers: Evidences from NYMEX. *Energy Economics*, *59*, 167–178.

Lo, A. W. (1991). Long-Term Memory in Stock Market Prices. *Econometrica*, *59*(5), 1279–1313.

López, J. L., & Contreras, J. G. (2013). Performance of multifractal detrended fluctuation analysis on short time series. *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, *87*(2), 1–8.

Lux, T. (2001). Turbulence in financial markets: The surprising explanatory power of simple cascade models. *Quantitative Finance*, *1*(6), 632–640.

Lux, T., & Alfarano, S. (2016). Financial power laws: Empirical evidence, models, and mechanisms. *Chaos, Solitons and Fractals*, *88*, 3–18.

Lv, X., & Shan, X. (2013). Modeling natural gas market volatility using GARCH with different distributions. *Physica A*, *392*(22), 5685–5699.

Mandelbrot, B. (1963). The Variation of Certain Speculative Prices. *The Journal of Business*, *36*(4), 394.

Mao, J. C. T. (1970). Essentials of Portfolio Diversification Strategy. *The Journal of Finance*, *25*(5), 1109–1121.

Matia, K., Ashkenazy, Y., & Stanley, H. E. (2003). Multifractal properties of price fluctuations of stocks and commodities. *Europhysics Letters*, *61*(3), 422–428.

McKeon, R., & Svetina, M. (2017). Protecting against Loss: Protective Put Strategies versus Stop-Loss Strategies. *The Journal of Investing*, *26*(3), 65–76.

Mitton, T., & Vorkink, K. (2007). Equilibrium underdiversification and the preference for skewness. *The Review of Financial Studies*, *20*(4), 1255–1288.

Moreira, A., & Muir, T. (2017). Volatility-Managed Portfolios. *The Journal of Finance*, *72*(4), 1611–1644.

Mukli, P., Nagy, Z., & Eke, A. (2015). Multifractal formalism by enforcing the universal behavior of scaling functions. *Physica A: Statistical Mechanics and its Applications*, *417*, 150–167.

Mylnikov, G. (2021). Volatility Targeting: It's Complicated! *The Journal of Portfolio Management*, *47*(8), 58–74.

Nelson, D. B. (1991). Conditional Heteroskedasticity in Asset Returns: A New Approach. *Econometrica*, *59*(2), 347–370.

Norouzzadeh, P., & Rahmani, B. (2006). A multifractal detrended fluctuation description of Iranian rial-US dollar exchange rate. *Physica A: Statistical Mechanics and its Applications*, *367*, 328–336.

O'Brien, J., & Szerszeń, P. J. (2017). An evaluation of bank measures for market risk before, during and after the financial crisis. *Journal of Banking and Finance*, *80*, 215–234.

Officer, R. R. (1972). The Distribution of Stock Returns. *Journal of the American Statistical Association*, *67*(340), 807–812.

Osborne, M. F. M. (1959). Brownian Motion in the Stock Market. *Operations Research*, *7*(2), 145–173.

Paluš, M. (2008). Bootstrapping multifractals: Surrogate data from random cascades on wavelet dyadic trees. *Physical Review Letters*, *101*(13), 1–4.

Pamuła, G., & Grech, D. (2014). Influence of the maximal fluctuation moment order q on multifractal records normalized by finite-size effects. *EPL*, *105*, 50004.

Peters, E. E. (1991). *Chaos and Order in the Capital Markets: A New View of Cycles, Prices, and Market Volatility*. John Wiley & Sons.

Peters, E. E. (1994). *Fractal market analysis: Applying chaos theory to investment and economics*. John Wiley & Sons.

Poon, S. H., Rockinger, M., & Tawn, J. (2004). Extreme value dependence in financial markets: Diagnostics, models, and financial implications. *Review of Financial Studies*, *17*(2), 581–610.

Rachev, S. T., Weron, A., & Weron, R. (1999). CED model for asset returns and fractal market hypothesis. *Mathematical and Computer Modelling*, *29*(10-12), 23–36.

Rak, R., & Grech, D. (2018). Quantitative approach to multifractality induced by correlations and broad distribution of data. *Physica A: Statistical Mechanics and its Applications*, *508*, 48–66.

Resnick, S., & Stărică, C. (1997). Smoothing the Hill Estimator. *Advances in Applied Probability*, *29*(1), 271–293.

Rubinstein, M. (1973). The Fundamental Theorem of Parameter-Preference Security Valuation. *Journal of Financial and Quantitative Analysis*, *8*(1), 61–69.

Schadner, W. (2022). U.S. Politics from a multifractal perspective. *Chaos, Solitons and Fractals*, *155*, 111677.

Schmitt, F., Schertzer, D., & Lovejoy, S. (2000). Multifractal Fluctuations in Finance. *International Journal of Theoretical and Applied Finance*, *3*(3), 361–364.

Schmitt, F., Schertzer, D., & Lovejoy, S. (1999). Multifractal analysis of foreign exchange data. *Applied Stochastic Models and Data Analysis*, *15*(1), 29–53.

Schneider, P. G., Wagner, C., & Zechner, J. (2020). Low-Risk Anomalies? *The Journal of Finance*, *75*(5), 2673–2718.

Schreiber, T., & Schmitz, A. (1996). Improved surrogate data for nonlinearity tests. *Physical Review Letters*, *77*(4), 635–638.

Schreiber, T., & Schmitz, A. (2000). Surrogate time series. *Physica D: Nonlinear Phenomena*, *142*(3-4), 346–382.

Segnon, M., & Trede, M. (2018). Forecasting market risk of portfolios: copula-Markov switching multifractal approach. *European Journal of Finance*, *24*(14), 1123–1143.

Shahzad, S. J. H., Nor, S. M., Mensi, W., & Kumar, R. R. (2017). Examining the efficiency and interdependence of US credit and stock markets through MF-DFA and MF-DXA approaches. *Physica A: Statistical Mechanics and its Applications*, *471*, 351–363.

Shao, Y. H. (2020). Does Crude Oil Market Efficiency Improve After the Lift of the U.S. Export Ban? Evidence From Time-Varying Hurst Exponent. *Frontiers in Physics*, *8*, 551501.

Shao, Y. H., Gu, G. F., Jiang, Z. Q., Zhou, W. X., & Sornette, D. (2012). Comparing the performance of FA, DFA and DMA using different synthetic long-range correlated time series. *Scientific Reports*, *2*, 835.

Sockin, M. (2015). Not so Great Expectations : A Model of Growth and Informational Frictions. *UT Austin Working Paper*.

Statman, M. (1987). How Many Stocks Make a Diversified Portfolio? *The Journal of Financial and Quantitative Analysis*, *22*(3), 353–363.

Stosic, D., Stosic, D., Ludermir, T. B., & Stosic, T. (2019). Multifractal behavior of price and volume changes in the cryptocurrency market. *Physica A: Statistical Mechanics and its Applications*, *520*, 54–61.

Straub, L., & Ulbricht, R. (2015). Endogenous Uncertainty and Credit Crunches. *SSRN Electronic Journal*.

Suárez-García, P., & Gómez-Ullate, D. (2014). Multifractality and long memory of a financial index. *Physica A: Statistical Mechanics and its Applications*, *394*, 226–234.

Tseng, J. J., & Li, S. P. (2011). Asset returns and volatility clustering in financial time series. *Physica A: Statistical Mechanics and its Applications*, *390*(7), 1300–1314.

Tseng, J. J., & Li, S. P. (2012). Quantifying volatility clustering in financial time series. *International Review of Financial Analysis*, *23*, 11–19.

Turiel, A., & Pérez-Vicente, C. J. (2003). Multifractal geometry in stock market time series. *Physica A: Statistical Mechanics and its Applications*, *322*, 629–649.

Turiel, A., & Pérez-Vicente, C. J. (2005). Role of multifractal sources in the analysis of stock market time series. *Physica A: Statistical Mechanics and its Applications*, *355*(2-4), 475–496.

