Supplementary Information: data and methodology detail

At the heart of our analysis are daily U.S. large cap equity returns, 7/1/1926 – 6/29/2018, obtained from Ken French's website at http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html. We use his excess return of the market (B_m-B_f) as the benchmark return series in our paper. Our volatility forecast for period *t* of various lookback horizons *n* is then formed as the simple realized standard deviation over the trailing *n* trading days [*t-n, t-n+1, …, t-2, t-1*]. Our target volatility for period *t* is set as the prior realized daily volatility in the excess return series from the beginning of the sample, i.e., [*1, 2, 3, …, t-2, t-1*]. From this we calculate the target equity exposure as the ratio of volatility target to predicted volatility. The daily return of the managed volatility portfolio amounts to that target weight times the realized equity market excess return for the corresponding day *t*. Unless otherwise specified, we use the trailing 20-day realized volatility as the volatility forecast when computing the MV series.

Our analysis features three separate sample periods that amount to different starting points for evaluating managed volatility performance. The first and longest period starts on 11/11/1929 (after an arbitrarily picked burn-in period of exactly 1,000 trading days), with the other periods starting on 1/4/1960 and 1/2/1990. To make the daily series more directly comparable to each other, we re-normalize the daily managed volatility return to have the same realized daily standard deviation as the underlying equity market excess return over the sample period under consideration[1].

We examine performance via cumulative returns for holding periods of 1, 20, 60, 240, 720, and 2,400 trading days – loosely corresponding to investor evaluation horizons of 1 day, 1 month, 1 year, 3 years, and 10 years. All performance metrics are based on returns in excess of cash. They are calculated directly off the holding period returns and not annualized. Given our emphasis on holding period length, we had to contend with the fact that evaluating non-overlapping windows is statistically more robust, but exposes us to an arbitrary calendar convention: there are *n* different non-overlapping histories

[1] This is done with the benefit of hindsight (in light of the full sample) and represents the only look-ahead bias in this analysis that we're aware of.

when investigating the holding period of length *n*. We resolve this tension by calculating our results for all possible histories, and averaging the statistics across them. Since our confidence intervals are based on bootstrapping this very process (see below), they are ultimately based on non-overlapping windows, and thus do not suffer from the biases normally introduced when calculating standard errors based on overlapping observations. Our work with overlapping intervals follows Sun et al. (2009), who illustrate that this approach yields unbiased estimates. For the shortest time period (since 1990), metrics are only shown if at least 10 non-overlapping data points are available in each individual series (before averaging across series).

We directly follow Morningstar (2009) in calculating utility based certainty equivalents (CEV). Based on their formula (10), we define the CEV of a given return stream as

$$(1) \qquad CEV = \left[\frac{1}{T}\sum_{t=1}^{T}(1+ER_t)^{-\gamma}\right]^{-\frac{(\frac{252}{d})}{\gamma}} - 1$$

Here, γ is the degree of risk aversion, *ER* is the excess return above cash of the portfolio, and *d* is the holding period under investigation, in business days (with 252 being the average number of active trading days in a year). Based on this definition of utility, our scoring of benchmark vs. MV strategies is done as follows. As above for the direct return evaluation, we build up longer period holding period returns from the same daily benchmark and MV return using all the possible non-overlapping histories. But in the case of utility and risk aversion, we also want to span a range of portfolios with different risk levels. We do this by first creating a range of 20 benchmark portfolios that have 5%, 10%, …, 95%, 100% equity exposure[2]. Next, we create a corresponding set of daily MV return streams amounting to 5%, 10%, …, 95%, 100% of the MV return studied in earlier parts of the paper. We then aggregate all these daily returns to all the relevant holding periods, as before. For a given utility function (as parameterized by γ), and a given holding period, we then calculate the CEV for all 20 benchmark and MV

[2] Here, the 100% version corresponds to the benchmark used for the traditional performance evaluation. Since we score excess returns above cash (and do not model a fixed income alternative), the remainder of each portfolio is effectively in cash, earning zero excess return above cash.

portfolio candidates separately. We score the benchmark and MV by selecting the best (highest CEV) portfolio from each family. The difference between these two CEVs forms our basis for comparatively scoring benchmark and MV in a utility context[3]. This is similar to contemplating the "utility efficient frontier" spanned by the benchmark vs. MV portfolios. Our perspective is that of an investment manager offering a range of MV portfolios with different risk profiles, in addition to the benchmark portfolios available. Our scoring then asks the question, for a given investor (characterized by risk aversion and holding period) who self-selects into choosing the preferred portfolio from each lineup, how much more useful is the MV lineup vs. the benchmark lineup?

