

Cirulli, Antonello; De Nard, Gianluca; Traut, Joshua; Walker, Patrick

Working Paper

Low risk, high variability: Practical guide for portfolio construction

Working Paper, No. 463

Provided in Cooperation with:

Department of Economics, University of Zurich

Suggested Citation: Cirulli, Antonello; De Nard, Gianluca; Traut, Joshua; Walker, Patrick (2025) : Low risk, high variability: Practical guide for portfolio construction, Working Paper, No. 463, University of Zurich, Department of Economics, Zurich, <https://doi.org/10.5167/uzh-268927>

This Version is available at:

<https://hdl.handle.net/10419/313097>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



**University of
Zurich**^{UZH}

University of Zurich
Department of Economics

Working Paper Series
ISSN 1664-7041 (print)
ISSN 1664-705X (online)

Working Paper No. 463

Low Risk, High Variability: Practical Guide for Portfolio Construction

Antonello Cirulli, Gianluca De Nard, Joshua Traut and Patrick Walker

January 2025

Low Risk, High Variability: Practical Guide for Portfolio Construction

Antonello Cirulli ¹ Gianluca De Nard ² Joshua Traut ^{3 *} Patrick Walker ⁴

January 20, 2025

Abstract

The low-risk anomaly challenges traditional financial theory by stating that less volatile stocks generate higher risk-adjusted returns. This paper explores how various portfolio construction choices influence the performance of low-risk portfolios. We show that methodological decisions critically influence portfolio outcomes, causing substantial dispersion in performance metrics across weighting schemes and risk estimators. This can only be marginally mitigated by incorporating constraints such as short-sale restrictions and size or price filters. Our analysis reveals that volatility-based estimators yield the most favorable performance distribution, outperforming beta-based approaches. Transaction costs are found to significantly affect performance and are vitally important in identifying the most attractive portfolios, highlighting the importance of realistic implementation constraints. Through rigorous empirical analysis, this study bridges the gap between theoretical insights and practical applications, offering actionable guidance to investors. The findings advocate for a cautious approach to nonstandard errors in portfolio modeling and emphasize the necessity of robust strategies in low-risk investing.

Keywords: Low-Risk Investing, Methodology, Nonstandard Errors, Portfolio Construction.

JEL Classification: C52, G11, G12, G15, G17.

¹OLZ AG, Zurich, Switzerland | antonello.cirulli@olz.ch

²Department of Economics, University of Zurich, Switzerland; Volatility and Risk Institute, New York University, United States & OLZ AG, Zurich, Switzerland | gianluca.denard@econ.uzh.ch | <https://denard.ch>

³Swiss Institute of Banking and Finance, University of St. Gallen, Switzerland & OLZ AG, Zurich, Switzerland | joshua.traut@unisg.ch | <https://www.joshuatraut.com>

⁴Department of Finance, University of Zurich, Switzerland & OLZ AG, Zurich, Switzerland | patrick.walker@olz.ch

*Corresponding author: Joshua Traut. We are grateful for insightful comments from David Anderson, Merlin Bartel, Damjan Kostovic and Michael Wolf. All errors are our own.

1 Introduction

Methodological uncertainty has emerged as a critical topic in quantitative finance, particularly in the study of portfolio construction and performance evaluation. The growing complexity of financial datasets, the large number of reported factors that appear relevant (Cochrane, 2011), the p-hacking problem (Harvey, 2017), and the proliferation of portfolio design choices have led to concerns about the robustness and replicability of research findings (e.g., see Hou et al., 2020; Jensen et al., 2023). These issues are especially relevant in the context of factor investing, where methodological choices can significantly affect the performance of strategies that seek to capture anomalies such as value, quality, momentum, and low-risk premiums (Hasler, 2023; Walter et al., 2023; Soebhag et al., 2024).

The existing literature highlights the challenge posed by methodological uncertainty in financial research. Menkveld et al. (2024) propose the concept of “nonstandard errors”, a measure of variability introduced by arbitrary methodological decisions. They show that these errors could be as large as or even larger than traditional standard errors. Soebhag et al. (2024) apply the notion of nonstandard errors to factor sorting portfolios, suggesting that conclusions drawn from portfolio backtests may often depend more on design choices than on underlying economic phenomena. Similarly, Hasler (2023) and Walter et al. (2023) emphasize the sensitivity of factor portfolio results to seemingly minor methodological details. In this portfolio context methodological uncertainty arises from a variety of choices, including the selection of exclusion criteria (e.g., size, price, age, sector and characteristics filter), the rebalancing frequency, portfolio size or weighting scheme. These decisions often involve trade-offs between theoretical rigor and practical implementability, as highlighted by Novy-Marx and Velikov (2016).

While current literature emphasizes methodological uncertainty in the portfolio construction process, decision variables related to the definition of the sorting variable have been largely overlooked. We address this gap by examining the computation of a specific factor (family) and thus quantify the estimation error of the signal used to construct sorting portfolios. Contrary to existing literature that takes factor definitions provided by the corresponding original papers, we show that choices in factor computation—such as the specific estimator, parametrization, and lookback window length—are equally critical and should not be overlooked. For instance, replicating a low-volatility portfolio involves decisions about the type of volatility estimator to use, including variations in lookback periods and data weighting. In this study, we demonstrate that the im-

fact of these (low-risk) factor computation choices is substantial and comparable to the influence of portfolio sorting construction decisions highlighted by Hasler (2023), Walter et al. (2023), and Soebhag et al. (2024).

To demonstrate that the above mentioned extensions lead to substantial variation in factor portfolio performance, we illustrate the impact of construction choices on the cumulative (net of transaction costs) returns of low-risk portfolios in Figure 1. Each path corresponds to a specific long-only portfolio construction method, as detailed later. In orange we plot 1,620 equally-weighted portfolios, in blue 1,620 value-weighted portfolios, and in black a value-weighted market portfolio that serves as our benchmark. Despite our focus on one factor family, the figure reveals significant dispersion in portfolio performance, underscoring the critical influence of methodological decisions and the necessity for a deeper evaluation of decision criteria. It is worth noting that this figure captures only a subset of our results, as limited-short and long-short portfolios are excluded here.¹ Consequently, the extent of dispersion observed in our analysis is even greater.

In addition to extending methodological uncertainty to the sorting variable, we further expand current findings by focusing on practically relevant portfolios. While much of the existing literature relies on dollar-neutral long-short portfolios without considering transaction costs, we extend the analysis to include more realistic and implementable portfolios. These include long-only and fully-invested limited-short portfolios. Additionally, we incorporate turnover analysis and provide results both with (net) and without (gross) transaction cost considerations. To assess the dispersion of results comprehensively, we evaluate performance across multiple relevant metrics rather than relying on a single measure and compare them also with a benchmark, which is neglected in the current literature (Hasler, 2023; Walter et al., 2023; Soebhag et al., 2024). This approach provides a more holistic view of the influence of portfolio construction choices.