Upson, R. B., Jessup, P. F., & Matsumoto, K. (1975). Portfolio Diversification Strategies. *Financial Analysts Journal*, *31*(3), 86–88.

Vaseghi, S. V. (2008). Power Spectrum Analysis. In S. V. Vaseghi (Ed.), *Advanced digital signal processing and noise reduction* (4th ed., pp. 271–294). John Wiley & Sons.

Venema, V., Ament, F., & Simmer, C. (2006). Nonlinear Processes in Geophysics A Stochastic Iterative Amplitude Adjusted Fourier Transform algorithm with improved accuracy. *Nonlinear Processes in Geophysics*, *13*, 321–328.

Wang, H. Y., & Wang, T. T. (2018). Multifractal analysis of the Chinese stock, bond and fund markets. *Physica A: Statistical Mechanics and its Applications*, *512*, 280–292.

Wang, Y., Wu, C., & Pan, Z. (2011). Multifractal detrending moving average analysis on the US Dollar exchange rates. *Physica A: Statistical Mechanics and its Applications*, *390*(20), 3512–3523.

Wei, Y., & Wang, P. (2008). Forecasting volatility of SSEC in Chinese stock market using multifractal analysis. *Physica A: Statistical Mechanics and its Applications*, *387*(7), 1585–1592.

Weron, A., & Weron, R. (2000). Fractal market hypothesis and two power-laws. *Chaos, Solitons and Fractals*, *11*, 289–296.

Zhou, W. X. (2009). The components of empirical multifractality in financial returns. *Europhysics Letters*, *88*(28004).

Zhou, W. X. (2012). Finite-size effect and the components of multifractality in financial volatility. *Chaos, Solitons and Fractals*, *45*(2), 147–155.

Appendices

A Additional Tables and Figures

	Stocks	P.E.	Bonds	F.X.	Crypto	Comm.	Real Est.
				t-test			
statistic	7.204	2.513	33.249	14.739	12.881	12.052	2.073
p value	6.39e-12	6.39e-03	3.40e-82	3.43e-33	9.40e-28	1.66e-25	1.98e-02
				Wilcoxon Test			
statistic	7676	6025	9984	9257	8900	8847	5892
p value	3.13e-11	6.15e-03	2.07e-34	1.24e-25	8.02e-22	2.77e-21	1.47e-02

Table A.1: Test for statistical significance in drawdown differences between series without heavy tails and series without volatility clusters

The table holds whether the difference in maximum drawdowns between the series without heavy tails and the series without volatility clusters is significant for seven indices of different asset classes. Tested hypothesis: maximum drawdowns of surrogates without heavy tails are significantly larger than drawdowns of surrogates without volatility clusters. Result: Both, simple t-tests and non-parametric Wilcoxon tests indicate that volatility clustering causes significantly larger maximum drawdowns than heavy-tails do. Each surrogate series comprises 200 surrogates. The surrogates without heavy tails were created using the IAAWT_n method. The surrogates without volatility clusters were created using the IAAFT method.

	Original	Tail Hedge	Cluster Hedge
		Panel A	
Mean	8.14%	5.94%	7.65%
Volatility	19.57%	14.10%	14.71%
Sharpe Ratio	0.42	0.42	0.52
Max. Drawdown	-54.12%	-40.90%	-40.55%
Worst YoY Loss	-47.50%	-31.55%	-28.54%
Skewness	-0.38	0.05	-0.60
Kurtosis	10.79	4.38	2.81
Tail Index	0.34***	0.21***	0.26***
Nonlin. Dep.	0.20***	0.16***	0.27***
Cluster Index	2.09***	1.75***	1.32*
		Panel B	
Mean	8.14%	8.13%	8.14%
Volatility	19.57%	19.30%	15.64%
Sharpe Ratio	0.42	0.42	0.52
Max. Drawdown	-54.12%	-51.34%	-42.47%
Worst YoY Loss	-42.34%	-37.05%	-29.02%
Skewness	-0.38	0.05	-0.60
Kurtosis	10.79	4.38	2.81
Tail Index	0.34***	0.21***	0.26***
Nonlin. Dep.	0.21***	0.16***	0.26***
Cluster Index	2.09***	1.75***	1.32*

Table A.2: Summary statistics of alternative practical portfolios

The table compares descriptive statistics for the original S&P 500 index returns as well as for a tail hedged and a volatility cluster hedged portfolio of the S&P 500 index. The time series range from the beginning of 1999 until the end of 2021. The tail hedged portfolio is represented by the CBOE S&P 500 Put Protection Index at a level of 10%. The volatility hedged portfolio is represented by the S&P 500 Daily Risk Control 15% Index. *Panel A* holds the results for the unscaled portfolios. *Panel B* holds the results for the hedged portfolios that are scaled to have the same mean return as the original portfolio. The tail index is calculated for the lower tail of the return distribution using the reciprocal of the Hill (1975) estimator with a threshold of $k = 60$. Nonlinear dependence is measured by the excess spreads in Hölder exponents using the MF-DFA method of Kantelhardt et al. (2002). The cluster index is calculated following Tseng and Li (2011). Statistical significance at confidence levels of 10%, 5%, and 1% are indicated with *, **, and *** respectively.

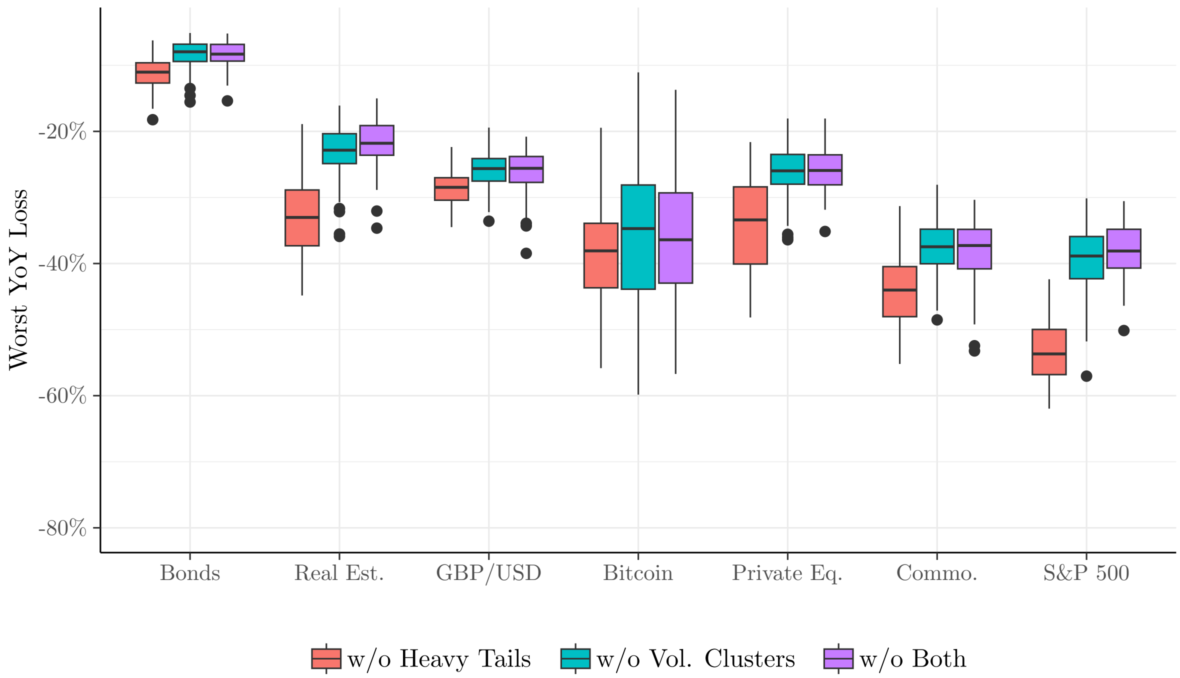

Figure A.1: Worst year-over-year returns of the different asset classes after disentangling stylized facts

The figure holds worst year-over-year returns for three surrogates series of seven indices. Each surrogate series comprises 200 surrogates. The surrogates without heavy tails were created using the IAAWT_n method. The surrogates without volatility clusters were created using the IAAFT method. The surrogates without both stylized facts were created using the IAAFT_n method.

