The Impact of Volatility Targeting

CAMPBELL R. HARVEY, EDWARD HOYLE, RUSSELL KORGAONKAR,

SANDY RATTRAY, MATTHEW SARGAISON, and OTTO VAN HEMERT[1]

This version: June 25th, 2018

ABSTRACT

Recent studies show that volatility-managed equity portfolios realize higher Sharpe ratios than portfolios with a constant notional exposure. We show that this result only holds for "risk assets", such as equity and credit, and link this to the so-called leverage effect for those assets. In contrast, for bonds, currencies, and commodities the impact of volatility targeting on the Sharpe ratio is negligible. However, the impact of volatility targeting goes beyond the Sharpe ratio: it reduces the likelihood of extreme returns, across all asset classes. Particularly relevant for investors, "left-tail" events tend to be less severe, as they typically occur at times of elevated volatility, when a target-volatility portfolio has a relatively small notional exposure. We also consider the popular 60-40 equity-bond "balanced" portfolio and an equity-bond-credit-commodity "risk parity" portfolio. Volatility scaling at both the asset and portfolio level improves Sharpe ratios and reduces the likelihood of tail events.

Keywords: Volatility, volatility targeting, balanced fund, risk parity, asset allocation, portfolio choice

JEL codes: E32, E44, G11, G12, G15, G17

---

[1] Harvey is at Duke University and is an advisor to Man Group plc. The other authors are at Man Group plc. The authors would like to thank Xinuo Chen, Antoine Forterre, Nick Granger, Keith Haydon, Graham Robertson, Darrel Yawitch, and participants at the 2018 J.P. Morgan Macro Quantitative & Derivatives Conference for valuable comments. Please direct correspondence to: Otto.VanHemert@man.com.

Introduction

One of the key features of volatility is that it is persistent, or "clusters". High volatility over the recent past tends to be followed by high volatility in the near future. This observation underpins Engle's (1982) pioneering work on ARCH models.[2] In this paper, we study the risk and return characteristics of assets and portfolios that are designed to counter the fluctuations in volatility. We achieve this by leveraging the portfolio at times of low volatility, and scaling down at times of high volatility. Effectively the portfolio is targeting a constant level of volatility, rather than a constant level of notional exposure.

Conditioning portfolio choice on volatility has attracted considerable recent attention. The financial media has zoomed in on the increasing popularity of risk parity funds.[3] In recent work, Moreira and Muir (2017) find that volatility-managed portfolios increase the Sharpe ratios in the case of the broad equity market and a number of dynamic, mostly long-short stock strategies.

While most of the research has concentrated on equity markets, we investigate the impact of volatility targeting across more than 60 assets, with daily data beginning as early as 1926. We find that Sharpe ratios are higher with volatility scaling for risk assets (equities and credit), as well as for portfolios that have a substantial allocation to these risk assets, such as a balanced (60-40 equity-bond) portfolio and a risk parity (equity-bond-credit-commodity) portfolio.

Risk assets exhibit a so-called leverage effect, i.e., a negative relation between returns and volatility, and so volatility scaling effectively introduces some momentum into strategies. That is, in periods of negative returns, volatility often increases, causing positions to be reduced, which is in the same direction as what one would expect from a time-series momentum strategy. Historically such a momentum strategy has performed well; see e.g. Hamill, Rattray, and Van Hemert (2016). For other assets, such as bonds, currencies, and commodities, volatility scaling has a negligible effect on realized Sharpe ratios.

We show that volatility targeting consistently reduces the likelihood of extreme returns (and the volatility of volatility) across our 60+ assets. Under reasonable investor preferences, a thinner left tail is much preferred (for a given Sharpe ratio).[4] Volatility targeting also reduces the maximum drawdowns for both the balanced and risk parity portfolio.

The outline of this paper is as follows. In Section 1, we discuss the data, volatility-scaling methods, and statistics used for comparing the performance of unscaled and volatility-scaled portfolios. In Section 2, we focus on US equities, for which we have data starting in 1926. In Section 3 we study US bonds and credit, and in Section 4 we look at 50 global equity indices, fixed income, currency, and commodity futures and forwards. The analyses for the multi-asset balanced and risk parity portfolios are covered in Section 5. In Section 6, we discuss the leverage effect to provide further insights as to why the Sharpe ratio of risk assets is improved by volatility scaling. We offer some concluding remarks in the final section and comment on methods other than volatility scaling that may improve the Sharpe ratio and left-tail risk of a long-only portfolio.

---

[2] ARCH is autoregressive conditional heteroscedasticity. Robert Engle shared the 2003 Nobel Prize in Economics "for methods of analyzing economic time series with time-varying volatility (ARCH)" (https://www.nobelprize.org).

[3] See, e.g., the August 6, 2017 Wall Street Journal article "What is risk parity?", https://www.wsj.com/articles/what-is-risk-parity-1502071260.

[4] Under the common assumption of concave utility, investors dislike the left tail more than they like the right tail. Hence, for a given Sharpe ratio, investors are willing to give up some of the right tail to reduce the left tail.

1. Preliminaries

There are three main inputs needed to study the impact of volatility scaling: securities return data, methods for volatility scaling, and performance statistics used to compare the unscaled and volatility-scaled investment returns.

1.1 Data

Our study relies on daily return data as a starting point. Often monthly data are available for longer histories, but these data are less suitable for obtaining responsive volatility estimates. Table 1 provides an overview.

Table 1: Securities and sample periods

In all cases, the frequency of the data used is at least daily. For S&P500 and 10-year Treasury futures we also use 5-minute intraday data. Data before the start of sample periods, where available, are used to initialize volatility measures.

	Asset class	Sample period		Source
Section 2	Equities (all US)	Full	1927-2017	K. French website
	Equities (all US)	I	1928-1957	K. French website
	Equities (all US)	II	1958-1987	K. French website
	Equities (all US)	III	1988-2017	K. French website
	Equities (S&P500 futures)	III	1988-2017	Man AHL data
	Equities (10 industries US)	Full	1927-2017	K. French website
Sec 3	Bonds (US, proxied from yield data)	II/III	1962-2017	FRB of St. Louis
	Bonds (10y Treasury future)	III	1988-2018	Man AHL data
	Credit (BoA ML US Corp. Credit Index, hedged with Treasuries)	III	1988-2018	Bloomberg, Man AHL data
Section 4	Commodity futures (6 ags, 6 energies, 7 metals)	III	1988-2018	Man AHL data
	Currencies forwards (9 crosses against the US dollar)	III	1988-2018	Man AHL data
	Equity index futures (10)	III	1988-2018	Man AHL data
	Fixed income futures (9 bonds, 3 interest rate)	III	1988-2018	Man AHL data
Sec 5	Balanced (60-40 equity-bond)	III	1988-2018	Man AHL data
	Risk parity (25-25-25-25 equities-bonds-credit-commds.)	III	1988-2018	Man AHL data

In Section 2, we will consider US equity data. The earliest daily return dataset available to us is from July 1st, 1926, and is obtained from Kenneth French's website.[5] It is the value-weighted returns of firms listed on the NYSE, AMEX, and NASDAQ; henceforth referred to as "Equities All US". We will also use the returns of the 10 industry portfolios, available from the same source and start date. We additionally use S&P500 futures data from 1988, which allows us to estimate volatility based on intraday data.

In Section 3, we focus our attention on fixed income. For US Treasury bonds, daily yields are available since 1962 from the Federal Reserve.[6] We construct proxy daily returns by assuming that 10-year

---

[5] See: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html. Until 1952 stocks on the NYSE traded on Saturday as well, and thus the data include the Saturday returns up to then.

[6] Federal Reserve Economic Data (FRED), see https://fred.stlouisfed.org. To illustrate the return dynamics over a longer period of time, we will use monthly data in Figure 6 obtained from Global Financial Data (GFD) from July 1926 (to match the start date of the equity data).

yields are par yields, and computing the return of par coupon bonds.[7] We additionally use 10-year Treasury futures data from 1988, which will again allow us to evaluate volatility estimates based on intraday data. We also explore credit returns, hedged with Treasuries, creating a long time series for an exposure that should resemble the synthetic CDX investment grade index that is available today. To this end, we use the Bank of America Merrill Lynch US Corporate Master Total Return index, and the hedging methodology follows Cook et al. (2017).

In Section 4, we use daily futures and forwards data for 50 liquid securities from Cook et al. (2017). This dataset covers commodities (6 agricultural, 6 energy, and 7 metal contracts), 9 currencies (all against the US dollar), 10 equities, 9 bonds, and 3 interest rate contracts.

1.2 Volatility scaling

We focus on excess returns, as they capture the compensation for bearing risk, not the time value of money. Excess returns are a type of "unfunded" returns; for example, a long equities position financed by borrowing at the risk-free T-bill rate. The unfunded nature of excess returns makes evaluating scaled position returns particularly straightforward. That is, volatility-scaled returns are simply inversely proportional to a conditional volatility estimate that is known a full 24-hours ahead of time, using returns up to t-2.[8] That is,

$$r_t^{\text{scaled}} = r_t \times \frac{\sigma^{\text{target}}}{\sigma_{t-2}} \times k^{\text{scaled}},$$

where we added a constant k (approximately 1), chosen such that ex-post, over the full sample period, the target volatility is realized. We do this to facilitate comparison across different securities, methods, and sample periods. We will set the volatility target to 10% annualized throughout.

Unscaled returns involve no conditional volatility estimate, just a constant to achieve the same ex-post 10% realized volatility as scaled returns:

$$r_t^{\text{unscaled}} = r_t \times k^{\text{unscaled}}.$$

Notice that futures and forwards trade on margin, so their returns are already essentially unfunded, and so the risk-free rate is not deducted. In addition, the Treasury-hedged credit returns are unfunded by construction.

To estimate volatility, we use the standard deviation of daily returns, with exponentially decaying weights to returns at different lags.[9] We find similar results when using equal weights to returns at over a rolling window of fixed length, or using estimates based on 3- or 5-day overlapping returns (not reported).

---

[7] Assuming that the 10-year yield is the par yield on a semi-annual coupon paying bond, we reprice the bond the following day using that day's 10-year yield, and assuming that all cash flows are now 1/261 years closer (with 261 the assumed number of weekdays per year). The return over the one-day period is new price minus one (par).

[8] Moreira and Muir (2017) take the vantage point of a mean-variance investor and thus scale by variance for most of their analysis. However, in a robustness check, Moreira and Muir (2017) show that scaling by volatility empirically performs equally well. Volatility scaling, the focus in this paper, leads to lower turnover than variance scaling does. Barroso and Santa-Clara (2015) and Daniel and Moskowitz (2016) study risk-managed momentum strategies.

[9] We compute the standard deviation with a stated zero mean (i.e., based on squared returns), to prevent relying on mean returns estimated with large error over short time windows. In all cases, we require 270 trading days of data before we form volatility-scaled returns. This ensures that the slowest volatility estimate (using exponential- decaying weights with a 90-day half-life) has at least three half-lives worth of data to warm-up on.

As volatility may be more precisely estimated with higher-frequency data, we also examine the effects of scaling by intraday volatility for the S&P500 and 10-year Treasury futures since 1988.[10] We obtain a volatility estimate from 5-minute returns over the liquid 9:15am to 2:00pm (Chicago time) time window.[11] We aggregate squared returns to a daily realized variance value, average these daily values with exponentially-decaying weights, and then take the square root.

For the Equities All US data, we work with calendar-day data, to account for Saturday returns before 1952.[12] For other assets, we work with weekday data. In all cases, we annualize the volatility estimate, adjusting for the number of data points, to ensure comparability.

1.3 Performance statistics

In Table 2, we list the performance statistics that we focus on. In most cases we evaluate these statistics at the monthly frequency, which we believe is more relevant to investors than, e.g., the daily frequency. More precisely, we use 30 calendar day (in the case of Equities All US) or 21-weekday overlapping returns. Only the mean and turnover of the notional exposure are evaluated using daily data.

Table 2: Performance statistics

As a default, we compute the Sharpe ratio, volatility of volatility (vol of vol), mean shortfall, and mean exceedance using a 1 month (21 weekdays or 30 calendar days) evaluation frequency. The mean and turnover of the notional exposure are evaluated using daily data.

Statistic	Description
Sharpe ratio	Ratio of the mean and standard deviation of the excess return (annualized)
Mean notional exposure	Mean daily exposure
Turnover notional exposure	Mean absolute daily exposure change, annualized, and divided by twice the mean exposure
Vol of vol	Standard deviation of the rolling 1-year standard deviation of 21-weekday or 30-calendar day overlapping returns
Mean shortfall (left tail)	Mean of returns below the p-th percentile (p = 1 and 5 will be considered)
Mean exceedance (right tail)	Mean of returns above the p-th percentile (p = 95 and 99 will be considered)

In the tables, we will report the Sharpe ratio both gross and net of transaction costs. We use the following transaction cost estimates, expressed as fraction of the notional value traded: 1.0bp (or 0.01%) for equities, 0.5bp for bonds, 0.5bp for credit, 1.0 for gold, 2.0bp for oil, and 3.5bp for copper.[13] In figures we will just show returns gross of transaction costs, but results are very similar on a net basis.

The amount of trading needed to implement the volatility scaling can be inferred from the mean notional exposure times the turnover of the notional exposure. The latter is obtained as the mean

---

[10] Andersen et al. (2003) show that realized intraday volatility predicts daily return volatility well for a number of currencies.

[11] This is the period with consistently liquid trading conditions over the full sample period and across both securities. Adding up the squared overnight return leads to slightly less persistence (not reported), which is consistent with Bollerslev, Hood, Huss, and Pedersen (2018), who find greater persistence of intraday volatility.

[12] Saturdays are half days before 1952. Volatility estimates will either count Saturday as a half day (in case of equal-weight fixed window) or use double the Saturday return (in case of exponentially decaying weights).