To investigate these extensions on methodological uncertainty and nonstandard errors in portfolio construction, we concentrate on the “factor zoo” within a single factor family rather than across factors. Specifically, we focus our empirical analysis on low-risk portfolios to examine the dispersion within this factor. For many fundamental factors, such as value or quality, certain proposed methodological enhancements, like varying lookback window lengths or employing alternative estimation techniques, are often not applicable or meaningful due to the low time series granularity of the data, stemming from infrequent updates in accounting cycles (e.g., income statements, balance

¹Including these portfolios would make the figure less readable without providing additional insights.

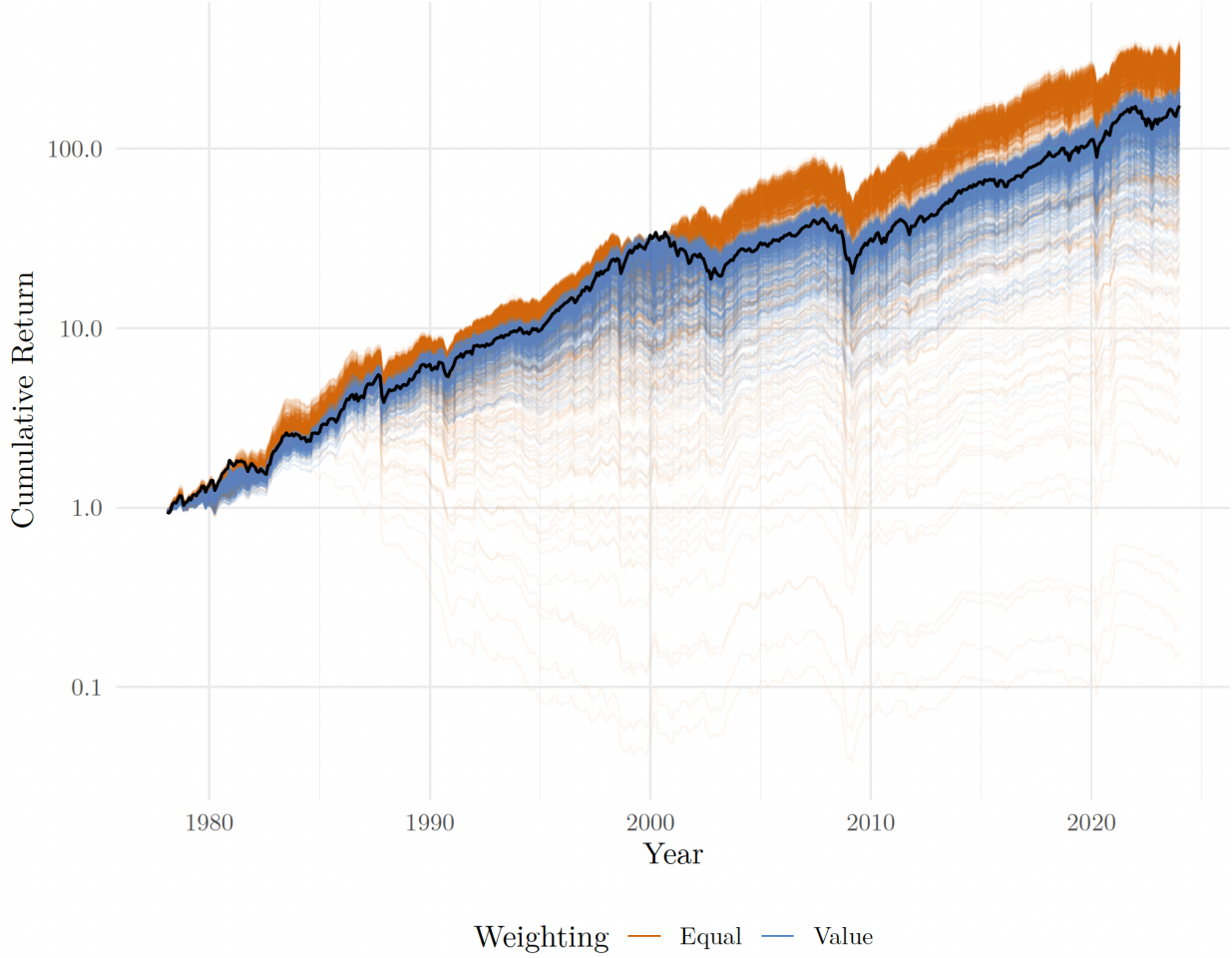


Figure 1: Cumulative returns of long-only low-risk portfolios with transaction costs

The figure shows the dispersion of cumulative returns of long-only portfolios after transaction costs across portfolio construction decisions. In total, the plot contains returns for 3,240 portfolios. We distinguish between portfolios with equally-weighted stocks (orange) and value-weighted stocks (blue). The cumulative returns are calculated based on 11,598 daily out-of-sample returns from January 1978 to December 2023 and presented on a log-10 y-axis.

sheets, and cash flow statements)². This renders all accounting-based factors unsuitable for our purposes.

The particular, the choice for the low-risk family, among other price-based ones like momentum, is motivated by three primary reasons. First, unlike momentum, which has been explicitly examined in prior studies, low-risk has not gained much attention, as it is either left out by current studies completely (Soebhag et al., 2024), examined jointly with all anomalies (Hasler, 2023), or grouped

²In this study we do not want to engage in methodological choices regarding the computation of fundamental variables, as this would introduce numerous other decision variables that are not comparable/consistent between factors.

within broader categories such as trading friction estimators (Walter et al., 2023). This relative scarcity of focused research highlights the need to explore low-risk further. Second, studies on portfolio construction choices including low-risk factors often focus narrowly on their potential to deliver a return premium (Hasler, 2023; Walter et al., 2023). Yet, the core purpose of low-risk factors is to optimize risk-adjusted returns, setting them apart from other anomalies, which predominantly target excess returns through long-short portfolios. This dual focus cannot be addressed with return premiums alone, but rather requires a more comprehensive performance assessment. Third, while there is relatively strong consensus on the definition of momentum as the prior 2-12 monthly return (Wiest, 2023), multiple definitions of low-risk strategies exist (Traut, 2023), making the analysis of low-risk signals particularly urgent.