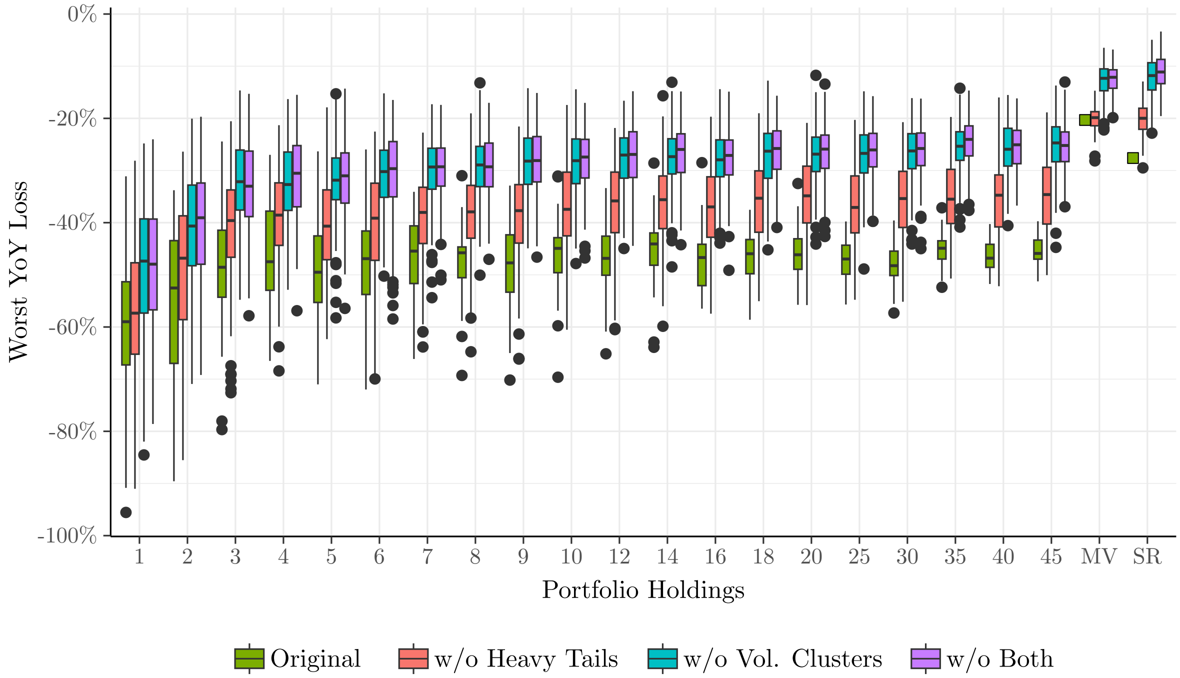

Figure A.2: Worst year-over-year returns of stock portfolios with different holdings after disentangling stylized facts

The figure holds worst year-over-year returns for the original portfolio and three surrogate series of stock portfolios with different holdings. The portfolios are composed using daily log-returns of randomly chosen constituents of the S&P 500 index for which return data is available between 01/01/1992 and 31/12/2021, resulting in an investible pool of 144 stocks. For each number of holdings, 40 equal-weighted portfolios as well as the minimum variance (MV) and maximum Sharpe ratio (SR) portfolios were formed. For each portfolio, we create five surrogates leading to a total number of 200 time series per portfolio holding for the equal-weighted portfolios. For the optimal MV and SR portfolios we create 200 surrogates. The surrogates without heavy tails were created using the IAAWT$_n$ method. The surrogates without volatility clusters were created using the IAAFT method. The surrogates without both stylized facts were created using the IAAFT$_n$ method.

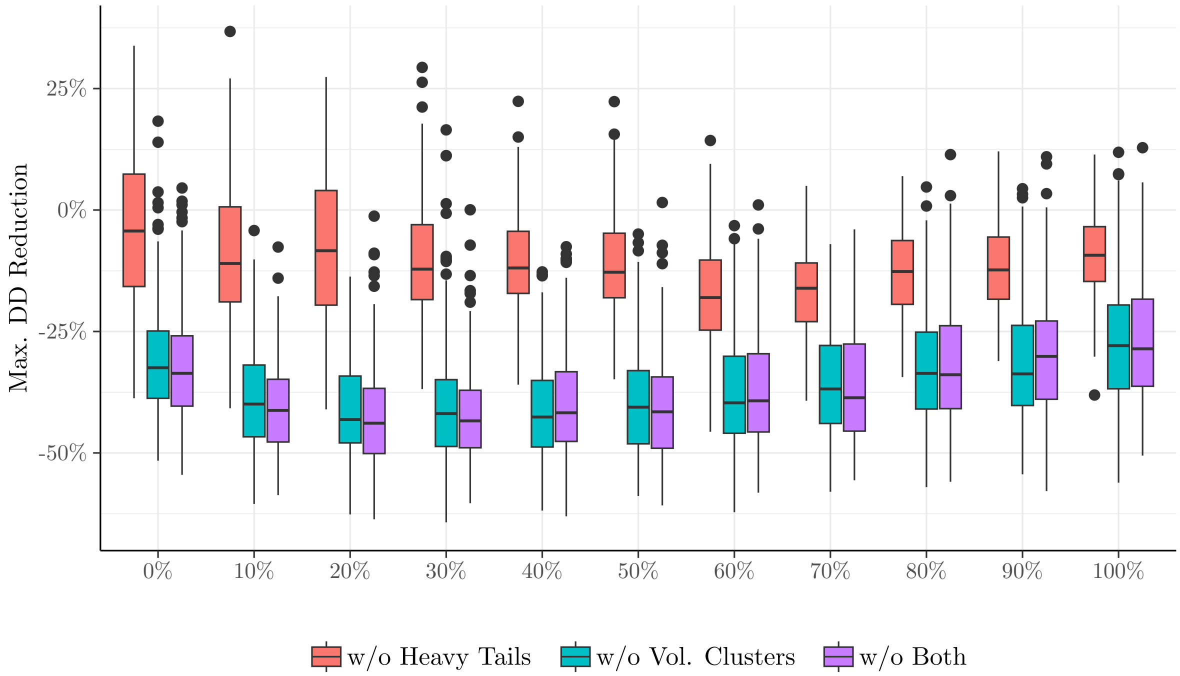

Figure A.3: Relative average maximum drawdown reduction of stock-bond index portfolios after disentangling stylized facts

The figure holds the relative maximum drawdown reductions for three surrogate series of portfolios that consist of a stock and bond index at different portfolio weights. The y-axis describes the portfolio weight of the stock index. Drawdown reductions were calculated compared to the original portfolio of the two indices. The portfolios were composed using daily log-returns between 01/03/1995 and 31/12/2021. Each surrogate series comprises 200 surrogates. The surrogates without heavy tails were created using the IAAWT$_n$ method. The surrogates without volatility clusters were created using the IAAFT method. The surrogates without both stylized facts were created using the IAAFT$_n$ method.