[13] We believe these estimates broadly reflect trading costs in these markets post the 2008 global financial crisis, but are not necessarily representative of trading cost further back in time. For credit, the estimate is reflective of trading the synthetic CDX investment grade index, which we noted before resembles a credit index exposure hedged with Treasuries. We note that while the trading costs for bonds and credit are lower than those of equities, for a given notional traded, we will show that one needs to trade larger notional quantities in these lower volatility fixed income assets in order to achieve a given volatility target, making equities ultimately the cheapest to trade of the assets considered on a "per unit of risk" basis.

absolute daily exposure change, annualized, and divided by twice the mean exposure. That is, turnover is expressed as the annual number of roundtrips of the mean exposure. We do not consider turnover incurred from rolling futures or forwards contracts. Notice that an unscaled position in a particular asset will have a zero turnover.

The volatility of the rolling 1-year realized volatility, i.e. “vol of vol”, is the statistic that most directly measures the extent to which the volatility scaling results in more constant risk exposure.

The “mean shortfall” is the realized counterpart of expected shortfall, also known as conditional value at risk. In contrast, the usual value-at-risk metric simply measures how bad the p-th percentile of the returns distribution is; that is, it ignores returns below the p-th percentile. The mean shortfall is preferred because it uses all the returns below the p-th percentile.

The mean shortfall measures left-tail behavior, which is most relevant for investors. However, we will also show the “mean exceedance”, the equivalent metric for the right tail, to illustrate how volatility scaling cuts both the left and the right tails.

While skewness and kurtosis are commonly reported, we have omitted them from our main analysis for two reasons. The first is that skewness and kurtosis are very sensitive to outliers, as their computation involves taking the third and fourth power of returns, respectively. Second, skewness is impacted by both left- and right-tail behavior, while investors are likely much more concerned with the left tail of the return distribution. Readers interested in more detail can refer to Appendix A, where we will also discuss tail skewness, tail kurtosis, and (maximum) drawdown.

2. US equities

In this section, we look at US equities, for which we have the longest history of daily data available.

2.1 Unscaled equity returns since 1926

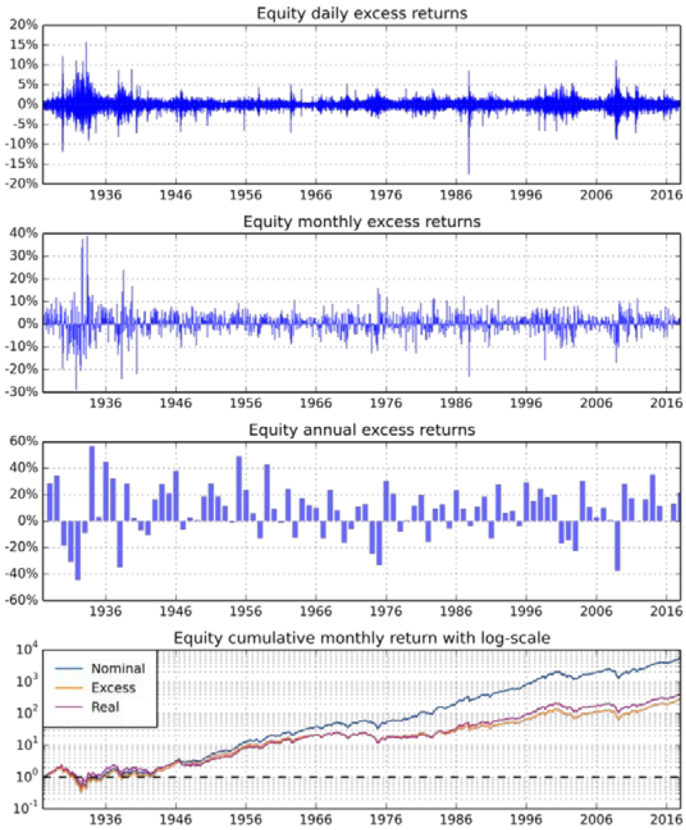
The top three panels of Figure 1 present daily, monthly, and annual excess equity returns. It is evident that volatility tends to cluster being persistently high during the 1930s (Great Depression), the early 2000s (following the bursting of the tech bubble) and 2007-2009 (Global Financial Crisis). The most negative day, October 19th, 1987 (Black Monday) is also clearly visible.

The bottom panel of Figure 1 contrasts nominal, excess, and real cumulative returns, with the nominal return markedly higher during the high-inflation 1970s and 1980s. Notice that the excess returns (the focus in this paper) are slightly below real returns. This is intuitive, as the short-rate deducted to arrive at excess returns captures both an inflation component (the larger effect empicirally) and a real rate component.[14]

[14] We use Consumer Price Index (all urban consumers) data from U.S. Department of Labor, Bureau of Labor Statistics. See: https://fred.stlouisfed.org/graph/?id=CPIAUCSL,CPIAUCNS.

Figure 1: Equities All US returns (1926-2017)

The first three panels of the figure are daily, monthly, and annual US equity returns in excess of the T-bill rate for the 1926-2017 period. No volatility scaling has been applied. The bottom panel shows cumulative (nominal, excess, and real) returns on a log-scale.

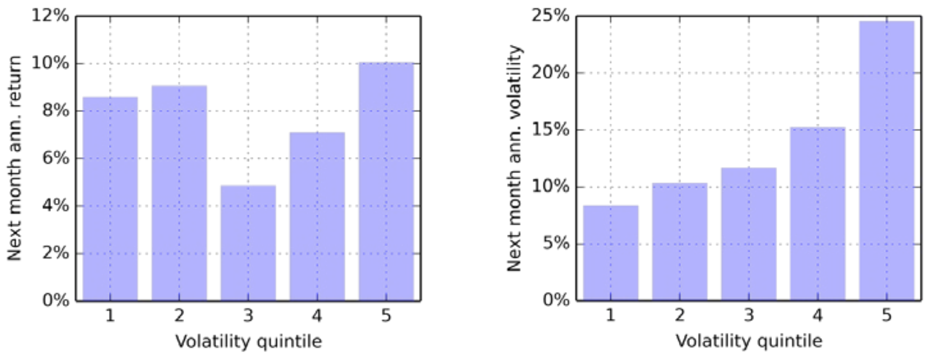


2.2 Persistence of equity volatility

In Figure 2, we sort returns into quintiles based on the previous month's volatility. The left panel shows the mean excess return and the right panel the volatility (both annualized) for the subsequent month. The persistence of volatility is evident in the right panel. However, the mean return shows no clear pattern across different quintiles (left panel). So expected returns do not seem to reflect the persistence in volatility; that is, they do not provide a substantially higher reward in the case of predictably higher volatility. This is a first indication that volatility scaling may improve the Sharpe ratio of a long equities investment, as we will establish in the next subsection.

Figure 2: Quintile analysis for Equities All US (1926-2017)

The left panel shows the mean excess return and the right panel the volatility (both annualized) when sorting on the previous month's return for Equities All US over the 1926-2017 period.



To further illustrate that equity volatility clusters, we show in Appendix B the autocorrelation of the monthly squared volatility (i.e., variance) of daily returns.

2.3 Performance of volatility-scaled equity returns

In Table 3, we show performance statistics for unscaled (top row) and volatility-scaled (other rows) Equities All US investments (1927-2017). We use exponentially-decaying weights for the volatility estimate with a half-life indicated in parentheses in the first column of the table.

Table 3: Performance statistics Equities All US (1927-2017)

The table reports the performance statistics detailed in Table 2 for Equities All US (1927-2016). The gross and net (of estimated costs) Sharpe ratio, vol of vol, mean shortfall (left tail), and mean exceedance (right tail) use a rolling 1 month (30 calendar day) evaluation frequency. To facilitate comparison, both the unscaled and volatility-scaled returns are shown at 10% full-sample volatility.

	Sharpe ratio		Notional exposure		Vol of	Left tail		Right tail	
Scaling	Gross	Net	Mean	Turnover	**vol**	1%	5%	95%	99%
Unscaled	0.40	0.40	52%	0.00	4.6%	-11.4%	-6.9%	6.3%	11.8%
Scaled (10-day half life)	0.48	0.48	71%	4.66	1.7%	-8.3%	-6.1%	5.8%	7.3%
Scaled (20-day half life)	0.49	0.49	70%	2.39	1.8%	-9.0%	-6.4%	5.8%	7.3%
Scaled (40-day half life)	0.50	0.50	69%	1.22	1.9%	-9.6%	-6.5%	5.8%	7.4%
Scaled (60-day half life)	0.51	0.51	68%	0.82	2.1%	-9.9%	-6.6%	5.9%	7.5%
Scaled (90-day half life)	0.51	0.51	67%	0.56	2.2%	-10.1%	-6.7%	5.9%	7.7%

The Sharpe ratio improves from 0.40 (unscaled) to between 0.48 and 0.51 (volatility scaled) and is not very sensitive to the choice of volatility estimate.[15], [16] The gross and net Sharpe ratio are the same for the reported precision, which we caveat with the statement that we use transaction cost estimates reflective of the current environment and apply this to the full history.[17] As described in Section 1, we use a rolling 1 month (30 calendar days) evaluation frequency. We find very similar results for a rolling 3-month (90 calendar days) evaluation frequency (not reported).

The mean exposure is higher with volatility scaling, in order to achieve the same 10% full sample realized volatility, as larger exposures are taken during low volatility episodes. The turnover is zero for the unscaled investment and ranges from about 5 times a year for the most responsive and reactive volatility estimates, to less than once a year for the least responsive volatility estimates.

Both the vol of vol and left tail (mean shortfall) materially improve with volatility scaling, and the improvement is greatest for the most responsive volatility estimates.[18] The right tail (mean exceedance) is also, not surprisingly, reduced with volatility scaling. Hence, volatility scaling cuts both the left and right tails. Consistent with this, in Appendix A, we show that kurtosis is much reduced when volatility scaling. The effect on skewness is more ambiguous, as both the left and right tails are cut. Also the maximum drawdown is lower with volatility scaling.

---

[15] To test for the statistical significance of this improvement, we run a regression of volatility-scaled daily returns (20-day half-life) on unscaled returns. We find an intercept of 0.64bp with a t-stat of 3.05 (Newey-West corrected with 30 lags). The R-squared of the regression is 0.73.

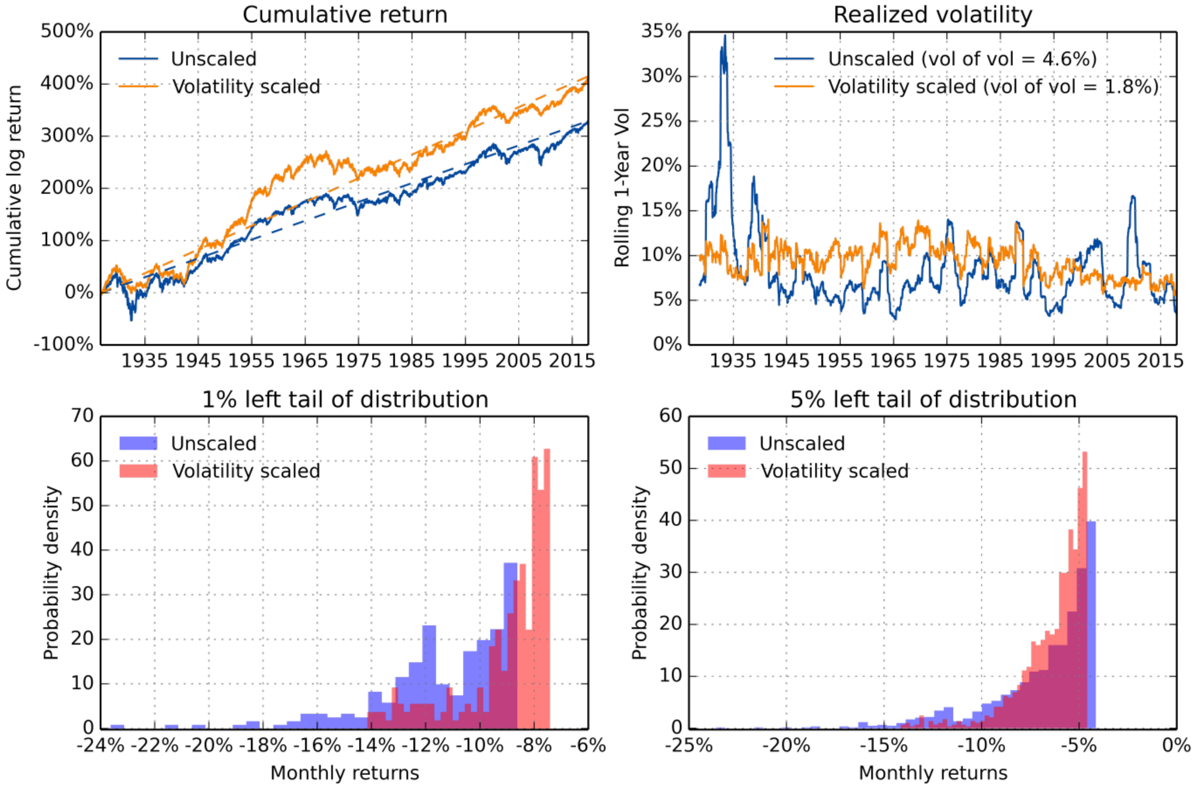
[16] Dopfel and Ramkumar (2013) and Moreira and Muir (2017) also find that volatility targeting improves the Sharpe ratio for equities since 1927.

[17] For example, for the case of a 10-day half-life, the costs are about $2 \times 1\text{bp} \times 71\% \times 4.66 = 0.066\%$ per year for a 10% volatility strategy. The unrounded gross and net Sharpe ratio are then 0.4831 and 0.4766, but both are 0.48 after rounding.

[18] See also Hocquard, Ng, and Papageorgiou (2013), who study how volatility targeting changes the tail risk properties of an equity portfolio since 1990.

Figure 3: Cumulative returns, realized volatility, left and right tail for Equities All US (1927-2017)

This figure compares unscaled and volatility-scaled (exponential weighted, 20-day half-life) Equities All US excess returns for the 1927-2017 period. The top-left panel shows the cumulative return. The top-right panel shows the rolling 1-year standard deviation of 1-month (30 calendar days) overlapping returns. The standard deviation of the rolling 1-year standard deviation is reported in parentheses in the legend. The bottom-left and bottom-right panels show the lowest 1% and 5% of the rolling 1-month (30 calendar days) return distribution. To facilitate comparison, both the unscaled and volatility-scaled returns are shown at 10% full-sample volatility.



