Stockholm School of Economics

Department of Finance
Master of Science, 2019

An Empirical Evaluation of Improved Volatility-Based Trading Strategies

Authors*
Raphael Bierschenk & Stefan Wennemar

Supervisor
Michael Halling

ABSTRACT

In 2017, Moreira and Muir published their paper “Volatility-Managed Portfolios”, showing that investors can beat the market, purely by choosing their risk exposure based on the inverse of last month’s realized variance. While their results are influential in nature, suggesting, against common belief, investors should take less risk in recessions, they singularly rely on realized variance as a risk measure and a fixed monthly rebalancing period. Our paper, therefore, analyzes the effect of altering these prepositions by using more sophisticated variance estimation methods and by varying rebalancing periods on a fixed and a flexible basis. We show, that doing so can not only increase risk-adjusted outperformance, but also elicit desirable characteristics, like a lower tail risk, decreased transaction costs, and higher cumulative performance. Furthermore, we find that simpler methods of variance estimation seem to be on par with more complicated models and the predictive power of autoregressive models deteriorates with the length of the estimation period. Finally, we conclude, that the general strategy’s alpha strongly depends on the occurrence of sustained market downturns like the Great Recession and that controlling for business cycles can explain its existence.

Keywords: *Volatility-Managed Portfolios, Volatility Timing, Moreira and Muir, Portfolio Choice, Variance Estimation*

Acknowledgments: *We would like to express our gratitude to our supervisor Michael Halling for his support and insightful comments.*

**Author contact information*

Raphael Bierschenk, 41502@student.hhs.se
Stefan Wennemar, 41508@student.hhs.se

Table of Contents

1. Introduction & Motivation **1**

2. Literature Review **4**

2.1. The Context of Volatility Timing 4

2.2. The Approach of Moreira and Muir 6

3. Theoretical Framework **9**

3.1. Variance Estimation 9

3.1.1. Realized Variance 9

3.1.2. Exponential Smoothing Average 10

3.1.3. Autoregressive Integrated Moving Average 11

3.1.4. Exponentially Weighted Moving Average Model 12

3.1.5. Generalized Autoregressive Conditional Heteroskedasticity Model 13

3.2. Parameter Optimization 13

4. Data and Methodology **16**

4.1. Data Description 16

4.2. Return Calculation and Performance Evaluation 16

4.3. Implementation of Varying Period Lengths 21

5. Empirical Results and Discussion **24**

5.1. Different Methods of Variance Estimation 24

5.1.1. Performance 24

5.1.2. Return and Weight Distribution 27

5.1.3. Transaction Costs and Leverage Constraints 29

5.1.4. Business Cycles 30

5.1.5. Relative Outperformance of Alternative Strategies 32

5.1.6. Subsample Analysis 34

5.1.7. Non-Performing Periods 36

5.2. Different Time Horizons 39

5.2.1. Fixed Periods 39

5.2.2. Flexible Periods 41

6. Conclusion and Future Research **46**

7. References **49**

8. Appendices **53**

List of Figures

Figure 1: Annualized Volatility.. 18

Figure 2: Cumulative Performance 26

Figure 3: Rolling One-Year Return.. 27

Figure 4: Alpha Depending on Holding Period 40

Figure 5: Alpha of Selected Strategies by Average Rebalancing Period 44

Figure 6: ACF and PACF of Time Series of Differenced Monthly Realized Variance........... 53

Figure 7: Example of Months with Irregularly Distributed Variance 54

Figure 8: Example of Sudden Volatility Spikes 54

Figure 9: Drawdown... 56

Figure 10: Density Functions of Returns and Weights 57

Figure 11: Appraisal Ratio Depending on Holding Period 59

Figure 12: Cumulative Performance Depending on Holding Period 60

Figure 13: Average Absolute Change in Weights Depending on Estimation Period 63

Figure 14: Alphas for Selected Strategies Considering Transaction Costs.............................. 63

List of Tables

Table 1: Estimated Parameters for Variance Estimation Models. 17

Table 2: Predictive Regressions on Variance Estimates. 19

Table 3: Performance of Volatility-Based Strategies. 25

Table 4: Summary Statistics of Returns and Weights. 28

Table 5: Alpha Considering Transaction Costs and Leverage Constraints. 30

Table 6: Performance Controlling for Recession Indicator. 32

Table 7: Performance of Alternative Strategies Controlling for Original Strategy. 33

Table 8: Relative Alpha Depending on Transaction Costs. 34

Table 9: Subsample Analysis. 35

Table 10: Analysis of Non-Performing Periods Based on Weights. 37

Table 11: Analysis of Non-Performing Periods Based on Correlations. 38

Table 12: Alphas for Flexible Rebalancing Periods. 43

Table 13: Performance Controlling for Fama French Three Factors. 55

Table 14: Appraisal Ratios for Flexible Rebalancing Periods. 61

Table 15: Root Mean Squared Errors for Flexible Rebalancing Periods. 62

1. Introduction & Motivation

"Often, the wisest to do during periods of extreme market volatility is to stick with the investment plan that you've already devised"

Bill McNabb; CEO Vanguard (Vanguard, 2015)

Among practitioners, it is conventional wisdom that investors in times of high volatility, especially recessions, financial crises, and market crashes, should not reduce their positions in the market, but rather stick to it or even embrace the situation as a buying opportunity. In October 2008 for example, right after Lehman Brothers filed for bankruptcy and the market eventually collapsed, Warren Buffet expressed this view. During a month where the S&P 500 lost almost 17%[1], he recommended America's population to go into equities, instead of holding on to their cash, based on the positive long-term outlook stocks offer (Buffet, 2008).

Moreover, this notion of investing when times are bad is not only common belief in the financial market, but also reflected within academic research. While French et. al. (1987) conclude that there is, in general, a positive relationship between volatility and expected market risk premia, Muir (2017) specifically discovers that "expected returns […] increase substantially in financial crises". Furthermore, Fama and French (1989) generalize this conception to business cycles, stating that expected returns are being "low near peaks […] [and] high near troughs".

In 2017, however, Alan Moreira and Tyler Muir published their paper "Volatility-Managed Portfolios" (Moreira and Muir, 2017), posing a puzzle to this generally accepted idea. Starting from the notion of the mean-variance investor, they devise a simple trading strategy that acknowledges expected returns not to be forecastable but produces large alphas in the market merely by choosing factor weights based on the previous month's realized variance. Concretely, the strategy scales factor excess returns once per month with the inverse of the previous month's realized variance, i.e. increases the risk exposure when variance was recently low and decreases it when variance was recently high. This implicates, that investors can outperform the market, by, for example, decreasing their stake once a crisis materializes through volatile

[1] Source: Thomson Reuters Eikon, accessed 05.12.2019.

price movements. Naturally, the idea, due to its far-reaching and controversial indications, despite its simplicity, received wide-spread attention, even in the public media (see for example CNBC (Rosenberg, 2016)). In their paper, Moreira and Muir (2017) can show how the application of their strategy to the market would have indeed been able to largely avoid the drawdowns of notable times of high volatility, like the Great Depression in the 1930s or the Great Recession in 2007 to 2009. However, as the authors acknowledge, the strategy seems to fare less well in periods without these sustained market downturns. At times, it can even underperform the simple buy-and-hold portfolio.

This becomes apparent when extending the analyzed data to 2019. For example, in October 2018, the market showcased a sudden drop in prices by 7.5%. In the same month, due to indicating a weight of around 3, the implemented volatility-managed portfolio lost 23%. Subsequently, the exposure of the strategy shrank, thereby largely missing out on the following rebound of prices. The main problem was, that the underlying assumption of persistence in variance temporarily broke down through the sudden negative return during otherwise steady times, leading to relatively isolated spikes in volatility. The backward-looking nature of the strategy is not appropriately able to capture these spikes, which poses a problem if they are not the initiation of a sustained crisis, i.e. tendentially a prolonged clustering of high volatility, but rather a unique event. Looking at the whole data range, a situation like October 2018 does not seem to be a singular event. The average return of the volatility-managed portfolio in its worst 12 months equals -19.8% - at the same time, the market only lost 7.5%. While this in itself does not question the overall performance or the underlying reasoning of the strategy, it reveals an additional risk in implementing it that can considerably harm investors.

Thus, the idea of this paper is to start from the simplistic nature of the theory of Moreira and Muir (2017) and empirically analyze whether adjusting the premises of their interpretation of volatility-managed portfolios can generally improve performance and help overcome aforementioned drawbacks of non-performing periods. Moreover, our aim is to study further implications for investors, both desirable and undesirable, and address questions of implementability. Specifically, we approach the alteration from two angles:

I. *Changing the method of variance estimation from realized variance to more sophisticated models.*

II. *Adjusting the rebalancing frequency from monthly to both a continuous number of fixed periods and to a flexible, case-dependent approach.*

Looking at the underlying idea of scaling excess returns by the inverse of last month's realized variance, these levers modify the two most relevant implicit assumptions and we hope to assess a continuum of results, that also sheds light on one's general understanding of volatility timing.

The general rationale behind the first adjustment is, that more sophisticated models might be better able to estimate the conditional variance, thereby improving the performance of the strategy by choosing the optimal market exposure more precisely. The rationale behind the second idea is, that a different, especially shorter, rebalancing period might make the strategy more flexible in adjusting to current trends in the underlying factors, thereby being better able to time volatility and, thus, the risk-return tradeoff.

The remainder of this paper is organized as follows. In chapter two, we review the current literature on volatility timing and the aforementioned paper of Moreira and Muir with a focus on their methodology in greater depth. In chapter three, we present the theoretical framework of the employed models of variance estimation. Subsequently, in chapter four we describe how we methodologically applied the priorly defined theories and implemented varying period lengths. In chapter five, we present and discuss the empirical results, before we conclude our main findings in chapter six.

2. Literature Review

The following chapter first reviews the literature on portfolio allocation based on mean-variance optimization with a focus on volatility timing, before diving deeper into the approach by Moreira and Muir (2017).

2.1. The Context of Volatility Timing

Already in 1952, Harry Markowitz (1952) laid the theoretical groundwork for what we now know as modern portfolio theory (Bodie et al., 2018). It stipulates, that it maximizes the utility of a risk-averse investor to choose a portfolio based on the risk-return tradeoff it offers, described as expected return divided by variance as a proxy for risk. However, while the underlying idea is theoretically appealing, the question of practical implementations becomes how to estimate the needed parameters.

Subsequently, especially the issue of how to predict returns became a widely discussed topic in financial research. Several models were devised, stipulating, that expected returns can be estimated using, among others, dividend yields (Fama and French, 1988, Campbell and Shiller, 1988, Rozeff, 1984), bond spreads (Keim and Stambaugh, 1986), term structures (Campbell, 1987) or book-to-market ratios (Pontiff and Schall, 1998). While this is only a subset of suggested predictors, an inherent problem of estimating them seems to be their persistence that can lead to biased coefficients (Stambaugh, 1999). A further issue is data mining, i.e. finding significant predictors by chance, purely based on considering a large enough number of potential instruments (Ferson et al., 2003). Consistent with this notion, Welch and Goyal (2007) find, that the models found by academia largely perform purely in out-of-sample tests.

These difficulties in forecasting returns question the meaningfulness in applying mean-variance optimization. Michaud (1989) states that mean-variance optimizers are in essence "estimation-error maximizers", and can yield results inferior to equal-weighted portfolios. Therefore, the practical implementation proves difficult (Green and Hollifield, 1992) and indeed, investors rarely employ mean-variance optimization in practice (Fisher and Statman, 1997).

However, the problem primarily originates in the estimation error of forecasting returns. For volatility, on the other hand, there is a broad consensus in empirical research, that it is to some extent predictable (Fleming et al., 2001). This is particularly comprehensively shown based on autoregressive conditional heteroskedasticity models (Bollerslev et al. (1992), Bollerslev et al. (1994), and Andersen and Bollerslev (1998)). Another, more practical example is for example devised by Graham and Harvey (1996), who find that the dispersion in market forecasts among investment newsletters can predict future volatility. Therefore, the question becomes, whether focusing on timing volatility instead of the interplay between excess return and risk can prove to be advantageous.

The research on timing volatility in the market is not as comprehensive as the work on timing returns and the mean-variance trade-off, however, several models have been devised over the past decades. Copeland and Copeland (1999) for example propose two strategies that alternate between investments in value and growth stocks, as well as between large-cap and small-cap stocks, based on changes in the market volatility index (VIX), producing positive excess returns. Underlying is the notion of a lagged correlation between increases in the VIX and outperformance of value and large-cap portfolios and between a decrease in the VIX and outperformance of growth and small-cap portfolios. The underlying idea of their strategy is based on a paper from French et al. (1987), stating that increases in volatility lead to increases in expected volatility, thereby, leading to increases in discount rates, causing stock prices to fall.

Fleming et al. (2001) generalize the idea and find that volatility timing has economic value. Based on the notion that variances and covariances can typically be estimated with far greater precision than the expected returns (Merton, 1980), they show that volatility timing strategies outperform unconditionally mean-variance efficient static portfolios with the same target expected return and volatility.

In 2009, Thomas and Shapiro (2009), then head and vice president of alternative investments at State Street Global Advisors[2], provided insights into volatility-based strategies from a more practical angle. Based on mean-variance optimization, they introduce an "Enhanced Managed Volatility (EMV) strategy". The idea is to maximize returns while controlling for volatility, by taking a stock index (they choose the Russell 3000) and optimizing its asset weights, subject to practical constraints, while keeping the volatility "at least 30% […] [below] the estimated volatility of the cap-weighted [portfolio]." As a result, the EMV strategy outperforms its index, yielding a higher return while having a lower volatility. They conclude that volatility-managed equity strategies "are likely to gain greater acceptance in the marketplace over time" and that they are especially relevant for institutional investors, that are particularly concerned about the risk of their portfolio, such as pension plans, endowments, and insurance companies.

[2] One of the largest asset management companies in the world, measured by assets under management (see https://www.advratings.com/top-asset-management-firms, accessed 05.12.2019).

As opposed to devising a specific strategy, Moreira and Muir (2019) try to answer the general question of whether long-term investors should time volatility and if, by how much they should adjust their portfolio. In order to do so, they create a framework that is able to capture volatility shocks, as well as expected and realized returns, and solve for a utility maximization problem. Their main finding is that "investors should substantially decrease their risk exposure after an increase in volatility and that ignoring variation in volatility leads to large utility losses."

2.2. The Approach of Moreira and Muir

After describing the general context and specific examples of volatility timing, the following paragraphs elaborate on the idea of volatility-managed portfolios by Moreira and Muir (2017). Their approach is novel in the sense that it focuses on the time series of several aggregate price factors, allowing them to draw economic conclusions. All mentions of Moreira and Muir in this subchapter refer to Moreira and Muir (2017).

As mentioned before, the basic idea is to scale factor excess returns by the inverse of their conditional variance. The determination of the induced weights and subsequent rebalancing happens at the end of each month. The monthly excess return of the strategy is therefore calculated as follows:

$$f_{t+1}^{\sigma} = \frac{c}{\hat{\sigma}_t^2(f)} f_{t+1}, \tag{1}$$

where f_{t+1} represents the excess return of the buy-and-hold portfolio and $\hat{\sigma}_t^2(f)$ denotes the estimator for the conditional variance. c is a constant that serves the purpose of controlling the average risk exposure of the strategy. To ensure comparability, it is set to equalize the unconditional variance of the volatility-managed portfolio with the variance of the buy-and-hold portfolio. Since, according to the authors, its magnitude has no effect on the Sharpe Ratio of the strategy, the full sample is being used to estimate it. Moreira and Muir motivate equation (1) starting from the portfolio optimization problem of a mean-variance investor, who can invest in a risky and a riskless asset. For such an investor, the optimal weight on the risky asset is proportional to the attractiveness of its mean-variance tradeoff: $\frac{E_t(f_{t+1})}{\hat{\sigma}_t^2(f)}$. The weight determination in equation (1) builds upon this, even though it refrains from timing expected returns, due to the incapacitation of reliably forecasting their movement. It rather focuses on timing volatility, substantiated by empirical evidence of its variability and persistency. In other words, since an investor cannot accurately foresee the returns of the next period, but volatility seems to be similar from one period to the next, a high volatility indicates an unfavorable conditional risk-return tradeoff, inducing mean-variance investors to lower their exposure, and vice versa.

To approximate the conditional variance, the authors use the previous month's realized variance. Under the assumption of 22 trading days per month, the formula is as follows:

$$\hat{\sigma}_t^2(f) = RV_t^2(f) = \sum_{d=1/22}^{1} \left(f_{t+d} - \frac{\sum_{d=1/22}^{1} f_{t+d}}{22} \right)^2. \tag{2}$$

Altogether, this enables them to form a volatility-managed portfolio based on any underlying factor of their choice. To then assess the performance of each application of the strategy, they run a time-series regression of the volatility-managed portfolio excess return on the buy-and-hold factor excess return:

$$f_t^{\sigma} = \alpha + \beta f_t + \epsilon_t \tag{3}$$

A positive intercept indicates that the volatility timing increases the Sharpe Ratio of the portfolio compared to the simple buy-and-hold case. When applied to systematic factors (e.g. the market), the alpha indicates that the volatility timing expands the mean-variance frontier.

In total, Moreira and Muir implement the strategy for nine factors: market, size, value, momentum, profitability, return on equity, investment, betting-against-beta, and a currency carry trade. Additionally, they form multifactor portfolios, using various combinations of these base factors. For each factor, they use return data from 1926 to 2015, or, if in between, from inception to termination. In almost all cases, volatility timing does produce a large, positive, significant alpha, even when controlling for the Fama-French three factors (Fama and French, 1996). Hence, the Sharpe Ratio increases, leading to considerable percentage utility gains for the mean-variance investor, calculated as a percentage increase in the squared Sharpe Ratio.

Subsequently, the authors further investigate their strategy from different angles to better assess applicability, as well as the source of its alpha. First, they analyze the effect of transaction costs of up to 14bps and leverage constraints down to 0% leverage and conclude that both do drive down the annualized alpha of the volatility-managed portfolio but cannot eliminate it. Furthermore, they show individual investors can alternatively embed leverage using option portfolios on the market to overcome potential leverage constraints. Together, this strengthens the argument, that the idea of volatility timing can be implemented in real-time. Next, Moreira and Muir show, that the outperformance of their idea does not stem from taking on business cycle risk since its market beta decreases during recessions. Moreover, it cannot be explained by strategies that exploit weak risk-return tradeoffs in the cross-section of stocks, namely risk parity and betting-against-beta. Finally, they prove, that the strategy can also be employed for most factors based on a common volatility factor[3] and that increasing the rebalancing period, i.e. decreasing the frequency, does gradually lower alphas, but does not immediately eliminate them.

Lastly, Moreira and Muir try to conceptualize a theoretical framework to explain the alpha of the strategy. Additionally, they contrast its high Sharpe Ratio with leading macro-finance models. Since detailed discussions of both theories go beyond the scope of this paper, the proceeding paragraphs focus on the respective main ideas.

First, the theoretical framework conceptualizes the relation between alpha and the price of risk μ_t/σ_t^2 in the following way:

$$\alpha = -cov\left(\frac{\mu_t}{\sigma_t^2}, \sigma_t^2\right)\frac{c}{E[\sigma_t^2]} \tag{4}$$

The stronger the conditional variance and the mean-variance tradeoff are negatively correlated, the higher alpha is. This intuitively makes sense, since the goal of timing volatility by the inverse of the realized variance is to increase the risk exposure in times of favorable risk-return tradeoffs and vice versa.