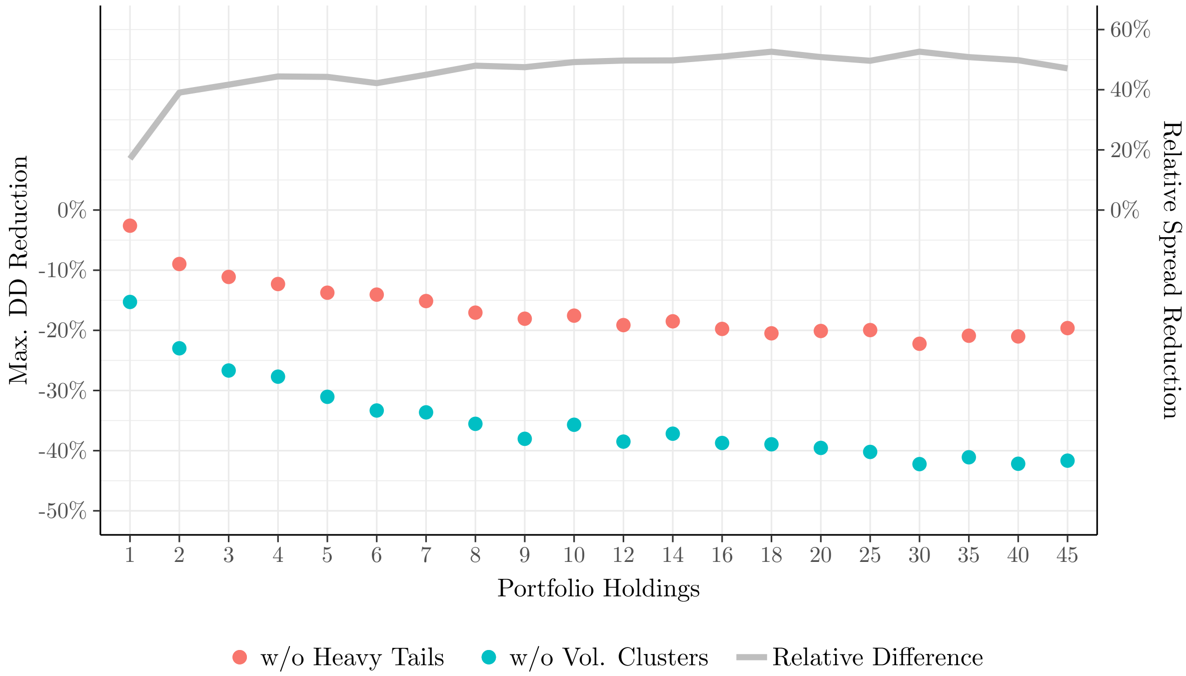

Figure A.4: Relative maximum drawdown reduction of stock portfolios with different holdings after disentangling stylized facts

The figure holds mean maximum drawdown reductions for two surrogate series of stock portfolios with different holdings. The portfolios are composed using daily log-returns of randomly chosen constituents of the S&P 500 index for which return data is available between 01/01/1992 and 31/12/2021, resulting in an investible pool of 144 stocks. For each number of holdings 40 equal-weighted portfolios were formed. For each portfolio, we create five surrogates leading to a total number of 200 time series per portfolio holding. The surrogates without heavy tails were created using the IAAWT$_n$ method. The surrogates without volatility clusters were created using the IAAFT method. The dots show the relative reduction of maximum drawdowns of the surrogate time series relative to the drawdown of the original time series on the left axis. The grey line shows the ratio between the two time series on the right axis. It thus measures the ratio between the relative reductions of spreads.

B Introduction to Nonlinear Dependence

While most readers with a background in economics are familiar with heavy-tailed distributions, the concept of nonlinear dependence may require a little more explanation. A time series is said to exhibit nonlinear dependence whenever the autocorrelation between observations varies in time. For example, imagine you observe the return series of an asset. You find that ten years ago the return of a given day was always positively correlated to the return of the previous day. Also, you find that this pattern is not observable for the past year. This exemplary asset is exposed to nonlinear dependence because its autocorrelation structure was not constant over time. Note, that the presence of this time-varying autocorrelation structure does not violate the stylized fact that (linear) autocorrelations in returns do not exist over longer periods.

In general, the serial dependence in time series measures the series' self-similarity and nonlinear dependence measure how those similarity patterns change over time. Before we get to the measurement of nonlinear dependence, we will first cover some basic characteristics of series that can be exposed to nonlinear dependence— so-called fractal systems. Fractal systems usually fulfill the properties of being self-similar and having no characteristic scale. We will describe both concepts briefly below. However, a fractal system is not necessarily exposed to nonlinear dependence.

To understand the concept of self-similarity, take, for example, a tree. It can be seen as a complex network of branches that become continually smaller. You will notice that if you zoom in, the shape of successive branches resembles that of the previous generation so the structure of the whole tree exhibits self-similarity. We can take the analogy of examining a tree from different levels of resolution to financial markets. Similarly, we can study the similarity of financial return series at different time scales. An example is to examine whether long-term economic cycles exhibit the same degree of self-similarity as individual trading days or months. We will notice that there are indeed patterns that appear to be similar. However, the measurement of the degree of self-similarity is not a simple task to do. This is where the property of not having a characteristic scale comes into play.

If we think about self-similar patterns we will notice that the series under investigation exhibits some level of persistence. A way to measure that persistence is to quantify the distance traveled between two points. For illustrative purposes, think about the coastline of the United Kingdom. Same as with the tree, it exhibits a degree of self-similarity as certain sections of the coastline may look very similar to others and they are not formed completely random. If you were to measure the length of that coastline between two points with an imaginary mile-long ruler, you would end up with a certain length. If you now take a meter-long ruler, the resulting length would grow as you can measure small twists and turns more accurately. Consequently, the length you measure depends on your ruler length (or scale).

In fact, it grows as a power-law function of your ruler length. Also, it is hard to argue which ruler is the correct one to use. Hence, the coastline has no characteristic scale.[26] Applying this example to financial markets, the outcome of the persistence estimation strongly depends on the time scaling of returns (the ruler). We can be certain that if we calculate the sum of absolute returns (the distance traveled) for any time interval of a financial return series using daily returns the result would be smaller as if we used 5-minute returns instead. Same as with the coastline, there is no 'true' or 'characteristic' scale to use for financial time series to quantify the distance traveled and different scales lead to different results.

Having covered the basic two properties of fractal systems, we can conclude that financial return series are of fractal nature. With this foundation, we can now formalize the concept of nonlinear dependence in financial return series. A way of measuring the persistence (distance between two points) in a financial series is to compute the volatility of returns over that time interval defined as

$$V_t(s) = \sqrt{\mathbb{E}\left[(R_t(s) - \mathbb{E}[R_t(s)])^2\right]} \tag{B.1}$$

for a log-return series $R_t(s)$ over a time window of s from time t to $t+s$. $\mathbb{E}$ denotes expectations to enhance readability. According to Hurst (1951), the variation of fractal time series can be described through a power-law relationship, where the range of variation increases in proportion to a power of the horizon (scale) over which the time series is viewed. Following this insight we can expect $V(s)$ to exhibit the power-law relation

$$V(s) = V_0 s^H \tag{B.2}$$

where V_0 denotes the base level of the volatility. Here, we are only interested in the scaling relation with respect to the time scale s so we can reduce the equation to a proportional relation:

$$V(s) \sim s^H \tag{B.3}$$

[26] This discovery was first made by Mandelbrot (1967).

In the above equations, H denotes the scaling exponent that is also known as the Hurst exponent (Hurst, 1956). There is vast empirical evidence that the power-law scaling properties of Equation B.3 hold for financial returns (see Jiang et al., 2019; Lux & Alfarano, 2016, for a profound review). For price-like series, H has the property of $H \in (0, 1)$ and is a measure of the returns' autocorrelation or persistence.

If returns are iid they are uncorrelated and $H = 0.5$. If $H > 0.5$ returns are persistent whereas if $H < 0.5$ they are anti-persistent. In finance, the iid assumption is widely employed which is why the volatility of returns is commonly scaled via the square-root of time (i.e., $\sqrt{s}$ or $s^{0.5}$). Nonlinear dependence refers to the circumstance that H is not constant but rather varies in s. This is also called 'multifractality' or 'multiscaling'. For financial return series, this means that returns in small time scales may exhibit higher (or lower) autocorrelation than in larger time scales. The degree of nonlinear dependence measures the fluctuation of H in s (i.e., the range of H across time scales). In general, the degree of autocorrelation has a strong influence on volatility and thus drawdowns and tail risk. If there is a high degree of autocorrelation, returns tend to trend and deviate strongly from their mean, leading to high volatility. Whereas low to negative autocorrelation will cause returns to cancel out each other and stay closer to their mean, leading to lower volatility. This is where the link between nonlinear dependence and volatility clusters is established.