In Figure 3, we further compare unscaled and volatility-scaled returns, where the latter uses a volatility estimate based on a half-life of 20 days. In the top-left panel, we plot the cumulative return, which shows that the volatility-scaled investment generally outperformed, except during the middle part of the sample period. The impact of volatility scaling is illustrated in the top-right panel, where we depict the rolling 1-year realized volatility for both unscaled and volatility-scaled 30-day overlapping returns. The realized volatility of volatility-scaled returns is much more stable over time. This is also evident from the vol of vol metric (i.e., the standard deviation of the rolling 1-year realized volatility) reported in the legend: 4.6% for unscaled returns versus 1.8% for volatility-scaled returns. Finally, in the bottom-left and bottom-right panels we show the lowest 1% and 5% of the 1-month (30-calendar days) return distribution.[19] Very negative returns of, say, -10% or worse are more common for unscaled returns.

To summarize, Figure 3 illustrates the two main ways volatility scaling has helped an Equities All US investment: first, it improves the risk-adjusted performance, and second, it reduces the left tail.

2.4 Performance of volatility-scaled equity returns, robustness across subsamples and industries

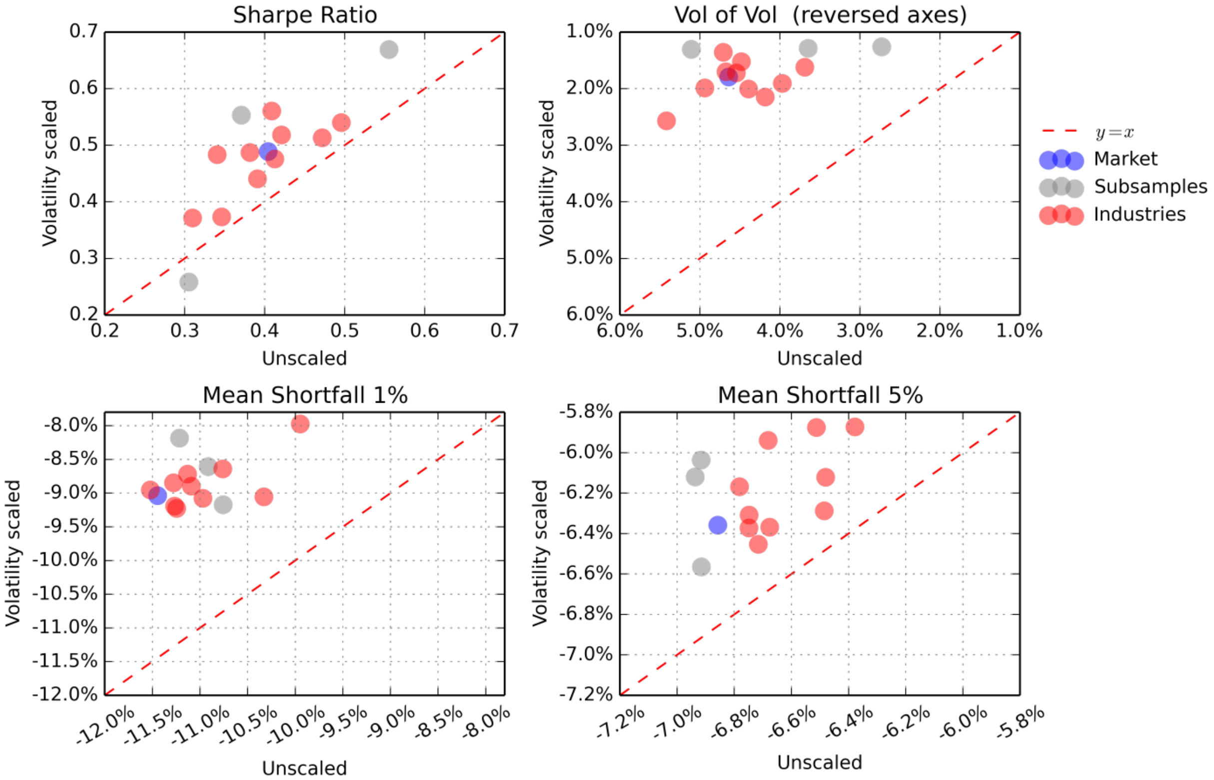
In Figure 4, we show the key statistics visually. We include equities broadly over the full sample period (1927-2017), equities broadly over three 30-year subsample periods, and 10 industry portfolios over

[19] The vertical axis for the histograms is the normalized frequency (frequency / (number of observations × bar width)). In this way, the areas of the bars sum to unity, hence the "probability density" labels. Rescaling in this way eases comparison between the two histograms, particularly in cases where the bar widths of the two histograms differ.

the full sample period. Subplots are such that in all cases observations above the dashed diagonal correspond to situations where volatility scaling improves the statistic.

Figure 4: Performance statistics for Equities All US (1927-2017)

This figure compares unscaled and volatility-scaled (exponential weighted, 20-day half-life) excess returns for Equities All US full sample (blue dots), subsamples (grey dots), and 10 industries full sample (red dots). For all four statistics, observations above the dashed diagonal line correspond to a situation where volatility scaling improves the statistic (using reversed axes for vol of vol). All statistics are based on a rolling 1 month (30 calendar days) evaluation frequency. To facilitate comparison, both the unscaled and volatility-scaled returns are shown at 10% full-sample volatility.



The Sharpe ratio improves in all cases, except during the 1957-1987 subsample period. The vol of vol and mean shortfall consistently and materially improve with volatility scaling.

Table 4: Performance statistics S&P 500 futures (1988-2017) using intraday vol estimates

The table reports the performance statistics detailed in Table 2 for S&P500 futures (1988-2016). We consider volatility estimates based on daily data (top panel) and 5-minute intraday data (bottom panel). The gross and net (of estimated costs) Sharpe ratio, vol of vol, mean shortfall (left tail), and mean exceedance (right tail) use a rolling 1 month (21 weekdays) evaluation frequency. To facilitate comparison, both the unscaled and volatility-scaled returns are shown at 10% full-sample volatility.

	Sharpe ratio		Notional exposure			Left tail		Right tail	
Scaling	Gross	Net	Mean	Turnover	**Vol of vol**	**1%**	**5%**	**95%**	**99%**
Unscaled	0.50	0.50	68%	0.00	3.8%	-11.2%	-6.9%	6.2%	9.1%
Volatility used for scaling based on daily data									
Scaled (10-day half life)	0.57	0.56	85%	4.56	1.1%	-7.8%	-6.0%	5.9%	7.6%
Scaled (20-day half life)	0.59	0.59	84%	2.31	1.2%	-8.4%	-6.1%	6.0%	7.7%
Scaled (40-day half life)	0.60	0.60	83%	1.17	1.5%	-9.0%	-6.3%	6.0%	7.9%
Scaled (60-day half life)	0.60	0.59	82%	0.78	1.8%	-9.3%	-6.4%	6.0%	8.1%
Scaled (90-day half life)	0.59	0.59	80%	0.53	2.1%	-9.6%	-6.5%	6.1%	8.2%
Volatility used for scaling based on 5-minute data									
Scaled (intraday, 5-day)	0.60	0.59	86%	3.96	1.3%	-7.5%	-5.7%	6.0%	7.4%
Scaled (intraday, 15-day)	0.62	0.62	85%	1.52	1.4%	-8.3%	-6.1%	5.9%	7.5%
Scaled (intraday, 25-day)	0.63	0.63	84%	0.97	1.5%	-8.7%	-6.2%	6.0%	7.6%
Scaled (intraday, 40-day)	0.63	0.62	83%	0.64	1.6%	-9.1%	-6.3%	6.0%	7.8%
Scaled (intraday, 60-day)	0.62	0.62	81%	0.45	1.8%	-9.4%	-6.4%	6.0%	8.0%

2.5 Performance of volatility-scaled S&P500 futures returns and the use of intraday data

Table 4 explores the benefit of using higher frequency S&P 500 future data (5-minute intervals from 1988) to estimate volatility. The top panel used daily data and the bottom panel uses the 5-minute bars. The Sharpe ratio, vol of vol, and left tail (mean shortfall) all slightly improve when using intraday data (and comparing versions with a similar turnover value).

3. US bonds and credit

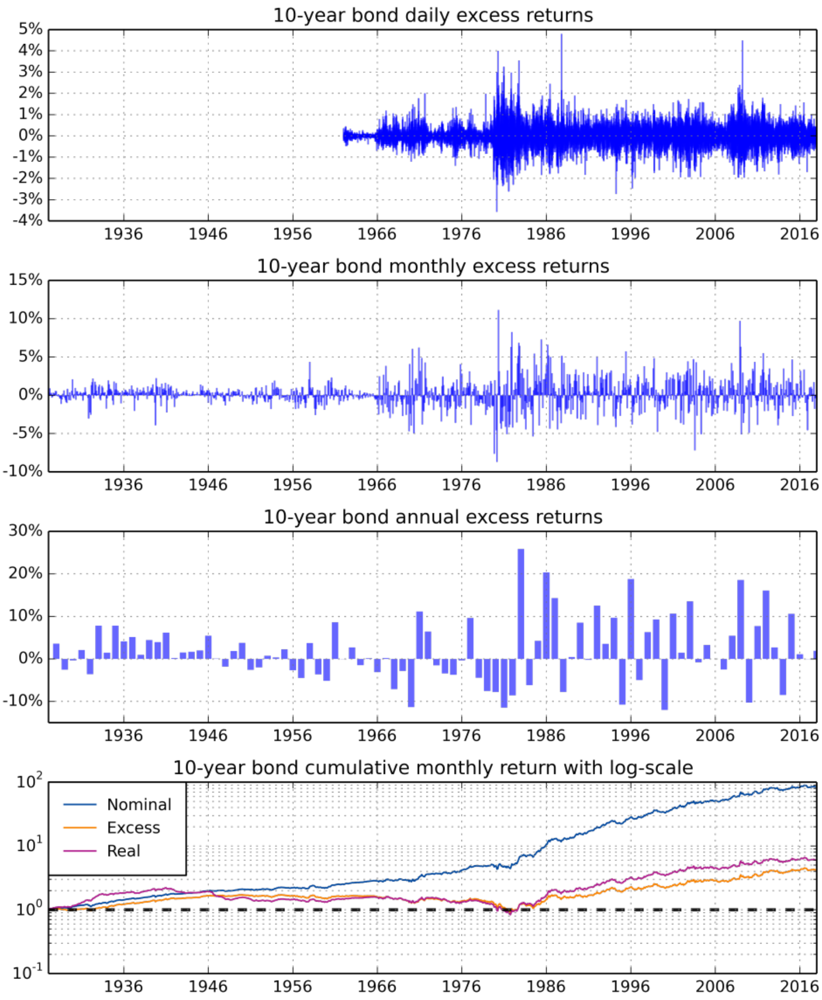
In this section we study US bonds and long credit risk positions.

3.1 Unscaled bond returns since 1926

As we did for equities, we start by examining bond returns since 1926, with (proxy) daily returns starting in 1962. In Figure 5, top three panels, we plot the daily, monthly, and annual excess returns.

Figure 5: US bond returns (1926-2017)

The top three panels of the figure show (proxy) daily, monthly, and annual US bonds returns in excess of the T-bill rate for the 1926-2017 period. No volatility scaling has been applied. The bottom panel shows cumulative (nominal, excess, and real) returns against a log-scale.



Returns were less volatile pre-1980, and much less so pre-1967. Hence it seems that bond markets have gone through different volatility regimes, lasting multiple decades. This is important to note, as structural breaks may render the evaluation of a bond volatility targeting strategy that includes data from before the mid-1980s less appropriate. In contrast, equity markets experienced clusters of volatility, but without clear structural breaks. From the bottom panel of Figure 5, we can see that excess returns were flat for the first 55-years of our sample period, and experienced a 40-year drawdown, ending in the 1980s.

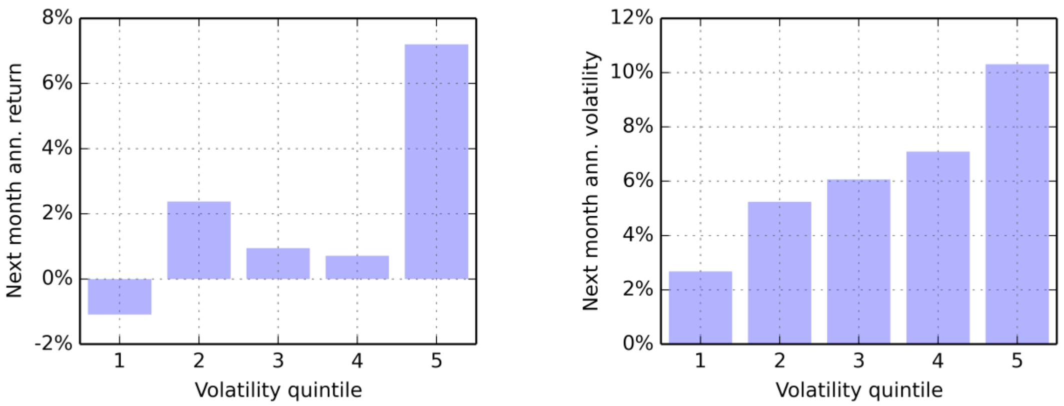
3.2 Persistence of bond volatility

As we did for Figure 2, in Figure 6 we sort returns into quintiles based on the previous month's volatility. The left panel shows the mean excess return and the right panel the volatility (both annualized) for the subsequent month. Volatility is persistent (right panel). However, in contrast to equities in Figure 2, the mean bond returns are not similar across different quintiles (left panel), but rather the returns are much higher in the high-volatility quintile. So it is not obvious that volatility scaling will impact the Sharpe ratio of a long bond investment.

To further illustrate that bond volatility clusters, we show in Appendix B the autocorrelation of the monthly squared volatility (i.e., variance) of daily returns.