Second, the central outcome of the micro-financial analysis is that four of the leading equilibrium asset pricing models, namely the habits model, the long-run risk model, the time-varying rare disasters model, and the intermediary-based asset pricing model, are not able to reproduce the alphas and Appraisal Ratios of volatility-managed portfolios, thus being challenged in their explanatory power. In turn, the authors deliberate alternative explanations to their findings, however, without exploring them empirically, or settling down on a final answer. The idea they assess to be theoretically feasible and consistent with their findings is based on the notion of slow trading behavior by investors, leading to lagged increases in expected returns after volatility shocks. This would explain an unfavorable risk-return tradeoff immediately after a volatility spike, when the strategy decreases its risk exposure, since the expected return in the market did not adjust yet, and a more favorable tradeoff, once the strategy indicates an increase of the risk exposure when the volatility shock ceases and the expected return is still elevated.

[3] The common volatility factor is estimated using the first principal component of realized variance across all factors. This means, each factor is normalized by the same variable, instead of the respective past realized variance.

3. Theoretical Framework

As described in chapter 1, our goal is to analyze the impact of adjusting the two main presumptions of the implementation of volatility-managed portfolios by Moreira and Muir (2017): the method of estimating the conditional variance and the rebalancing period. The following chapter focuses on the statistical rationale and background of the employed models of variance determination. Since we also vary the rebalancing periods, the descriptions refer terminologically to periods as unit of time. Furthermore, the chapter explains the chosen method of optimizing the necessary input parameters. All mentions of Moreira and Muir in this chapter refer to Moreira and Muir (2017).

3.1. Variance Estimation

Due to the importance of estimating variance, and thereby volatility, in financial theory, a multitude of models with varying degrees of complexity and sophistication have been devised to explain and forecast its movement over time. However, no clear consensus has been formed yet as to what the general best model is. Poon and Granger (2003) have conducted a meta-study of almost 100 individual studies, concluding that "as a rule of thumb, historical volatility methods work equally well compared with more sophisticated ARCH class and SV [stochastic volatility model forecasts] models". Since we try to analyze whether adjusting the method can improve the estimation of the conditional variance, our aim in model selection is to reflect a diversified range of approaches utilized in academia. For similar models, we decide among them based on their degree of generalization and their applicability and feasibility in the context of volatility timing. To ensure comparability over the whole data set, our only restriction is, that the data needed to calculate the volatility estimate is available from 1926 on, ruling out possibilities like option-implied volatility.

3.1.1. Realized Variance

To replicate the strategy by Moreira and Muir and as a basis for the following methods, we first calculate the realized variance per period. For that, we use the formula for population variance, since the complete set of daily return data per time period is available:

$$\hat{\sigma}_t^2(f) = \sum_{d=1}^{n} \left(f_{t+d} - \frac{\sum_{d=1}^{n} f_{t+d}}{n} \right)^2 . \tag{5}$$

f denotes the factor excess return and n represents the number of days per period. The formula is, in essence, a generalization of equation (2), the formula used by Moreira and Muir to calculate the realized variance per month. Similarly, equation (5) calculates the realized variance on a per period level.

3.1.2. Exponential Smoothing Average

Moreira and Muir (2017) assume one can approximate the conditional variance with the realized variance of last period, based on its persistence and variability. However, the drawback is, that this model disregards any long-term trends and assumes investors' expectations adjust timely after changes in realized volatility. This makes the chosen weights highly susceptible to outliers and volatility spikes, which can make sense if a volatility jump precedes a sustained market downturn but can hurt performance if it is solely a singular market correction. Taking a higher number of past values of variance into account can mitigate this problem, only adjusting the weight considerably if there is a continuous trend in volatility, at the expense of reacting slower at the outset actual crises.

The simplest option to incorporate a larger number of periods is to take the historical simple average. This, however, would diminish the influence of the latest period with an increasing number of periods. Hence, it would marginalize the incremental changes in weights and the portfolio would converge to the buy-and-hold portfolio, instead of capturing time-dependent trends. Another option is the moving average, having the advantage that only a specified number of past periods is being taken into account. Setting aside the problem of determining the number of periods, this, however, would go against the employed idea of persistence in variance, stipulating that shocks affect the future expectations of variance (Ding and Granger, 1996), which in practice seems to occur at a decaying rate (Poterba and Summers, 1984). Therefore, it seems most sensible to utilize an exponential smoothing average (ESA) that assigns geometrically declining weights to each observation. Thereby, each observation of realized variance has a diminishing impact on the current variance estimate and only a limited number of periods is meaningfully taken into account. In the following, we use the formula for the exponential smoothing average expressed by Brooks (2019):

$$S_t = \alpha y_t + (1-\alpha) S_{t-1}, \tag{6}$$

where α denotes the smoothing constant, y_t the current realized value, in our case the realized variance of last period, and S_t the current, smoothed value. When implemented, only one parameter, α, has to be estimated, increasing the practical applicability of the model.

3.1.3. Autoregressive Integrated Moving Average

A more general approach to model the time series of realized variance is the autoregressive integrated moving average (ARIMA) model. It combines the autoregressive (AR) with the moving average (MA) model while allowing for a differencing of the data to achieve more desirable properties. Thereby, the ARIMA model enables the estimate of a variable or a differenced variable to be dependent on its own previous values and the current and previous values of a white noise disturbance term, both with flexible order and flexible weights. By that, the model is more tailored to the data than the exponential smoothing model, however, it also partly loses the advantage of simplicity by requiring the estimation of more parameters.

Following, the individual components of the ARIMA are described in successive order, based on the notation of Brooks (2019). The first component is the MA model. It originates in the idea that a variable depends on the current and previous values of a white noise disturbance term, representing the forecasting error of previous periods. White noise describes a process that has a constant mean and variance and zero autocovariance, meaning that each value is uncorrelated with all other values in the series. Therefore, the MA model, with order q, can be expressed as:

$$y_t = \mu + \sum_{i=1}^{q} \theta_i u_{t-i} + u_t, \tag{7}$$

where u_t denotes the white noise process, i.e. the forecast error. θ_i represents the weights on the lagged white noise variables and has to be estimated as an input parameter.

The AR model, on the other hand, rests on the idea that a variable depends on its own previous values plus an error term. Hence, the AR model, with order p, can be represented as follows:

$$y_t = \mu + \sum_{i=1}^{p} \phi_i y_{t-i} + u_t. \tag{8}$$

ϕ_i denotes the weights on the previous values and has to be estimated as an input parameter.

Both models taken together form the ARMA model, with order q and p:

$$y_t = \mu + \sum_{i=1}^{q} \theta_i u_{t-i} + \sum_{i=1}^{p} \phi_i y_{t-i} + u_t. \tag{9}$$

A necessary prerequisite of the ARMA model is that the data used is weakly stationary, i.e. has a constant mean, constant variance, and constant covariance structure. In case the data is not stationary, it is usually being differenced. The ARIMA model is a generalization of the ARMA model that allows for differencing within the model and, therefore, introduces the additional order d. An ARIMA(p, d, q) model is equal to an ARMA(p, q) model, where the data is differenced d times.

3.1.4. Exponentially Weighted Moving Average Model

As opposed to the previous models, the exponentially weighted moving average (EWMA) model is not a general univariate time series model, applied to the series of realized variances[4], but rather specifically designed to model volatility. Therefore, it directly takes the underlying random variable. i.e. returns, as input, under the assumption of a zero-mean, but allowing for heteroskedasticity, that is time-varying volatility. Under these assumptions, the conditional variance of the random variable becomes (Brooks, 2019):

$$\sigma_t^2 = var(u_t|u_{t-1}, u_{t-2}, \dots) = E[u_t^2|u_{t-1}, u_{t-2}, \dots]. \tag{10}$$

To approximate the desired distribution, log returns are usually being used as input. Taking the logarithm of returns brings the mean approximately to zero.

The EWMA model estimates the conditional variance with an approach similar to the ESA in equation (6), where the estimated conditional variance depends on the past realizations, weighted with a geometrically declining weight (Hull, 2018):

$$\hat{\sigma}_n^2 = \lambda\hat{\sigma}_{n-1}^2 + (1-\lambda)u_{n-1}^2, \tag{11}$$

where λ stands for the constant weighting parameter[5]. As for the ESA, the advantage of the EWMA model lies in its simplicity and straightforward application.

---

[4] The idea of Moreira and Muir can essentially be seen as the implementation of a random walk model, where the value of each period equals the value of last period plus an i.i.d. random variable: $y_t = y_{t-1} + e_t$ (Wooldridge, 2018). Hence, the expected value of the conditional variance equals the realized variance of last period.

[5] Due to the way the models are defined, α in the ESA and λ in the EWMA model are reversed. In other words, in the ESA, α scales the current realized value and $(1-\alpha)$ the estimate of last period, whereas, in the EWMA model, $(1-\lambda)$ scales the current realization and $(1-\lambda)$ the estimate of last period.

3.1.5. Generalized Autoregressive Conditional Heteroskedasticity Model

The Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model is a generalization of the EWMA model, but similar in its prerequisites. The equation is analogous to (9), but additionally considers a long-run average variance rate V_L to reflect that the variance has the tendency to get pulled back to its long-run average over time (Hull, 2018):

$$\sigma_n^2 = \gamma V_L + \alpha u_{n-1}^2 + \beta \sigma_{n-1}^2, \tag{12}$$

where γ, α, and β denote the three different weights on V_L, u_{n-1}^2, and σ_{n-1}^2, whereby all have to sum up to unity. Since return data for the whole time period is available, V_L can be calculated by taking the average over all observations of u_i^2. For a stable process, α and β must be below one, otherwise, the weight applied to the long-term variance would be negative (Hull, 2018).

Strictly speaking, the depicted equation showcases a GARCH(1,1) model, since only the values of the previous period are taken into account. Higher-order models are theoretically possible, but in general not necessary and “rarely […] even entertained in the academic finance literature” (Brooks, 2019). Therefore, we focus on the GARCH(1,1) case.

3.2. Parameter Optimization

To employ the discussed models of variance estimation, several input parameters are needed. The following chapter describes the theoretical background behind their optimization. For all methods, parameters are determined in-sample in order to be able to comprehensively assess the advantages and disadvantages of each methodology.

Exponential Smoothing Average

For the exponential smoothing average, we approximate the smoothing constant α by the parameter $(1 - \lambda)$ from the EWMA model, since both models are based on the same underlying idea of scaling past values with an exponentially declining weight. While we are aware, that this is only an approximation, it enables us to find the optimal parameter flexibly dependent on the chosen period length.

ARIMA

For the ARIMA model, it is necessary to specify the three order parameters: p, d, and q. p denotes the order of the AR model, q the order of the MA model, and d the order of differencing.

First, d is being determined, since it ensures, that the data used is stationary, a prerequisite for the ARMA model. To find the order of differencing, we test the original time series of realized variance for the null hypothesis of stationarity against the alternative hypothesis of unit root, using the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test (Kwiatkowski et al., 1992). By defining stationarity as the null hypothesis, the KPSS test overcomes the shortcoming of other tests that define unit root as the null hypothesis, like the Augmented Dickey-Fuller (ADF) tests (Dickey and Fuller, 1981) or the Phillips-Perron (PP) tests (Phillips and Perron, 1988). They fail to reject the null hypothesis for many economic series. However, not rejecting does not equal accepting the null hypothesis, but can originate in insufficient information in the data (Brooks, 2019). Therefore, we employ the KPSS test, defining stationarity as the null hypothesis, to only difference our data if the test rejects the null hypothesis, instead of differencing it, if a test fails to reject it.

Subsequently, the order p and q of the AR and the MA model have to be specified. The idea is to fit an ARIMA model to different values for both orders and choose the one which features the lowest information criterion. The idea of information criteria is to select the model, which minimizes the residual sum of squares while accounting for a penalty term, penalizing the number of added terms to find a parsimonious model. The two most common information criteria are Akaike's (1974) Information Criterion (AIC) and Schwarz's (1978) Bayesian Information Criterion (BIC). While no criterion seems to be definitely superior to the other (Brooks, 2019), AIC appears to be better suited for the context of predictions (Kuha, 2004). Therefore, we employ the AIC to find the optimal ordering of the ARIMA model. The AIC is defined as follows (Brooks, 2019):

$$AIC = ln(\hat{\sigma}^2) + \frac{2k}{T}, \tag{13}$$

where $\hat{\sigma}^2$ stands for the residual variance, k equals the number of parameters estimated $(p + q + 1)$, and T denotes the sample size. Thereby, an additional term is only being added, if the reduction in the residual sum of squares outweighs the increase in the penalty term.

To estimate the specific parameters of the ARIMA model, the maximum likelihood (ML) method is being used. The idea of ML is to choose the parameter in a way, that, when the model is applied to the given data, it maximizes the likelihood of the actual observations appearing. In practice, for the ARIMA model, this is done using statistical software.

EWMA and GARCH

To estimate the parameters in the EWMA and GARCH model based on the available data, we use the maximum likelihood method. The following description is based on Hull (2018).

At the outset, as before, it is assumed, that the observations u_i, conditional on their variance, are normally distributed with a mean of zero. Furthermore, v_i depicts the estimated variance for period i, dependent on the chosen respective parameters. The likelihood of observing u_i, therefore, is defined as the probability density function for X when $X = u_i$. Accumulated over all m periods, the likelihood of all observations occurring in the order in which they are equals:

$$\prod_{i=1}^{m}\left[\frac{1}{\sqrt{2\pi v_i}}exp\left(\frac{-u_i^2}{2v_i}\right)\right]. \tag{14}$$

The best parameters are the ones that maximize the equation, dependent on v_i. Taking logarithms and ignoring constant multiplicative factors, the maximization problem becomes:

$$\sum_{i=1}^{m}\left[-\ln(v_i) - \frac{u_i^2}{v_i}\right]. \tag{15}$$

Using this equation and an iterative search, the optimal parameters for both, the EWMA model and the GARCH model, can be found. Since EWMA requires only one parameter, an iterative process, testing incremental values between the boundaries of zero and one can be employed. For GARCH, since two parameters have to be estimated, we use the Nelder-Mead method (Nelder and Mead, 1965), a search method for multidimensional unconstrained optimizations.

4. Data and Methodology

4.1. Data Description

In the original paper, Moreira and Muir (2017) apply their envisioned strategy to nine different single-factor portfolios and several multifactor portfolios. Due to the applicability for investors, the relevance in financial research, and to be able to focus on the analysis and comparison between the varying methods of implementation, we concentrate on the market factor. The data of monthly and daily returns stems from Kenneth French's website (French, 2019). The market there is a value-weighted index and constitutes all CRSP firms incorporated in the US and listed on the New York Stock Exchange (NYSE), the American Stock Exchange (AMEX) and the Nasdaq Stock Market (NASDAQ). To control for Fama French three factors, we also employ French's return data on the size (SMB) and the value (HML) factor. We use data from July 1926 to August 2019. The original paper is based on data from July 1926 to April 2015.

Additionally, to analyze the behavior of the strategies during business cycles, we use the NBER based Recession Indicator (FRED, 2019), a binary monthly variable, indicating recessionary periods with a one and expansionary periods with a zero. The indicator stems from the Federal Reserve Bank of St. Louis and is based on data from the National Bureau of Economic Research. All mentions of Moreira and Muir in this chapter refer to Moreira and Muir (2017).

4.2. Return Calculation and Performance Evaluation

Variance

Based on the theoretical framework presented in chapter 3, we optimize the relevant parameters and calculate the variance estimates. For fixed periods, including monthly, we omit the use of the EWMA model, since the GARCH(1,1) model is similar and the more general one, incorporating mean reversion, thereby being "theoretically more appealing" (Hull, 2018). We only use the EWMA model to approximate the smoothing parameter for the implementation of the exponential smoothing average. For flexible rebalancing periods, which are based on daily variance estimates, we employ the EWMA model to have a broader range of methods to contrast, since the period length of one day permits the, in our context, meaningful application of ESA and ARIMA. The reason is, that the calculation of a time series of daily realized variance based on equation (5) would only be possible if intraday data was available[6] for the whole time period.

[6] To calculate some measure of realized variance, at least two observations are needed. Therefore, having the cumulative return over one day is not enough to meaningfully calculate the realized variance for that day.

Table 1: Estimated Parameters for Variance Estimation Models.

In this table, we report the estimated parameters for the ESA, ARIMA and GARCH model using the maximum likelihood optimization described in chapter 3.2. To ensure comparability among parameters, we estimate $1-\alpha$ instead of α for the ESA model because $1-\alpha$ corresponds methodologically to λ in the EWMA model.

Variance Model	Estimated Parameters
Exponential Smoothing Average	• $(\mathbf{1}-\boldsymbol{\alpha})$: 0.8930
Autoregressive Integrated Moving Average	• **Order**: (1,1,2) • $\boldsymbol{\phi_1}$: 0.2307 • $\boldsymbol{\theta_1}$: -0.6616 • $\boldsymbol{\theta_2}$: -0.2119
Generalized Autoregressive Conditional Heteroske-dasticity	• $\boldsymbol{\gamma}$: $7.289e^{-5}$ • $\boldsymbol{\alpha}$: 0.1185 • $\boldsymbol{\beta}$: 0.8588

For the exemplary period length of one month, the one Moreira and Muir used in their original paper, the estimated parameters are reported in Table 1 above. For the ARIMA model, the KPSS test rejects stationarity at the 5% significance level with a p-value of 0.012. For the differenced time series of realized variance, the test fails to reject the null hypothesis with a p-value of greater than 0.1, suggesting a differencing of order one. The AIC suggests an ARIMA(1,1,2) model, examining orders for p and q of up to ten. For further reference, auto-correlation function and partial autocorrelation function of the time series of differenced realized volatility can be found in Appendix A. The parameters for all other period lengths are optimized the same way.

Using the respective parameters, we can calculate the variance estimates for each model. In the very first period (July 1926), where no variance estimate is available to initiate the series, we approximate it with the realized value. In other words, for example for the exponential smoothing average, we use y_1 for S_1, in order to be able to calculate S_2.

All variance estimates for the monthly level are plotted in Figure 1 below. At first glance, the course of the models based on realized variance looks similar, without extreme outliers. As expected, last month's realized variance is the most volatile measure, since it only takes one period into account. On the other hand, ESA has the most stable measure, since it, in theory, takes all previous periods into account. A consequence of this approach is that outliers, i.e. volatility spikes, have a lasting influence on the estimated values. Therefore, the ESA estimates tend to be higher than the estimates only based on last month's variance after times of high volatility. The ARIMA measure is somewhat in the middle, following last month's variance more closely, but also often overstating it after volatility spikes. Lastly, the GARCH estimates are less closely related to the other measures. This is a natural consequence of the different method of estimation, based on asset returns instead of realized variance. It also takes all previous periods into account, comparable to ESA, thereby being relatively stable.