In the presence of high nonlinear dependence, scaling volatility via a constant factor can lead to estimation errors as each time scale requires its unique scaling exponent. A simple example of these errors can be found in Table B.1. The table holds the volatilities for log-returns of the S&P 500 index at different time scales. Note, that the underlying return series are daily returns and higher return intervals were generated using accumulation. Thus, the sample size for the volatility estimates is constant for each time interval but samples overlap. The bold numbers are the calculated volatilities using the return series of the same time scale. The remaining entries in each row hold volatility estimates from a different time scale that were scaled up or down following the common convention of scaling via the square-root of time, $\sqrt{s}$. The values in parentheses underneath the scaled estimates show the respective estimation error. The results show that using the common scaling convention can lead to large errors. For example, estimating yearly return volatility using the scaled daily return volatility yields an overestimation of 19%. We demonstrate that the scaling errors can not be traced back to simple linear dependencies by fitting an ARMA(1,1) model to the daily return series and doing the same scaling evaluation for its residuals in Table B.2. It can be seen that the values are almost identical so controlling for a linear model does not improve scaling errors.

	Daily	Weekly	Monthly	Yearly	10 Years
Daily	**1.21%**	1.08%	1.06%	1.02%	0.92%
		(-11%)	*(-13%)*	*(-16%)*	*(-24%)*
Weekly	2.71%	**2.42%**	2.73%	2.23%	2.01%
	(12%)		*(13%)*	*(-8%)*	*(-17%)*
Monthly	6.64%	4.19%	**4.72%**	4.65%	4.18%
	(41%)	*(-11%)*		*(-2%)*	*(-11%)*
Yearly	19.17%	17.45%	16.36%	**16.10%**	14.48%
	(19%)	*(8%)*	*(2%)*		*(-10%)*
10 Years	60.62%	55.19%	51.73%	50.92%	**45.80%**
	(32%)	*(21%)*	*(13%)*	*(11%)*	

Table B.1: Volatility scaling errors of the S&P 500 index

This table reports volatilities of the S&P 500 index log returns at different return intervals between 1 January 1996 and 31 December 2021. The returns are scaled to the respective intervals via accumulation. Each row holds the volatility of returns for each period. The bold numbers are the actual volatilities as calculated from the respective return interval. The remaining entries in each row hold the volatility estimates from a different interval that are scaled down/up following the common convention of multiplying the volatility estimate with the square-root of the respective time interval, $s^{0.5}$, to which the volatility should be scaled. The numbers in parentheses underneath each row report the estimation error of the scaled volatility estimates.

	Daily	Weekly	Monthly	Yearly	10 Years
Daily	**1.22%**	1.07%	1.04%	1.00%	0.90%
		(-12%)	*(-14%)*	*(-18%)*	*(-26%)*
Weekly	2.72%	**2.39%**	2.69%	2.20%	1.97%
	(14%)		*(12%)*	*(-8%)*	*(-18%)*
Monthly	6.66%	4.14%	**4.65%**	4.58%	4.10%
	(43%)	*(-11%)*		*(-2%)*	*(-12%)*
Yearly	19.21%	17.24%	16.11%	**15.85%**	14.21%
	(21%)	*(9%)*	*(2%)*		*(-10%)*
10 Years	60.76%	54.51%	50.96%	50.12%	**44.95%**
	(35%)	*(21%)*	*(13%)*	*(12%)*	

Table B.2: ARMA adjusted volatility scaling errors of the S&P 500 index

This table reports volatilities of the residuals of an ARMA(1,1) model that was fit to S&P 500 index log returns between 1 January 1996 and 31 December 2021. The residuals are handled like returns and scaled to different intervals via accumulation. Each row holds the volatility of residuals for each period. The bold numbers are the actual volatilities as calculated from the respective residual interval. The remaining entries in each row hold the volatility estimates from a different interval that are scaled down/up following the common convention of multiplying the volatility estimate with the square-root of the respective time interval, $s^{0.5}$, to which the volatility should be scaled. The numbers in parentheses underneath each row report the estimation error of the scaled volatility estimates.

C Regression analysis

In order to gain further insights into the characteristics preserved by our surrogate series, we employ a regression approach that controls for different characteristics of financial time series. Specifically, we regress the original series as well as our surrogates on lagged values of itself and/or transformations thereof. Subsequently, we examine the residuals derived from this regression to uncover valuable information regarding their distribution and dependence structure. We conduct the analysis using the S&P 500 series as an illustrative example. Throughout the procedures we account for linear autocorrelation, nonlinear momentum/reversals, time-varying means, and linear autocorrelation in volatility. The regression equations utilized in our investigation are as follows:

$$r_t = r_{t-1} + \cdots + r_{t-L} + \epsilon_t \tag{C.1}$$

$$r_t = D_{t-1} r_{t-1} + (1 - D_{t-1}) r_{t-1} + \cdots + D_{t-L} r_{t-L} + (1 - D_{t-L}) r_{t-L} + u_t \tag{C.2}$$

$$r_t = \dot{\mu}_{t-1} + \ddot{\mu}_{t-1} + \dddot{\mu}_{t-1} + v_t \tag{C.3}$$

$$(r_t - \dddot{\mu}_t)^2 = (r_{t-1} - \dddot{\mu}_{t-1})^2 + \cdots + (r_{t-L} - \dddot{\mu}_{t-L})^2 + w_t \tag{C.4}$$

where r_t denotes the observation at time t, and L represent the number of lags considered. D_t is a binary dummy, which takes the value 1 if $r_t > 0$ and 0 if $r_t < 0$. $\dot{\mu}$, $\ddot{\mu}$, and $\dddot{\mu}$ correspond to the mean return of the previous week, month, and year, respectively. ϵ, u_t, v_t, and w_t are the error terms in our models. To ensure consistency, we conduct the regressions with $L = 100$. Our primary focus lies on the four error terms, as we are interested in observing their distribution and their interdependencies after having controlled for various features in the regression setting.

In our analysis, Equation C.1 explores the extent to which r_t can be explained by its linear autocorrelation structure. The objective of Equation C.2 is to examine the role of reversals and non-linear momentum in explaining the behavior of the time series. Equation C.3 assesses the influence of time-varying means on the time series. Specifically, this model examines whether the mean of the data displays changes on a weekly, monthly, or yearly basis. Lastly, Equation C.4 investigates the volatility of r_t by exploring the linear dependence that can explain the variability in the time series.

The results of the analysis are captured in Figure C.1, Figure C.2, and Table C.1. Figure C.1 plots the error terms of the four regression models for the four different time series across time, thus capturing the clustering of residuals. It can be observed that for all regression models, we can see strong clustering behavior of the residuals for the original series and the series without heavy tails. For the series without volatility clusters and without both stylized facts, such clustering cannot be observed. This means that the clustering cannot be explained by our four controls. Also, this shows that the clustering behavior of the original series is still preserved in our surrogate series without heavy tails. The last two columns of

Table C.1 confirms these visual observations. It holds the degree of nonlinear dependence and the clustering index by Tseng and Li (2011).