The need for a comprehensive study of low-risk strategies is further underscored by the growing attention they have received over the past two decades. This interest is driven by the apparent paradox that less volatile stocks often deliver higher risk-adjusted returns, a phenomenon widely referred to as the low-risk anomaly (see Blitz et al., 2023; Traut, 2023, for an overview). Recent investment manager surveys highlight that low-risk investing is among the most commonly employed price-based anomaly strategies, on par or even surpassing momentum (Invesco, 2023, 2024). While extensive academic literature supports the existence of this anomaly across global (equity) markets, there remains a gap between theoretical insights and their practical implementation.

This paper bridges that gap by focusing on the (impact of) construction and evaluation (choices) of realistic and implementable low-risk portfolios. Most existing studies focus on simplistic sorting mechanisms and overlook the profound influence of methodological choices, such as risk estimator selection and configuration, transaction costs handling, rebalancing frequency, and portfolio constraints. These decisions are pivotal in shaping portfolio performance, yet their impact remains largely overlooked both in the context of low-risk investing and factor investing more broadly. Especially the neglect of transaction costs is shown to significantly influence the performance of factor portfolios and thus requires consideration in the context of evaluating methodological choices (Detzel et al., 2023). Another distinctive feature of this study is its focus on long-only and limited-short portfolios, which align with the risk-averse preferences of many (low-risk) investors and regulatory constraints in several jurisdictions. Furthermore, the long-short methodology takes inherent leverage, introducing significant risk to such portfolios and thus counteracting the fundamental objective of a low-risk strategy—to invest with low risk.

Overall, this paper contributes to the literature in several key ways. First, we systematically examine the impact of risk estimators on portfolio performance, highlighting that the choice of estimation methodology (risk estimator and its lookback window length) is one of the main performance variability sources with a comparable impact to the weighting scheme and transaction costs considerations. Second, we analyze the role of transaction costs and performance evaluation metrics, offering insights into the trade-offs between return enhancement and cost efficiency. Consequently, we demonstrate that while the portfolio construction choices outlined in Hasler (2023), Walter et al. (2023), and Soebhag et al. (2024) provide a solid foundation for investigating methodological uncertainty, they overlook two critical factors: the computation of the sorting variable and the consideration of transaction costs. Third, this paper not only advances the academic understanding of low-risk investing, but also provides actionable insights for practitioners seeking to implement effective and robust low-risk portfolio strategies.

The remainder of this paper is organized as follows. [Section 2](#) presents the data and methodology, detailing the empirical framework and the portfolio construction decisions. [Section 3](#) provides a comprehensive empirical evaluation, discussing the implications of construction choices and performance outcomes. Finally, [Section 4](#) concludes with key findings, practical recommendations, and avenues for future research. Appendices contain additional results and auxiliary materials.

2 Data and Methodology

2.1 Data

We construct our sample using the CRSP daily dataset, spanning from January of 1973 to December of 2023. The dataset is restricted to primary stocks listed on AMEX, NYSE, and NASDAQ, excluding stock-days with missing observations for returns or shares outstanding. Since the longest lookback period in our analysis is five years, we start our out-of-sample analysis in January 1978, providing $T = 11,598$ daily returns per portfolio and benchmark. This starting point is selected to avoid the structural changes introduced in the early 1970s, as the CRSP universe tripled in size in 1973. The eligible universe for the out-of-sample analysis ranges in size from 3,569 to 7,465 stocks over time (see [Figure A.1](#)). On average, it has a size of 5,090 stocks, resulting in an average size of 1,697 stocks for our long-only tercile portfolios, 1,018 stocks for our long-only quintile portfolios, and 509 stocks for our long-only decile portfolios. The corresponding quantile long-short portfolios are twice as large.

Daily data for the Fama and French (2015) five-factor model are sourced from the WRDS database, while NBER recession identifiers are obtained from the Federal Reserve Bank of St. Louis.

Transaction costs are approximated using half spreads, as proposed by Novy-Marx and Velikov (2016). When quote data are available, spreads are calculated as the quoted bid-ask spread divided by twice the contemporaneous mid-point, averaged over month t . In cases where quote data are unavailable, spreads are estimated using the CHL spread estimator by Abdi and Ranaldo (2017). For any remaining missing values, we impute spreads using the methodology described in Novy-Marx and Velikov (2016). Additional details regarding the spread estimation process are provided in [Appendix B](#).

2.2 Portfolio Construction and Evaluation Choices

In our empirical analysis we focus on some of the most prominent estimators in the low-risk literature. First, we use the total volatility of historical returns (*hist*) as advocated by Blitz and van Vliet (2007), Baker and Haugen (2012), and Blitz et al. (2013). This measure is arguably the simplest one in our analysis as it is based on the sample volatility, more precisely the equally-weighted volatility of simple returns, $\hat{\sigma}_i^{hist}$, for all stocks in the eligible universe $i = 1, \dots, N$.

Secondly, we employ the seminal RiskMetrics method to estimate the risk of a stock (RiskMetrics, 1996; Mina & Xiao, 2001; Zumbach, 2007). It is closely related to the *hist* methodology but puts more weight on more recent observations in the calculation of volatility. Henceforth, $\sigma_{i,t}^{ewma}$ is estimated using an exponential weighted moving average (*ewma*) as

$$\hat{\sigma}_{i,t}^{ewma} = \sqrt{\lambda(\hat{\sigma}_{i,t-1}^{ewma})^2 + (1 - \lambda)r_{i,t-1}^2}, \quad (1)$$

where $\lambda \in [0, 1]$ is the decay factor or alternatively $\phi = -\log(2)/\log(\lambda)$ is the half-life parameter. It is worth noting that the *ewma* methodology is the equivalent of an I-GARCH(1) estimator and thus represents the broad literature that forecasts volatility with GARCH models (e.g., Bollerslev, 1986; Engle, 2001).

Another refinement of the *hist* methodology is to use idiosyncratic volatility of residual returns instead of total returns to determine risk. This is commonly done relative to the Fama and French (1993) three-factor model as proposed by Ang et al. (2006), which was later on validated by Ang et al. (2009) and Dimson et al. (2017), and can be computed by running the following regression

$$r_{i,t}^e = \hat{\alpha}_i + \hat{\beta}_{i,m} r_{m,t}^e + \hat{\beta}_{i,SMB} r_{SMB,t} + \hat{\beta}_{i,HML} r_{HML,t} + \hat{\epsilon}_{i,t}^{ff3}, \quad (2)$$

where m , SMB , and HML refer to the three factors of the Fama and French (1993) three-factor model and the superscript e denotes returns in excess of the risk-free rate, r_f . The standard deviation of all $\hat{\epsilon}_{i,t}^{ff3}$, $\hat{\sigma}_i^{ff3}$, is then taken as the measure for risk. Thus, we refer to this risk estimator as $ff3$. We further include the idiosyncratic volatility of residuals to the Fama and French (2015) five-factor model ($ff5$) as it represents a more contemporaneous methodology that is likely employed by practitioners to isolate idiosyncratic risk.