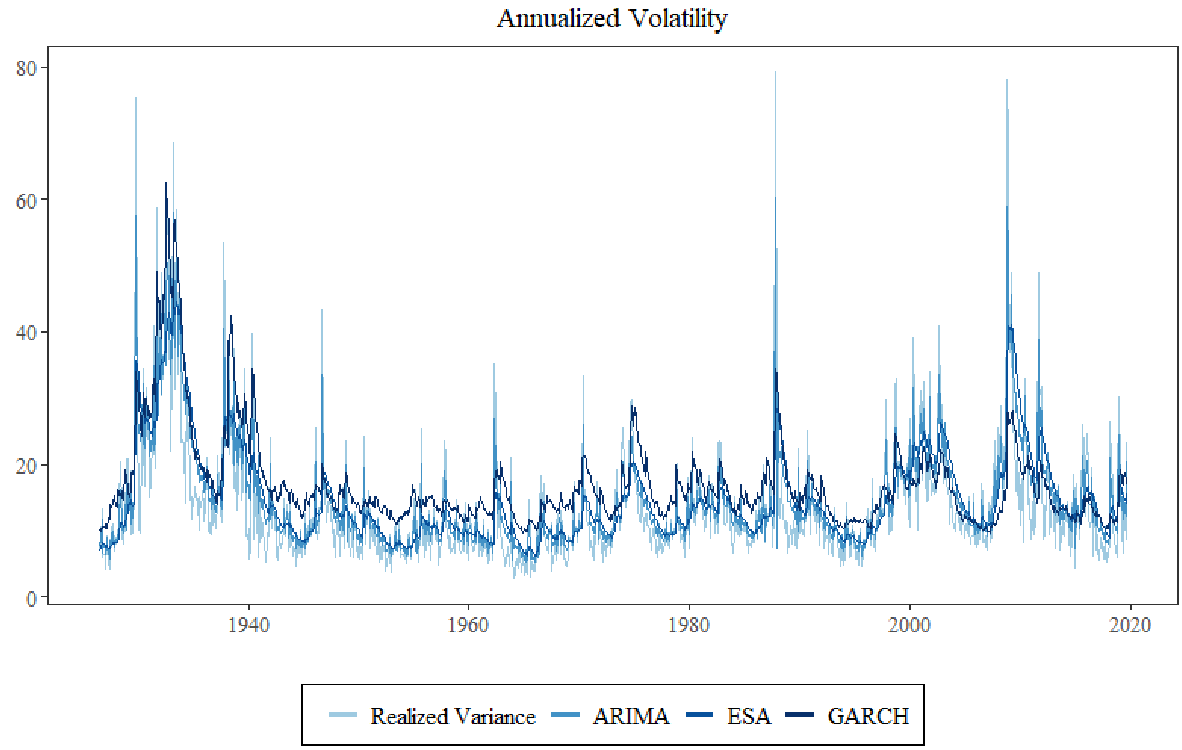


Figure 1: Annualized Volatility

The plot depicts the annualized volatility (based on the assumption of 264 trading days) for all four measures of variance determination over the whole time period. Since GARCH is based on returns expressed in percent, whereas realized variance is calculated using absolute numbers, as in the paper by Moreira and Muir, the GARCH estimates are multiplied by 10,000 to make them comparable in magnitude.

To assess how well suited the different models are to time favorable risk-return tradeoffs, we regress the realized variance, the realized return, and the risk-return tradeoff of next month on the model variance estimates. The results are shown in Table 2. As can be seen, every variance estimator is a statistically highly significant predictor for next month's realized variance. The prediction seems to be the strongest for realized variance and ARIMA and weakest for GARCH. With regards to returns, as already noted by Moreira and Muir, no estimator seems to be able to have significant predictive power on the 5% level. For the return per unit of variance, the variance estimators all have significant predictive power, even though the R^2 is very low. This means, for all four models, that a low variance estimate indicates a favorable risk-return tradeoff next period and a high variance estimate the opposite. Therefore, based on this preliminary analysis, all four methods seem to be sensible to use in volatility-managed portfolios.

Table 2: Predictive Regressions on Variance Estimates.

In this table we regress the next month's realized variance [(1) to (4)], the next month's return [(5) to (8)] and the next month's return per unit of variance [(9) to (12)] on the variance estimates determined by the realized variance, ARIMA, ESA, and GARCH models on a monthly level. The sample period ranges from July 1926 to August 2019. Standard errors are in parentheses and adjust for heteroskedasticity. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

	Variance Next Month				Return Next Month				Return per Unit of Variance			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Realized Variance	0.608***				-0.002				-0.001***			
	(0.088)				(0.009)				(0.000)			
ARIMA		0.897***				-0.001				-0.002***		
		(0.104)				(0.012)				(0.000)		
ESA			0.880***				0.005				-0.003***	
			(0.087)				(0.013)				(0.000)	
GARCH				0.718***				0.008				-0.002***
				(0.065)				(0.012)				(0.000)
Constant	9.098***	2.458	2.900*	2.600	0.700***	0.679***	0.543**	0.431	0.247***	0.271***	0.287***	0.273***
	(1.709)	(1.977)	(1.612)	(1.694)	(0.195)	(0.254)	(0.250)	(0.289)	(0.021)	(0.024)	(0.026)	(0.024)
Observations	1,117	1,117	1,117	1,117	1,117	1,117	1,117	1,117	1,117	1,117	1,117	1,117
R^2	0.370	0.397	0.261	0.280	0.000	0.000	0.001	0.003	0.009	0.015	0.018	0.012

Weights and Return

Based on calculating the different variance estimates, we determine the chosen weights of the different strategy implementations and subsequently the implied returns. For the weights, consistent with equation (1), the variance estimate of each period represents the denominator and only the respective constant c hast to be determined to find the full scaling vector. As described in subchapter 2.2, c purely serves the purpose of equaling the standard deviation of each strategy to ensure comparability and has no influence on the Sharpe Ratio. This implies, that by equating the variance of each strategy with the variance of the market, we can back out the constant:

$$Var(f_{t+1} + r_f) = Var\left(\frac{c}{\hat{\sigma}_t^2(f)} f_{t+1} + r_f\right). \tag{16}$$

Since every parameter except c is known, the equation can be rewritten as follows:

$$c^2 * Var\left(\frac{f_{t+1}}{\hat{\sigma}_t^2(f)}\right) + c * 2 * Cov\left(\frac{f_{t+1}}{\hat{\sigma}_t^2(f)}, r_f\right) + Var(r_f) - Var(f_{t+1} + r_f) = 0. \tag{17}$$

Solving the quadratic equation yields:

$$c_{1,2} = \frac{-Cov\left(\frac{f_{t+1}}{\hat{\sigma}_t^2(f)}, r_f\right) \pm \sqrt{\left(Cov\left(\frac{f_{t+1}}{\hat{\sigma}_t^2(f)}, r_f\right)\right)^2 - Var\left(\frac{f_{t+1}}{\hat{\sigma}_t^2(f)}\right) * \left(Var(r_f) - Var(f_{t+1} + r_f)\right)}}{Var\left(\frac{f_{t+1}}{\hat{\sigma}_t^2(f)}\right)}. \tag{18}$$

Since the risk-free rate and the excess return per unit of realized variance are almost uncorrelated, the first term of the nominator is expected to be close to zero and smaller than the second term of the nominator[7]. Therefore, we always employ solution c_1 to ensure non-negativity in weights and for simplicity further on refer to it as c.

Having the excess market return, the variance estimates and the constant c enables us to determine the returns of each strategy per period.

[7] For the original implementation of the strategy on monthly basis, the first term equals -0.02 and the second one -28.26.

Performance

Using these returns, we can assess the performance of each of the implementations of the idea of Moreira and Muir. For that, we run the same time series regression of volatility-managed excess returns on factor excess returns, as described in the original paper, depicted in equation (3). This gives us the beta of the strategy in relation to the market, i.e. the systematic risk the respective strategy takes on, and the alpha, that is the outperformance. To be able to better compare the alpha, we contrast it with the root mean squared error (RMSE) of the regression, which equals the square root of the average of the squared errors. In our context, it represents the idiosyncratic risk the strategy, additionally to its systematic risk, takes on. Dividing alpha by the RMSE gives the Appraisal Ratio: $\frac{\alpha}{\sigma(\varepsilon)}$ (Treynor and Black, 1973). It describes the risk-adjusted alpha and can help assess the added value compared to a benchmark (Goodwin, 1998). To evaluate the risk-adjusted overall performance, we use the Sharpe Ratio defined as the excess return divided by the standard deviation of the portfolio return $\left(SR = \frac{E[r_p - r_f]}{\sigma_p}\right)$. To shed further light on the performance of the implementations of the strategy, we repeat the time series regression in equation (3), controlling for the Fama French three factors (Fama and French, 1996). All performance measures are ultimately being annualized to ensure comparability and ease of interpretation. Annualization is done under the previously mentioned assumption of 22 trading days per month and 12 months per year.[8]

4.3. Implementation of Varying Period Lengths

Besides adjusting the method of variance estimation, we also analyze changing the defined period length, both on a fixed and flexible basis. As discussed, Moreira and Muir primarily consider a monthly rebalancing, meaning that once per month the realized variance is calculated, the new weight determined, and the necessary trading conducted. Additionally, they test the possibility of longer holding periods[9], but conclude, that the alphas “gradually decline in magnitude” (Moreira and Muir, 2017). However, they do not examine a rebalancing more frequent than monthly, or a flexible rebalancing, even though this could offer the advantage of better adjustments to volatility trends and thereby more accurate estimates of the conditional

[8] We are aware, using 264 trading days per year differs from commonly employed assumptions (e.g. 252). However, it ensures consistency with the original paper of Moreira and Muir.
[9] The terms “holding period” and “rebalancing period” are used interchangeably, since the time of holding the portfolio equals the time between two rebalancing points.

variance. On the other hand, more frequent rebalancing most likely increases trading costs, indicating, that the magnitude of both effects has to be weighed against each other.

The potential drawback of monthly rebalancing is, that the differences in daily returns can be irregularly distributed and clustered over the course of a month. In other words, it is possible, that returns are highly volatile at one point, but otherwise rather steady. If, for example, the returns are volatile in the first week of the month, the overall realized monthly variance probably overestimates the actual conditional variance at the rebalancing point and decreases the risk exposure, even though the last three weeks have been rather steady. On the other hand, if the returns are volatile in the last week, the realized monthly variance probably underestimates the conditional variance. Two actual examples of such patterns are showcased in Appendix B.

Fixed Rebalancing Periods

To analyze the impact of changing the rebalancing frequency, we first assess the effect of using fixed rebalancing periods on a continuum between five to 50 trading days (equaling one and ten calendar weeks, not accounting for holidays), with a focus on the influence of the adjustments on performance in terms of alpha and Appraisal Ratio. Since the specific number of days per month varies, the fixed rebalancing period of 22 days only approximates the period length of one month used by Moreira and Muir. However, the alterations are helpful to visualize both the granular effect of changing the rebalancing frequency on the performance of the strategy and the varying impact of trading costs. We conduct the analysis for all four previously used implementations of volatility timing.

Flexible Rebalancing Periods

The main drawback of every implementation of a fixed rebalancing period is, that it induces changes in weights and therefore trading costs at every rebalancing point, no matter the magnitude or reasonableness. This increases the implementation costs of higher frequencies through more active trading. Furthermore, shorter fixed periods can primarily mitigate the aforementioned problem of irregularly distributed variance. We try to overcome these problems by setting up a decision rule, that re-evaluates at every decision point, whether adjusting the weight, thereby causing trading costs, is appropriate or not. An example to visualize the potential advantage is provided in Appendix B, depicting a prolonged period of several months, showcasing almost no movement in trailing 22-day variance, before it suddenly spikes. In such periods, it does not seem to be necessary to continuously change the weight in incremental steps, but rather seems better to time the actual jump in variance. In other words, the idea is to utilize the advantage of frequent decision points, being able to flexibly respond to current trends in volatility, while avoiding disproportionate trading, leading to higher costs, in order to still ensure applicability for the individual investor in real-time.

We implement the idea of flexible rebalancing for the case of daily decision points. To estimate the conditional variance, as discussed in subchapter 4.2, we employ the EWMA and GARCH model and, inspired by the original approach by Moreira and Muir, the realized variance over rolling time windows. Intervals of varying lengths are being considered: one week, two weeks, one month, two months, three months, six months, one year, and two years (subsumed under "Var model"). We start implementing the methods at the end of the fifth trading day to have a minimum of one week of return data available. For the realized variance based on rolling windows longer than five days, we first expand the window of data used until the desired length is reached, thereafter dropping the last observation to move the rolling window one day forward. For instance, in the first two years, the realized variance over two years is being calculated using all data available, and after two years, the rolling time window is being employed. The EWMA and GARCH model, on the other hand, always take all values into account.

Next, we calculate the relative change of the variance estimate for each model and interval from the last trading day to the current day. As a decision rule, this percentage change is being compared to a prespecified interval, indicating a weight adjustment if the value lies outside the interval. Consequently, the definition of the interval affects the total number of reallocations, leading to a tradeoff between more refined adjustments to volatility changes and higher trading costs. To elicit any trends in the interplay between the components, we employ varying thresholds, corresponding to average rebalancing periods of one month, two weeks, one week, two days, and the extreme case of daily rebalancing. The thresholds are determined based on the quantiles of the respective distribution of daily percentage changes in estimated variance. The previously mentioned average rebalancing periods correspond to the following two-sided intervals: 95.5%, 90%, 80%, 50%, and 0%[10]. For example, a 90% interval means, that a reallocation is indicated if the percentage deviation lies within the bottom or top 5% of the total distribution. Determining the rebalancing points enables us to subsequently calculate weights and returns per period.

[10] Exemplary, for average rebalancing every month (22 trading days) the interval equals: 1 - 1/22 ~ 95.5%

5. Empirical Results and Discussion

In the following chapter, we present and discuss the empirical results based on the previously presented theoretical framework and methodology. Subsection 5.1. focuses on the implementation of the strategy from Moreira and Muir using four different methods of variance estimation and assesses their respective performance and real-world applicability from different angles. Furthermore, it describes, whether adjusting the estimation method can help overcome or mitigate the at the beginning introduced risk of strong, temporary market underperformance. Lastly, subsection 5.2. describes and discusses our implementation of both, varying fixed and flexible period lengths. All mentions of Moreira and Muir in this chapter refer to Moreira and Muir (2017).

5.1. Different Methods of Variance Estimation

Our first step, as previously discussed, is to replicate the approach of Moreira and Muir for the market factor of the Kenneth French library, both for the original scaling factor realized variance and the three additional methods of variance estimation: ESA, ARIMA, and GARCH.

5.1.1. Performance

All four implementations of the volatility-managed portfolio yield results substantially better than the simple buy-and-hold portfolio. As can be seen from Table 3, the alpha is always above 3% and, apart from GARCH, even above 4%. When controlling for all three Fama French factors, the alpha is even higher. This is due to the negative factor loadings on the additional factors HML and SMB. In other words, the outperformance does not originate in size or value premia, but rather has a negative exposure to these risk factors, leading to lower expected returns and, thereby, to a higher alpha. The outputs of the regressions controlling for the Fama French three factors can be found in Appendix C.

Among the implementations, the alpha differs. It is the highest for the original strategy based on the realized variance and for the strategy based on the ARIMA model (4.7%). The ESA strategy shows a slightly lower alpha with 4.3% and the GARCH strategy the lowest alpha with 3.3%. However, to fully comprehend the performance of each strategy, we have to account for the root mean squared error to consider the idiosyncratic risk each strategy takes on. Here, the original strategy has the highest risk, leading to an Appraisal Ratio of 0.32. ARIMA performs the best with an Appraisal Ratio of 0.35, followed by ESA with 0.33. Last comes GARCH with 0.30. Looking at the overall performance in terms of the Sharpe Ratio, the original strategy actually performs worst due to its low beta of 0.6, leading to an overall lower

Table 3: Performance of Volatility-Based Strategies.

In this table we run time-series regressions of volatility-managed returns on the market factor: $f_t^\sigma = \alpha + \beta f_t + \epsilon_t$. The managed returns, f^σ, are calculated by scaling the market returns by the inverse of the variance estimate. The methods used for variance estimation are Realized Variance (1), ARIMA (2), ESA (3) and GARCH (4). The data is monthly and the sample period ranges from July 1926 to August 2019. Standard errors are in parentheses and adjust for heteroskedasticity. All factors are annualized in percent per year by multiplying monthly factors by 12. Additionally, we report the Sharpe ratio of volatility-managed returns, the Appraisal Ratio and the Fama French three factor alpha. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

	(1) Var	(2) ARIMA	(3) ESA	(4) GARCH
Mkt-RF	0.61***	0.70***	0.72***	0.80***
	(0.05)	(0.06)	(0.06)	(0.06)
Alpha (α)	4.69***	4.67***	4.28***	3.34***
	(1.50)	(1.35)	(1.32)	(1.12)
Observations	1,117	1,117	1,117	1,117
R^2	0.37	0.48	0.51	0.63
Residual Std. Error	50.70	46.00	44.67	38.72
Sharpe Ratio	0.52	0.55	0.54	0.52
Appr Ratio	0.32	0.35	0.33	0.30
Alpha FF3	5.19***	5.40***	5.08***	4.31***
	(1.50)	(1.35)	(1.30)	(1.09)

expected return, while having, by definition, the same standard deviation. The Sharpe Ratio of the ARIMA strategy is the highest, followed by ESA and GARCH. Overall, it seems like the three alternative strategies take on a higher systematic risk in exchange for a lower idiosyncratic risk (by construction, one has to be higher and one has to be lower since the overall standard deviation is equal). However, in total, all strategies are relatively similar with GARCH falling a bit behind with respect to the outperformance. Changing the method of variance estimation appears to have some advantages, even though the presented results do not offer a clear cut rationale for doing so.

To further understand the performance of all strategies, we plot their cumulative performance in contrast to the buy-and-hold portfolio. This replicates the hypothetical case of an individual investor investing $1 in 1926 and always reinvesting her proceeds. As can be seen, all four volatility-managed strategies outperform the simple buy-and-hold portfolio. Furthermore, even though the disparity varies over time, all strategies show a similar course. However, the hypothetical end amount of money differs. ARIMA is the most successful strategy, yielding $53,000 in the last period, followed by ESA with $43,000, GARCH with $32,000, the original strategy with $29,000 and the buy-and-hold strategy with $6,600. Overall, this substantiates the previous notion that all implementations offer a similar alpha, but the actual returns vary due to different risk loadings, with the alternative strategies yielding higher expected returns.

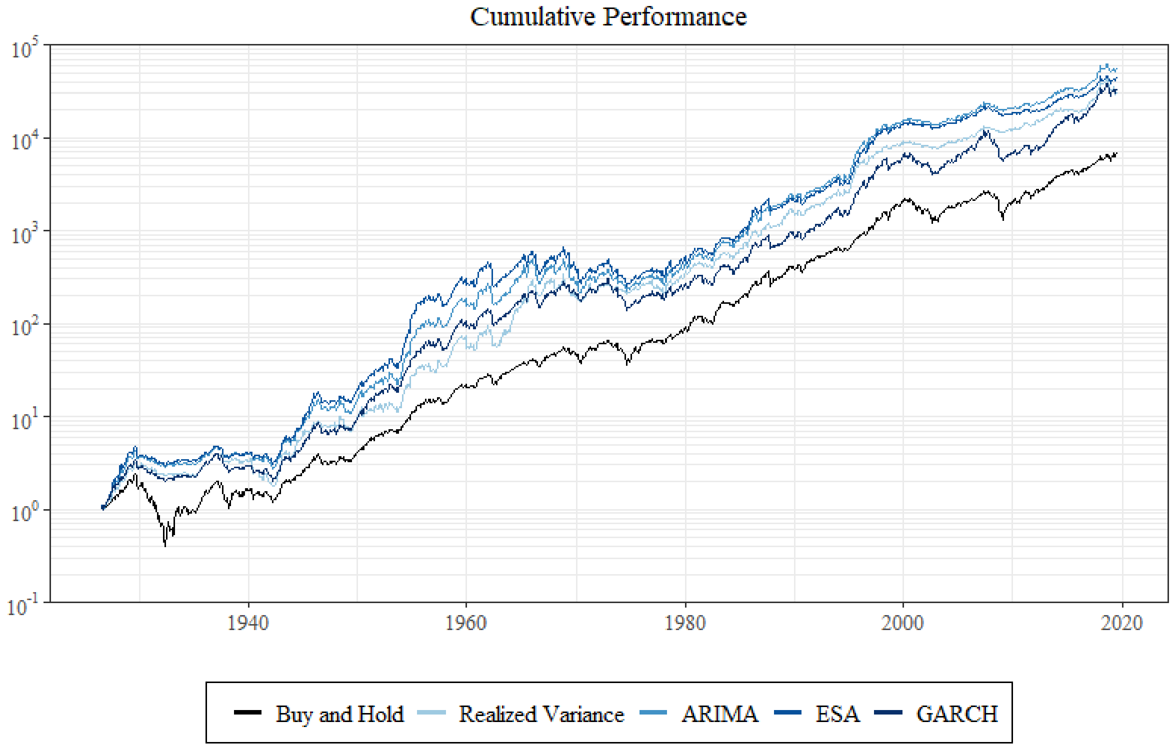


Figure 2: Cumulative Performance

The plot depicts the cumulative performance of $1 invested in July 1926 until August 2019 for the Buy-and-Hold portfolio, Moreira and Muir's realized variance strategy and the ARIMA, ESA, and GARCH strategies. The y-axis is on a log-scale and all strategies have the same unconditional monthly standard deviation.