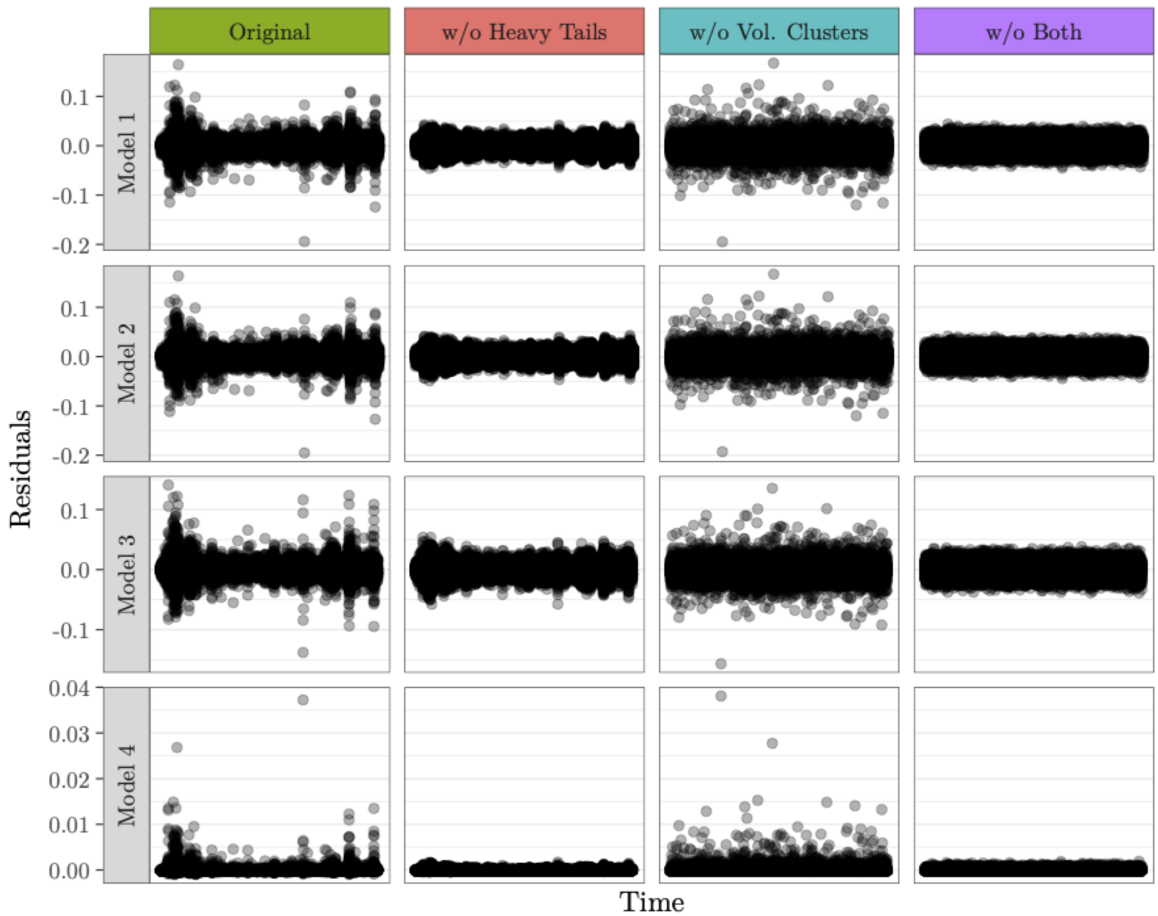

Figure C.1: Regression errors over time by series and model

The figure plots the residuals of four regression models of the original S&P 500 index returns and for three surrogate series that do not exhibit heavy tails, volatility clusters, or both. The time series range from the beginning of 1926 until the end of 2021. The surrogate series without heavy tails was created using the IAAWT$_n$ method. The surrogate series without volatility clusters was created using the IAAFT method. The surrogate series without both stylized facts was created using the IAAFT$_n$ method.

Figure C.2 holds QQ-plots of the error terms of the four regression models for the four different time series. It thus captures the distribution of residuals. We can easily identify strong heavy tails for the original series as well as for the series without volatility clusters for all regression models where r_t is the dependent variable. For the series without heavy tails and without both stylized facts, no tails can be observed. Again, this means that the heavy-tailedness can not be explained by the three controls and that the original tails are preserved by the series without volatility clusters. Because the fourth model focuses on volatility and thus only yields positive results, the plots look different than the other. However, we can observe that there are large outliers for the original series that are preserved by the surrogate without volatility clusters and no longer existent for the series without heavy tails or without both stylized facts. The first three columns of Table C.1 again confirm these visual observations. It holds excess skewness and kurtosis as well as a tail index, calculated as the reciprocal of the Hill (1975) estimator.

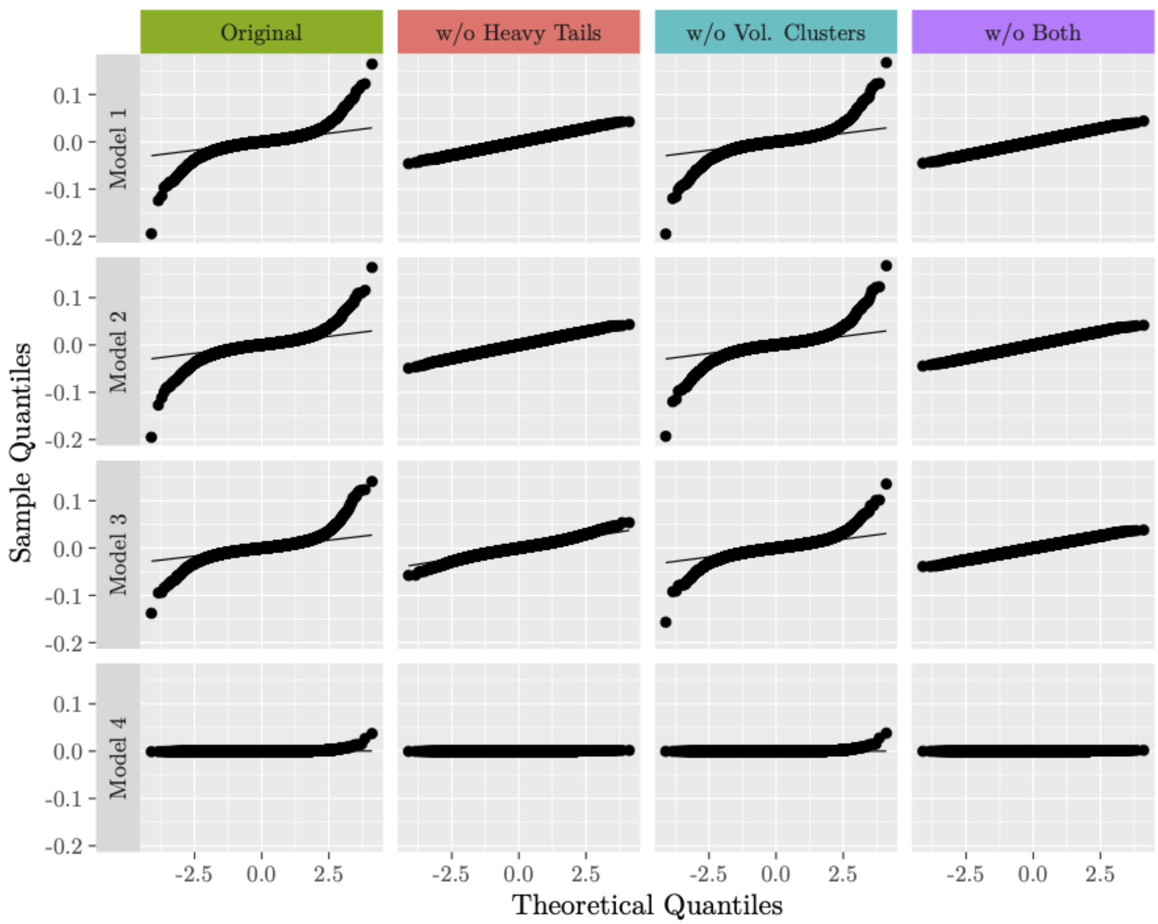

Figure C.2: Q-Q Plot of regression errors series and model

The figure holds Q-Q plots for the residuals of four regression models of the original S&P 500 index returns and for three surrogate series that do not exhibit heavy tails, volatility clusters, or both. The time series range from the beginning of 1926 until the end of 2021. The surrogate series without heavy tails was created using the IAAWT$_n$ method. The surrogate series without volatility clusters was created using the IAAFT method. The surrogate series without both stylized facts was created using the IAAFT$_n$ method.

Overall, the regression analysis shows that none of the four investigated characteristics can explain the presence of heavy tails and volatility clusters in the time series. It further shows that the two stylized facts are sufficiently preserved by the surrogate series as none of the four investigated characteristics alters their presence in the respective surrogates.