Figure 6: Quintile analysis for US bonds (1962-2017)

The left panel shows the mean excess return and the right panel the volatility (both annualized) when sorting on the previous month's return for US bonds over the 1962-2017 period.



3.3. Performance of volatility-scaled bond returns (since 1963)

In Table 5, we report the performance statistics for US bonds over the 1963-2017 period. Consistent with the quintile analysis displayed in Figure 6, volatility scaling decreases the Sharpe ratio over this period. The reason is straightforward: the 1960-1980 period was characterized by both negative returns and low volatility. So a volatility targeting approach would lead to relatively large exposures during this extended bond bear market. Volatility targeting does lead to a lower vol-of-vol during this period, as it did for equities.

Table 5: Performance statistics US bonds (1963-2017)

The table reports the performance statistics detailed in Table 2 for US bonds, proxied from daily yield data (1963-2016). The gross and net (of estimated costs) Sharpe ratio, vol of vol, mean shortfall (left-tail), and mean exceedance (right-tail) use a rolling 1 month (21 weekdays) evaluation frequency. To facilitate comparison, both the unscaled and volatility-scaled returns are shown at 10% full-sample volatility.

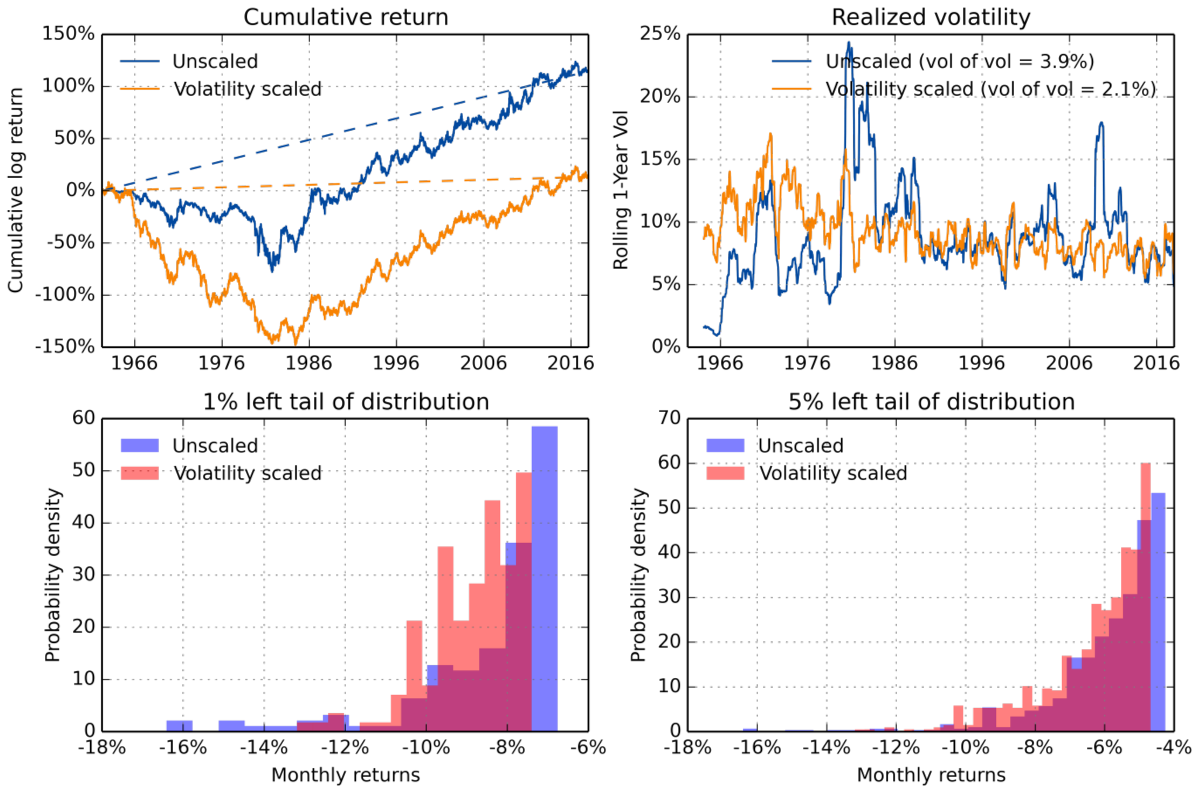
	Sharpe ratio		Notional exposure		Vol of	Left tail		Right tail	
Scaling	Gross	Net	Mean	Turnover	**vol**	1%	5%	95%	99%
Unscaled	0.25	0.25	127%	0.00	3.9%	-8.4%	-5.9%	7.1%	11.7%
Scaled (10-day half life)	0.05	0.04	180%	4.96	2.1%	-8.6%	-6.3%	6.0%	8.2%
Scaled (20-day half life)	0.06	0.06	179%	2.51	2.1%	-8.9%	-6.3%	6.1%	8.4%
Scaled (40-day half life)	0.08	0.08	177%	1.27	2.2%	-9.0%	-6.4%	6.2%	8.5%
Scaled (60-day half life)	0.08	0.08	174%	0.86	2.4%	-9.2%	-6.4%	6.2%	8.5%
Scaled (90-day half life)	0.09	0.09	170%	0.58	2.6%	-9.5%	-6.4%	6.2%	8.6%

In Figure 7, we contrast the cumulative return, realized volatility, and 1% and 5% left tail the return distribution for an unscaled and volatility-scaled bond investment. In all cases, the volatility-scaling is done using exponentially-decaying weights with a half-life of 20 days. Visible from the top-right panel is that the unscaled bond investment indeed has a low realized volatility during the 1964-1980 period. The underperformance of the volatility scaled invesment is solely due to the pre-1980 period.

One could argue that bond markets underwent a structural change in the mid-1980s, with monetary policy more geared towards inflation targeting. Hence the post-1988 sample period considered in the next subsection may be more representative for today's bond markets.

Figure 7: Cumulative returns, realized volatility, left and right tail for US bonds (1963-2017)

This figure compares unscaled and volatility-scaled (exponential weighted, 20-day half-life) US bond returns (proxied from daily yield data) for the 1963-2017 period. The top-left panel shows the cumulative return. The top-right panel shows the rolling 1-year standard deviation of 1-month (21 weekdays) overlapping returns. The standard deviation of the rolling 1-year standard deviation is reported in parentheses in the legend. The bottom-left and bottom-right panel show the lowest 1% and 5% of the rolling 1-month (21 weekdays) return distribution. To facilitate comparison, both the unscaled and volatility-scaled returns are shown at 10% full-sample volatility.



3.4. Performance of volatility-scaled bond returns (since 1988)

In Table 6, we report the performance statistics over 1988-2017, for which we have US 10yr Treasury futures data, both daily and intraday. We see that in general the vol of vol is much lower with volatility scaling, but the Sharpe ratio and mean shortfall (left tail) are similar. Using intraday data for the volatility estimate produces a slight improvement**.**

In Figure 8, we contrast the cumulative return, realized volatility, and 1% and 5% left tail the return distribution for an unscaled and volatility-scaled bond investment. In all cases, the volatility-scaling is done using exponentially-decaying weights with a half-life of 20 days. Consistent with Table 6, the main difference between unscaled and scaled returns for Treasuries over the 1988-2017 period is the lower vol of vol when volatility scaling (top-right panel).

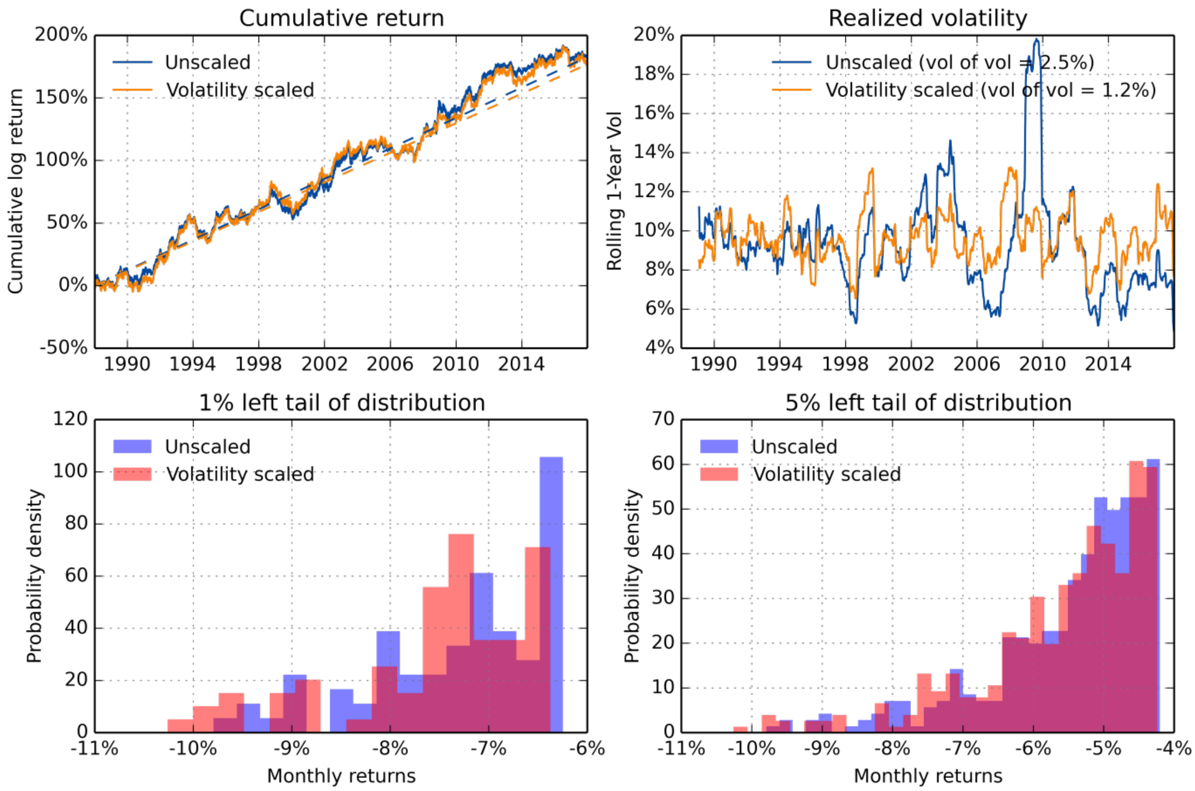
Table 6: Performance statistics 10-year Treasury futures (1988-2017), also using intraday vol estimates

The table reports the performance statistics detailed in Table 2 for US 10-year Treasury futures (1988-2016). We consider volatility estimates based on daily data (top panel) and 5-minute intraday data (bottom panel). The gross and net (of estimated costs) Sharpe ratio, vol of vol, mean shortfall (left tail), and mean exceedance (right tail) use a rolling 1 month (21 weekdays) evaluation frequency. To facilitate comparison, both the unscaled and volatility-scaled returns are shown at 10% full-sample volatility.

	Sharpe ratio		Notional exposure			Left tail		Right tail	
Scaling	Gross	Net	Mean	Turnover	**Vol of vol**	1%	5%	95%	99%
Unscaled	0.64	0.64	171%	0.00	2.5%	-7.3%	-5.5%	6.7%	10.1%
Volatility used for scaling based on daily data									
Scaled (10-day half life)	0.63	0.62	182%	4.33	1.2%	-7.5%	-5.6%	6.3%	7.9%
Scaled (20-day half life)	0.63	0.63	182%	2.19	1.2%	-7.5%	-5.6%	6.3%	8.1%
Scaled (40-day half life)	0.63	0.63	182%	1.11	1.4%	-7.5%	-5.6%	6.4%	8.4%
Scaled (60-day half life)	0.63	0.63	181%	0.74	1.5%	-7.5%	-5.6%	6.5%	8.6%
Scaled (90-day half life)	0.63	0.63	180%	0.49	1.6%	-7.5%	-5.6%	6.5%	8.8%
Volatility used for scaling based on 5-minute data									
Scaled (intraday, 5-day)	0.66	0.65	186%	3.80	1.1%	-7.4%	-5.4%	6.2%	7.6%
Scaled (intraday, 15-day)	0.64	0.64	186%	1.39	1.1%	-7.4%	-5.5%	6.3%	7.7%
Scaled (intraday, 25-day)	0.64	0.64	186%	0.87	1.2%	-7.5%	-5.5%	6.3%	7.9%
Scaled (intraday, 40-day)	0.63	0.63	185%	0.56	1.3%	-7.5%	-5.5%	6.4%	8.1%
Scaled (intraday, 60-day)	0.63	0.63	184%	0.38	1.4%	-7.6%	-5.5%	6.4%	8.3%

Figure 8: Cumulative returns, realized volatility, left and right tail for Treasury futures (1988-2017)

This figure compares unscaled and volatility-scaled (exponential weighted, 20-day half-life) 10-year Treasury futures returns for the 1988-2017 period. The top-left panel shows the cumulative return. The top-right panel shows the rolling 1-year standard deviation of 1-month (21 weekdays) overlapping returns. The standard deviation of the rolling 1-year standard deviation is reported in parentheses in the legend. The bottom-left and bottom-right panel show the lowest 1% and 5% of the rolling 1-month (21 weekdays) return distribution. To facilitate comparison, both the unscaled and volatility-scaled returns are shown at 10% full-sample volatility.



3.5. Performance of volatility-scaled credit returns

Moving on to credit, we see in Table 7 a substantial increase in the Sharpe ratio when using a relatively fast volatility estimate. For slower estimates (longer half-lives) the situation reverses. Also, the mean shortfall (left tail) is similar to unscaled for fast volatility estimates, but worse otherwise. The vol of vol is reduced for all volatility scaling cases considered. Credit is related to equities in the sense that both are exposed to firms' cash flow risk (i.e., both are risk assets), and so it is intuitive we see some similarities with the previously-discussed results for equities (e.g., the improvement of the Sharpe ratio for relatively fast volatility estimates).

Table 7: Performance statistics for US Credit (1988-2017)

The table reports the performance statistics detailed in Table 2 for US credit, hedged with Treasuries (1988-2017). The gross and net (of estimated costs) Sharpe ratio, vol of vol, mean shortfall (left-tail), and mean exceedance (right-tail) use a rolling 1 month (21 weekdays) evaluation frequency. To facilitate comparison, both the unscaled and volatility-scaled returns are shown at 10% full-sample volatility.