Additionally, we use bootstrapping to perform inference on the difference in strategy performance between benchmark and MV. We utilize the bootstrap methodology because the true distribution of the test statistics is unknown. Recall that for holding periods of *n* days ($n>1$), we report average statistics for the *n* possible ways of defining non-overlapping histories from our historical data. Our bootstrapping process mimics this approach. For each non-overlapping history, we bootstrap periods of the same length as our historical sample from the data (1,000 scenarios, with replacement). The same time index values of draws are taken from the MV and benchmark samples within each bootstrap, in order to ensure that test statistics for benchmark and MV are based on the same historical returns within each scenario. We obtain a distribution for the relevant test statistic (difference in Sharpe Ratio or CEV between MV and benchmark), using the average across all histories within each same bootstrap iteration, consistent with the computation of test statistics. We then compute the 5^{th} and 95^{th} percentiles of the distribution of the statistic, which we report as a confidence interval in the relevant tables in the paper.

[3] Note that the optimal benchmark portfolio can feature a different risk level than the optimally chosen MV portfolio. E.g., for gamma = 7, the optimal benchmark portfolio may be the 45% version, but the optimal MV portfolio may be the version featuring 35% (rather than 45%) exposure to the original MV portfolio. Such gaps of 1-2 "notches" on the risk spectrum are quite common, but beyond that the optimal portfolios tend to be in the same region of the risk spectrum for both portfolio families.

Supplementary Information: Portfolio return normalization

Another way to appreciate the robustness of the improvement in return normalization provided by MV is shown in Figure A1. Rather than focusing on variability in SR with different backtest starting points in Figure 1, Figure A1 focuses on kurtosis and skewness for MV and BM. Additionally, this analysis measures the kurtosis and skewness of non-overlapping 60-day holding periods in every historical 10-year period, rolling forward daily. In direct contrast to the variability of SR with the sample starting point, the kurtosis improvement from MV for shorter (in this case, 60 day) holding periods remains persistent throughout, while it is quite variable for BM.