Model	Series	Skewness	Kurtosis	Tail Index	Nonlin. Dep.	Cluster Index
Model 1	Original	-0.09	16.37	0.28***	0.18***	2.12***
Model 1	w/o Heavy Tails	0.01	-0.02	0.11***	0.19***	2.66***
Model 1	w/o Vol. Clusters	-0.07	16.90	0.28***	0.01	1.00
Model 1	w/o Both	0.02	-0.02	0.11***	0.00	0.99
Model 2	Original	-0.17	15.82	0.28***	0.17***	2.12***
Model 2	w/o Heavy Tails	0.01	0.04	0.11***	0.13***	2.68***
Model 2	w/o Vol. Clusters	-0.06	16.80	0.28***	0.04	1.00
Model 2	w/o Both	0.01	-0.00	0.10***	0.00	0.98
Model 3	Original	0.75	16.53	0.30***	0.14***	2.20***
Model 3	w/o Heavy Tails	0.01	1.29	0.16***	0.15***	2.27***
Model 3	w/o Vol. Clusters	-0.03	11.18	0.28***	0.02*	1.22
Model 3	w/o Both	0.00	-0.01	0.11***	-0.00	1.02
Model 4	Original	24.75	1174.70	0.20***	-0.07	2.47***
Model 4	w/o Heavy Tails	2.60	10.32	0.16***	0.12	2.55***
Model 4	w/o Vol. Clusters	25.30	1191.29	0.13***	0.01	1.16
Model 4	w/o Both	2.70	10.26	0.07***	-0.02	1.02

Table C.1: Summary statistics of regression term by model and series

The table compares descriptive statistics for the residuals of four regression models of the original S&P 500 index returns and for three surrogate series that do not exhibit heavy tails, volatility clusters, or both. The time series range from the beginning of 1926 until the end of 2021. The surrogate series without heavy tails was created using the IAAWT_n method. The surrogate series without volatility clusters was created using the IAAFT method. The surrogate series without both stylized facts was created using the IAAFT_n method. The tail index is calculated for the lower tail of the return distribution using the reciprocal of the Hill (1975) estimator with a threshold of $k = 100$. Nonlinear dependence is measured by the excess spreads in Hölder exponents using the MF-DFA method of Kantelhardt et al. (2002). The cluster index is calculated following Tseng and Li (2011). Statistical significance at confidence levels of 10%, 5%, and 1% are indicated with *, **, and *** respectively.

D Company Names

The company names according to Datastream of the constituents in our investment pool for the equity portfolios are:

ABBOTT LABORATORIES, ADVANCED MICRO DEVICES, AFLAC, AIR PRDS CHEMS, ALTRIA GROUP, AMERICAN EXPRESS, AMERICAN INTL GP, ANALOG DEVICES, AON CLASS A, APA, APPLE, APPLIED MATS, AUTOMATIC DATA PROC, AVERY DENNISON, BALL, BANK OF AMERICA, BANK OF NEW YORK MELLON, BATH AND BODY WORKS, BECTON DICKINSON, BOEING, BRISTOL MYERS SQUIBB, BROWN FORMAN B, CAMPBELL SOUP, CATERPILLAR, CHEVRON, CHURCH DWIGHT CO, CINCINNATI FINL, CLOROX, COCA COLA, COLGATE PALM, COMCAST A, COMERICA, CONAGRA BRANDS, CONSOLIDATED EDISON, CORNING, CSX, CUMMINS, CVS HEALTH, DANAHER, DEERE, DOVER, DXC TECHNOLOGY, EATON, EDISON INTL, ELI LILLY, EMERSON ELECTRIC, ENTERGY, EQUIFAX, EXELON, EXXON MOBIL, FEDEX, FIFTH THIRD BANCORP, FIRSTENERGY, FMC, FORD MOTOR, FRANKLIN RESOURCES, GENERAL ELECTRIC, GENERAL MILLS, GLOBE LIFE, HALLIBURTON, HASBRO, HERSHEY, HESS, HOME DEPOT, HOWMET AEROSPACE, HP, HUMANA, ILLINOIS TOOL WORKS, INTEL, INTERNATIONAL BUS MCHS, INTERNATIONAL PAPER, INTERPUBLIC GROUP, INTL FLAVORS FRAG, JACOBS SOLUTIONS, JOHNSON JOHNSON, JP MORGAN CHASE CO, KELLOGG, KEYCORP, KIMBERLY CLARK, KLA, L3HARRIS TECHNOLOGIES, LENNAR A, LINCOLN NATIONAL, LOEWS, LOWE S COMPANIES, LUMEN TECHNOLOGIES, M T BANK, MARSH MCLENNAN, MASCO, MCCORMICK COMPANY NV, MCDONALDS, MEDTRONIC, MERCK COMPANY, MOTOROLA SOLUTIONS, NEWELL BRANDS XSC, NEWMONT, NIKE B, NORTHROP GRUMMAN, NUCOR, OMNICOM GROUP, PACCAR, PARKER HANNIFIN, PEPSICO, PFIZER, PNC FINL SVS GP, PPG INDUSTRIES, PROCTER GAMBLE, PROGRESSIVE OHIO, PULTEGROUP, RAYTHEON TECHNOLOGIES, REGIONS FINL NEW, ROCKWELL AUTOMATION, S P GLOBAL, SCHLUMBERGER, SEALED AIR, SHERWIN WILLIAMS, SOUTHWEST AIRLINES, STANLEY BLACK DECKER, STATE STREET, STRYKER, SYSCO, TARGET, TELEFLEX, TERADYNE XSC, TEXAS INSTRUMENTS, TEXTRON, THERMO FISHER SCIENTIFIC, TJX, TRANE TECHNOLOGIES, TRAVELERS COS, TYSON FOODS A, UNION PACIFIC, V F, VIATRIS, W R BERKLEY, WALGREENS BOOTS ALLIANCE, WALMART, WALT DISNEY, WELLS FARGO CO, WEYERHAEUSER, WHIRLPOOL, WILLIAMS, WW GRAINGER, X3M

E MF-DFA Model Specification

The MF-DFA algorithm is one of the most widely employed workflows to analyze nonlinear dependence of time series. It slices the time series into sub-series of equal length. This process is repeated for varying lengths, which are also referred to as the model's scales, s. For each iteration, the trend within each sub-series is removed and fluctuations are measured using the error function $F(q, s)$ defined as

$$F(q,s) = \left\{ \frac{1}{2N} \sum_{k=1}^{2N} [\hat{F}^2(s,k)]^{q/2} \right\}^{1/q} \tag{E.1}$$

for $q \neq 0$, and

$$F_0(s) = exp\left\{ \frac{1}{4N} \sum_{k=1}^{2N} ln[\hat{F}^2(s,k)] \right\} \tag{E.2}$$

for $q = 0$, where $\hat{F}^2(s,k)$ is the variance of the signal $x_i (i = 1, \ldots, N_s)$ around its local trend, defined as

$$\hat{F}^2(s,k) = \frac{1}{s} \sum_{k=1}^{s} \left\{ x_{(k-1)s+j} - P_{k,j} \right\}^2 \tag{E.3}$$

where N is the finite length of the time series and $P_{k,j}$ is the polynomial trend subtracted for jth data in kth sub series $(k = 1, ..., N)$. The error function $F(q,s)$ can be interpreted as an extended version of a root-mean-squared error function whose exponents change in q. The overall goal of the analysis is to determine whether the fluctuations around the local trend depend on the time scale s of the sub-series for different values of q. The strength of nonlinear dependence can be quantified using the power-law

$$F(q,s) \sim s^{h(q)} \tag{E.4}$$

The strength of nonlinear dependence is then measured as the maximum Δh of the generalized Hurst exponent profile defined as

$$\Delta h = \max_q h(q) - \min_q h(q) \tag{E.5}$$

This means that if $h(q)$ varies with the change of q, the time series is exposed to nonlinear dependence. For series that only exhibit linear autocorrelation $h(q)$ is a constant and thus $\Delta h = 0$. Another measure of nonlinear dependence is the maximum Hölder spread $\Delta\alpha$ which measures the width of the so-called Hölder or singularity spectrum $f(\alpha)$. This spectrum can be retrieved with the help of a Legendre transform (Feder, 1988; Peitgen et al., 2004) of the spectrum of generalized Hurst exponents as

$$\alpha = h(q) + qh'(q), \qquad f(\alpha) = q[\alpha - h(q)] + 1 \tag{E.6}$$

The Legendre transform is a way to display the spectrum of generalized Hurst exponents differently. It does not influence the results of the analysis. The width of the singularity spectrum, $\Delta\alpha$ is the most widely adopted measure for the strength of nonlinear dependence (Jiang et al., 2019). We also use this indicator to measure nonlinear dependence to make our work comparable to that of others.