	Sharpe ratio		Notional exposure			Left tail		Right tail	
Scaling	Gross	Net	Mean	Turnover	**Vol of vol**	1%	5%	95%	99%
Unscaled	0.30	0.30	273%	0.00	6.7%	-14.5%	-7.2%	6.9%	13.5%
Scaled (10-day half life)	0.49	0.46	510%	4.79	4.1%	-11.8%	-7.1%	5.8%	7.2%
Scaled (20-day half life)	0.41	0.39	486%	2.49	4.3%	-13.3%	-7.6%	5.6%	7.1%
Scaled (40-day half life)	0.30	0.30	458%	1.30	4.6%	-14.9%	-8.0%	5.3%	6.7%
Scaled (60-day half life)	0.24	0.23	439%	0.89	4.9%	-15.7%	-8.2%	5.1%	6.6%
Scaled (90-day half life)	0.18	0.17	415%	0.61	5.2%	-16.4%	-8.4%	5.1%	6.9%

4. Futures and forwards

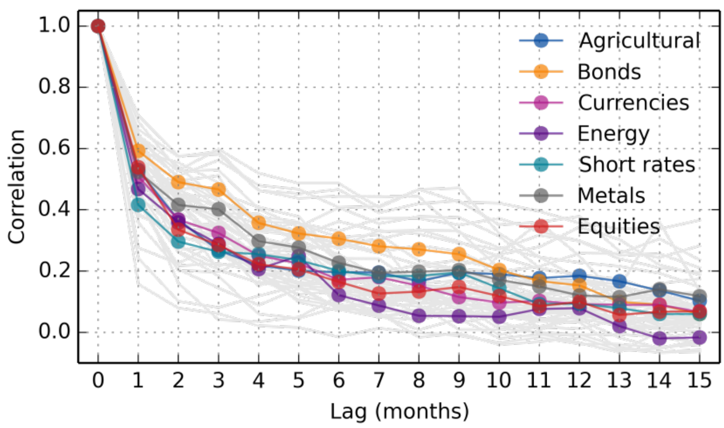
In this section, we study 50 futures and forwards across global equities, fixed income, currencies (all against the USD), and commodities.

4.1 Persistence of futures volatility

In Figure 9, we show the autocorrelation of the monthly variance of daily returns for the 50 futures and forwards markets in light grey. The average for different sectors is superimposed in color. Persistence in variance is ubiquitous with a remarkably similar autocorrelation pattern across the 50 markets and seven sectors considered. Each of the seven sectors has an autocorrelation of around 0.5 for consecutive monthly variances, which then gradually decreases to 0.1-0.2 for the autocorrelation at a lag of 12 months.

Figure 9: Autocorrelation of futures and forwards variance (1988-2017)

The figure shows the autocorrelation of the monthly variance of 50 daily future and forwards returns for the 1988-2017 period.

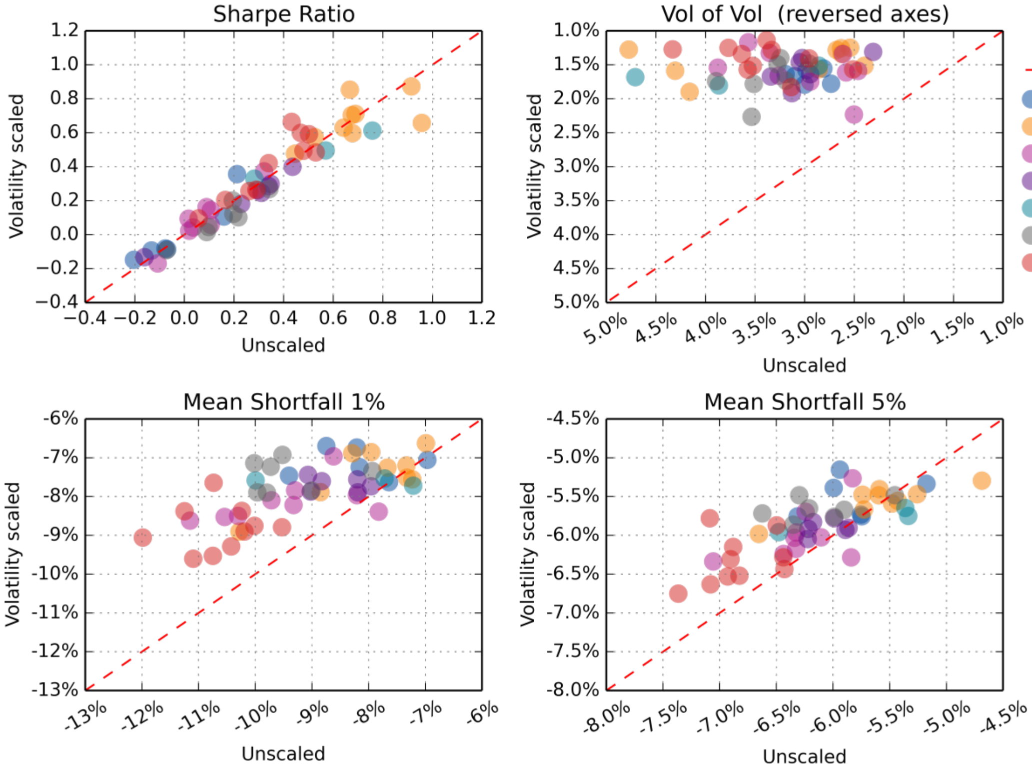


4.2 Performance of volatility-scaled futures returns

The performance of the 50 futures and forwards markets is depicted in Figure 10. The Sharpe ratio improves slightly for equity indices when volatility scaling, but is similar for other assets. The vol of vol and mean shortfall improve materially for almost all assets with volatility scaling.

Figure 10: Performance statistics for futures & forwards (1988-2017)

This figure compares unscaled and volatility-scaled (exponential weighted, 20-day half-life) returns for 50 global futures and forwards markets, where different sectors are represented by different colors. For all four statistics, observations above the dashed diagonal line correspond to a situation where volatility scaling improves the statistic (using reversed axes for vol of vol). All statistics are based on a rolling 1 month (21 weekdays) evaluation frequency. To facilitate comparison, both the unscaled and volatility-scaled returns are shown at a 10% full-sample volatility.



5. Portfolios

So far, we have considered single asset investments. In this section, we turn our attention to two popular multi-asset portfolios: the 60-40 equity-bond "balanced" portfolio and a 25-25-25-25 equity-bond-credit-commodity "risk parity" portfolio.

We will contrast three ways to implement such a portfolio:

1. Unscaled at both the asset and portfolio level,
2. Volatility scaling at the asset level only,
3. Volatility scaling at both the asset and portfolio level.

In all cases the asset level returns are subject to the full-sample scaling to 10%, as discussed in subsection 1.3, which means that as a starting point the allocation to the different asset classes is in proportion to full-sample volatility, and thus we can sensibly compare the different cases.

For simplicity we assume portfolios are rebalanced to the target asset allocation mix each day.[20]

5.1 Balanced 60-40 equity-bond portfolio

In Table 8, we report the performance statistics for the balanced 60-40 equity-bond portfolio, based on S&P500 and 10-year Treasury futures return data.[21] Because of the abovementioned asset-level

---

[20] For a discussion on the effect of the rebalancing frequency for a 60-40 balanced portfolio, see Granger et al. (2014).

scaling to 10% volatility in all cases, the 60-40 split here is in risk terms. The Sharpe ratio, vol of vol, and expected shortfall (left tail) all improve from asset-level volatility scaling and further improve from a second volatility scaling step at the portfolio level, which essentially adjusts for time variation in the correlation between different assets. Also, the improvement in left-tail returns is greater than the reduction in right-tail returns.

Table 8: Performance statistics for the balanced portfolio (1988-2017)

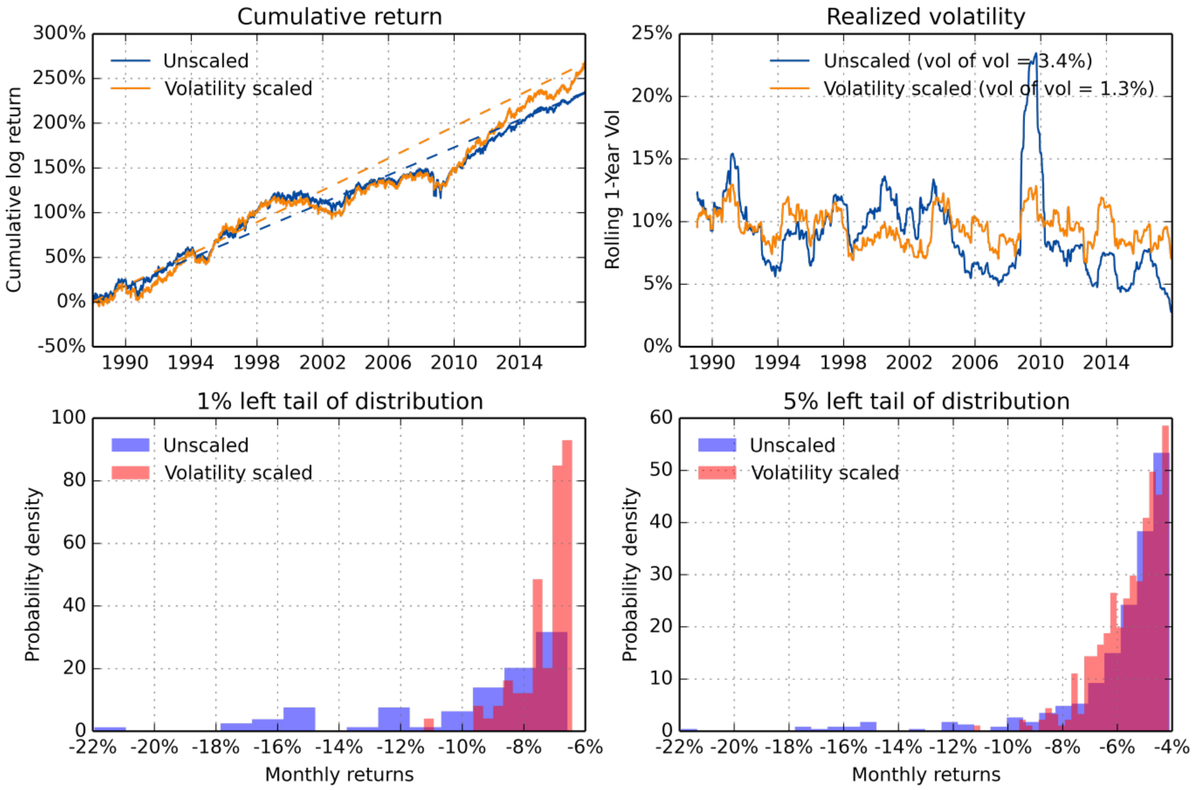
The table reports the gross and net (of estimated costs) Sharpe ratio, vol of vol, mean shortfall (left tail), and mean exceedance (right tail) statistics described in Table 2 for the 60-40 equity-bond "balanced" portfolio. We contrast an unscaled portfolio with a portfolio with volatility scaling at the asset level only and a portfolio with volatility scaling at both the asset and portfolio level. The volatility-scaling is done using exponentially-decaying weights with a half-life of 20 days. We use a rolling 1 month (21 weekdays) evaluation frequency. To facilitate comparison, both the unscaled and volatility-scaled returns are shown at 10% full-sample volatility.

		Sharpe ratio		Notional exposure		Vol of	Left tail		Right tail	
Asset scaling	**Portfolio scaling**	Gross	Net	Mean	Turnover	**vol**	1%	5%	95%	99%
Unscaled	Unscaled	0.80	0.80	153%	0.00	3.4%	-9.7%	-6.0%	6.8%	9.7%
Scaled (20-day half life)	Unscaled	0.87	0.87	179%	2.24	2.2%	-8.0%	-5.6%	6.6%	8.7%
Scaled (20-day half life)	Scaled (20-day half life)	0.91	0.90	183%	3.95	1.3%	-7.3%	-5.5%	6.5%	8.0%

In Figure 11, we contrast the cumulative return, realized volatility, and 1% and 5% left tail of the return distribution for an unscaled and volatility-scaled (at both the asset and portfolio level) balanced portfolio. Volatility-scaling is done using exponentially-decaying weights with a half-life of 20 days. Consistent with the results of Table 8, both the cumulative return and left tail improve with volatility scaling.

Figure 11: Cumulative returns, realized volatility, left and right tail returns for balanced portfolio (1988-2017)

This figure compares unscaled and volatility-scaled (exponential weighted, 20-day half-life) 60-40 equity-bond "balanced" portfolio returns for the 1988-2017 period. The top-left panel shows the cumulative return. The top-right panel shows the rolling 1-year standard deviation of 1-month (21 weekdays) overlapping returns. The standard deviation of the rolling 1-year standard deviation is reported in parentheses in the legend. The bottom-left and bottom-right panels show the lowest 1% and 5% of the rolling 1-month (21 weekdays) return distribution. To facilitate comparison, both the unscaled and volatility-scaled returns are shown at 10% full-sample volatility.



[21] Asvanunt, Nielsen and Villalon (2015) consider various strategies to reduce the size of tail events for 60/40 equity/bond portfolios. These include options-based approaches, and shifting to a risk parity asset class allocation based on risk exposures.

5.2 Risk parity portfolio

We now turn our attention to the 25-25-25-25 equity-bond-credit-commodity, or "risk parity" portfolio.[22] The equity, bond, and credit assets are for the US (S&P500, 10-year Treasury, and the credit index hedged with Treasuries used in Section 3). The commodity component is equally split between gold, copper, and crude oil, which one may consider "macro commodities" or securities that may serve as a (partial) inflation hedge. The diversifying potential of commodities is particularly relevant in this context, as a main motivation to invest in a risk parity portfolio is that its returns may be more consistent across different macro environments, including inflationary environments, than a more traditional balanced portfolio.