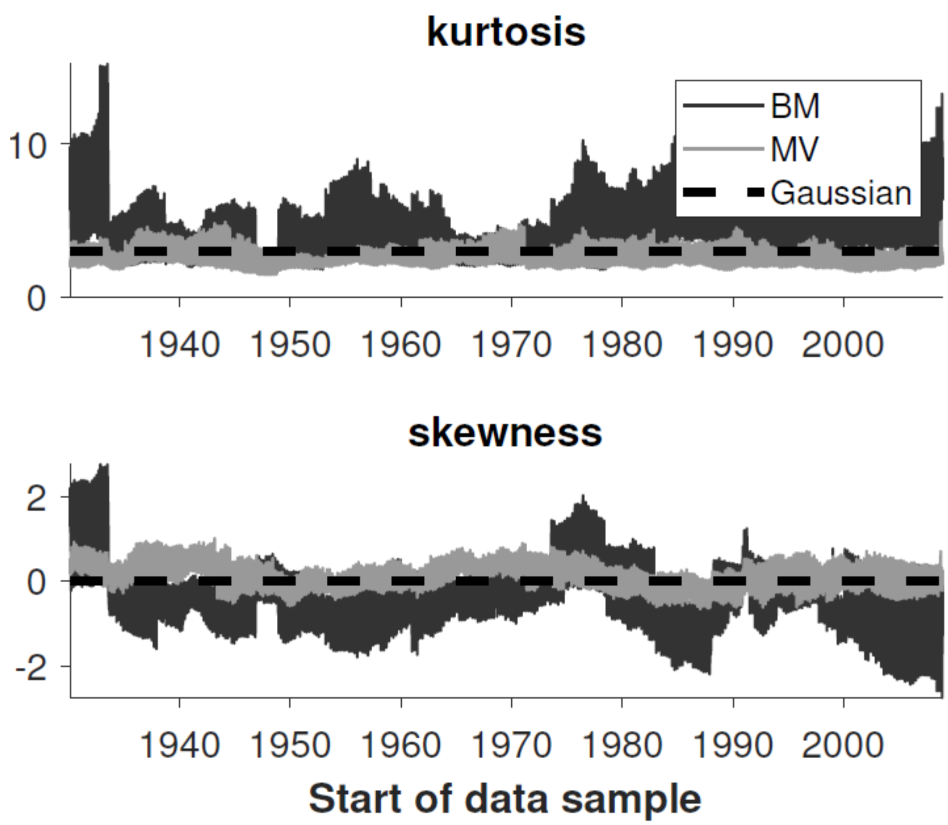

Figure A1. Comparison of benchmark and MV skewness and kurtosis over non-overlapping 60-day holding periods over daily rolling 10-year periods.

Supplementary Information: Implementation considerations

Our unconstrained MV strategy serves as an illustration of the basic benefits of volatility management. We used a near-term forecast of volatility, and modified exposures, without constraints, in response to changes in that forecast. In this section, we examine the impact of realistic implementation considerations on the utility improvements we illustrated earlier in the paper. We only consider implementation constraints that do not dilute the volatility management value proposition. For example, many managed volatility strategies in the marketplace cap equity exposure or do not employ leverage due to regulatory considerations or client preferences. Partial allocations to MV are also possible. In many cases, these additional constraints serve to regulate the extent to which MV can depart from the benchmark, reducing the tracking error of MV compared to a static benchmark – as opposed to transaction cost concerns. They are explicit or implicit ways of regulating the portfolio allocation to MV. While these are valid and important considerations, they lead to exposure modifications that are no longer wholly consistent with the expectations for an MV strategy.

Instead, we focus on implementation considerations that retain the full range of allowable exposures, by allowing for constraints on daily trade sizes. To reduce the impact of small trades, we arbitrarily impose a minimum daily trade floor of 10%. When optimal trades result in exposure changes lower than the minimum trade floor, we carry over the exposure from the previous day. On the other end of the spectrum, an unconstrained MV may lead to very large daily changes in exposure. At times, these changes signal a change in regime, and at other times, they revert. To limit the turnover impact of these large exposure changes, we also arbitrarily select a maximum trade size of 50% per day. When the optimal trade size exceeds the maximum, the daily exposure only changes by the maximum allowable amount. While the minimum and maximum trade size constraints permit cumulative exposure levels similar to those of unconstrained MV, they decelerate trading activity, thereby reducing turnover and transaction costs.

We also consider the utility benefit remaining after incorporating reasonable transaction costs. A rapidly changing near-term volatility forecast, without any trading constraints or penalties, can lead to large daily changes in exposures, potentially incurring high transaction costs. To examine their impact, we re-calculated utility improvements, while penalizing trades for a reasonable bid-ask spread[4].

Figure A1 illustrates a modest reduction in utility gains when adding transaction costs to the baseline (uncapped) MV strategy. Focusing on data since 1929, transaction costs cause a 10-20% reduction in the utility improvement. Importantly, a meaningful utility improvement relative to the benchmark remains even after incorporating transaction costs. When we additionally impose the trading caps described above, the utility improvement remains intact, even rises moderately for the shortest holding periods. We interpret this improvement in utility as a result of removing nuisance trades that ultimately do not impact the efficacy of the MV strategy. In both cases, with and without daily trading caps, the utility improvement relative to the benchmark is meaningful and remains largely intact. We draw similar conclusions when the sample begins in 1990. The picture starting in 1960 is somewhat different. The 1960 starting point represents an inopportune time for volatility management, and consequently, we only observe a small baseline improvement in utility in that sample. When factoring in transaction costs, the utility improvement degrades almost completely.

[4] To match the large cap US equity exposure in the benchmark, the hedge could be implemented using S&P Index futures. Futures are a cheap way to apply leverage to a portfolio. We assumed a 3 bps bid-ask spread on the future, incurred for every trade. We did not factor in roll costs, margin requirements or collateral management.