Another dominant research stream focuses on measuring risk with market beta (Fama & MacBeth, 1973; Asness et al., 2014; Frazzini & Pedersen, 2014). We follow this stream and calculate our *beta* risk measure as

$$\hat{\beta}_{i,m} = \hat{\rho}_{i,m}^e \frac{\hat{\sigma}_i^e}{\hat{\sigma}_m^e}, \quad (3)$$

where $\hat{\rho}_{i,m}^e$ is the sample correlation of r_i^e and r_m^e and $\hat{\sigma}_i^e$ and $\hat{\sigma}_m^e$ are the sample standard deviations of r_i^e and r_m^e respectively.

For our signal calculation described above and for portfolio construction, we require a minimum of 90% of daily returns for the respective lookback period, and we filter out stocks with missing return on the last day of our lookback period. Furthermore, we exclude stocks that have missing volume observations for the last month in the portfolio formation.³

With our choices regarding the methodological setup of the portfolio construction, we are strongly guided by Hasler (2023), Walter et al. (2023), and Soebhag et al. (2024). We exclude decision forks that are commonly used in studies on factors that are based on book variables (e.g., the exclusion of financial or utility companies) as the rationale for these decisions does not apply to low-risk

³The missing data filtration is included to keep the lookback period decision variable clean. For example, if data for a stock was only available for the past four years, it shall not be included in the 5-year lookback portfolio. Note that the literature is usually less strict on missing data filters, but as we focus on practically relevant portfolios we raise the bar on data coverage; e.g., Hou et al. (2020) and the references therein use a minimum of just 15 daily return observations for total volatility, 100 daily return observations for idiosyncratic volatility and from 120 to 750 daily return observations for beta estimation. The volume filter is included because this paper focuses on realistic portfolios and it is known that especially at the beginning of our sample the CRSP daily file is prone to data errors (Bryzgalova et al., 2024). To avoid biases due to data errors such data are excluded. Also, it is realistic to assume that investors in practice would not trade stocks with insufficient data or for which they must assume their data is flawed.

investing. Furthermore, we expand their decision variables by choices that may not be used in academic research yet, but are relevant for practitioners.

1. **Portfolio Type:** Current studies that examine the influence of methodological choices in the portfolio construction context solely focus on studying long-short factor portfolios (Hasler, 2023; Walter et al., 2023; Soebhag et al., 2024). These portfolios are not only difficult to obtain for many market participants but also involve taking on leverage, which is arguably ‘contrary’ to the goal of investing with low risk. However, the recent work by Blitz et al. (2024) shows that taking on some leverage can significantly improve the performance of low-risk portfolios. Because of this, we consider three portfolio types in our analysis, a long-only portfolio, which we denote as *100-0*, a fully invested limited-short portfolio, that takes on a leveraged position of 130% on the least risky stocks and finances this by short selling 30% of the riskiest stocks (*130-30*), and the standard dollar-neutral long-short portfolio that goes long on the least risky stocks and short on their most risky counterparts with equal investment amounts (*100-100*). It is worth noting that, while return calculations for long-only and limited-short portfolios are straightforward, there is greater flexibility when calculating returns for dollar-neutral long-short portfolios. Following Detzel et al. (2023), we adopt an interest-free cash account methodology, which is further detailed in [Appendix C](#).
2. **Size Exclusion:** To mitigate the effects of small and potentially illiquid stocks on the results, we employ the two most commonly used size filters as in Walter et al. (2023) and Soebhag et al. (2024). These filters either exclude stocks that are smaller than the 10% or 20% thresholds based on the market capitalization of NYSE stocks.
3. **Price Exclusion:** Price exclusions aim to reduce the impact of trading frictions like short sale constraints on the outcome of the portfolios. Again, we focus on the most popular choices as identified by Hasler (2023), Walter et al. (2023), and Soebhag et al. (2024) and exclude stocks with a price below 1\$ or 5\$ in the portfolio formation.
4. **Lookback Window:** The lookback for the case of low-risk portfolios specifies how much past data is taken to estimate the respective signals for the portfolio formation. This is an important choice as one needs to balance the trade-off between more accurate estimates and including potentially more outdated information. The low-risk literature shows considerable variation in the choice of lookback windows, spanning from as short as one month (e.g., Ang et al., 2006) to as long as five years (e.g., Frazzini & Pedersen, 2014). To keep the analysis

manageable, we limit our focus to three specific settings, using data from the past one, three, and five years to compute our risk estimators.⁴

5. **Rebalancing Frequency:** We include monthly and annual rebalancing frequencies, as these are the two most common choices in academic literature (Hasler, 2023; Walter et al., 2023). Additionally, we consider quarterly rebalancing, a frequency that is often employed in style ETFs, and weekly rebalancing to capture a broader spectrum of investor activity levels.
6. **Portfolio Buckets:** The number of sorting portfolios used to distinguish between riskier and less risky stocks should, in theory, correlate with performance if low-risk investing is effective. We include groupings of three (tercile), five (quintile), and ten (decile) portfolios, with the latter two also employed in Hasler (2023) and Walter et al. (2023). Tercile portfolios are included as well, as they are used in replication studies (e.g., Jensen et al., 2023) and in some low-risk strategies (see Traut, 2023).
7. **Weighting Scheme:** The weighting scheme is among the most frequently controlled decision criteria. Although equally-weighted portfolios tend to incur in higher trading costs, value-weighted portfolios often underperform, as many anomalies are more pronounced in smaller stocks. Therefore, same as in Hasler (2023), Walter et al. (2023), and Soebhag et al. (2024), we consider both equal and value weighting.

Unlike current studies that primarily address methodological uncertainty in portfolio construction, we expand our analysis to include decisions related to portfolio evaluation. While existing studies often focus solely on average returns (Hasler, 2023; Walter et al., 2023) or Sharpe ratios (Soebhag et al., 2024), we evaluate our portfolios using a range of criteria (based on daily frequency t):

I **Cumulative Return (Cum.):** That we also denote as C and is defined as

$$C = \prod_{t=1}^T (1 + r_{p,t}) - 1 , \quad (4)$$

where $r_{p,t}$ is the return of portfolio p at time t .