To dive deeper into the performance of all strategies over time, we plot the rolling one-year average returns of all strategies over the whole time period in Figure 3 below. As can be seen, all of them, including the market, have a relatively similar course, especially in the second half of the data. However, all strategies, partly apart from GARCH, are able to avoid notable market downturns like the Great Depression (1929 - 1933), the burst of the dot-com bubble (2001), or the Great Recession (2007 - 2009) (NBER, 2019). Nevertheless, all of them, in general, miss out on the initial rebound in prices subsequent to such downturns. Furthermore, the strategies seem to profit from sustained market rallies without much volatility, like in 1995, leading them to overweight the market, earning excess returns. The risk, however, as described initially, is a sudden drop in prices, which then is also being magnified by the high weights, potentially eliminating the previous excess return. All in all, it is difficult to clearly differentiate the performance of the strategies, with GARCH slightly falling behind, due to its incapability to time some of the notable market downturns.

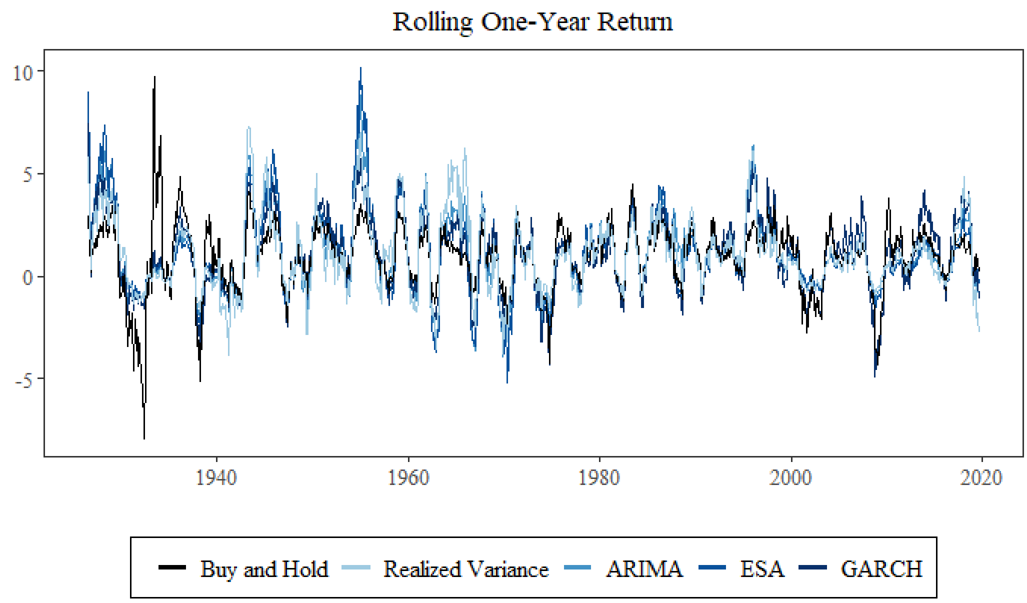


Figure 3: Rolling One-Year Return

The plot depicts rolling one-year returns from the Buy-and-Hold portfolio, Moreira and Muir's realized variance strategy, and the ARIMA, EWMA, and GARCH strategies. All strategies have the same unconditional monthly standard deviation.

To assess the performance over time from another angle, we also plot the drawdowns of each strategy in Appendix D. The drawdown is defined based on a hypothetical investment of $1 in 1926 and depicts, if negative, the percentage change from the historical peak to the current value of the investment. Otherwise, it is zero. It underpins the previous notion, that all strategies, partly except from GARCH, are able to avoid large losses in notable market downturns.

5.1.2. Return and Weight Distribution

To further our analysis of the performance of each respective strategy, we calculate descriptive statistics for the distributions of returns and weights, reported in Table 4.

As previously expected, due to the similar alpha but higher beta, the average returns of the alternative strategies are higher than for the original strategy by Moreira and Muir. However, all means are relatively close together. The original strategy is the only one with positive skewness. While this generally indicates, that returns lower than the mean are more common than returns higher than the mean, it is nevertheless often preferred by investors. This is because risk-averse investors are reluctant to accept the higher probability of large losses, however small, combined with the limited gain, exhibited by return distributions with negatives skewness (Arditti, 1967). However, the difference in skewness is again relatively small. Furthermore, the kurtosis of the original strategy is considerably higher than the kurtosis of the other strategies. A higher kurtosis indicates more extreme values in the distribution. This is further

Table 4: Summary Statistics of Returns and Weights.

In Panel A we report several statistical measures for the distribution of volatility-managed returns for Moreira and Muir's realized variance, ARIMA, ESA, and GARCH strategy. In Panel B we report several statistical measures for the distribution of the respective weights.

Panel A: Returns				
	Var	ARIMA	ESA	GARCH
Mean	1.07	1.12	1.10	1.08
SD	5.32	5.32	5.32	5.32
Skewness	0.17	-0.19	-0.24	-0.52
Kurtosis	9.28	5.76	6.24	4.49
Min	-30.62	-24.85	-23.69	-26.25
P25	-1.12	-1.35	-1.40	-1.83
P50	0.90	1.07	1.04	1.31
P75	3.04	3.77	3.83	4.36
Max	35.59	26.98	28.89	17.33
Panel B: Weights				
	Var	ARIMA	ESA	GARCH
Mean	1.25	1.25	1.20	1.26
SD	1.32	0.91	0.79	0.61
Skewness	3.50	1.35	0.90	0.31
Kurtosis	23.53	5.62	3.76	2.67
P50	0.92	1.07	1.07	1.22
P75	1.57	1.72	1.69	1.64
P90	2.62	2.40	2.28	2.11
P99	6.73	4.15	3.56	2.71
Max	14.50	5.85	4.36	2.98

substantiated by the minimum and maximum return, which are farther away from the mean for the original strategy. The alternative measures of computing the variance, therefore, seem to lower the number of extreme cases in returns, a desirable property to reduce the tail risk of the strategy and make the returns more predictable. Lastly, the standard deviation of the strategies is by construction the same.

Since the return in each period depends on the weights chosen, we additionally look into their statistics to better understand the distribution of returns. And while the mean of the weight is similar for all strategies, the standard deviation and especially kurtosis of the original implementation is distinctively the highest. The apparent occurrence of extreme weights is confirmed by the quantiles. While the median of the original strategy is actually the lowest, the 99% quantile is the highest.

In Appendix E, additionally, the density function of returns and weights is showcased. Both further underpin the notion, that the original strategy is less dispersed around its mean, but has more extreme outliers than the alternative implementations, leading to the same standard deviation, as defined. This increases the tail risk for investors and can potentially help explain the at the beginning introduced observation of occasional significant market underperformance. The topic of non-performing is further analyzed in chapter 0.

5.1.3. Transaction Costs and Leverage Constraints

All of the aforementioned results are based on the presumption that trading does not incur any costs and investors can freely leverage their investment, i.e. increase the weight on the market above one when needed. However, in practice, both assumptions are not realistic and might substantially alter the actual performance of implementing the strategy and potentially even eradicate the outperformance. Therefore, we study the effect of incorporating both aspects to a varying degree.

First, we assess the impact of trading costs, following the approach implemented by Moreira and Muir. They use trading costs of 1bps, 10bps, and 14pbs. 1pbs comes from Fleming et al. (2003), using the price of S&P 500 futures contracts, 10 bps comes from Frazzini et al. (2012), depicting average historical costs for a long-short trade, and 14bps accounts for additional costs of trading in times of high volatility. Additionally, we report break-even costs, i.e. the costs that would drive alpha down to zero. The results are shown in Panel A of Table 5.

As can be seen in the table, incorporating trading costs does drive down the alpha of each strategy. However, under given assumptions, all implementations still offer a considerable alpha over the normal buy-and-hold portfolio. Especially ESA and GARCH seem to be largely independent of trading costs. The reason is that their variance estimates change slower since they meaningfully incorporate more periods looking backward. This leads, on average, to smaller weight adjustments at each rebalancing date. Nevertheless, especially for ESA, the smaller adjustments do not seem to hurt the overall overperformance. For example, the ESA strategy still yields a positive alpha even with trading costs of over 4%. Moreira and Muir (2017) also attempt to make their strategy less dependent on trading costs by incorporating leverage constraints and using volatility and expected variance as a scaling factor instead of realized variance. However, they only achieve break-even trading costs of 161bps, while sacrificing overall overperformance compared to the original strategy. All in all, the alternative strategies and especially ESA seem to offer the advantage of achieving an alpha similar to the original strategy, while considerably driving down absolute transaction costs due to smaller absolute weight adjustments at each rebalancing point. This is especially desirable for smaller investors since they often face comparably higher transaction costs (Baker et al., 2015).

Table 5: Alpha Considering Transaction Costs and Leverage Constraints.

In Panel A, we evaluate the volatility timing strategies when including transaction costs. We report the average absolute change in monthly weights $|\Delta w|$ and the alpha of each strategy, depending on the implemented costs. The last column backs out the implied trading costs in basis points needed to drive the alphas to zero. In Panel B, we evaluate the volatility timing strategies when accounting for various leverage constraints. We report the alpha of each strategy for the unconstrained case and for the cases if an investor faces a leverage constraint of 100%, 50% or 0% (i.e. no leverage) constraint.

Panel A: Alpha Considering Transaction Costs						
	Alpha	$\|\Delta w\|$	1bps	10bps	14bps	Break-Even
Var	4.69	0.73	4.61	3.82	3.47	54bps
ARIMA	4.67	0.40	4.62	4.19	3.99	97bps
ESA	4.28	0.08	4.27	4.18	4.14	446bps
GARCH	3.34	0.12	3.33	3.20	3.14	238bps

Panel B: Alpha Considering Leverage Constraints				
	No Constraint	100% Leverage	50% Leverage	0% Leverage
Var	4.69	3.56	3.03	2.07
ARIMA	4.67	4.32	3.60	2.28
ESA	4.28	3.95	3.37	2.13
GARCH	3.34	3.15	2.64	1.58

Second, we incorporate leverage constraints of varying degrees. The reason is, that for private investors, leverage might not be readily available and institutional investors might be restricted by their investment mandate. We consider constraints of 100%, 50%, and 0%. In other words, we incorporate weight constraints of 2, 1.5, and 1. The resulting alphas are shown in Table 5.

As can be seen, imposing a leverage constraint can have a strong impact on the respective alpha of the strategy. In theory, the effect is ambiguous since constraints limit extreme weights, thereby most likely limiting extreme returns of unknown size. However, the alpha is impacted negatively, showing that the effect of constraining the strategy in reacting to market volatility is stronger than the potentially positive one of limiting outliers. For a leverage constraint of 100%, particularly the original strategy suffers. With tighter leverage constraints, the impact evens out, until, for a constraint of 0%, the alpha of all strategies approximately halves.

5.1.4. Business Cycles

After assessing the practical implementability of volatility timing based on four different variance estimation techniques, we examine how the strategies behave during business cycles to better understand their risk exposure and potentially their source of alpha. As mentioned in the data description, we use the NBER based recession indicator from the Federal Reserve Bank of St. Louis, a binary variable denoting one in recessionary and zero in expansionary periods. The basis of our analysis forms the time series regression in equation (3), additionally adjusted for an interaction term between market excess return and recession indicator and the indicator itself. The results are shown in Table 6.

As can be seen, the interaction term of market excess return and recession indicator is highly significant on the 1% level. This means that during recessions, the market beta of each strategy drops - looking at the data roughly by half. In other words, by scaling the excess returns by the inverse of the estimate of conditional variance, the strategy is able to time business cycles, taking on less risk during recessions, usually associated with a downturn in the market. During expansionary periods, the risk exposure is higher, for GARCH even around 1. Additionally, controlling for recessions makes the alpha of all strategies in expansionary times insignificant. During recessionary times, though, it is significant for ARIMA and ESA on the 10% level. All in all, this implies that reducing the risk exposure during recessions and increasing it during expansions might explain the alpha generated by volatility timing. However, this observation does not question the meaningfulness of volatility-managed portfolios in itself, since the recessionary classification of a period can only be done ex-post, thereby necessitating a decision rule based on backward-looking data. Volatility timing does provide such a rule. Furthermore, it complements the previously expressed notion, that decreasing the granularity of weight adjustments might not considerably alter performance while showcasing other desirable properties like lowering transaction costs – be it through a less sensitive variance estimation technique like ESA or simplified adjustments only based on business cycles.

Table 6: Performance Controlling for Recession Indicator.

In this table, we run time-series regressions of volatility-managed returns on the market factor, the binary NBER recession indicator and an interaction term of market factor and recession dummy $f_t^{\sigma} = \alpha_0 + \alpha_1 1_{rec,t} + \beta_0 f_t + \beta_1 f_t 1_{rec} + \epsilon_t$. The managed returns, f^{σ}, are calculated by scaling the market returns by variance. The methods used for variance estimation are Realized Variance (1), ARIMA (2), ESA (3) and GARCH (4). The data are monthly and the sample period ranges from July 1926 to August 2019. Standard errors are in parentheses and adjust for heteroskedasticity. All factors are annualized in percent per year by multiplying monthly factors by 12. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

	(1) Var	(2) ARIMA	(3) ESA	(4) GARCH
Mkt-RF	0.83***	0.90***	0.90***	1.01***
	(0.08)	(0.08)	(0.08)	(0.08)
Mkt-RF * Recession	-0.51***	-0.47***	-0.43***	-0.49***
	(0.09)	(0.10)	(0.10)	(0.10)
Recession	4.68	6.80*	7.22*	1.45
	(3.50)	(3.91)	(4.14)	(3.52)
Alpha (α)	1.62	1.39	1.09	0.92
	(1.65)	(1.38)	(1.32)	(1.06)
Observations	1,117	1,117	1,117	1,117
R^2	0.44	0.54	0.56	0.69

5.1.5. Relative Outperformance of Alternative Strategies

Based on the previous analysis, the three alternative strategies, in terms of performance, seem to be mostly on par or better than the original implementation. However, to better comprehend whether they actually outperform the original strategy, we repeat the time series regression from equation (3), while also controlling for the excess return of the realized variance implementation as an additional factor.

The regressions suggest, that all three alternative implementations have highly significant factor loadings (on the 1% level) on both the market and the original variance strategy. The performance of the strategy based on ESA seems to be less dependent on the realized variance strategy, however, the factor is still highly significant. At the same time, the constant of all alternative strategies is not significant anymore. This means, when controlling for the market and the original strategy, the three alternatives do not offer an additional, significant excess return. On the contrary, in large parts, their movement can be explained by the factor loadings, indicated by all R^2 being higher than 85%. Hence, all in all, it seems changing the way of determining the variance might offer several desirable properties, as mentioned before, but cannot significantly beat the strategy by Moreira and Muir in terms of performance.

Table 7: Performance of Alternative Strategies Controlling for Original Strategy.

In this table, we run time-series regressions of volatility-managed returns on the market factor and the realized variance strategy returns $f_{t+1}^{\sigma} = \alpha + \beta_1 f_{t+1} + \beta_2 f_{t+1}^{Var} + \epsilon_{t+1}$. The managed returns, f^{σ}, are calculated by scaling the market excess returns by the inverse of the variance estimates. The methods used for variance estimation are ARIMA (1), ESA (2), and GARCH (3). The data is monthly and the sample period ranges from July 1926 to August 2019. Standard errors are in parentheses and adjust for heteroskedasticity. All factors are annualized in percent per year by multiplying monthly factors by 12. *, **, and *** indicate significance at the 10%, 5%, and 1% level.

	(1) ARIMA	(2) ESA	(3) GARCH
Mkt-RF	0.78***	0.90***	0.72***
	(0.01)	(0.02)	(0.03)
Var	0.22***	0.09***	0.28***
	(0.03)	(0.02)	(0.04)
Alpha (α)	1.00	0.08	0.24
	(0.72)	(0.49)	(0.65)
Observations	1,117	1,117	1,117
R^2	0.87	0.93	0.89

Since the alternative strategies seem to especially offer advantages when considering transaction costs, we repeat mentioned regressions while incorporating the previously used set of costs, plus the additional case of 50bps (Thomas and Shapiro, 2009).

As presumed, with transaction costs, the alternative strategies seem to offer an alpha over the original one that increases in magnitude with the costs. This is especially the case for ESA, which is, as discussed before, the least influenced one by costs. It offers relatively the highest alpha, which becomes significant on the 10% level for 10bps. The same is true for ARIMA, even though the alpha is lower. The 50bps are intended as an extreme case to further assess the pattern. Indeed, for 50bps, the alpha of all three alternative strategies becomes significant on the 1% level. All in all, these results substantiate our previous notion that the alternative strategies are similar to the original ones in terms of performance without considering transaction costs but can offer an additional advantage when incorporating them. Depending on the type of investor and the costs she faces, this, especially over time, can make a substantial difference. For example, reconsidering the hypothetical case of $1 invested in 1926, incorporating transaction costs of 10bps diminishes the final value of the original strategy by 55%. For ESA, however, it only lowers the final value by 8% (with 36% for ARIMA and 12% for GARCH).[11]

[11] Considering costs of 10bps, the original strategy yields $13,000, ARIMA $34,000, ESA $39,000, and GARCH $28,000. The numbers without costs can be bound in chapter 5.1.1.

Table 8: Relative Alpha Depending on Transaction Costs.

In this table we show alphas of the regressions described in the caption of Table 8, accounting for transaction costs of 1bps, 10bps, 14bps, and 50bps. Standard errors are in parentheses and adjust for heteroskedasticity. All factors are annualized in percent per year by multiplying monthly factors by 12. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

	ARIMA	ESA	GARCH
0pbs	1.00	1.32	1.04
	(0.72)	(0.97)	(0.90)
1bps	1.01	1.36	1.07
	(0.72)	(0.97)	(0.90)
10bps	1.18*	1.76*	1.32
	(0.71)	(0.97)	(0.90)
14bps	1.26*	1.94**	1.43
	(0.71)	(0.96)	(0.90)
50bps	2.00***	3.58***	2.47***
	(0.70)	(0.95)	(0.89)

5.1.6. Subsample Analysis

Lastly, since the whole time period from 1926 to 2019 spans almost 100 years, we look at the performance of each strategy in sub-periods of roughly 31 years: 1926 - 1957, 1957 - 1988, and 1988 - 2019. The goal is to better understand how dependent the outperformance of the strategy on the specific time frame is. All regression results are shown in Table 9.

The table shows, that the outperformance of the strategy does indeed depend on the considered time period. In the first period, the strategies generate the largest alpha, with almost 10% up to 15%, depending on the specific implementation. All alphas are significant on the 1% level. The betas, on the other hand, are the lowest at around 0.5 to 0.7. Both observations can be related back to Figure 3, showing the rolling average returns. As discussed, there we can see, that all strategies are able to time the considerable market downturn in the Great Depression, leading to a lower exposure and a high alpha compared to the buy-and-hold portfolio. The first period is also the most volatile one with regard to the market returns.

The second period, from 1957 - 1988 is the least successful one. The alphas are either slightly positive or slightly negative, but never significantly different from zero, not even on the 10% level. The beta, on the other hand, is the highest. In general, the period was relatively stable, compared to the other two. In such times, the strategy more or less mirrors the buy-and-hold portfolio, but potentially suffers from major setbacks like initially discussed.