Our 25% allocation to commodities is a simplification of what asset managers do in practice. First, they often augment the commodity exposure with inflation-indexed bonds. Second, because commodity futures may not earn a passive risk premium over the long run (it is *a priori* not obvious if a premium is earned on the long or the short side), some asset managers supplement the passive long commodity exposure with dynamic overlays based on momentum or carry, for example.

In Table 9, we report the performance statistics for the illustrative risk parity portfolio. We again find that volatility scaling is useful at both the asset and portfolio level, and generally all the performance statistics improve compared to an unscaled portfolio.

Table 9: Performance statistics for the risk parity portfolio (1988-2017)

The table reports gross and net (of estimated costs) Sharpe ratio, vol of vol , mean shortfall (left tail), and mean exceedance (right tail) statistics described in Table 2 for the 25-25-25-25 equity-bond-credit-commodity "risk parity" portfolio. We contrast an unscaled portfolio with a portfolio with volatility scaling at the asset level only and a portfolio with volatility scaling at both the asset and portfolio level. The volatility-scaling is done using exponentially-decaying weights with a half-life of 20 days. We use a rolling 1 month (21 weekdays) evaluation frequency. To facilitate comparison, both the unscaled and volatility-scaled returns are shown at 10% full-sample volatility.

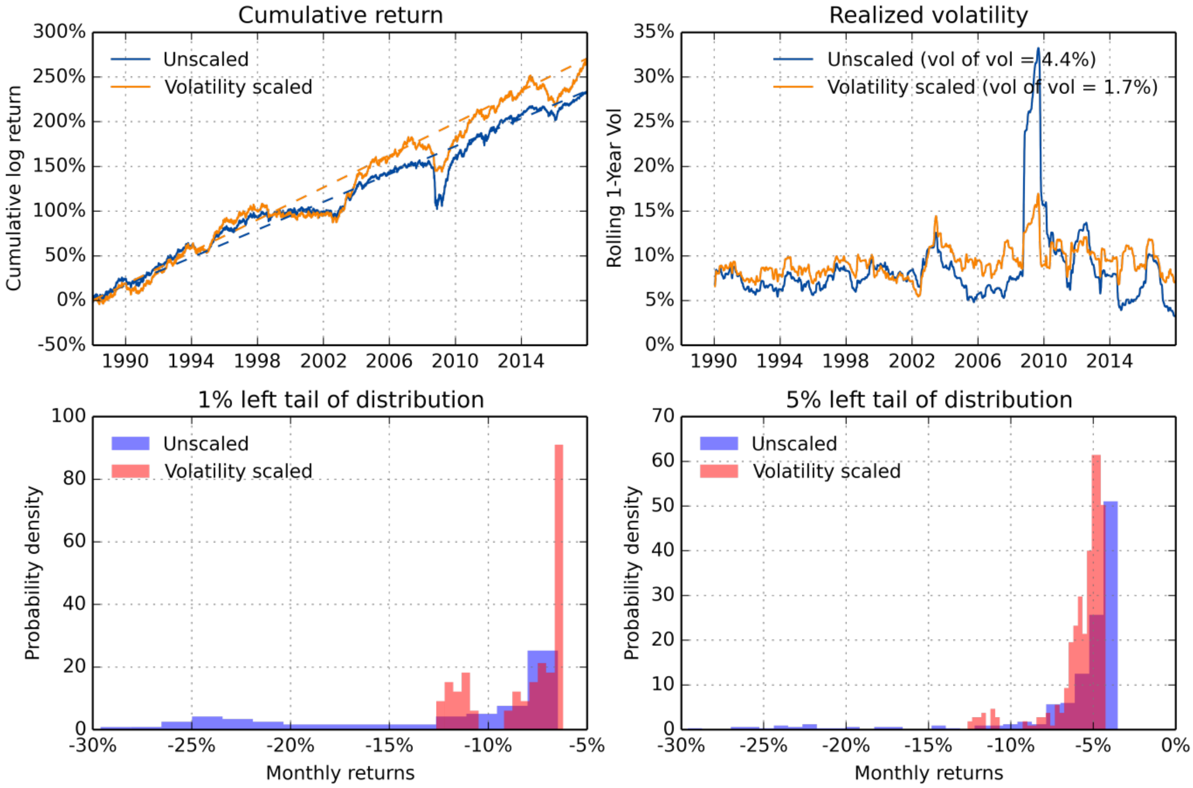
		Sharpe ratio		**Notional exposure**			**Left tail**		**Right tail**	
Asset scaling	**Portfolio scaling**	Gross	Net	Mean	Turnover	**Vol of vol**	1%	5%	95%	99%
Unscaled	Unscaled	0.80	0.80	268%	0.00	4.3%	-12.8%	-6.1%	6.5%	9.4%
Scaled (20-day half life)	Unscaled	0.89	0.88	412%	2.39	1.7%	-8.8%	-6.0%	5.9%	7.0%
Scaled (20-day half life)	Scaled (20-day half life)	0.91	0.89	402%	3.95	1.7%	-8.2%	-5.7%	6.2%	7.9%

In Figure 12, we contrast the cumulative return, realized volatility, and 1% and 5% left tail of the return distribution for an unscaled and volatility-scaled (at both the asset and portfolio level) risk parity portfolio. Volatility scaling is done using exponentially-decaying weights with a half-life of 20 days. Both the cumulative return and left tail improve with volatility scaling.

[22] Fleming, Kirby, and Ostdiek (2001, 2003) study the allocation across stocks, bonds, and gold using the conditional covariance matrix. Asness, Frazzini, and Pedersen (2012) compare the performance of a balanced and risk parity portfolios using monthly data.

Figure 12: Cumulative returns, realized volatility, left and right tail for the risk parity portfolio (1988-2017)

This figure compares unscaled and volatility-scaled (exponential weighted, 20-day half-life) 25-25-25-25 equity-bond-credit-commodities "risk parity" portfolio returns for the 1988-2017 period. The top-left panel shows the cumulative return. The top-right panel shows the rolling 1-year standard deviation of 1-month (21 weekdays) overlapping returns. The standard deviation of the rolling 1-year standard deviation is reported in parentheses in the legend. The bottom-left and bottom-right panels show the lowest 1% and 5% of the rolling 1-month (21 weekdays) return distribution. To facilitate comparison, both the unscaled and volatility-scaled returns are shown at 10% full-sample volatility.



6. Why does volatility scaling particularly improve the SR of risk assets?

In this section, we examine possible explanations for why volatility scaling improves the Sharpe ratio for risk assets, such as equities and credit, but has no effect on the Sharpe ratio of other assets. Our analysis suggests an answer that can be split into three parts: (1) only risk assets empirically display a so-called leverage effect, (2) the leverage effect effectively introduces some momentum, and (3) such a momentum overlay is beneficial for the Sharpe ratio. Indeed, we will show that the tendency of volatility scaling to introduce some momentum empirically explains much of the cross-sectional variation in the Sharpe ratio improvement when volatility scaling.

6.1 Leverage effect is confined to risk assets

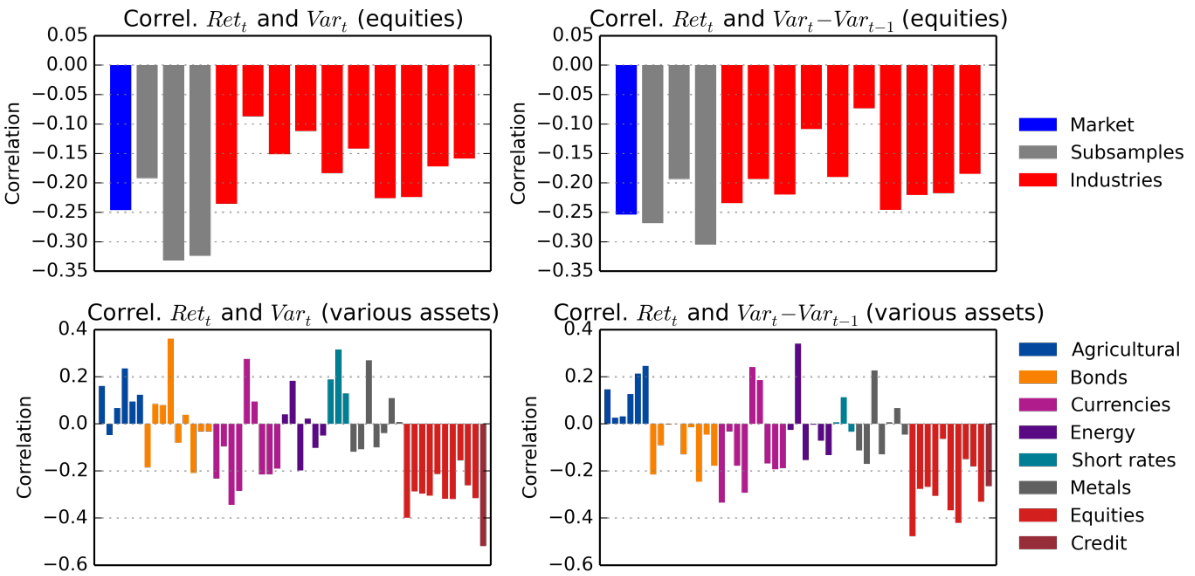
Equities and credit display a leverage effect, which is the tendency of returns to have a negative contemporaneous correlation to changes in volatility. The classic explanation by Black (1976) is that a negative equity return leads to a higher firm value-to-equity ratio (more leverage in the capital structure of the firm), which in turn means equity volatility should increase (holding constant the firm's cash flow volatility).[23]

[23] See also Christie (1982) for a discussion of the leverage effect. Bekaert and Wu (2000) argue that there is also a volatility feedback effect where the causality is reversed compared to the leverage effect; volatility increases give rise to higher risk premia and so negative returns.

In Figure 13, we indeed observe this leverage effect empirically for equities and credit, but not for other assets. The top panels show results for Equities All US (1926-2017), the three subsamples, and the 10 industry portfolios considered before in Section 2. The bottom panels show results for credit (Section 3) and the 50 futures and forwards (Section 4), for the 1988-2017 sample period. The right panels show the leverage effect: a negative correlation between monthly observations of the return and the change in variance. The left panels show a very similar picture for the correlation between monthly observations of return and the level of the variance.

Figure 13: Leverage effect for various assets

This figure shows the contemporaneous correlation of monthly observations for asset returns and variance (left panels) and change in variance (right panels). The variance estimate is based on intra-month daily data. The top panels show results for Equities All US (1926-2017), the three subsamples, and the 10 industry portfolios considered before in Section 2. The bottom panels show results for credit (Section 3) and the 50 futures and forwards (Section 4), for the 1988-2017 sample period.



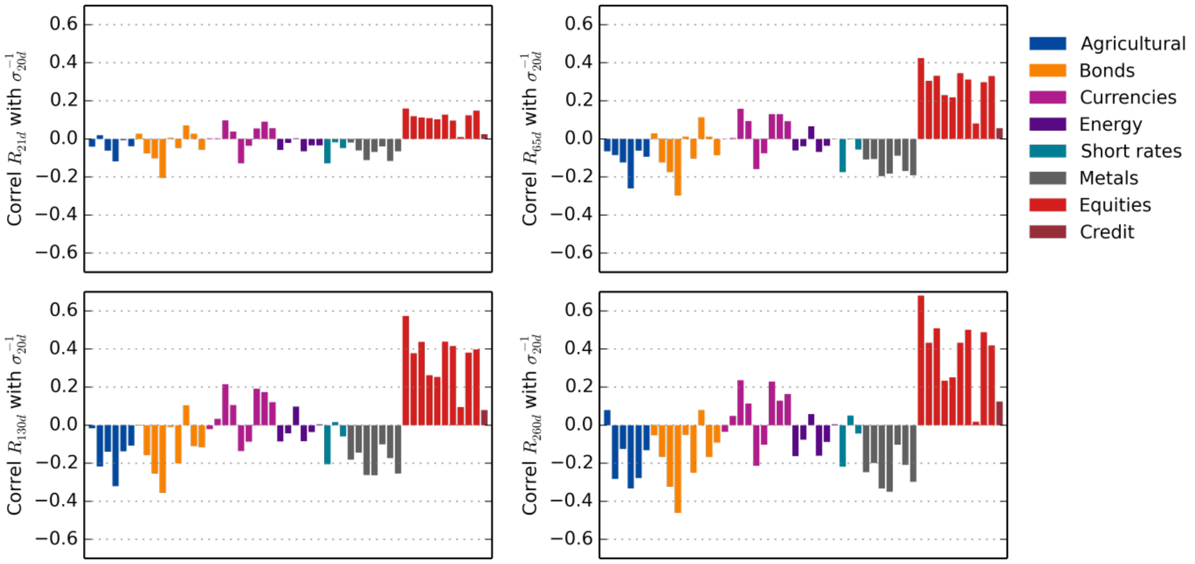
6.2 Leverage effect introduces some momentum

When applied to assets exhibiting the leverage effect, volatility scaling effectively introduces some time-series momentum into strategies. That is, negative returns tend to be followed by a reduction in the position size (as volatility is higher in that case) and positive returns tend to be followed by an increase in the position size (as volatility is lower in that case).

In Figure 14, we show more explicitly for which assets volatility scaling leads to changes in position sizes that are in the momentum direction (i.e., smaller after negative returns, bigger after positive returns). Specifically, we show the correlation between the reciprocal of the volatility estimate (which is proportional to position sizing when volatility scaling is applied) and the past 21-, 65-, 130-, and 260-day returns (1-, 3-, 6-, 12-month momentum) in the four panels respectively. Here we consider a 20-day half-life for the volatility estimate. Mirroring the results for the leverage effect in Figure 13, we see in Figure 14 that only risk assets consistently show a positive correlation between position sizing (reciprocal volatility) and momentum (past returns). In fact, for other assets the correlation is predominantly negative, introducing a bet on mean reversion.

Figure 14: correlation past returns and 1/vol for various assets

This figure shows the correlation between the reciprocal of the volatility estimate (which is proportional to position sizing when volatility scaling is applied) and the past 21-, 65-, 130-, and 260-day returns (1-, 3-, 6-, 12-month momentum) in the four panels respectively. The volatility estimate is based on exponential weighted returns with a 20-day half-life. We consider credit (Section 3) and the 50 futures and forwards (Section 4), for the 1988-2017 sample period.