⁴It is important to note that for the *ewma* estimator, the decay factor, λ effectively sets the lookback window for the analysis. Because of this, we choose λ for the respective lookback windows such that 99.9% of the probability mass falls within one, three and five years, leading to λ of 0.97, 0.99 and 0.995 respectively. These decay factors are much larger than 0.94 that is recommended for daily data by RiskMetrics (RiskMetrics, 1996; Mina & Xiao, 2001; Zumbach, 2007). In Appendix D we formalize the selection of the respective λ for the different lookback windows and show that the average performance of *ewma* portfolios with $\lambda = 0.94$ is inferior to that of portfolios with $\lambda = 0.97$, making the former infeasible from a practical perspective.

II **Average Return (Ret.)**: That we also denote as \bar{r}_p and is defined as

$$\bar{r}_p = \frac{1}{T} \sum_{t=1}^T r_{p,t} , \quad (5)$$

III **Alpha**: Defined as the estimated intercept, $\hat{\alpha}$, from the regression against the Fama and French (2015) five-factor model

$$\begin{aligned} r_{p,t}^e = & \hat{\alpha} + \hat{\beta}_{p,m} r_{m,t}^e + \hat{\beta}_{p,SMB} r_{SMB,t} + \hat{\beta}_{p,HML} r_{HML,t} \\ & + \hat{\beta}_{p,RMW} r_{RMW,t} + \hat{\beta}_{p,CMA} r_{CMA,t} + \hat{\epsilon}_{p,t} , \end{aligned} \quad (6)$$

where $r_{p,t}^e$ is the return in excess of the risk free rate for portfolio p and $\hat{\epsilon}$ is the estimated error term.

IV **Standard Deviation (Std.)**: Defined as

$$\hat{\sigma}_p = \sqrt{\frac{1}{T-1} \sum_{t=1}^T (r_{p,t} - \bar{r}_p)^2} , \quad (7)$$

V **Sharpe Ratio (SR)**: Defined as

$$SR = \frac{\frac{1}{T} \sum_{t=1}^T r_{p,t}^e}{\sqrt{\frac{1}{T-1} \sum_{t=1}^T (r_{p,t}^e)^2}} , \quad (8)$$

VI **Information Ratio (IR)**: Defined as

$$IR = \frac{\frac{1}{T} \sum_{t=1}^T r_{p,t} - r_{m,t}}{\sqrt{\frac{1}{T-1} \sum_{t=1}^T (r_{p,t} - r_{m,t})^2}} , \quad (9)$$

where $r_{m,t}$ is the return of the market taken from the Fama and French (2015) five-factor model that serves as our benchmark.

VII **Maximum Drawdown (m. DD)**: Defined as the maximum observed loss from a peak to a trough of a portfolio before a new peak is attained:

$$\text{m. DD} = \max_{t \in (0,T)} \left[\max_{\tau \in (0,t)} \frac{C(\tau) - C(t)}{C(\tau) + 1} \right] . \quad (10)$$

We calculate all performance measures for the total period of our sample, that is, we include all the $T = 11,598$ daily (stock, benchmark and portfolio) returns from January 1978 to December 2023. Also, we compare them against a benchmark based on the value-weighted market return of the Fama and French (2015) five-factor model. It is worth noting that roughly 7% of our long-short portfolios default over our sample period. For these portfolios, the performance measures are calculated for the time before default. More details about the defaulting long-short portfolios and how we calculate their returns can be found in [Appendix C](#).

Last, building on findings by Detzel et al. (2023), who demonstrate that the performance of factor models is significantly affected by the inclusion or exclusion of transaction costs, we consider two cases for each portfolio configuration. The first excludes transaction costs, as is common in low-risk investing literature, while the second includes them to evaluate whether low-risk strategies hold up under realistic implementation assumptions.⁵ To our knowledge, we are the first to incorporate transaction costs in assessing portfolio modeling choices, allowing us to examine the robustness of low-risk investing across various portfolio decisions in a realistic market environment. Considering all feasible portfolio constructions and performance evaluation combinations, we examine a total of 136,080 performance metrics for 9,720 portfolios. Our complete empirical setup is illustrated as a decision tree in [Figure 2](#).

3 Empirical Results

3.1 General Overview of the Portfolio Strategies

To obtain a general understanding of the impact of construction choices on low-risk (sorting) portfolios, we first focus on the dispersion and similarity of all the investigated strategies. We begin by examining whether Sharpe ratios are robust to varying construction choices. Note that the results for other performance measures, such as cumulative returns, mean returns, standard deviation, and maximum drawdown, are comparable. Therefore, we restrict our attention to arguably the most important risk-adjusted performance measure: the Sharpe ratio. However, dispersion results for other performance measures are available upon request.

In [Figure 3](#), we plot the (annualized) gross (without transaction costs) and net (with transaction costs) Sharpe ratios for all the $9,720 \times 2$ portfolios. These include the $3,240 \times 2$ long-only (100-0),

⁵The details about how we estimate transaction costs can be found in [Appendix B](#).