Table 9: Subsample Analysis.

In this table we run time-series regressions of volatility-managed returns on the market factor: $f^{\sigma}_{t+1} = \alpha + \beta f_{t+1} + \epsilon_{t+1}$. The managed returns, f^{σ}, are calculated by scaling the market returns by variance. The methods used for variance estimation are Realized Variance (1), ARIMA (2), ESA (3) and GARCH (4). The data are monthly. For regressions (1) - (4) we apply data from 1926 - 1957. For regressions (5) - (8) we use data from 1926 - 1957. For regressions (9) - (12) we apply data from 1926 - 1957. Standard errors are in parentheses and adjust for heteroskedasticity. All factors are annualized in percent per year by multiplying monthly factors by 12. Additionally, we report the Sharpe ratio of volatility-managed returns, the Appraisal Ratio and the alpha controlling for Fama French three factors. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

	1926 - 1957				1957 - 1988				1988 - 2019			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Var	ARIMA	ESA	GARCH	Var	ARIMA	ESA	GARCH	Var	ARIMA	ESA	GARCH
Mkt-RF	0.56***	0.61***	0.61***	0.71***	0.71***	0.85***	0.89***	0.93***	0.65***	0.75***	0.80***	0.90***
	(0.09)	(0.08)	(0.08)	(0.08)	(0.05)	(0.05)	(0.04)	(0.04)	(0.06)	(0.05)	(0.05)	(0.04)
Alpha (α)	9.51***	13.34***	15.51***	12.08***	1.86	-0.12	-1.86	-1.62	4.14**	4.14**	3.41**	1.49
	(3.49)	(3.20)	(3.15)	(2.79)	(1.98)	(1.51)	(1.32)	(1.04)	(2.10)	(1.79)	(1.63)	(1.20)
Observations	372	372	372	372	371	371	371	371	372	372	372	372
R^2	0.31	0.37	0.38	0.51	0.51	0.72	0.79	0.86	0.42	0.57	0.64	0.81
Residual Std. Error	68.75	65.52	65.47	58.19	37.81	28.45	24.54	20.26	38.49	33.06	30.27	22.25
Standard Deviation	23.90	23.90	23.90	23.90	15.42	15.42	15.42	15.42	14.57	14.57	14.57	14.57
Sharpe Ratio	0.65	0.83	0.92	0.82	0.36	0.28	0.18	0.21	0.65	0.71	0.68	0.60
Appraisal Ratio	0.48	0.71	0.82	0.72	0.17	-0.01	-0.26	-0.28	0.37	0.44	0.39	0.23
Alpha FF3	9.82***	13.72***	15.92***	12.52***	0.90	-0.73	-2.53*	-2.12*	4.08*	4.26**	3.60**	1.62
	(3.48)	(3.21)	(3.15)	(2.77)	(2.00)	(1.67)	(1.43)	(1.13)	(2.10)	(1.79)	(1.63)	(1.20)

In the last period, the original strategy, as well as the strategies based on ARIMA and ESA variance estimates, again generate an alpha of around 3% to 4%, significant on the 5% level. GARCH does not generate a significant alpha. This can also be explained in Figure 3, where we can see, as described, that the first three strategies are able to avoid the market downturns of the burst of the dot-com bubble and the Great Recession, while GARCH is not.

Overall, we can conclude that the actual outperformance of the strategies does depend on the respective times period. The subsample analysis implies, that the strategies especially outperform the market if there is a major crisis or sustained market downturns that they are able to avoid through volatility timing. In rather stable times, the advantages of volatility-managed portfolios seem limited.

5.1.7. Non-Performing Periods

As can be seen, the strategy of Moreira and Muir yields significant alphas and Appraisal Ratios in all four of our considered implementations. However, as previously discussed, the original one suffers shortcomings in periods of high volatility changes. If a period with stable returns is followed by a period of large losses, the variance managed portfolio is not able to adequately adjust the weight and considerably underperforms the buy-and-hold portfolio. Subsequently, it reduces the risk exposure, thereby often missing the following rebound in returns. Such constellations do not happen often over the whole sample period, but it noteworthily happened two times in recent years: in October 2018 and in May 2019. These instances are outside the considered timeframe in the original paper. Extending the analyzed timeframe, therefore, reduces the reported alphas from 4.86% to 4.69%. To better understand the shortcomings of the strategy and whether implementing our alternative estimations of variance can mitigate them, we employ several ways to specifically compare their performance. Namely, we calculate the average weights, the correlation of returns, and the root mean squared deviation for the periods of interest.

We define a non-performing period as one, where the market yields an unfavorable return, but the strategy indicates to choose a weight above one (i.e. over-weights the market). Over the whole timeframe, this happened 185 times, corresponding to 17% of all months and 45% of the months with negative market returns. At first glance, these numbers seem to be quite high. However, the idea of volatility-managed portfolios does not claim to be able to predict returns or avoid negative returns altogether but is based on the preposition of scaling down when the risk-return tradeoff is unfavorable.

We employ different measures of non-performing periods with varying severity in the constraints, namely:

- First set of months: Mkt < 0%, $w > 1$,
- Second set of months: Mkt < -5%, $w > 1$,
- Third set of months: Var < -10%, $w > 1$,
- Fourth set of months: Var < -15%, $w > 1$,

where Mkt is the market return, Var is the realized variance strategy's return, and w is the weight that the realized variance strategy assigns to the market. We define bad periods not only by extraordinary low market returns but also by low strategy returns to be specifically able to capture the extreme cases, where the strategy does not perform well, instead of just the cases where the whole market was down. However, it is relevant to keep in mind, that a negative strategy return always indicates a negative market return and vice versa, since weights are defined as above one and also, in general, never negative. For the respectively chosen set of months, we compare all implemented strategies to the buy-and-hold portfolio and the three alternative strategies to the original implementation of the strategy.

First, we look at the average chosen weight of each strategy during times when the original one did not perform.

Table 10: Analysis of Non-Performing Periods Based on Weights.

In Panel A we report the average weights for the realized variance, ARIMA, ESA, and GARCH strategy in the four previously defined sets of months as well as for the subsample of months where the original weights are above one, independent of the returns. In Panel B we report the relative deviation in weights of the ARIMA, ESA and GARCH strategy from the realized variance strategy for the same sets of months.

Panel A: Average Weights					
	$w > 1$	Mkt < 0% $w > 1$	Mkt < -5% $w > 1$	Var < -10% $w > 1$	Var < -15% $w > 1$
Var	2.14	1.19	1.93	3.02	2.96
ARIMA	1.92	1.19	1.92	2.55	2.54
ESA	1.68	1.14	1.71	2.06	2.12
GARCH	1.57	1.21	1.57	1.63	1.73
Panel B: Relative Deviation in Average Weights					
	$w > 1$	Mkt < 0% $w > 1$	Mkt < -5% $w > 1$	Var < -10% $w > 1$	Var < -15% $w > 1$
Var	0.00	0.00	0.00	0.00	0.00
ARIMA	-0.10	0.00	0.00	-0.16	-0.14
ESA	-0.22	-0.04	-0.11	-0.32	-0.28
GARCH	-0.27	0.01	-0.18	-0.46	-0.41

When only taking the periods into account when the original strategy indicates a weight above one, the average weight is around 2.1. This is higher than for all three other strategies. However, this suffers from the bias that the periods for all strategies are chosen based on a constraint only referring to the strategy based on realized variance. In other words, potentially, periods exist where the alternative strategies non-perform while the original strategy does perform. Therefore, we have to be cautious in interpreting the results. Nevertheless, in the bad periods of the original strategy, the alternatives all seem to be more conservative, rightfully choosing a lower weight, leading to less negative returns. In other words, compared to the strategy by Moreira and Muir, they are better at lowering the market exposure in periods where the original strategy suffers from indicating too high weights. The difference tends to become more pronounced the more severe the constraints are.

Next, we calculate the correlation of all strategies with the market and of the three alternative strategies with the original strategy during the non-performing periods.

Table 11: Analysis of Non-Performing Periods Based on Correlations.

In this table, we report the correlations between the returns of the realized variance, ARIMA, ESA, and GARCH strategy and the market returns as well as the correlations between the returns of the realized variance strategy and the ARIMA, ESA, and GARCH strategy. In addition, we state the number of observations of each set of months.

	Whole Sample	Mkt < 0% w > 1	Mkt < -5% w > 1	Var < -10% w > 1	Var < -15% w > 1
Observations	1,117	415	24	21	12
Mkt-Var	0.61	0.32	0.62	0.74	0.73
Mkt-ARIMA	0.69	0.41	0.20	0.30	-0.01
Mkt-ESA	0.72	0.46	0.06	0.20	-0.10
Mkt-GARCH	0.80	0.63	0.35	0.57	0.40
Var-ARIMA	0.91	0.86	0.64	0.43	-0.10
Var-ESA	0.83	0.71	0.42	0.28	-0.22
Var-GARCH	0.79	0.61	0.64	0.65	0.41

The correlation between the strategies and the market can essentially be seen as the beta of the strategy. It slightly deviates but would be the same if we used excess returns to calculate the correlations, not overall returns. The reason is, that c and therefore the weights are defined in a way that equalizes the variance of the returns for all strategies with the variance of the market return. Therefore, the variance cancels out of the beta calculation with the correlation remaining. The derivation is included in Appendix F. For the whole sample, the results are therefore similar to those already reported. Looking at the non-performing periods, we can see that the original strategy actually becomes closer correlated with the market the more it underperforms. The other strategies, however, show the reversed trend. Their correlation with the market decreases in non-performing periods. This is a desirable property since it is preferable for investors to decrease the risk exposure during bad periods.

This reversed trend is being validated by the correlation of the alternative strategies with the original strategy. During bad periods, the correlation decreases, even becomes negative for ARIMA and ESA. GARCH showcases a similar trend, but not as pronounced. These results show, that the alternative strategies do behave differently from the original strategy during its non-performing periods and apparently decrease the exposure. Therefore, employing alternative methods for variance estimation seems to be a factor mitigating the risk of periods of underperformance.

All in all, the three alternative strategies seem to behave preferable during non-performing periods from an investor's point of view, even though one has to be cautious about the biased definition of the non-performing periods. Not only are their chosen weights lower, leading to better returns during considered market downturns, but they are also less correlated with market movements. Among the alternative strategies, ESA seems to be the one with the most desirable properties.

5.2. Different Time Horizons

After analyzing the implementation of the strategy of Moreira and Muir based on four different variance estimates and comparing their behavior during non-performing periods, the next chapters describe our examination of the second employed lever: the impact of changing the rebalancing period, both on a fixed and a flexible basis.

5.2.1. Fixed Periods

The idea of a more frequent rebalancing is to increase the performance of the strategy by furthering the flexibility of the strategy in adapting to changes in the market and by possibly enhancing the precision of the variance estimates. To achieve a broad spectrum of results in order to elicit overall trends, while getting reasonable estimates for the calculated variance, we consider rebalancing periods between five and 50 trading days, corresponding to a minimum of one week and a maximum of ten weeks (not considering holidays). Changing the rebalancing period correspondingly changes the estimation period for the realized variance.

We then calculate performance measures on an annualized basis. Figure 4 depicts the alpha of each respective strategy, depending on the holding period.

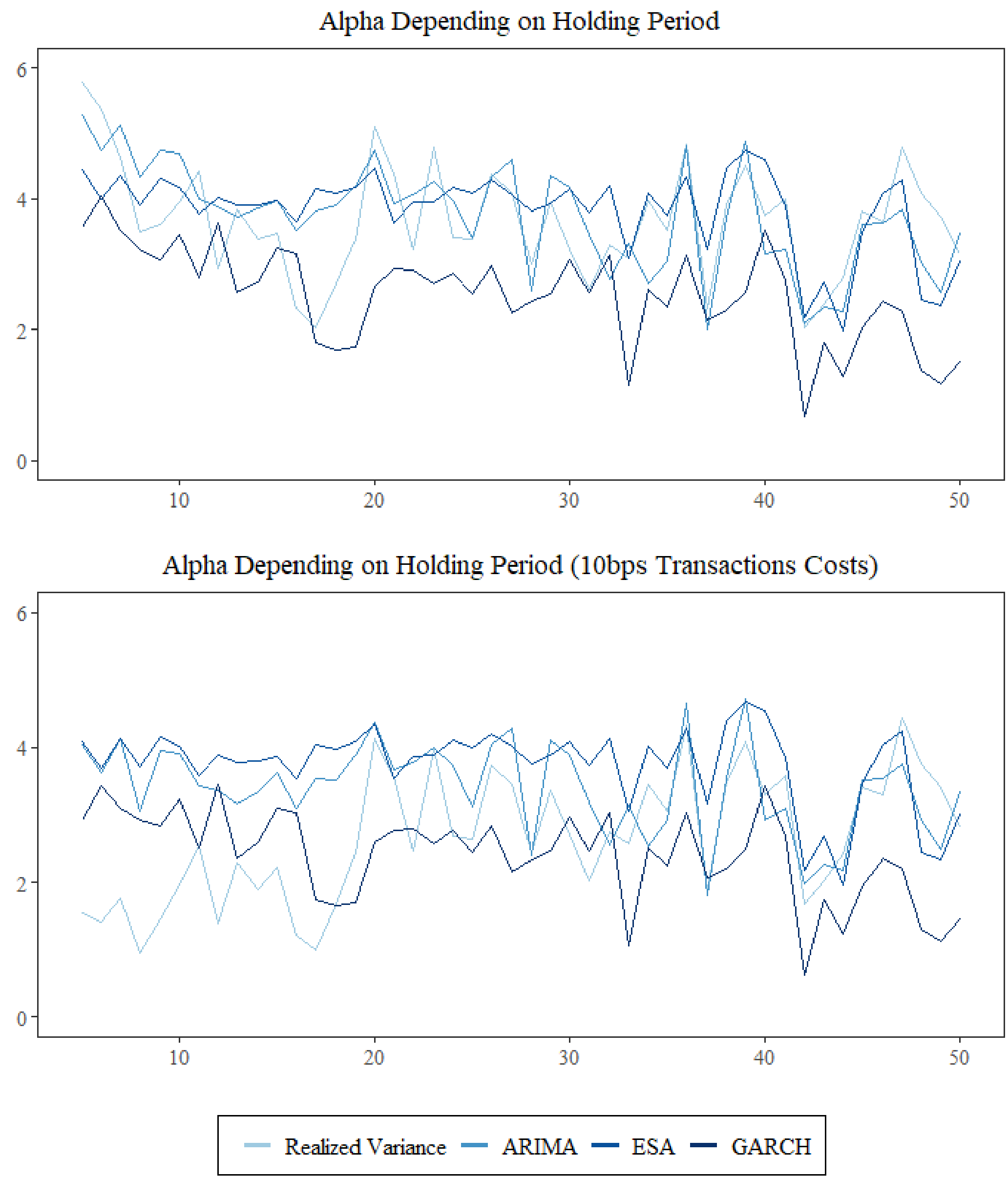


Figure 4: Alpha Depending on Holding Period

The top panel plots the alphas of the realized variance, ARIMA, ESA, and GARCH strategy depending on re-balancing periods between five and 50 trading days. The bottom panel shows the alphas of respective strategies depending on the holding period when additionally accounting for transaction costs of 10bps. Alphas are annualized and stated in percent.

As can be seen, the alpha varies considerably over time with a general downward trend with the length of the rebalancing period. Relatively, ESA seems to offer the most stable alpha, mostly slightly below 4%, especially up until a period length of 30 days. ARIMA is on average on a similar level, but varies more, depending on the period length. The same applies to realized variance, which, however, seems to be even more sensitive to the rebalancing frequency. GARCH is for short period lengths comparable with the other strategies, however, overall has the most volatile alpha and is from period length eighteen on almost consistently inferior. This suggests that GARCH is less utilizable for longer forecasting periods, which compares to the general notion that GARCH is made to forecast volatility on shorter horizons.

Compared to before, intuitively, trading costs should play a bigger role for shorter periods due to the more frequent rebalancing and therefore trading activity. To see whether this might eradicate the seemingly slight advantage of shorter periods, we incorporate transaction costs of 10bps in all strategies and again calculate the alpha dependent on the period length. The alpha of all strategies decreases, compared to before, however, the change is most pronounced for the strategy based on realized variance, especially for shorter holding periods. Up to almost twenty days, it becomes the inferior strategy, before getting closer again to the alternative ones. The alpha of ARIMA also goes down, even though not as strong, and is now slightly below the ESA alpha for most period lengths. The impact of trading costs on ESA and GARCH is almost negligible, even for frequent rebalancing. Overall, the level of alpha is similar over all period lengths, however, shorter lengths showcase less volatile alphas, at least for ARIMA and ESA.

In Appendix G and Appendix H, we also plot the Appraisal Ratio and the cumulated performance of $1 invested from 1926 to 2019 for all four strategies, both without transaction costs and considering costs of 10bps. The course of Appraisal Ratio and alpha are overall similar. However, the function of realized variance moved slightly downwards, while the function of ARIMA moved slightly upwards, both due to the differing RMSE. Therefore, ESA and ARIMA become the almost consistently superior strategy for period lengths of up to twenty days. The graph of cumulative performance is considerably more volatile than the graphs for alpha and Appraisal Ratio. Again, ARIMA and ESA seem to be the most appealing strategies. Realized Variance is clearly inferior for rebalancing lengths of up to around fifteen days. However, in general, it is difficult to elicit a clear trend.

All in all, the impact of the rebalancing period on the overall performance is not clear-cut, but the strategies tend to produce higher alphas for shorter periods. Furthermore, the estimates tend to be less sensitive to adjustments in length for shorter periods. Among the orders, ARIMA and ESA seem to be the best performing strategies, with GARCH and realized variance, depending on the measure, falling behind.

5.2.2. Flexible Periods

After considering varying fixed rebalancing frequencies, we implement our devised strategy based on flexible periods. The idea is to be able to adjust the risk exposure following large changes in volatility while decreasing the frequency of trading in more stable times to save transaction costs. To elicit trends, we consider varying decision rules and variance estimation methods. Additionally, we consider the impact of transaction costs, alongside the previous assumptions of 1bps, 10bps, and 14 bps. After calculating the returns of each strategy, we run a time series regression on the market excess-return. Table 12 reports the alphas of all strategies, depending on the decision rule, method of variance estimation, and transaction costs. Additionally, it shows the average absolute weight changes.

Regarding the decision rule, trading more often on average seems to be preferable. The alpha of the strategies increases almost consistently from an average rebalancing period of 22 days to an average of two days, before falling again for the case of daily trading. This indicates, that while alpha seems to rise with the average rebalancing frequency, there appears to be a limit, above which the advantage of increasing the frequency ceases.

Considering the different techniques of variance estimation, EWMA seems to perform slightly better than GARCH, however, the performance of both strategies is overall similar, which can be explained by the closely related underlying model. The performance-based on realized variance is strongly dependent on the chosen estimation period. The shortest interval of one week seems to be too short for meaningful estimates, showcasing a negative alpha for all decision rules. Extending the estimation periods to two weeks, however, immensely improves the performance, especially for the case of more frequent trading. For an average holding period of two days, it even achieves the overall highest alpha of all implementations. However, the alpha is highly susceptible to trading costs. This originates in the comparably high average absolute change in weights, a number that consistently decreases with the length of the period that is being considered to calculate the realized variance. A greater length lowers the influence of one data point dropping out and one being added in moving the rolling window, leading to lower absolute weight changes. Regarding performance, extending the length has mixed effects. In general, between estimation periods of two weeks and one year, the alpha first goes down before increasing again. For less frequent rebalancing, the initial drop is less pronounced than for more frequent rebalancing. However, the subsequent increase is more pronounced for more frequent rebalancing. For two years, the performance in general drops again. EWMA and GARCH are in their performance generally similar to a cross-section of the strategies based on realized variance, performing better than some of their realizations and worse than others.