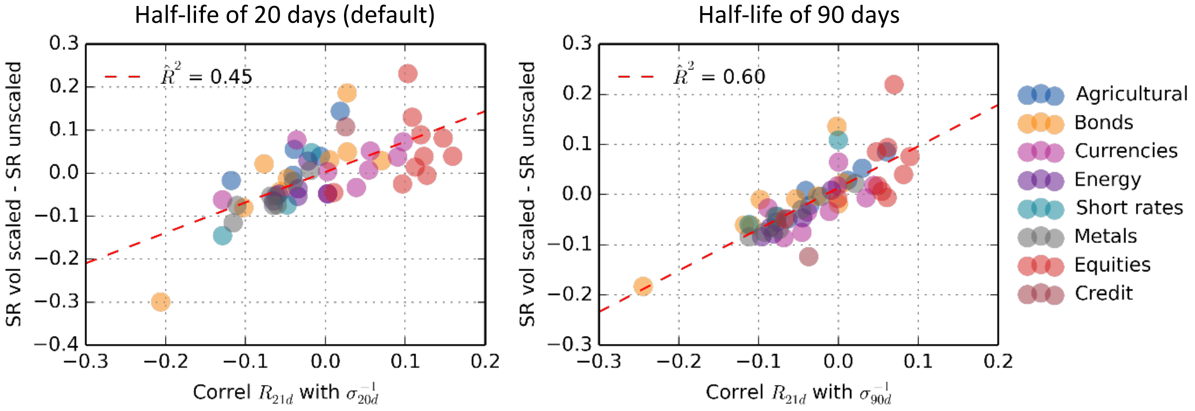
6.3 Linking momentum-ness of returns and the impact of volatility scaling on the Sharpe ratio

The final part of our investigation is to link directly cross-sectional differences in the impact of volatility scaling on the Sharpe ratio and asset return properties.

The evidence suggests that time-series momentum strategies have historically performed well; see, e.g., Hamill, Rattray, and Van Hemert (2016).[24] In Figure 15, we show that it is indeed the "momentum-ness" of volatility scaling that explains a large part of the cross-sectional variation in the Sharpe ratio improvement when volatility scaling for the various assets considered. We find the shorter-term, 1-month, momentum of returns to be most relevant here. Using a 20-day half-life for volatility scaling, the R-squared is 45%. For a slower volatility estimate using a 90-day half-life, the R-squared is even higher at 60%.

Figure 15: improvement SR when volatility scaling vs correlation (past returns, 1/vol) for various assets

This figure plots the improvement from volatility scaling (vertical axis) versus the "momentum-ness" of volatility scaling, determined as the correlation between the past 21-day returns and the reciprocal of the volatility estimate, on the (horizontal axis). The volatility volatility estimate is based on exponential weighted returns with a 20-day half-life (left panel) and 90-day half-life (right panel). We consider credit (Section 3) and the 50 futures and forwards (Section 4), for the 1988-2017 sample period.



[24] Dachraoui (2018) also argues there is a link between the presence of a leverage effect and the extent to which volatility targeting improves the performance for an asset.

7. Concluding remarks

In this paper, we contrasted the performance of both individual assets and portfolios with a constant notional exposure (unscaled) to strategies that target a constant level of volatility (scaled). Our initial evidence, consistent with recent studies, indicates that volatility scaling helps to boost Sharpe ratios. However, most recent research has focused on equities. Our results show that this boost is specific to so-called "risk assets" (e.g., equity and credit) or portfolios that have a sizable allocation to these risk assets. That is, for other assets, such as fixed income, currencies and commodities, the effect of a simple volatility scaling on the Sharpe ratio is negligible.

While the Sharpe ratio is important, most investors have broader investment objectives. We show that volatility scaling has one unambiguous effect across assets and asset classes: it reduces the likelihood of extreme returns (and the volatility of volatility). In particular the lower probability of very negative returns (left-tail events) is valuable for investors.

While we provided a detailed historical account of the impact of volatility targeting across 60+ assets and two multi-asset portfolios, some topics are beyond the scope of this paper. We will comment on three.

First, the detailed analysis for equity and bonds was done for US assets, for which we have the longest daily return history. A caveat of this approach is that the US is an ex-post winner in the sense that over the past century it had robust economic growth and no major war on its own soil. This may particularly matter for bonds, which can start to resemble a credit investment when the creditworthiness of a government is questioned by investors. As such, our finding that volatility scaling does not meaningfully improve the Sharpe ratio of a bond investment should also be caveated, and going forward volatility scaling may improve the Sharpe ratio of bonds that unexpectedly start to behave in a more "credit-like" manner.

Second, while the focus in this paper was on volatility scaling, there are other methods with the potential to improve the risk management of a long portfolio. Hamill, Rattray, and Van Hemert (2016) show that trend-following strategies tend to work particularly well at times of equity and bond market sell-offs. Hence a trend-following overlay may further improve the risk and return of a long portfolio. Indeed, in another Man Group paper, Haydon (2018) illustrates the benefits of such an overlay for a balanced 60-40 equity-bond portfolio.

Finally, while we explored intraday data for S&P500 and Treasury futures and found some benefits vis-à-vis daily data, we believe in this paper we only scratched the surface of this topic. More assets now have good quality intraday data, and for more hours of the day. In addition, advances in statistical modelling may help us to use the intraday data to get more timely estimates, see, e.g., Noureldin, Shephard, and Sheppard (2012) for a discussion on multivariate high-frequency-based volatility (HEAVY) models.

References

Andersen, T., T. Bollerslev, F. Diebold, and P. Labys (2003), "Modeling and forecasting realized volatility", *Econometrica*, 71(2), 579-625.

Asness, C., A. Frazzini, and L. Pedersen (2012), "Leverage aversion and risk parity", *Financial Analysts Journal*, 68(1), 47-59.

Asvanunt, A., L. Nielsen, and D. Villalon (2015), "Working Your Tail Off: Active Strategies Versus Direct Hedging", *Journal of Investing,* 24(2), 134-145.

Barroso, P., and P. Santa-Clara (2015), "Momentum has its moments", *Journal of Financial Economics*, 116(1), 111–120.

Bekaert, G., and G. Wu (2000), "Asymmetric volatility and risk in equity markets", *Review of Financial Studies*, 13(1), 1–42.

Black, F. (1976), "Studies of stock price volatility changes" *Proceedings from the American Statistical Association, Business and Economic Statistics Section*, 177-181.

Bollerslev, T., B. Hood, J. Huss, and L. Pedersen (2018), "Risk everywhere: modeling and managing volatility", working paper.

Cook, M., E. Hoyle, M. Sargaison, D. Taylor, and O. Van Hemert (2017), "The best strategies for the worst crises", Man Group working paper.

Christie, A. (1982), "The stochastic behavior of common stock variances: Value, leverage and interest rate effects", *Journal of Financial Economics*, 10(4), 407-432.

Dachraoui, K. (2018), "On the optimality of target volatility strategies", *Journal of Portfolio Management*, 44(5), 58-67.

Daniel, K., and T. Moskowitz (2016), "Momentum crashes", *Journal of Financial Economics*, 122(2), 221-247.

Dopfel, F., and S. Ramkumar (2013), "Managed Volatility Strategies: Applications to Investment Policy", *Journal of Portfolio Management*, 40(1), 27-39.

Engle, R (1982), "Autoregressive conditional heteroskedasticity with estimates of the variance of UK inflation", *Econometrica*, 50, 987-1008.

Fleming, J., C. Kirby, and B. Ostdiek (2001), "The economic value of volatility timing.", *Journal of Finance*, 56(1), 329-352.

Fleming, J., C. Kirby, and B. Ostdiek (2003), "The economic value of volatility timing using 'realized' volatility", *Journal of Financial Economics*, 67(3), 474-509.

Granger, N., D. Greenig, C. Harvey, S. Rattray, D. Zou (2014), "Rebalancing risk", AHL working paper.

Hamill, C., S. Rattray, and O. Van Hemert (2016), "Trend following: equity and bond crisis alpha", Man AHL working paper.

Haydon, K. (2018), "Volatility is back: better to target returns or target risk?", Man Group working paper.

Hocquard, A., S. Ng, and N. Papageorgiou (2013), “A constant-volatility framework for managing tail risk”, *Journal of Portfolio Management*, 39(2), 28-40.

Hogg, R. (1972), “More light on the kurtosis and related statistics”, *Journal of the American Statistical Association*, 67(338), 422-424.

Moreira, A., and T. Muir (2017), “Volatility managed portfolios”, *Journal of Finance*, 72(4), 1611-1644.

Noureldin, D., N. Shephard, and K. Sheppard (2012), “Multivariate high-frequency-based volatility (HEAVY) models”, *Journal of Applied Econometrics*, 27(6), 907-933.

Appendix A: Other risk metrics

In this appendix, we explore the following additional risk metrics to contrast unscaled and volatility-scaled returns: skewness, kurtosis, tail skewness, tail kurtosis, and (maximum) drawdown. Throughout, we use a volatility estimate based on exponentially-decaying weights (20-day half-life). As before, we evaluate these statistics for 1-month (30 calendar days/21 weekdays) returns.

Skewness and kurtosis are the third and fourth central moment. Kurtosis is reported in excess of 3, so a normal distribution has a kurtosis value of zero. We can also define *tail skewness* and *tail kurtosis* based on mean shortfalls and exceedances. Writing U_α for the mean exceedance of the $(1-\alpha)$-quantile, and L_α as the mean shortfall of the α-quantile, Hogg (1972) proposed the following as a measure of kurtosis:

$$\frac{U_{0.05}-L_{0.05}}{U_{0.5}-L_{0.5}}.$$

We define tail kurtosis to be the excess version of this statistic. That is, we subtract a constant so that normal random variables have zero kurtosis:

$$\text{Tail kurtosis} = \frac{U_{0.05}-L_{0.05}}{U_{0.5}-L_{0.5}} - 10\exp\left\{-\frac{1}{2}\Phi^{-1}(0.05)^2\right\},$$

where $\Phi^{-1}(x)$ is the standard normal quantile function.

Along similar lines, we define tail skewness as the following measure of asymmetry between tails:

$$\text{Tail skewness} = \frac{U_{0.05}+L_{0.05}-2M}{U_{0.5}-L_{0.5}},$$

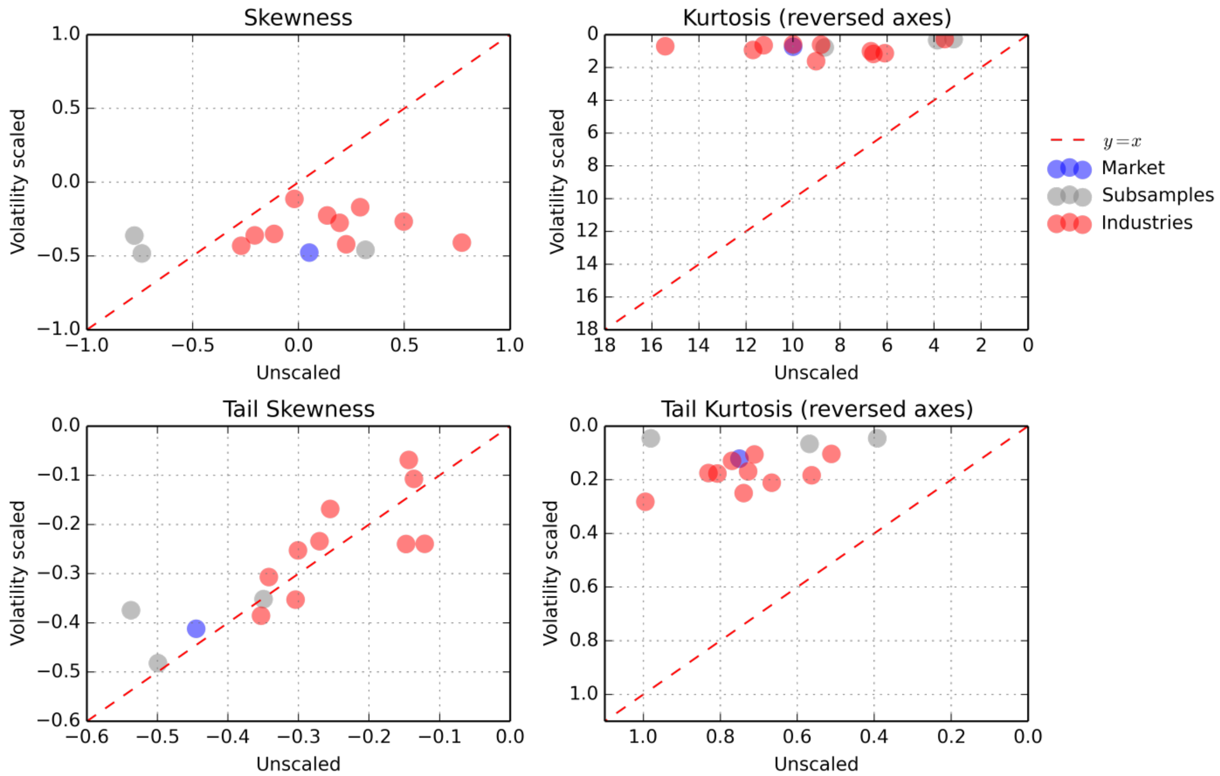
where M is the median.

In Figure A1, we show the kurtosis (top-right panel) and tail kurtosis (bottom-right panel) for Equities All US over the full 1927-2017 sample period, for three subsamples, and for 10 industry portfolio for the 1927-2016 period also. Both kurtosis and tail kurtosis are much lower with volatility scaling, which is expected since we have already shown that both the left and right tails are thinned by volatility scaling.

The results for skewness (top-left panel) are more mixed, with better (less negative) skewness in two of the three subsample periods, but worse skewness in the remaining subsample, the full sample, and the 10 industry portfolios. The tail skewness (bottom-left panel) is fairly similar for unscaled and volatility-scaled returns, in line with earlier findings that the left and right tail are reduced similarly when volatility scaling. As we argued before, we believe investors likely care about the left tail much more than the right tail (for a given Sharpe ratio), rendering skewness or tail skewness (which are impacted by both) less useful risk metrics.