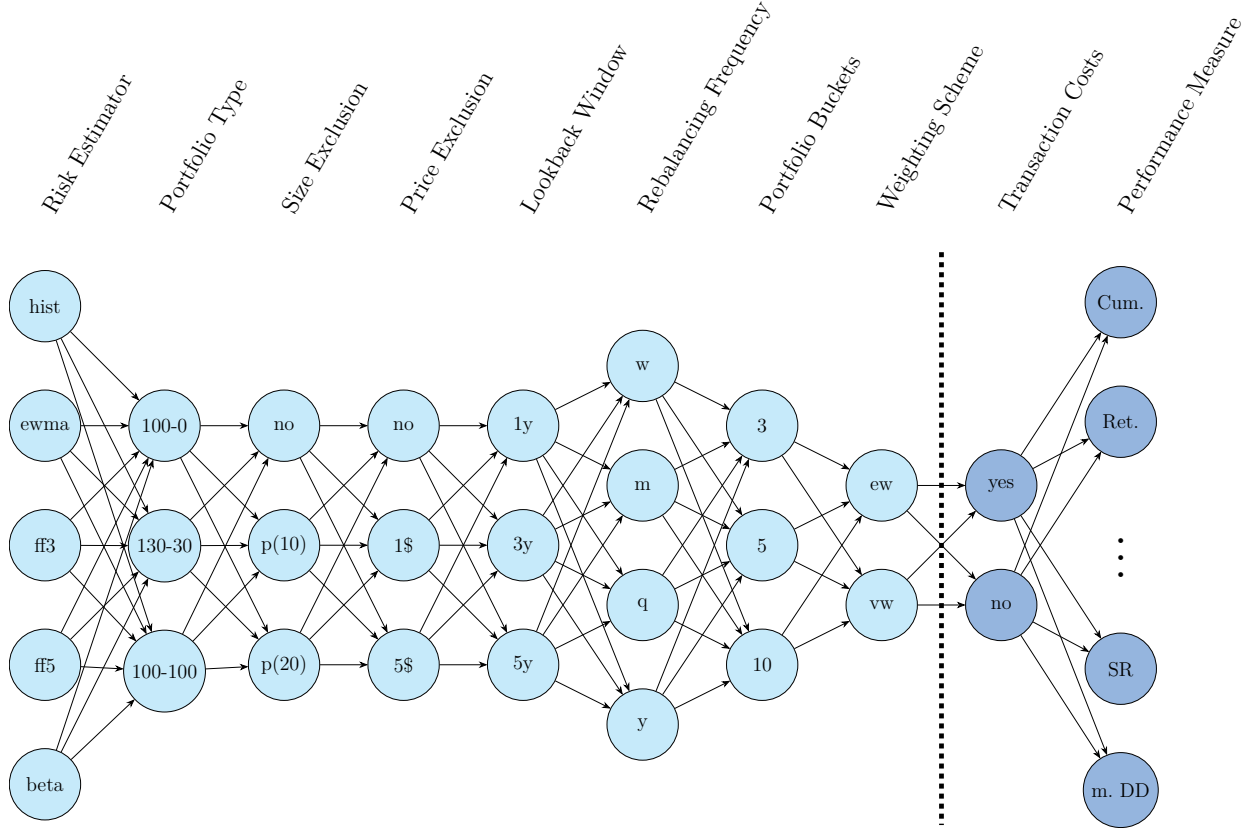


Figure 2: Decision tree of portfolio construction and evaluation choices

The figure illustrates all possible decision paths for our portfolio construction and evaluation. We consider the paths of eight portfolio construction decision forks and two performance evaluation forks for our sorting portfolios. The portfolio construction forks include the estimation of risk (historical volatility, exponentially weighted moving average, idiosyncratic volatility with respect to the Fama and French 3-factor and 5-factor model, or market beta), portfolio type (long-only (100-0), limited-short (130-30), or long-short (100-100)), small stocks exclusion dependent on market equity quantiles (none, smaller than $p(10)$ or $p(20)$), stock price exclusion (none, smaller than \$1 or \$5), lookback window for the risk estimators (past returns of the previous one, three, or five years), the number of sorting portfolio buckets (three, five, and ten), and weighting scheme (equal- or value-weighting). The performance evaluation forks include transaction costs (with or without) and the performance metric (cumulative return, average return, alpha, standard deviation, Sharpe ratio, information ratio, or maximum drawdown). In total we construct 9,720 portfolios for which we calculate 136,080 performance metrics.

3, 240 \times 2 limited-short (130-30), and 3, 240 \times 2 dollar-neutral (100-100) portfolios.⁶ The x-axis shows the 6,480 different portfolio variations for any portfolio type, ordered from low to high Sharpe ratios. For the fully-invested portfolios (100-0 and 130-30), the benchmark (BM) is the Sharpe ratio of the market-cap-weighted portfolio, whereas for the dollar-neutral portfolio (100-100), it is a Sharpe ratio of zero.

The figure reveals substantial variation in the obtained annualized Sharpe ratios. Depending on

⁶For a detailed analysis of the impact of transaction costs, refer to [Section 3.3](#), and for the methodology used to

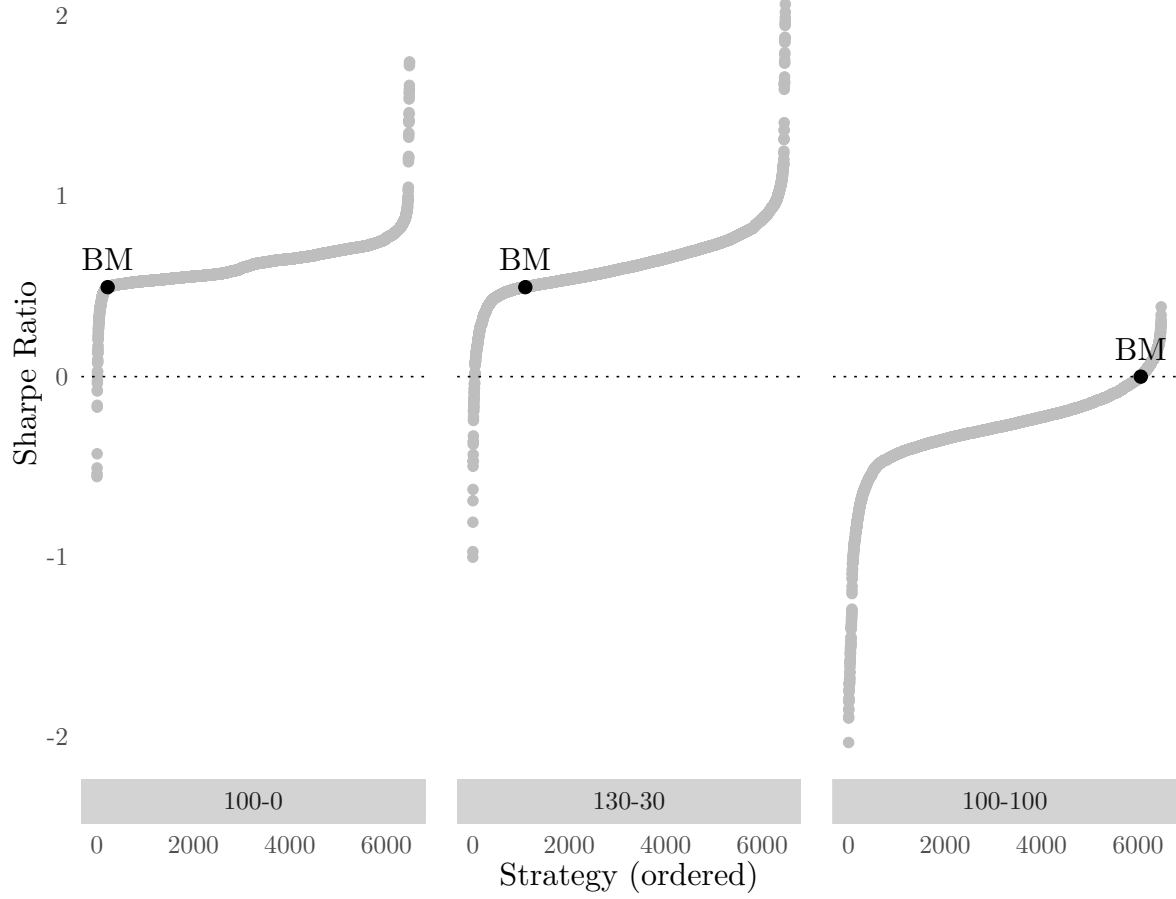


Figure 3: Sharpe ratios of all portfolios across portfolio construction choices

This figure plots the (annualized) gross (without transaction costs) and net (with transaction costs) Sharpe ratios for all the $9,720 \times 2$ portfolios, respectively for the $3,240 \times 2$ long-only (100-0), $3,240 \times 2$ limited-short (130-30), and the $3,240 \times 2$ dollar-neutral (100-100) portfolios across varying portfolio construction choices. The x-axis shows the 6,480 different versions for any portfolio type ordered from low to high Sharpe ratio. For the fully-invested portfolios (100-0 and 130-30) the benchmark (BM) is the Sharpe ratio of the value-weighted market portfolio, whereas for the dollar-neutral portfolio (100-100) it is a Sharpe ratio of zero. All the portfolios are based on 11,598 daily out-of-sample returns from January 1978 to December 2023.