Table 12: Alphas for Flexible Rebalancing Periods.

We run time-series regressions of volatility-managed returns on the market excess return for all strategies with flexible time periods. We show alphas for the case of no transaction costs and when including costs of 1bps, 10bps, and 14bps. Furthermore, we report the average absolute change in weights ($|\Delta w|$). Panels A to E differ by the quantile that determines when the portfolio is to be reallocated. The sample period is July 1926 to August 2019. All alphas are annualized and stated in percent per year.

Panel A: 95.5% quantile										
	1 week	2 weeks	1 month	2 months	3 months	6 months	1 year	2 years	EWMA	GARCH
$\|\Delta w\|$	1.19	1.04	0.67	0.39	0.30	0.17	0.10	0.06	0.48	0.28
Default	-0.16	4.12	3.12	3.61	3.60	3.80	4.22	3.96	2.63	1.91
1 bps	-0.30	3.99	3.04	3.56	3.57	3.78	4.20	3.95	2.57	1.88
10 bps	-1.59	2.86	2.31	3.13	3.24	3.59	4.10	3.89	2.05	1.58
14 bps	-2.16	2.36	1.99	2.94	3.10	3.51	4.05	3.86	1.82	1.44
Panel B: 90% quantile										
	1 week	2 weeks	1 month	2 months	3 months	6 months	1 year	2 years	EWMA	GARCH
$\|\Delta w\|$	0.85	0.87	0.48	0.27	0.19	0.10	0.06	0.03	0.36	0.24
Default	-0.53	3.43	3.99	4.18	3.97	3.80	4.89	3.96	3.33	2.87
1 bps	-0.76	3.20	3.87	4.11	3.92	3.77	4.88	3.95	3.24	2.81
10 bps	-2.77	1.13	2.72	3.47	3.48	3.54	4.75	3.88	2.39	2.24
14 bps	-3.66	0.21	2.21	3.18	3.28	3.44	4.69	3.84	2.02	1.99
Panel C: 80% quantile										
	1 week	2 weeks	1 month	2 months	3 months	6 months	1 year	2 years	EWMA	GARCH
$\|\Delta w\|$	0.54	0.69	0.34	0.17	0.12	0.06	0.03	0.02	0.25	0.21
Default	-0.62	4.31	4.10	3.98	4.10	4.19	4.91	4.45	3.96	3.79
1 bps	-0.91	3.95	3.92	3.89	4.03	4.16	4.89	4.44	3.83	3.68
10 bps	-3.50	0.68	2.28	3.06	3.46	3.87	4.73	4.35	2.65	2.68
14 bps	-4.65	-0.77	1.55	2.70	3.20	3.74	4.66	4.31	2.12	2.24
Panel D: 50% quantile										
	1 week	2 weeks	1 month	2 months	3 months	6 months	1 year	2 years	EWMA	GARCH
$\|\Delta w\|$	0.29	0.41	0.19	0.10	0.07	0.03	0.02	0.01	0.16	0.16
Default	-0.33	5.26	4.23	4.69	4.57	4.64	5.12	4.79	4.52	5.03
1 bps	-0.70	4.72	3.97	4.56	4.48	4.60	5.10	4.78	4.32	4.81
10 bps	-4.09	-0.17	1.68	3.39	3.70	4.21	4.90	4.66	2.46	2.86
14 bps	-5.60	-2.35	0.67	2.87	3.34	4.04	4.81	4.62	1.64	1.99
Panel E: 0% quantile										
	1 week	2 weeks	1 month	2 months	3 months	6 months	1 year	2 years	EWMA	GARCH
$\|\Delta w\|$	0.15	0.23	0.11	0.06	0.04	0.02	0.01	0.01	0.10	0.12
Default	-0.31	5.31	4.17	4.71	4.39	4.50	4.96	4.54	4.58	4.91
1 bps	-0.71	4.70	3.88	4.56	4.29	4.45	4.92	4.52	4.31	4.59
10 bps	-4.27	-0.78	1.32	3.22	3.38	3.97	4.64	4.33	1.96	1.72
14 bps	-5.85	-3.21	0.18	2.62	2.97	3.75	4.52	4.25	0.91	0.44

To visualize the previously described pattern, we plot the alpha of selected strategies in Figure 5, depending on the average rebalancing period. For realized variance we exemplary show calculation periods of one month and one year.

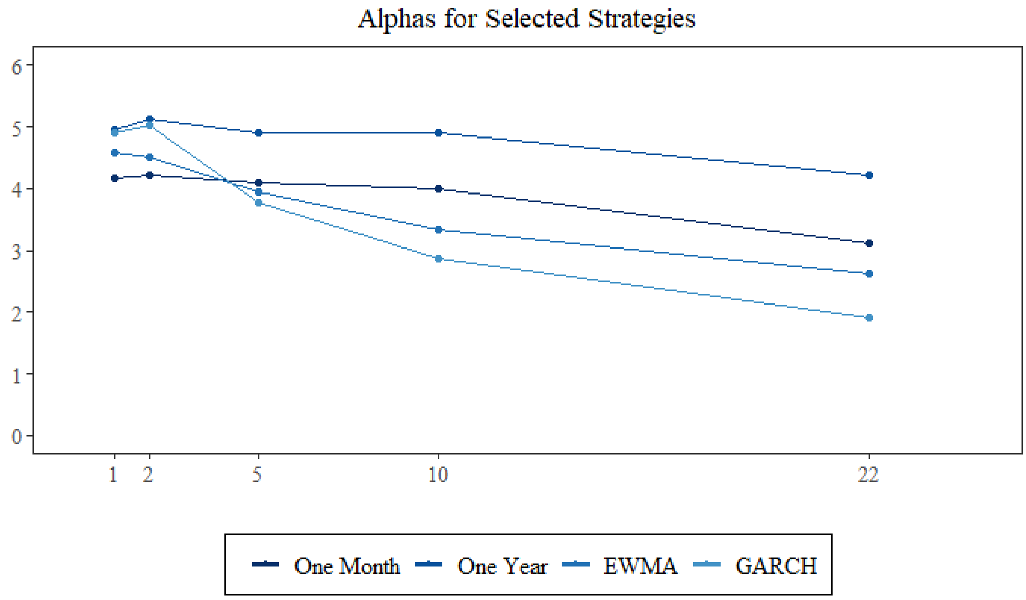


Figure 5: Alpha of Selected Strategies by Average Rebalancing Period

The plot shows the alphas of the strategy based on realized variance using estimation periods of one month and one year and the EWMA and GARCH strategy depending on the average rebalancing period.

As can be seen, the alpha of the strategies differs most for less frequent rebalancing. With more frequent rebalancing, the alphas tend to raise and converge up until an average rebalancing period of two days. Afterward, three out of four alphas decrease, EWMA being the exception.

For further reference, in Appendix J, we also plot the average absolute change in weights for the strategies based on realized variance, depending on the length of the calculation period. Furthermore, we plot the alpha of selected strategies, depending on transaction costs. The first graph showcases, that expanding the calculation period decreases the average absolute weight changes exponentially, due to the diminishing impact of dropping and adding one data point in the rolling window. This explains why the strategy based on a longer calculation period is less influenced by transaction costs. Furthermore, as can be seen in the second graph, decreasing the average trading frequency lowers the dependency on transaction costs.

Overall, Table 12 shows that the alpha tends to increase with more frequent rebalancing, with the length of the calculation period for the strategies based on realized variance and with lower transaction costs. In Appendix I, we also include tables of Appraisal Ratio and root mean square deviation for the same set of parameters. For the Appraisal Ratio, the pattern is similar, however, the increase in Appraisal Ratio with rebalancing length and calculation period is slightly dampened since the root mean squared error is also increasing with these statistics.

Concluding, the results are strongly dependent on the input parameters. The strategy most comparable to the original idea of Moreira of Muir, with an average rebalancing period of 22 days, based on realized variance estimated over one month, has considerably lower performance measures than the static case. On the other hand, increasing the rebalancing frequency to every two days, on average, and decreasing the calculation period length to two weeks, increases the alpha considerably to 5.3%, strongly outperforming the base idea. When considering transaction costs, the best strategy seems to be based on realized variance with a calculation period of one year, being largely independent of their inclusion.

6. Conclusion and Future Research

In our paper, we analyze how adjusting the two main presumptions of volatility-managed portfolios, based on the idea of Moreira and Muir (2017), alters the performance and applicability of the strategy for investors. Specifically, we approach the modification of the underlying approach by two levers: changing the method of variance estimation to more sophisticated models and adjusting the rebalancing period on a varying fixed and a flexible basis.

Answering these questions can be relevant to both practitioners and researchers. First, volatility-managed portfolios seem to generally showcase desirable properties for private and institutional investors due to their overall outperformance and ability to avoid market downturns during recessions. However, the implementation of Moreira and Muir (2017) suffers from tail risks and occasional outliers in return, leading to sudden large losses in selected periods. Furthermore, the strategy is susceptible to the inclusion of trading costs and leverage constraints. Therefore, improving the strategy with regards to these drawbacks would increase real-world applicability.

Second, since the authors are able to explain the implementation of their idea from the vantage point of the mean-variance estimator, even simplifying the optimization problem by refraining from estimating expected returns, the showed outperformance poses a puzzle to accepted economic and financial theory. By analyzing the dependency of the strategy on its underlying assumptions, we can further the general understanding of volatility timing and help to assess its scope with regards to how universal and hence fundamental the original findings are.

Our first lever is to adjust the method of variance estimation to more sophisticated methods, namely an Exponential Smoothing Average, an Autoregressive Integrated Moving Average, and a General Autoregressive Heteroskedasticity model. Based on monthly rebalancing, as in the original paper, we show that all strategies except GARCH are similar in their performance in terms of alpha, with GARCH slightly falling behind. However, while the standard deviation is by construction the same, the alternative strategies seem to take on a higher systematic risk in exchange for a lower idiosyncratic risk, thereby increasing their Appraisal Ratio and Sharpe Ratio in contrast to the realized variance strategy. This is confirmed by the average backtested return and cumulative performance. While the overall course over time is similar, all the alternative strategies would have yielded a higher final value when hypothetically applied by an investor.

Looking at the distribution of returns indicates that the original strategy is less dispersed around the mean, but has a greater number of extreme outliers, all in all leading to the same standard deviation. These extreme outliers suggest a greater tail risk and make the periodical performance less predictable for the individual investor, decreasing her utility. This notion is also reflected within the respective distribution of weights.

Furthermore, we consider the implementation of trading costs and leverage constraints to assess real-world applicability. For trading costs, the difference is striking, especially the implementations of ESA and GARCH are almost unsusceptible regarding their inclusion. This can be particularly relevant for smaller investors, on average facing higher transaction costs. Regarding leverage constraints, all strategies behave similar, in a sense that more severe constraints drive down the respective alpha. We also assess the performance of the alternative strategies by regressing their returns on the market, controlling for the return of the original strategy, to elicit relative outperformance. For the case of no transaction costs, this eradicates the alpha, implying that in terms of performance, the alternative strategies do not offer a clear-cut advantage. However, when including transaction costs, they are actually able to generate a significant alpha above the base strategy that increases with the magnitude of costs.

To better understand the overall performance over time, we analyze how the strategies fare during the business cycle, i.e. during recessionary and expansionary periods. All implementations are able to reduce the exposure during the former periods and increase the exposure during the latter ones. Furthermore, accounting for cycles makes the alpha insignificant, implying that the outperformance might not originate in granularly responding to changes in estimated variance, but rather in generally adjusting the exposure based on overall market trends. Our subsample analysis goes in a similar direction, showing that the strategies especially do well in times of crises and sustained market downturns, but offer limited advantages in overall stable times.

Lastly, we compare how the different implementations of volatility-managed portfolios fare specifically in times when the original one fails to adequately respond to market changes, i.e. in periods of considerable underperformance. Here, the alternative strategies act preferably, rightfully indicating lower weights and behaving more independent from the original strategy and the market. However, one has to be cautious about potential biases through the definition of non-performing periods.

Considering our second lever, changing the rebalancing frequency, the effect is not clear-cut. For shorter fixed periods, the alpha seems to be higher, but the overall volatility of performance depending on period length is considerable. While all alphas are, nevertheless, positive, this indicates the originally reported magnitude of alphas cannot necessarily be taken at face value. While monthly rebalancing is intuitively appealing, it is an exemplary assumption that influences the size of alpha, even though this does not question the existence or significance of it. As expected, the impact of transaction costs increases with more frequent rebalancing, even though ESA and GARCH are still largely invariant of their inclusion. The impact of considering flexible periods is mixed. Depending on the assumptions, it increases or decreases the overall alpha, ranging from negative values to the highest alpha reported in our paper. In general, it seems that for relatively longer average rebalancing periods of two weeks and one month, flexible periods do not offer an advantage. However, for an on average more frequent rebalancing, they do increase the alpha. Furthermore, even for more frequent trading, selected strategies are almost invariant to transaction costs.

All in all, we show, that volatility-managed portfolios produce a significant alpha, independent of the specific underlying assumptions. However, altering them does change the specific performance, even though it is difficult to infer recommendations for investors since the differences are relatively small and our analysis focuses on in-sample estimation. Nevertheless, when estimating the variance, it seems generally sensible to take more past values into account than only the realized variance of last month, while emphasizing the influence of more recent observation. This reduces the number of extreme observations and thereby extreme returns and decreases transaction costs through more subtle weight adjustments.

For future research, it especially seems valuable to further broaden the analyzed scope and apply the ideas of volatility-managed portfolios under varying assumptions to markets outside the US. Furthermore, while we tried to implement a range of variance estimation techniques, they only represent a subset of all possible methods. Therefore, the selection could be extended, for example, considering option implied volatility or stochastic volatility models. Moreover, to further the assessment of real-time applicability, the analysis can be repeated using out-of-sample parameter estimation. Lastly, it could be meaningful to examine more closely how business cycle timing affects volatility timing. The fact that it diminishes the alpha could imply that it is not necessary to adjust the risk exposure as refined as with the inverse of the estimated variance, but rather sufficient to alternate between a higher weight for expected expansionary periods and a lower weight for expected recessionary periods.

7. References

Akaike, H., 1974, A new look at the statistical model identification, Selected Papers of Hirotugu Akaike, 215-222.

Andersen, Torben G. and Bollerslev, Tim, 1998, Answering the skeptics: Yes, standard volatility models do provide accurate forecasts, International economic review 885-905.

Arditti, Fred D., 1967, Risk and the required return on equity, The Journal of Finance 22(1), 19-36.

Baker, H. Kent, Filbeck, Greg and Kiymaz, Halil, 2015, Mutual funds and exchange-traded funds: building blocks to wealth, (Oxford University Press, .

Bodie, Zvi, Kane, Alex and Marcus, Alan J., 2018, Investments, (McGraw-Hill Education, New York, NY).

Bollerslev, Tim, Chou, Ray Y. and Kroner, Kenneth F., 1992, ARCH modeling in finance: A review of the theory and empirical evidence, Journal of Econometrics 52(1-2), 5-59.

Bollerslev, Tim, Engle, Robert F. and Nelson, Daniel B., 1994, ARCH models, Handbook of econometrics 42959-3038.

Brooks, Chris, 2019, Introductory econometrics for finance, Fourth edition edn (Cambridge University Press, Cambridge).

Buffet, Warren E., Oct 16, 2008, Buy American. I Am. Available: https://www.nytimes.com/2008/10/17/opinion/17buffett.html [Dec 03, 2019].

Campbell, John Y., 1987, Stock returns and the term structure, Journal of Financial Economics 18(2), 373-399.

Campbell, John Y. and Shiller, Robert J., 1988, The dividend-price ratio and expectations of future dividends and discount factors, The Review of Financial Studies 1(3), 195-228.

Copeland, Maggie M. and Copeland, Thomas E., 1999, Market timing: Style and size rotation using the VIX, Financial Analysts Journal 55(2), 73-81.

Dickey, David A. and Fuller, Wayne A., 1981, Likelihood ratio statistics for autoregressive time series with a unit root, Econometrica: journal of the Econometric Society 1057-1072.

Ding, Zhuanxin and Granger, Clive W., 1996, Modeling volatility persistence of speculative returns: a new approach, Journal of Econometrics 73(1), 185-215.

Fama, Eugene F. and French, Kenneth R., 1996, Multifactor explanations of asset pricing anomalies, The journal of finance 51(1), 55-84.

Fama, Eugene F. and French, Kenneth R., 1989, Business conditions and expected returns on stocks and bonds, Journal of Financial Economics 25(1), 23-49.

Fama, Eugene F. and French, Kenneth R., 1988, Dividend yields and expected stock returns, Journal of Financial Economics 22(1), 3-25.

Ferson, Wayne E., Sarkissian, Sergei and Simin, Timothy T., 2003, Spurious regressions in financial economics? The Journal of Finance 58(4), 1393-1413.

Fisher, Kenneth L. and Statman, Meir, 1997, The Mean–Variance-optimization puzzle: Security portfolios and food portfolios, Financial Analysts Journal 53(4), 41-50.

Fleming, Jeff, Kirby, Chris and Ostdiek, Barbara, 2003, The economic value of volatility timing using “realized” volatility, Journal of Financial Economics 67(3), 473-509.

Fleming, Jeff, Kirby, Chris and Ostdiek, Barbara, 2001, The economic value of volatility timing, The Journal of Finance 56(1), 329-352.

Frazzini, Andrea, Israel, Ronen and Moskowitz, Tobias J., 2012, Trading costs of asset pricing anomalies, Fama-Miller working paper 14-05.

FRED, 2019, NBER based Recession Indicators for the United States from the Period following the Peak through the Trough. Available: https://fred.stlouisfed.org/series/USREC [Dec 02, 2019].

French, Kenneth R., 2019, Data Library. Available: https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html [Sep 21, 2019].

French, Kenneth R., Schwert, G. W. and Stambaugh, Robert F., 1987, Expected stock returns and volatility, Journal of Financial Economics 19(1), 3-29.

Goodwin, Thomas H., 1998, The information ratio, Financial Analysts Journal 54(4), 34-43.

Green, Richard C. and Hollifield, Burton, 1992, When will mean-variance efficient portfolios be well diversified? The Journal of Finance 47(5), 1785-1809.

Hull, John C., 2018, Options, Futures and Other Derivatives, (Pearson Education, Harlow, Essex).

Keim, Donald B. and Stambaugh, Robert F., 1986, Predicting returns in the stock and bond markets, Journal of Financial Economics 17(2), 357-390.

Kuha, Jouni, 2004, AIC and BIC: Comparisons of assumptions and performance, Sociological methods & research 33(2), 188-229.

Kwiatkowski, Denis, Phillips, Peter C., Schmidt, Peter and Shin, Yongcheol, 1992, Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? Journal of Econometrics 54(1-3), 159-178.

Markowitz, Harry, 1952, Portfolio selection, The journal of finance 7(1), 77-91.

Merton, Robert C., 1980, On estimating the expected return on the market: An exploratory investigation, Journal of Financial Economics 8(4), 323-361.

Michaud, Richard O., 1989, The Markowitz optimization enigma: Is ‘optimized’optimal? Financial Analysts Journal 45(1), 31-42.

Moreira, Alan and Muir, Tyler, 2019, Should Long-Term Investors Time Volatility? Journal of Financial Economics 131(3), 507-527.

Moreira, Alan and Muir, Tyler, 2017, Volatility-Managed Portfolios, The Journal of Finance 72(4), 1611-1644.

Muir, Tyler, 2017, Financial crises and risk premia, The Quarterly Journal of Economics 132(2), 765-809.

NBER, 2019, US Business Cycle Expansions and Contractions [Dec 04, 2019].

Nelder, John A. and Mead, Roger, 1965, A simplex method for function minimization, The computer journal 7(4), 308-313.

Phillips, Peter C. and Perron, Pierre, 1988, Testing for a unit root in time series regression, Biometrika 75(2), 335-346.