Figure A1: Skewness and kurtosis for Equities All US (1926-2017)

This figure compares unscaled and volatility-scaled (exponential weighted, 20-day half-life) excess returns for Equities All US full sample (blue dots), subsamples (grey dots), and 10 industries full sample (red dots). For both skewness (left panel) and kurtosis (right panel), observations above the dashed diagonal line correspond to a situation where volatility scaling improves the statistic (using reversed axes for kurtosis). Both statistics are based on a rolling 1 month (30 calendar days) evaluation frequency. To facilitate comparison, both the unscaled and volatility-scaled returns are shown at 10% full-sample volatility.



In Figure A2, we repeat the exercise for the 50 futures and forwards studied in Section 4. Kurtosis and tail kurtosis are much reduced with volatility scaling, while the results for skewness and tail skewness are mixed. This mirrors the results of the Equities All US shown in Figure A1.

Finally, in Figure A3, we look at drawdown plots for Equities All US (top panel), the balanced portfolio (middle panel), and the risk parity portfolio (bottom panel). The drawdown level at a given point in time is determined by comparing the total return index level (cumulative return since the start) to the maximum level achieved up to that point in time (the high-water mark):

$$\text{Drawdown}_t = \frac{\text{Index}_t}{\max\limits_{s \le t} \text{Index}_s} - 1.$$

In all three cases, the maximum drawdown (reported in the legend) is substantially reduced with volatility scaling: from to -60.9% to -42.8% for Equities All US (top panel), from -28.1% to -22.2% for the balanced portfolio, and from -42.1% to -32.1% for the risk parity portfolio. Volatility scaling particularly reduced the drawdown during the Great Depression in the 1930 and during the 2007-2009 Global Financial Crisis. That said (maximum) drawdown is not our preferred risk metric, because it is derived from a single realized return path, and as such is not a robust metric.

Figure A2: Skewness and kurtosis for futures & forwards (1988-2017)

This figure compares unscaled and volatility-scaled (exponential weighted, 20-day half-life) returns for 50 global futures and forwards markets, where different sectors are represented by different colors. For both skewness (left panel) and kurtosis (right panel), observations above the dashed diagonal line correspond to a situation where volatility scaling improves the statistic (using reversed axes for kurtosis). Both statistics are based on a rolling 1 month (21 weekdays) evaluation frequency. To facilitate comparison, both the unscaled and volatility-scaled returns are shown at 10% full-sample volatility

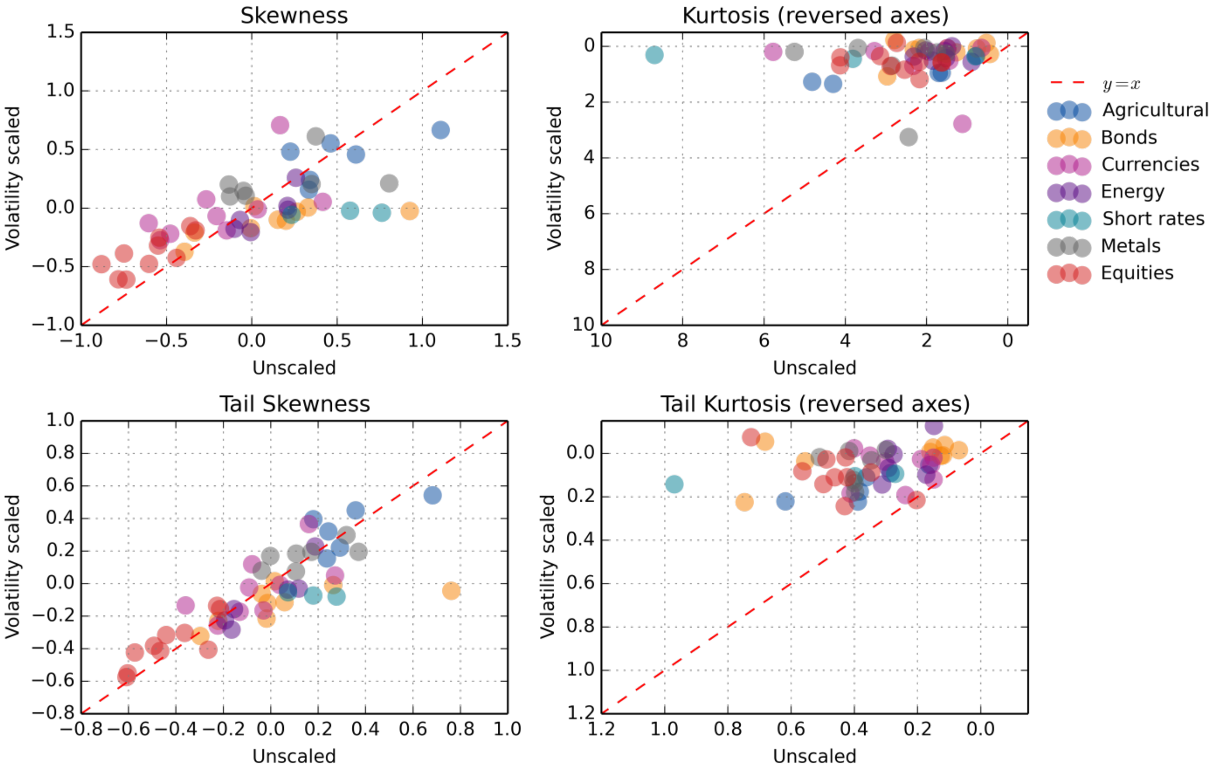
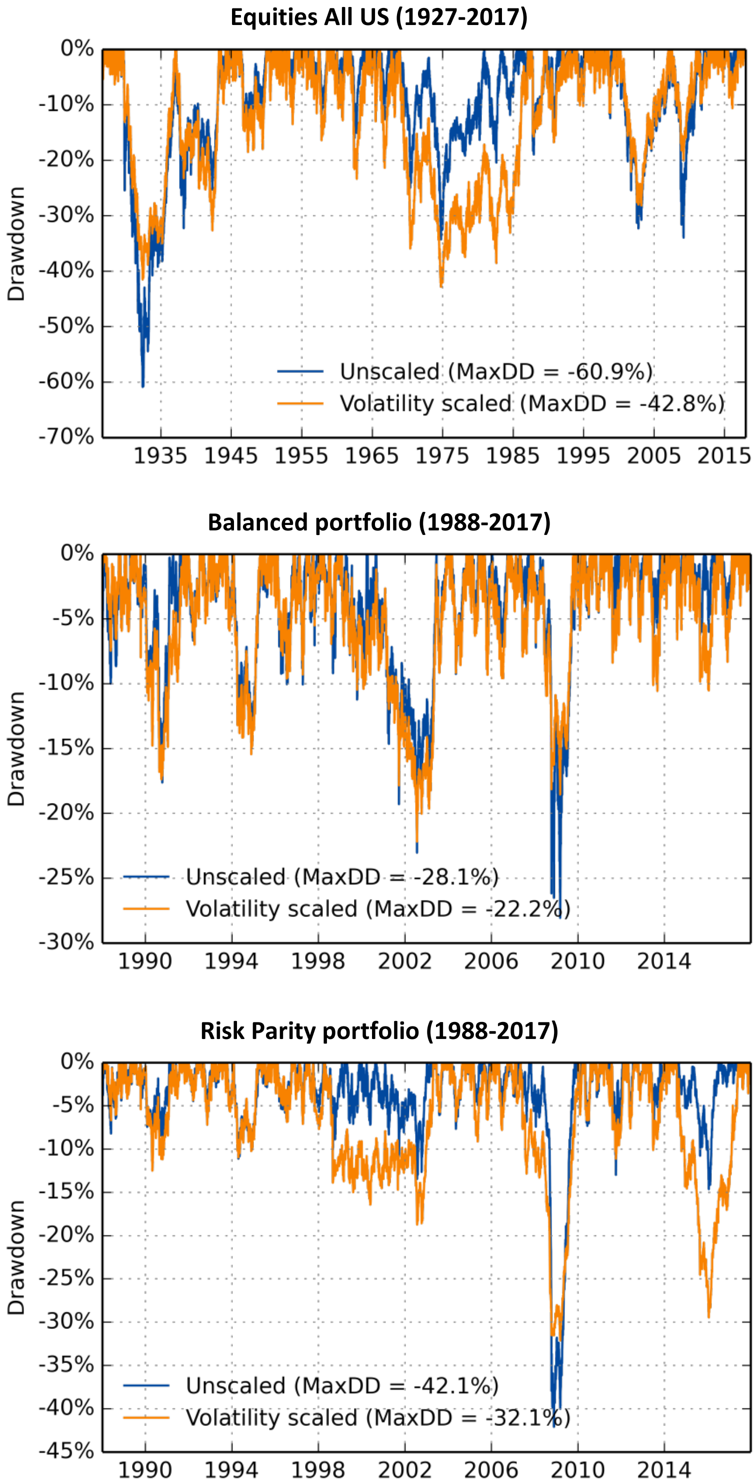


Figure A3: Drawdown plot for Equities All US, Balanced portfolio, and Risk Parity portfolio

This figure compares the unscaled and volatility-scaled (exponential weighted, 20-day half-life) drawdown plot for Equities All US (1927-2017), the 60-40 equity-bond Balanced portfolio (1988-2017), and the 25-25-25-25 equity-bond-credit-commodities Risk Parity portfolio (1988-2017).

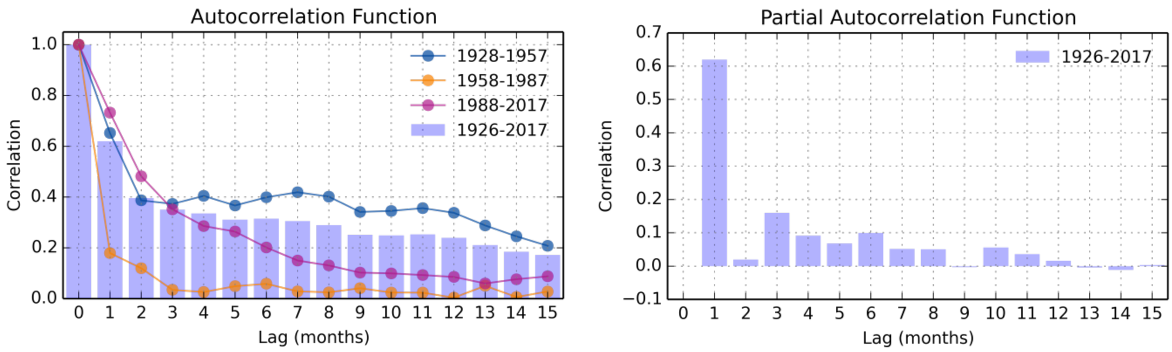


Appendix B: autocorrelation of variance

To further illustrate that equity volatility clusters, we show in Figure B1 (left panel) the autocorrelation of the monthly squared volatility (i.e., variance) of daily returns.[25] The variance of adjacent months are around 0.6 correlated over the full (1926-2017) sample period. The correlation slowly decays for additional lags, and is around 0.2 for months a year apart. Of the three sub-samples considered, the middle (1958-1987) stands out as having much less autocorrelation in the monthly variances. As was already visible in Figure 1, the middle sub-sample corresponds to a period with fewer extreme bursts in volatility in the first place. In Figure B1 (right panel), we show the partial autocorrelation, measuring how predictive the lag k variance is for the current month, after taking into account the effect of lags 1 to k-1. Most of the predictive power is captured by lag 1.

Figure B1: Persistence of monthly variance for Equities All US (1926-2017)

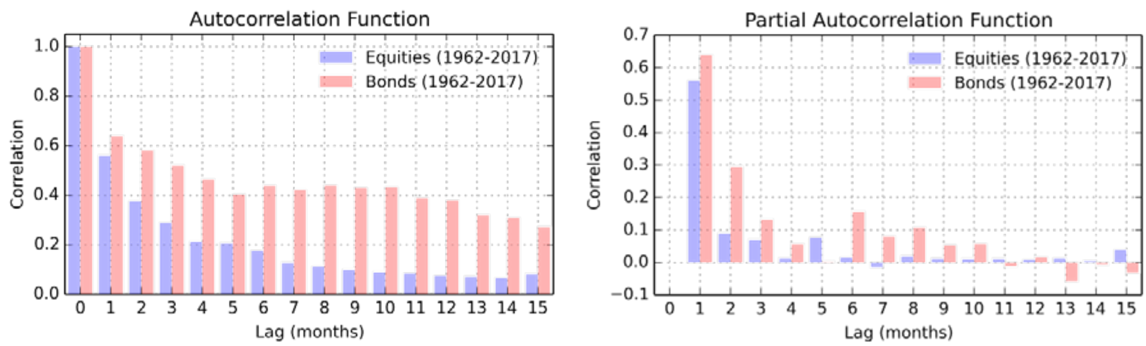
The figure shows the autocorrelation (left panel) and partial autocorrelation (right panel) of the non-overlapping monthly variance of daily Equities All US excess returns for the full sample period (1926-2017), and the autocorrelation also for the three 30-year sub-sample periods.



In Figure B2, we repeat the above exercise including bonds for the 1962-2017 period, for which we have (proxy) daily bond returns. Bond variance is much more persistent, and, in fact, the autocorrelation only slightly falls from lag 3 to 12. This is just another manifestation of the various prolonged volatility regimes the bond market has experienced, which we discussed earlier. In contrast, equity markets experience volatility clusters of, say, half a year, but do not exhibit the prolonged regimes the bond market experienced.

Figure B2: Persistence of monthly US equity and bond variance (1962-2017)

The figure shows the autocorrelation (left panel) and partial autocorrelation (right panel) of the monthly variance of daily bond (US, proxied from yield data) and Equities All US excess returns for the 1962-2017 period.



[25] Modelling of volatility, including autocorrelation analyses, is typically conducted with squared volatility, i.e., variance.