how the low-risk portfolio is constructed and evaluated, Sharpe ratios range between -0.55 and 1.75 for 100-0, between -1.00 and 2.07 for 130-30, and between -2.03 and 0.39 for 100-100. While the fully-invested portfolios exhibit similar average and median Sharpe ratios, slightly above 0.6 , **Figure 3** indicates that the uncertainty arising from construction choices is larger when short selling is permitted. Another notable finding is the dispersion of relative performance against the benchmark in terms of Sharpe ratio. While 97% of the 100-0 portfolios and 83% of the 130-30 portfolios outperform the market-cap benchmark Sharpe ratio of 0.5 , only 7% of the 100-100

calculate transaction costs, see [Appendix B](#).

portfolios achieve a positive Sharpe ratio.

Interestingly, our finding that only 7% of the dollar-neutral long-short portfolios achieve a positive Sharpe ratio is significantly weaker than results reported in the existing literature on the low-risk anomaly (see Traut, 2023, for an overview). There are two main reasons for this. First, unlike previous studies, we incorporate transaction costs into our portfolios, which significantly impacts performance (see Section 3.3). Second, while existing literature often calculates long-short portfolio returns as the difference between the returns on the long and short portfolios—a straightforward but simplifying approach—this method implicitly assumes that both portfolios maintain equal value at each time step, t . This assumption underestimates adverse movements, such as when the short portfolio grows larger than the long portfolio. In contrast, we employ a cash account methodology that avoids this oversimplification, as further outlined in Appendix C.

In Table A.1, we report the construction choices of the top and bottom five Sharpe ratio portfolios. From this perspective, there is a notable dispersion in the performance of the best and worst portfolios across portfolio types and transaction cost considerations. For fully-invested portfolios without transaction costs, the beta estimator produces both the highest and lowest Sharpe ratios. The best-performing portfolios are generally the weekly rebalanced equally-weighted portfolios, while the worst-performing portfolios are the yearly rebalanced value-weighted portfolios. However, when transaction costs are accounted for, the results become more heterogeneous. For the 100-0 portfolios, the equally-weighted portfolios based on ewma and hist perform the best, whereas the *higher-frequency* (weekly and monthly) rebalanced equally-weighted portfolios based on one-year beta perform the worst. For the 130-30 portfolios, the *lower-frequency* (annually and quarterly) rebalanced equally-weighted portfolios based on (one-year) ewma perform the best, while the *higher-frequency* (weekly and monthly) rebalanced equally-weighted portfolios based on one-year ewma perform the worst. For the 100-100 portfolios, no clear or consistent pattern emerges to classify which portfolio construction choices perform best or worst, except that weekly rebalanced equally-weighted portfolios consistently perform the worst. More detailed results on the construction choices of winner and loser portfolios are discussed in Section 3.5.

Another perspective on the portfolio construction uncertainty problem is to evaluate the similarity in performance across strategies. In Figure 4, we present the average correlation between risk estimators in the off-diagonal elements and the average correlation within a risk estimator on the diagonal. The reported correlation numbers are for portfolios that include transaction costs;

however, the gross results are nearly identical. When computing different portfolios within a factor family, specifically within the low-risk anomaly, we find that correlations are generally high. Nevertheless, several noteworthy patterns emerge across portfolio types and risk estimators.

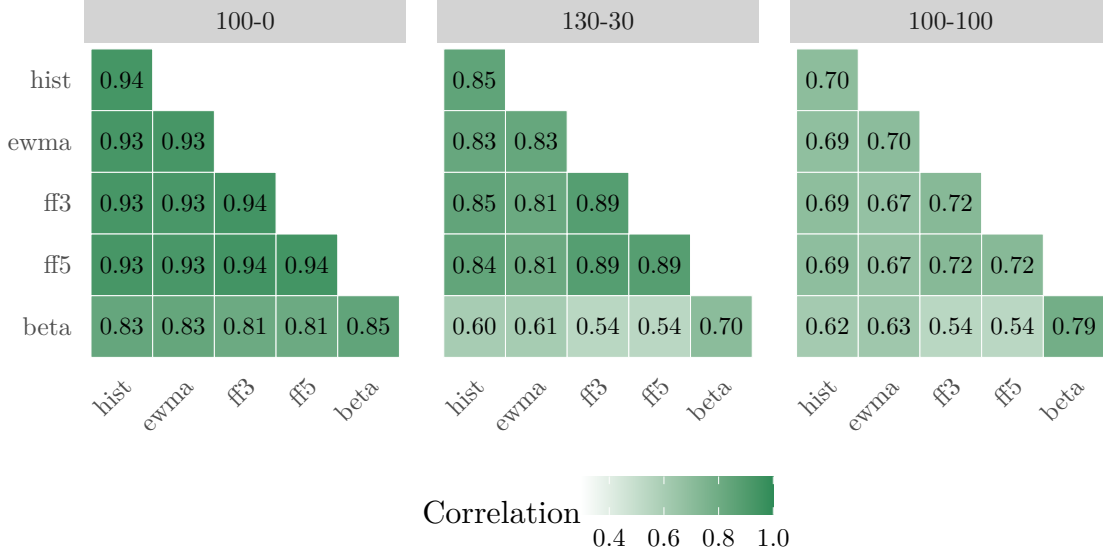


Figure 4: Average correlations between and within portfolios including transaction costs

This figure shows the average correlation between portfolios formed on different risk estimators in the off-diagonal elements and the average correlation within portfolios based on the same risk estimator on the diagonal. The analysis is carried out separately for long-only (100-0), limited-short (130-30), and long-short (100-100) portfolios. Correlations are calculated based on 11,598 daily out-of-sample returns including transaction costs from January 1978 to December 2023.