Pontiff, Jeffrey and Schall, Lawrence D., 1998, Book-to-market ratios as predictors of market returns, Journal of Financial Economics 49(2), 141-160.

Poon, Ser-Huang and Granger, Clive W., 2003, Forecasting volatility in financial markets: A review, Journal of economic literature 41(2), 478-539.

Poterba, James M. and Summers, Lawrence H., 1984, The persistence of volatility and stock market fluctuations, Working Paper No. 1462, National Bureau of Economic Research Cambridge, Mass., USA.

Rosenberg, Alex, Mar 23, 2016, When markets get scary, panicking is smart: Yale profs. Available: https://www.cnbc.com/2016/03/23/when-markets-get-scary-panicking-is-smart-yale-profs.html [Dec 03, 2019].

Rozeff, Michael S., 1984, Dividend yields are equity risk premiums, Journal of Portfolio management 68-75.

Schwarz, Gideon, 1978, Estimating the dimension of a model, The annals of statistics 6(2), 461-464.

Stambaugh, Robert F., 1999, Predictive regressions, Journal of Financial Economics 54(3), 375-421.

Thomas, Ric and Shapiro, Robert, 2009, Managed volatility: A new approach to equity investing, The Journal of Investing 18(1), 15-23.

Treynor, Jack L. and Black, Fischer, 1973, How to use security analysis to improve portfolio selection, The journal of business 46(1), 66-86.

Vanguard, 2015, What to do during market volatility? Perhaps nothing. Available: https://personal.vanguard.com/us/insights/guide/MARKET-VOLATILITY [Dec 03, 2019].

Welch, Ivo and Goyal, Amit, 2007, A comprehensive look at the empirical performance of equity premium prediction, The Review of Financial Studies 21(4), 1455-1508.

Wooldridge, Jeffrey M., 2018, Introductory Econometrics, Seventh edition edn (Cengage Learning, Boston, MA).

8. Appendices

Appendix A: Acf and Pacf of Differenced Monthly Realized Variance

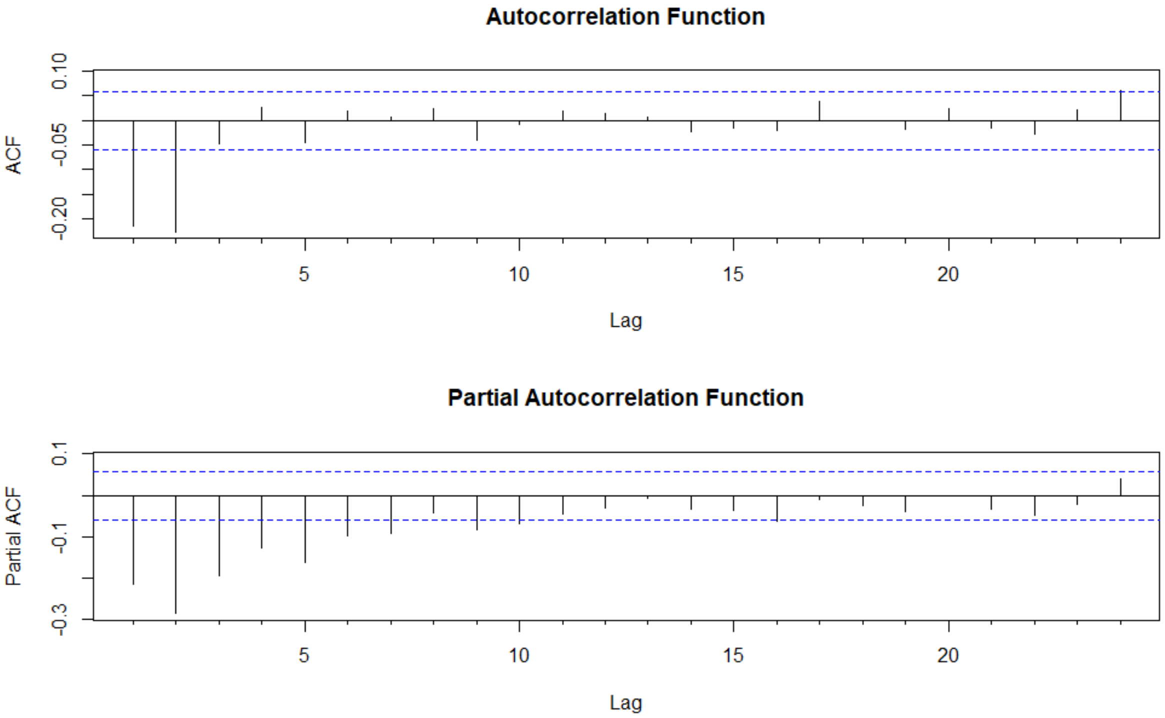


Figure 6: ACF and PACF of Time Series of Differenced Monthly Realized Variance

The plot shows the autocorrelation function and the partial autocorrelation function for the time series of the differenced monthly realized variance for a lag of up to 24 months.

Appendix B: Exemplary Periods of High Volatility of Volatility

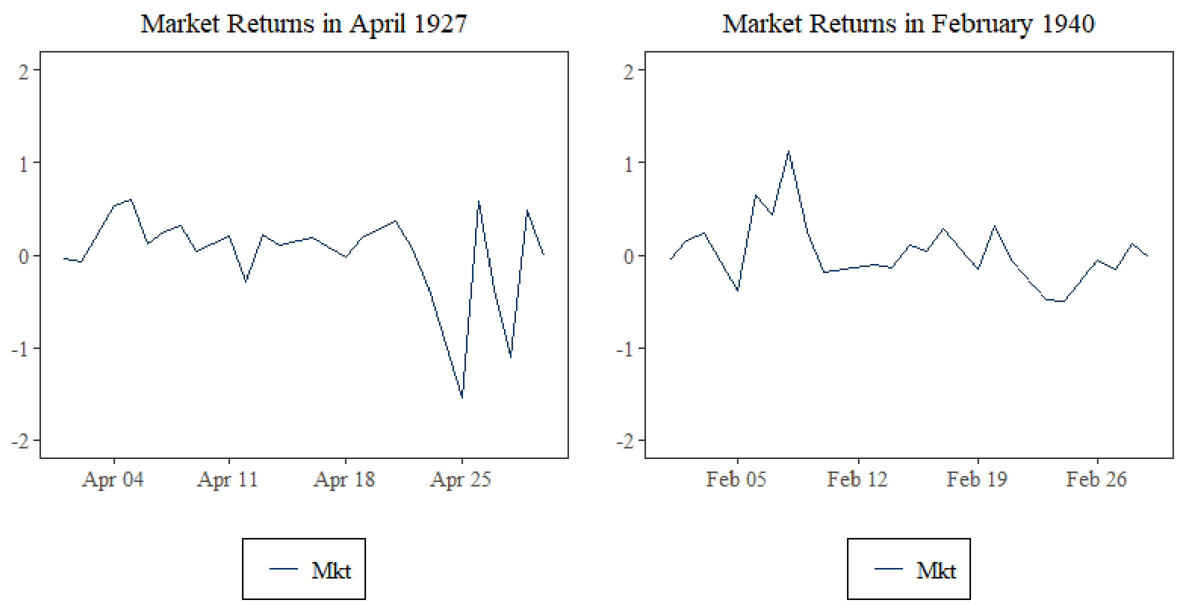


Figure 7: Example of Months with Irregularly Distributed Variance

The left panel plots the market returns for the month of April 1927. The right panel depicts the market returns for the month of February 1940.

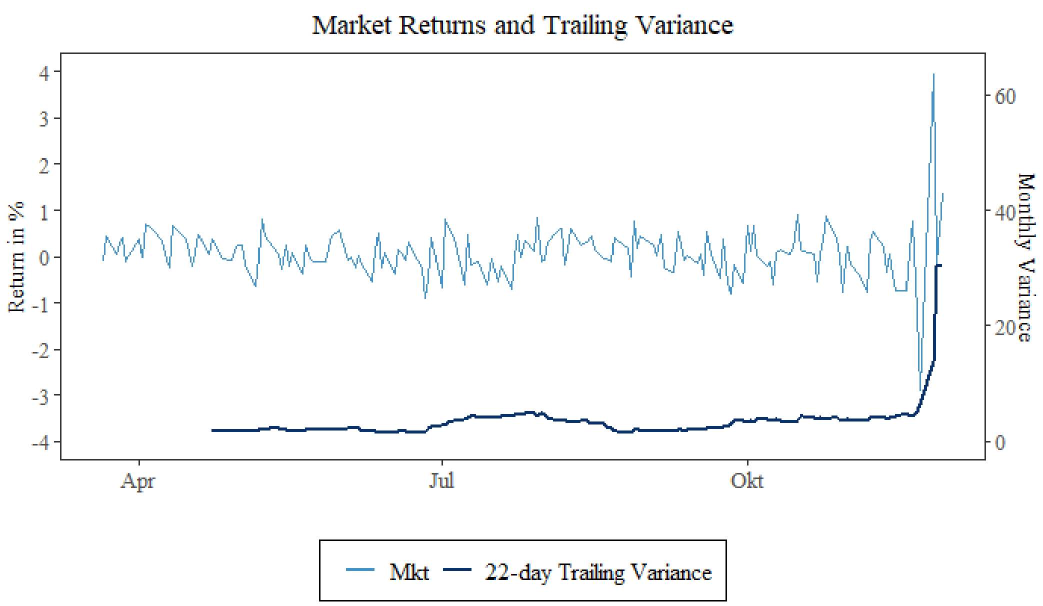


Figure 8: Example of Sudden Volatility Spikes

The plot shows the market returns from March 21st, 1963 to November 29th, 1963 as well as the 22-day trailing variance. The left y-axis depicts returns in percent and the left y-axis depicts monthly variance.

Appendix C: Fama French Three Factors Regression

Table 13: Performance Controlling for Fama French Three Factors.

In this table we run time-series regressions of volatility-managed returns on the Fama French three factors: $f_t^\sigma = \alpha + \beta_0 f_t + \beta_1 SMB_t + \beta_2 HML_t + \epsilon_t$. The managed returns, f^σ, are calculated by scaling the market excess returns by the inverse of the variance estimate. The methods used for variance estimation are Realized Variance (1), ARIMA (2), ESA (3), and GARCH (4). The data is monthly and the sample period ranges from July 1926 to August 2019. Standard errors are in parentheses and adjust for heteroskedasticity. All factors are annualized in percent per year by multiplying monthly factors by 12. Additionally, we report the Sharpe ratio of volatility-managed returns, the Appraisal Ratio and the Fama French three factor alpha. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

	(1) Var	(2) ARIMA	(3) ESA	(4) GARCH
Mkt-RF	0.64***	0.74***	0.76***	0.85***
	(0.05)	(0.05)	(0.05)	(0.04)
SMB	-0.05	-0.04	-0.02	-0.01
	(0.05)	(0.05)	(0.05)	(0.05)
HML	-0.15**	-0.23***	-0.26***	-0.32***
	(0.06)	(0.06)	(0.06)	(0.06)
Alpha (α)	5.19***	5.40***	5.08***	4.31***
	(1.50)	(1.35)	(1.30)	(1.09)
Observations	1,117	1,117	1,117	1,117
R^2	0.38	0.51	0.54	0.68

Appendix D: Drawdown

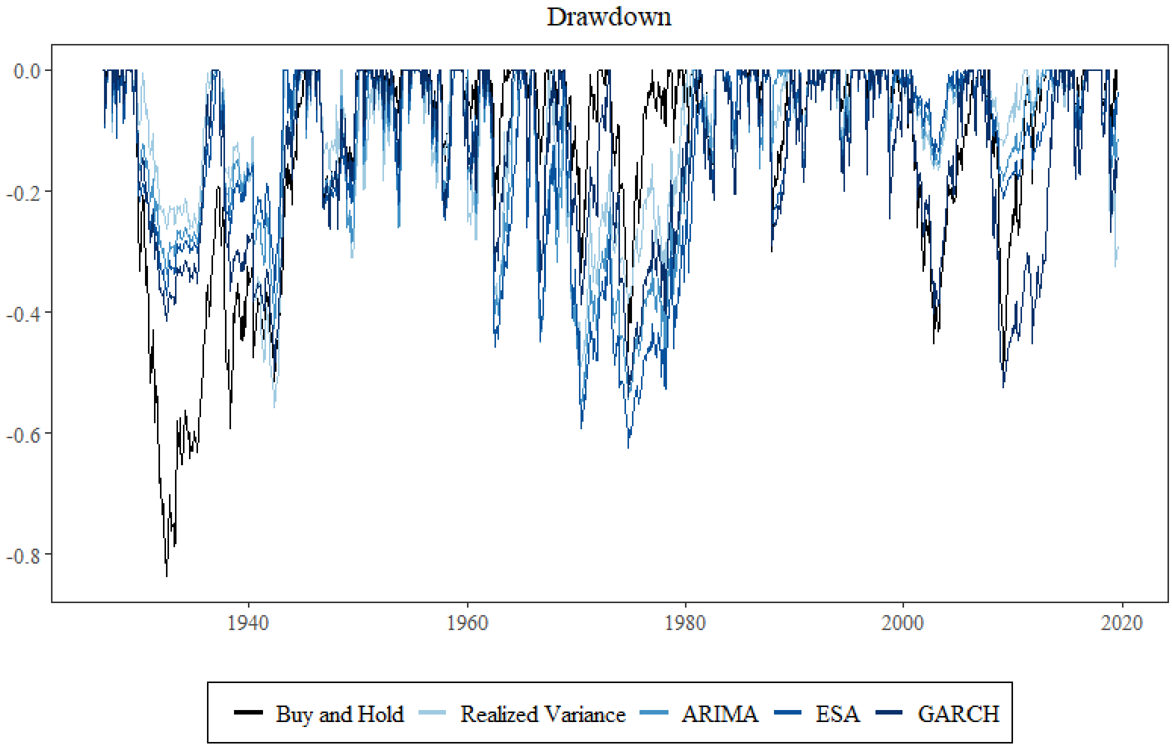
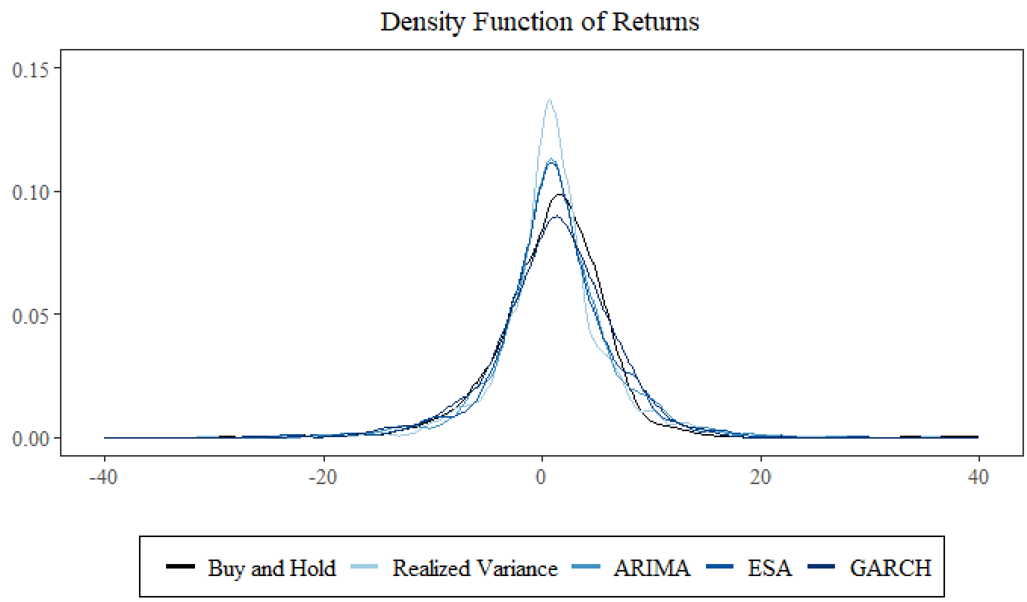


Figure 9: Drawdown

The plot depicts the historical drawdowns for the Buy-and-Hold portfolio, Moreira and Muir's realized variance strategy and the ARIMA, ESA, and GARCH strategies. The drawdown is defined based on a hypothetical investment of $1 in 1926 and depicts, if negative, the percentage change from the historical peak to the current value of the investment. Otherwise, it is zero. All strategies have the same unconditional monthly standard deviation.

Appendix E: Density Functions of Returns and Weights



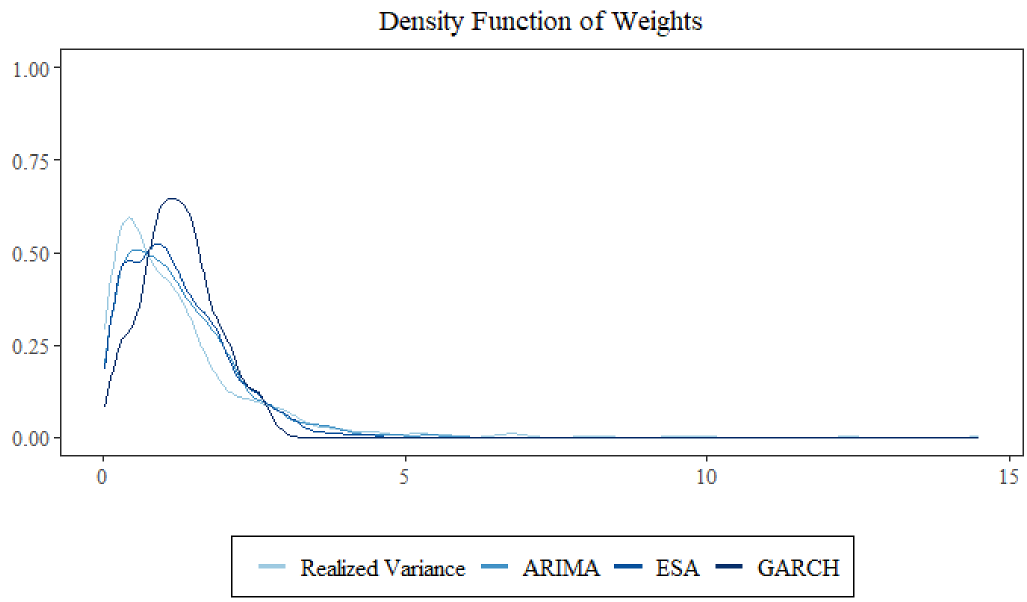


Figure 10: Density Functions of Returns and Weights

The top panel plots the density functions of returns of the Buy-and-Hold portfolio, Moreira and Muir's realized variance strategy and the ARIMA, EWMA, and GARCH strategies. The lower panel plot depicts the density functions of weights for the same strategies. All strategies have the same unconditional monthly standard deviation.