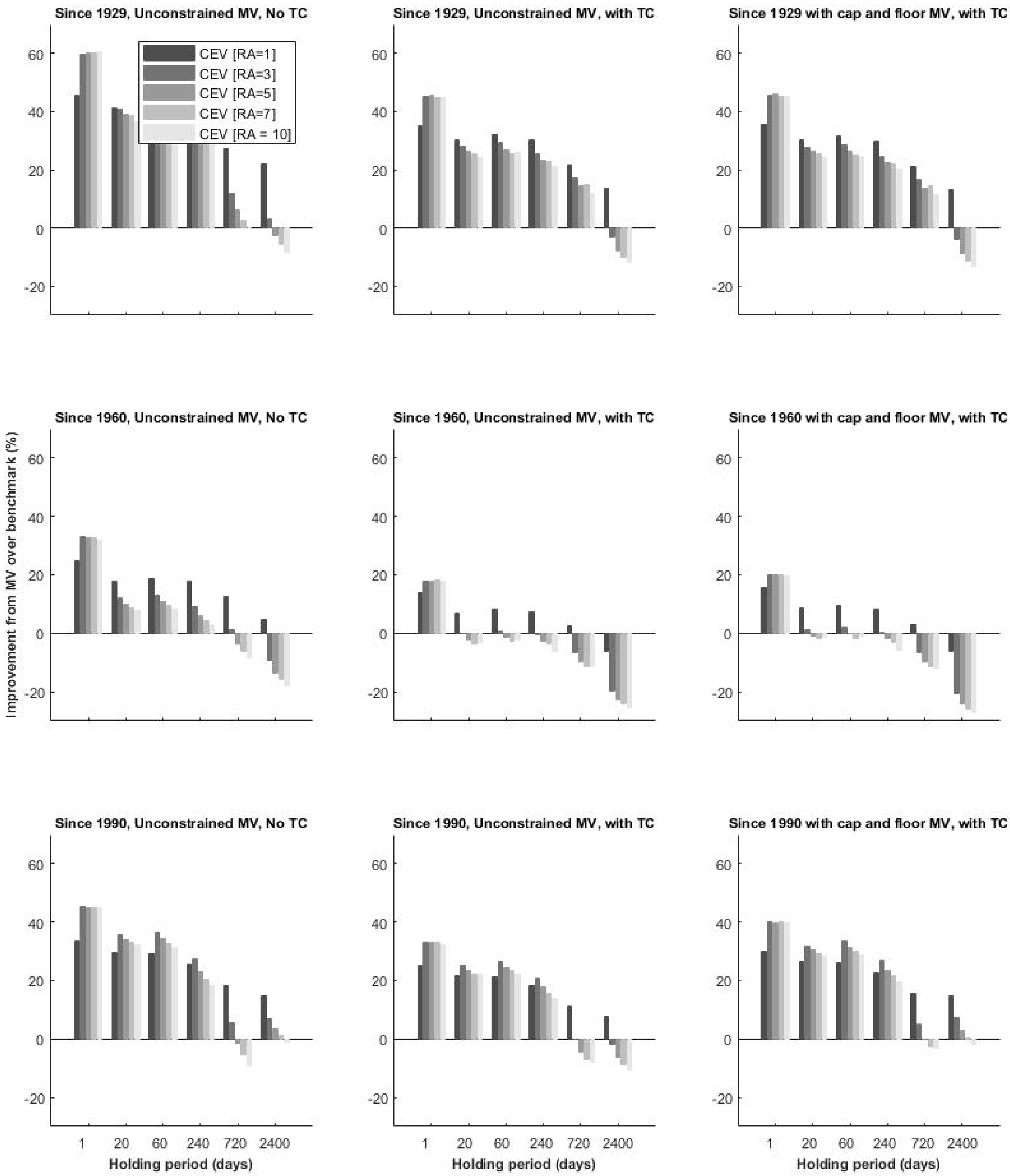

Figure A2 Percentage improvement of MV over benchmark for CEV with and without transaction costs.