In the model setup, determining optimal detrending polynomial $P_{k,j}$, scales s, and range for q is crucial for retrieving meaningful results from the analysis. However, finding optimal parameters is still a widely debated task in literature (Grech & Pamuła, 2013; Kantelhardt et al., 2002; Oswiecimka et al., 2013; Pamuła & Grech, 2014; Shao et al., 2012).

To tackle this task we follow the recommendations made by Ihlen (2012). To avoid overfitting we choose to let $P_{k,j}$ be of first order. Further, we use 20 loq-equally spaced scales starting at $s_{min} = 32$ to $s_{max} = N/10$, where N denotes the series' finite length. Log-equal spacing is recommended because of the power-law relation between $F(q, s)$ and s (Ihlen, 2012). We set $s_{min} = 32$ because literature finds that spurious nonlinear dependence are introduced for financial time series with less than 30 observations (Buonocore et al., 2016). This finding is further bolstered by literature focusing on the measurement of nonlinear dependence. One spurious source is short samples where fluctuations not related to nonlinear dependence may dominate over fluctuations that originate from nonlinear dependence due to small statistics of short data (Grech & Pamuła, 2012; Grech & Pamuła, 2013; López & Contreras, 2013; Mukli et al., 2015; Pamuła & Grech, 2014; Rak & Grech, 2018). We set $s_{max} = N/10$ because, in the fitting procedure, the number of subseries would be very small, making estimates statistically unreliable. These restrictions are more conservative than those by Kantelhardt et al. (2002) who proposed the MF-DFA method. They put an upper limit s_{max} to their scales at $N/4$ and a lower limit of $s_{min} > 10$. It is worth noting, however, that the choice of the scale does not appear to affect the results of a DFA analysis as much as it does for other analysis methods of nonlinear dependence (Shao et al., 2012). We set our moment order $q \in [-5, 5]$. This decision is guided by the fact that the potential range of scaling exponents becomes broader as the corresponding PDF deviates further from a Gaussian distribution. Thus, in choosing the range for our q we relate exponents of power-law tails in financial time series that are found to range between two and five (e.g., Ghashghaie et al., 1996; Gopikrishnan et al., 1998; Longin, 1996; Pagan, 1996; Rak & Grech, 2018). We chose to limit $|q|$ within that ballpark to avoid the introduction of spurious nonlinear dependence which is known to accelerate for larger ranges of q and is especially relevant for distributions exhibiting heavy tails (Ihlen, 2012; Pamuła & Grech, 2014; Rak & Grech, 2018). The decision to employ an equal width of the range of q around 0 is to keep estimates unbiased. This is recommended by Ihlen (2012) to make the analysis balanced because estimations in the MF-DFA procedure will be dominated by segments with small (large) fluctuations if $q < 0$ ($q > 0$) due to the way the error function (Equation E.1) is set up (Kantelhardt et al., 2002).

References

Buonocore, R. J., Aste, T., & Di Matteo, T. (2016). Measuring multiscaling in financial time-series. *Chaos, Solitons and Fractals*, *88*, 38–47.

Feder, J. (1988). *Fractals*. Plenum Press.

Ghashghaie, S., Breymann, W., Peinke, J., Talkner, P., & Dodge, Y. (1996). Turbulent cascades in foreign exchange markets. *Nature*, *381*(6585), 767–770.

Gopikrishnan, P., Meyer, M., Amaral, L. A., & Stanley, H. E. (1998). Inverse cubic law for the distribution of stock price variations. *European Physical Journal B*, *3*(2), 139–140.

Grech, D., & Pamuła, G. (2012). Multifractal background noise of monofractal signals. *Acta Physica Polonica A*, *121*(2 B), 34–39.

Grech, D., & Pamuła, G. (2013). On the multifractal effects generated by monofractal signals. *Physica A: Statistical Mechanics and its Applications*, *392*(23), 5845–5864.

Hill, B. M. (1975). A Simple General Approach to Inference About the Tail of a Distribution. *The Annals of Statistics*, *3*(5), 1163–1174.

Hurst, H. E. (1951). Long-Term Storage Capacity of Reservoirs. *Transactions of the American Society of Civil Engineers*, *116*(1).

Hurst, H. E. (1956). The Problem of Long-Term Storage in Reservoirs. *Hydrological Sciences Journal*, *1*(3), 13–27.

Ihlen, E. A. (2012). Introduction to multifractal detrended fluctuation analysis in Matlab. *Frontiers in Physiology*, *3*(141).

Jiang, Z. Q., Xie, W. J., Zhou, W. X., & Sornette, D. (2019). Multifractal analysis of financial markets: A review. *Reports on Progress in Physics*, *82*(12).

Kantelhardt, J. W., Zschiegner, S. A., Koscielny-Bunde, E., Havlin, S., Bunde, A., & Stanley, H. E. (2002). Multifractal detrended fluctuation analysis of nonstationary time series. *Physica A: Statistical Mechanics and its Applications*, *316*, 87–114.

Longin, F. M. (1996). The Asymptotic Distribution of Extreme Stock Market Returns. *The Journal of Business*, *69*(3), 383–408.

López, J. L., & Contreras, J. G. (2013). Performance of multifractal detrended fluctuation analysis on short time series. *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, *87*(2), 1–8.

Lux, T., & Alfarano, S. (2016). Financial power laws: Empirical evidence, models, and mechanisms. *Chaos, Solitons and Fractals*, *88*, 3–18.

Mandelbrot, B. (1967). How long is the coast of Britain? *Science*, *156*(3775), 636–638.

Mukli, P., Nagy, Z., & Eke, A. (2015). Multifractal formalism by enforcing the universal behavior of scaling functions. *Physica A: Statistical Mechanics and its Applications*, *417*, 150–167.

Oswiecimka, P., Drod, S., Kwapie, J., & Górski, A. Z. (2013). Effect of detrending on multifractal characteristics. *Acta Physica Polonica A*, *123*(3), 597–603.

Pagan, A. (1996). The econometrics of financial markets. *Journal of Empirical Finance*, *48*(1-3), 15–102.

Pamuła, G., & Grech, D. (2014). Influence of the maximal fluctuation moment order q on multifractal records normalized by finite-size effects. *EPL*, *105*, 50004.

Peitgen, H.-O., Jürgens, H., & Saupe, D. (2004). *Chaos and Fractals: New Frontiers of Science*. Springer.

Rak, R., & Grech, D. (2018). Quantitative approach to multifractality induced by correlations and broad distribution of data. *Physica A: Statistical Mechanics and its Applications*, *508*, 48–66.

Shao, Y. H., Gu, G. F., Jiang, Z. Q., Zhou, W. X., & Sornette, D. (2012). Comparing the performance of FA, DFA and DMA using different synthetic long-range correlated time series. *Scientific Reports*, *2*, 835.

Tseng, J. J., & Li, S. P. (2011). Asset returns and volatility clustering in financial time series. *Physica A: Statistical Mechanics and its Applications*, *390*(7), 1300–1314.

Swiss Finance Institute

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

swiss:finance:institute

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