Appendix F: Derivation of Regression Beta Approximately Equalling Correlation

Our regression beta is defined as follows:

$$\beta = \frac{Cov(f_{t+1}^{\sigma}, f_{t+1})}{Var(f_{t+1})} = \frac{\rho(f_{t+1}^{\sigma}, f_{t+1})\sigma(f_{t+1}^{\sigma})\sigma(f_{t+1})}{\sigma^2(f_{t+1})} = \frac{\rho(f_{t+1}^{\sigma}, f_{t+1})\sigma(f_{t+1}^{\sigma})}{\sigma(f_{t+1})}$$

By incorporating c, we set:

$$\sigma\left(f_{t+1}^{\sigma} + r_f\right) = \sigma\left(f_{t+1} + r_f\right)$$

Based on the notion of almost zero covariance, the equation approximately becomes:

$$\sigma(f_{t+1}^{\sigma}) \approx \sigma(f_{t+1})$$

Therefore, the regression beta approximately equals:

$$\beta \approx \rho(f_{t+1}^{\sigma}, f_{t+1}).$$

Appendix G: Appraisal Ratio Depending on Holding Period

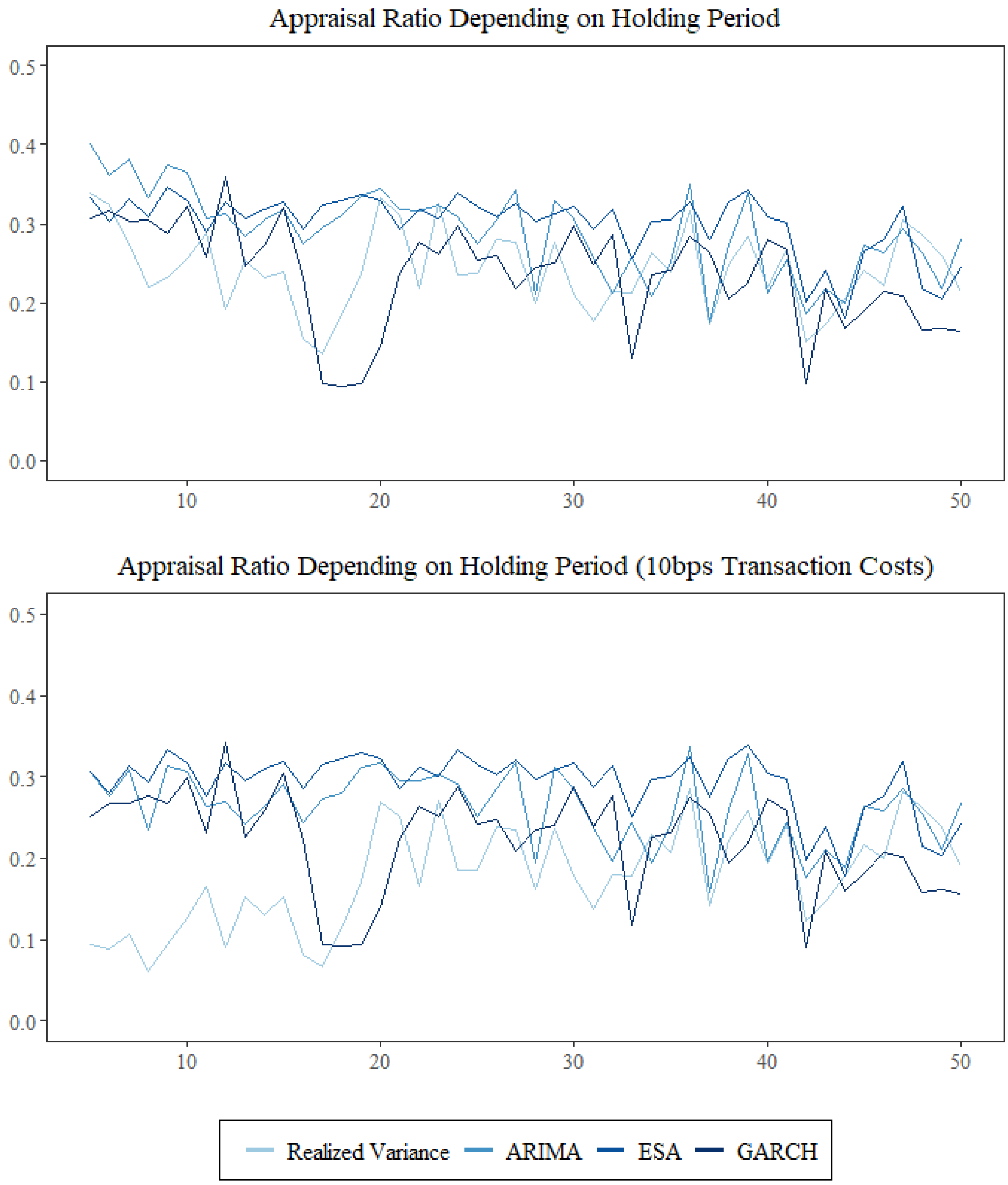


Figure 11: Appraisal Ratio Depending on Holding Period

The top panel plots the Appraisal Ratios of the realized variance, ARIMA, ESA, and GARCH strategy depending on the rebalancing period with values between five and 50 trading days. The bottom panel shows the Appraisal Ratios for the respective strategies depending on the rebalancing period when accounting for transaction costs of 10bps. Alphas are annualized and stated in percent per year.

Appendix H: Cumulative Performance Depending on Holding Period

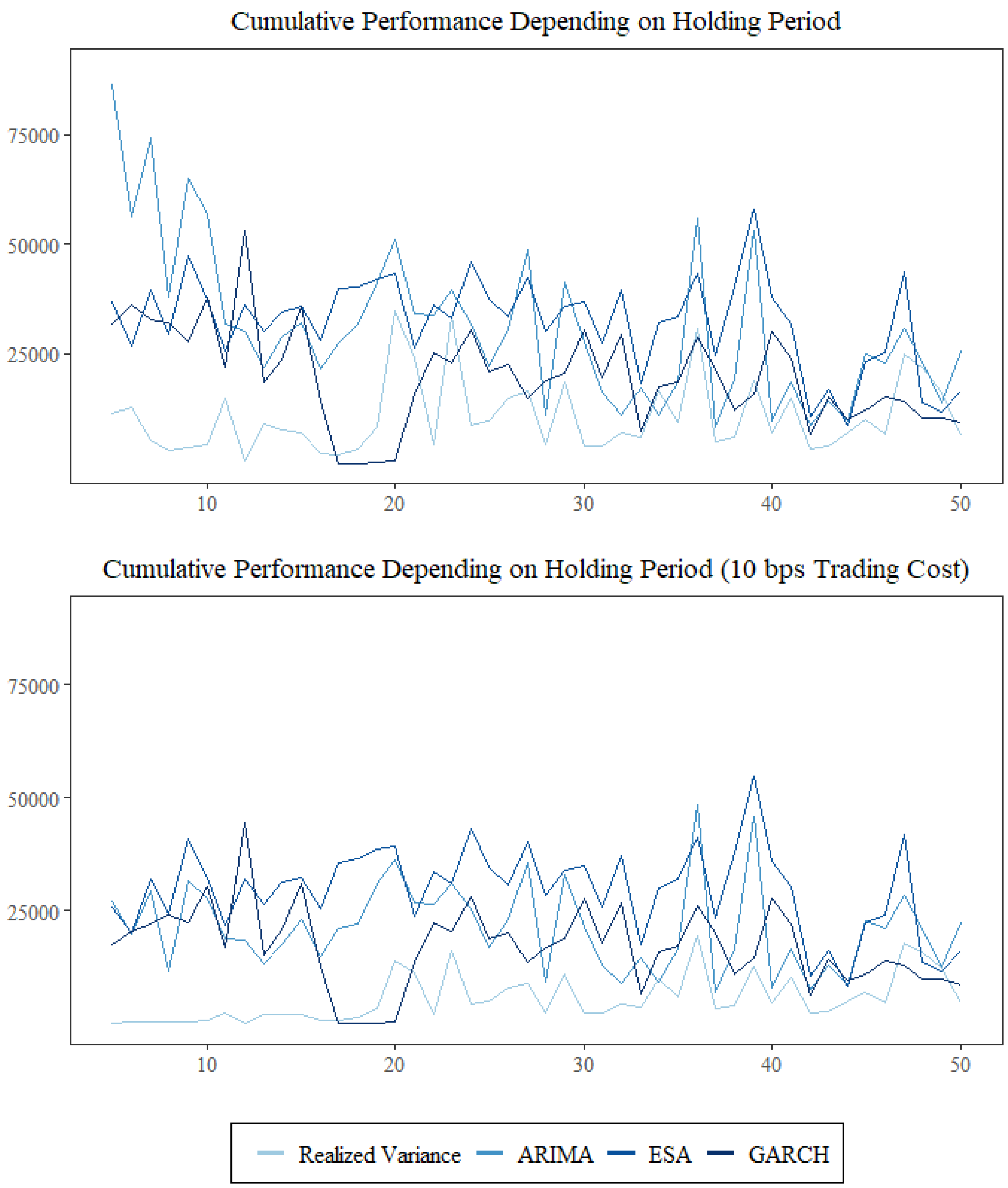


Figure 12: Cumulative Performance Depending on Holding Period

The top panel plots the cumulative performance of the realized variance, ARIMA, ESA, and GARCH strategy depending on the rebalancing period with values between five and 50 trading days. The bottom panel shows the cumulative performance of the respective performance depending on the period length when accounting for transaction costs 10bps. Alphas are annualized and stated in percent.

Appendix I: Appraisal Ratio and RMSE for Flexible Time Periods

Table 14: Appraisal Ratios for Flexible Rebalancing Periods.

We run time-series regressions of volatility-managed returns on the Fama-French market factor for all strategies with flexible time periods. We show Appraisal Ratios for the default case and when including transaction costs of 1bps, 10bps, and 14bps, respectively. Further, we report the average absolute change in weights ($|\Delta w|$). Panels A to E differ by the quantile that determines when the portfolio is to be reallocated. The sample period is July 1926 to August 2019 for Mkt.

	1 week	2 weeks	1 month	2 months	3 months	6 months	1 year	2 years	EWMA	GARCH
Panel A: 95.5% quantile										
Default	-0.01	0.23	0.19	0.26	0.25	0.26	0.30	0.27	0.19	0.15
1 bps	-0.02	0.23	0.19	0.26	0.24	0.26	0.30	0.27	0.19	0.15
10 bps	-0.08	0.16	0.14	0.23	0.22	0.25	0.30	0.27	0.15	0.13
14 bps	-0.11	0.13	0.12	0.21	0.21	0.24	0.29	0.26	0.13	0.12
Panel B: 90% quantile										
	1 week	2 weeks	1 month	2 months	3 months	6 months	1 year	2 years	EWMA	GARCH
Default	-0.03	0.23	0.26	0.31	0.29	0.29	0.34	0.31	0.23	0.23
1 bps	-0.04	0.21	0.25	0.30	0.28	0.28	0.34	0.31	0.22	0.23
10 bps	-0.14	0.07	0.18	0.25	0.25	0.27	0.33	0.31	0.17	0.18
14 bps	-0.18	0.01	0.15	0.23	0.24	0.26	0.33	0.30	0.14	0.16
Panel C: 80% quantile										
	1 week	2 weeks	1 month	2 months	3 months	6 months	1 year	2 years	EWMA	GARCH
Default	-0.04	0.29	0.27	0.29	0.30	0.31	0.37	0.36	0.29	0.32
1 bps	-0.05	0.26	0.26	0.29	0.30	0.31	0.37	0.36	0.28	0.31
10 bps	-0.19	0.05	0.15	0.23	0.25	0.29	0.36	0.35	0.20	0.22
14 bps	-0.25	-0.05	0.10	0.20	0.24	0.28	0.35	0.35	0.16	0.19
Panel D: 50% quantile										
	1 week	2 weeks	1 month	2 months	3 months	6 months	1 year	2 years	EWMA	GARCH
Default	-0.02	0.35	0.29	0.34	0.34	0.36	0.41	0.41	0.33	0.40
1 bps	-0.04	0.31	0.27	0.33	0.33	0.35	0.41	0.41	0.31	0.38
10 bps	-0.22	-0.01	0.11	0.25	0.27	0.32	0.39	0.40	0.18	0.23
14 bps	-0.30	-0.15	0.05	0.21	0.25	0.31	0.39	0.39	0.12	0.16
Panel E: 0% quantile										
	1 week	2 weeks	1 month	2 months	3 months	6 months	1 year	2 years	EWMA	GARCH
Default	-0.02	0.35	0.29	0.35	0.33	0.35	0.40	0.38	0.33	0.39
1 bps	-0.04	0.31	0.27	0.33	0.32	0.34	0.40	0.38	0.31	0.36
10 bps	-0.24	-0.05	0.09	0.23	0.25	0.30	0.37	0.36	0.14	0.13
14 bps	-0.32	-0.21	0.01	0.19	0.22	0.29	0.36	0.35	0.07	0.03

Table 15: Root Mean Squared Errors for Flexible Rebalancing Periods.

We run time-series regressions of volatility-managed returns on the Fama-French market factor for all strategies with flexible time periods. We show Root Mean Square Errors for the default case and when including transaction costs of 1bps, 10bps, and 14bps, respectively. Further, we report the average absolute change in weights ($|\Delta w|$). Panels A to E differ by the quantile that determines when the portfolio is to be reallocated. The sample period is July 1926 to August 2019 for Mkt.

	1 week	2 weeks	1 month	2 months	3 months	6 months	1 year	2 years	EWMA	GARCH
Panel A: 95.5% quantile										
Default	66.93	61.07	56.68	48.00	50.84	50.01	48.08	50.58	46.95	43.34
1 bps	67.19	61.06	56.68	47.99	50.84	50.01	48.08	50.58	46.95	43.33
10 bps	69.69	61.01	56.66	47.97	50.83	50.00	48.08	50.58	46.88	43.28
14 bps	70.87	61.00	56.66	47.97	50.83	50.00	48.08	50.58	46.85	43.26
Panel B: 90% quantile										
	1 week	2 weeks	1 month	2 months	3 months	6 months	1 year	2 years	EWMA	GARCH
Default	96.09	77.35	78.12	70.19	70.97	68.42	73.10	65.02	74.06	64.01
1 bps	96.50	77.35	78.11	70.19	70.97	68.42	73.10	65.01	74.05	64.00
10 bps	100.37	77.49	78.04	70.17	70.93	68.41	73.09	65.01	74.03	63.91
14 bps	102.21	77.59	78.02	70.16	70.92	68.41	73.09	65.01	74.03	63.87
Panel C: 80% quantile										
	1 week	2 weeks	1 month	2 months	3 months	6 months	1 year	2 years	EWMA	GARCH
Default	127.83	108.41	110.33	98.40	98.86	98.09	95.96	90.28	98.41	87.26
1 bps	128.39	108.40	110.32	98.39	98.85	98.08	95.96	90.28	98.40	87.25
10 bps	133.73	108.45	110.29	98.36	98.83	98.07	95.94	90.27	98.37	87.15
14 bps	136.26	108.56	110.29	98.35	98.82	98.06	95.94	90.27	98.36	87.11
Panel D: 50% quantile										
	1 week	2 weeks	1 month	2 months	3 months	6 months	1 year	2 years	EWMA	GARCH
Default	201.94	174.09	168.55	156.62	155.37	148.94	143.08	135.18	158.72	144.15
1 bps	202.85	174.08	168.55	156.61	155.37	148.94	143.07	135.18	158.72	144.15
10 bps	211.49	174.25	168.62	156.57	155.38	148.95	143.06	135.17	158.75	144.11
14 bps	215.59	174.48	168.68	156.56	155.39	148.95	143.06	135.17	158.78	144.10
Panel E: 0% quantile										
	1 week	2 weeks	1 month	2 months	3 months	6 months	1 year	2 years	EWMA	GARCH
Default	279.01	243.93	236.13	221.52	217.29	210.06	202.41	192.58	224.31	207.01
1 bps	280.29	243.98	236.20	221.61	217.39	210.19	202.55	192.75	224.36	207.02
10 bps	292.43	244.84	236.96	222.49	218.46	211.47	204.04	194.42	224.90	207.14
14 bps	298.19	245.51	237.38	222.96	219.00	212.12	204.80	195.27	225.19	207.22

Appendix J: Further Visualizations for Flexible Rebalancing Periods

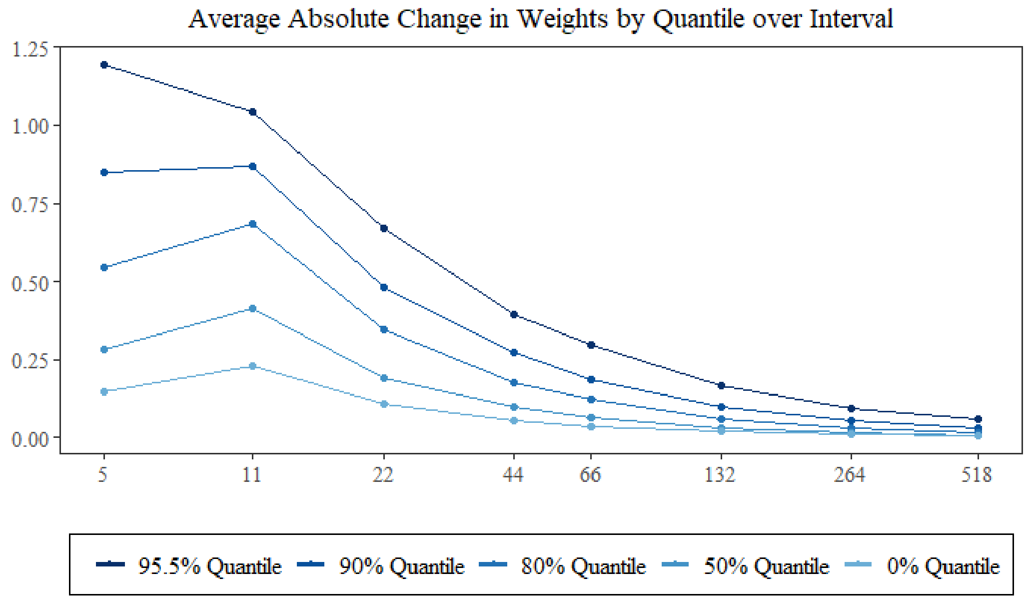


Figure 13: Average Absolute Change in Weights Depending on Estimation Period

The plot shows the average absolute change in weights for the strategy based on realized variance by average rebalancing period over the length of the estimation period. For ease of interpretation, the x-axis is depicted in log-scale.

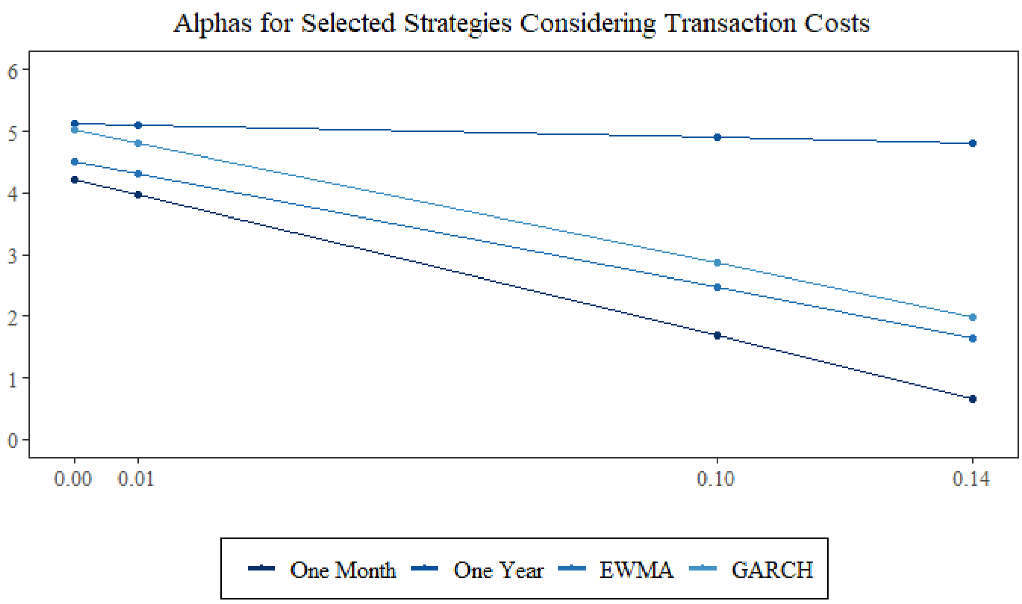


Figure 14: Alphas for Selected Strategies Considering Transaction Costs

The plot depicts the alphas of the realized variance strategy based on estimation periods of one month and one year and the alphas of the EWMA and GARCH strategy under the assumption of an average rebalancing every two days and accounting for transaction costs of 1bps, 10bps, and 14bps.