First, consistent with the findings on Sharpe ratio dispersion, we observe that correlation numbers decrease as more short selling is permitted. For the 100-0 portfolios, the average correlation is 0.9, for 130-30 it drops to 0.76, and for 100-100 it decreases further to 0.64. Second, the within risk estimator correlations are generally higher than the correlations between different risk estimators. Third, portfolios sorted on beta exhibit the lowest correlations, both with other beta portfolios and with portfolios based on other risk estimators. This highlights the instability of beta as a portfolio construction variable, given that it relies on an OLS regression coefficient, which is susceptible to potentially large estimation errors. However, beta-based portfolios also offer the largest diversification potential when combined with other low-risk portfolios.

Analyzing the out-of-sample distribution of various portfolio performance measures and the correlations between portfolio returns, we conclude that increasing the degrees of freedom in portfolio construction exacerbates the problem of methodological uncertainty. This further supports the idea that weight constraints in portfolio construction are detrimental to the robustness of a strat-

egy and can help mitigate result dispersion.⁷ To quantify the methodological uncertainty of various practically relevant low-risk portfolios, we focus on the concept of nonstandard errors in the next section.

3.2 Nonstandard Errors Versus Standard Errors

To evaluate the influence of construction choices on portfolio performance and to make our results comparable to existing literature, we calculate standard and nonstandard errors for our collection of low-risk portfolios. In our definition of nonstandard errors, we are guided by Soebhag et al. (2024) who use Sharpe ratios as their evaluation criterion. We define nonstandard errors as the cross-sectional standard deviation of Sharpe ratios across all portfolios (1,296 in total) for a given low-risk estimator and portfolio type. Standard errors, in contrast, are calculated as the average monthly Sharpe ratio (time-series) standard deviation across all construction choices for the same estimator and portfolio type.⁸ The annualized results of both error types are shown in Figure 5.

Similar to the findings of Menkveld et al. (2024) on nonstandard errors in research design choices and those of Soebhag et al. (2024) and Walter et al. (2023) on nonstandard errors in factor portfolio sorts, we find that nonstandard errors for low-risk portfolios are generally too large to be ignored and are substantial relative to standard errors. A key contribution of our analysis is the distinction between portfolio types. While the existing literature predominantly focuses on long-short dollar-neutral portfolios, we demonstrate that for long-only portfolios and many fully-invested portfolios with limited short selling, nonstandard errors tend to be smaller and are often lower than their standard error counterparts. Overall, our findings indicate that nonstandard errors are less stable across portfolio types and risk estimators compared to standard errors. Furthermore, as the degrees of freedom increase (i.e., more short selling allowed), the relative impact of nonstandard errors (compared to standard errors) becomes larger. This finding offers promising implications for investors who face various (regulatory or risk management) constraints on their portfolio holdings. For these investors, nonstandard errors pose less of a challenge than suggested by previous literature.

To evaluate how much each construction choice of our decision tree influences portfolio outcomes

⁷See, for example, Jagannathan and Ma (2003), who demonstrate that imposing (even incorrect) weight constraints reduces risk in large portfolios and prove that it is equivalent to using a shrinkage estimator for the covariance matrix, as in Ledoit and Wolf (2004) and De Nard (2022).

⁸For the average standard error definition see Menkveld et al. (2024) and Walter et al. (2023).

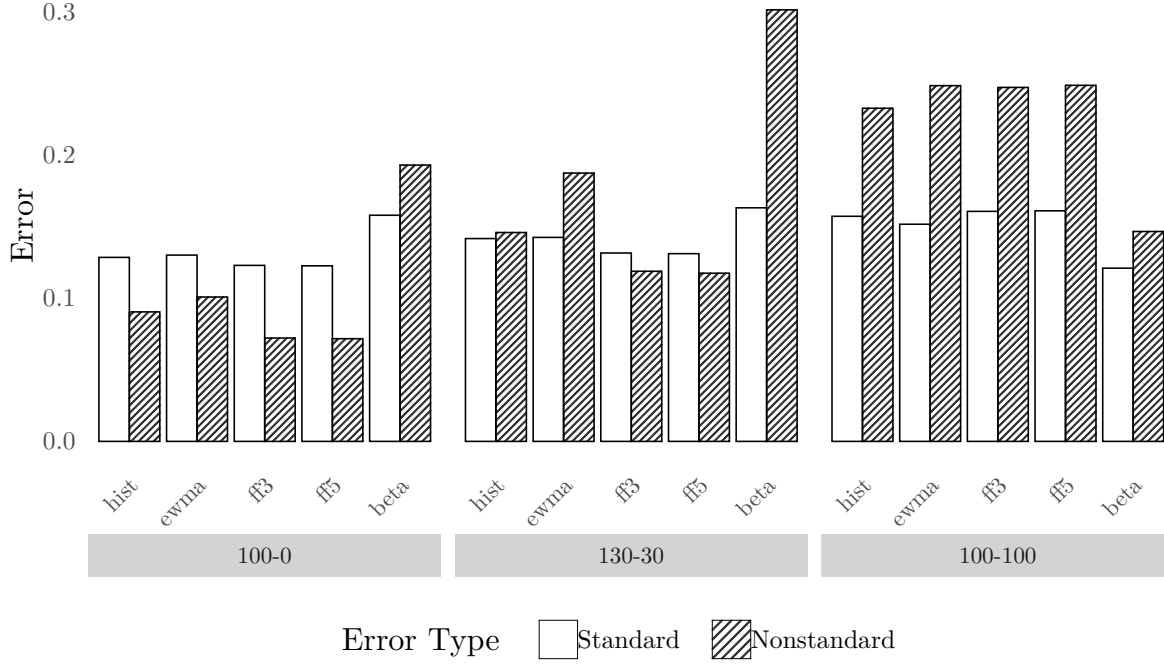


Figure 5: Standard vs. nonstandard errors

This figure plots the nonstandard error (stripes) and standard error (no pattern) for each risk estimator and portfolio type. The nonstandard error is defined as the cross-sectional standard deviation of yearly Sharpe ratios for all decision variables. The standard error is the mean time-series standard error of yearly Sharpe ratios across all decision variables. The results are grouped by portfolio type: long-only (100-0), limited-short (130-30), and long-short (100-100). Both error types are calculated based on 11,598 daily out-of-sample returns from January 1978 to December 2023.

we follow Walter et al. (2023) and calculate the mean absolute differences (MAD) for a pair of paths (i, j) that only differ in one construction choice made at one specific fork, f , of our decision tree. This set of matched paths for a combination of estimator e and portfolio type p that only differs in one fork f can be defined as

$$S_f^{e,p} = \{(i, j) | c_{i,n} = c_{j,n} \quad \forall n \in \{1, \dots, 7\}, c_{i,f} \neq c_{j,f}\}, \quad (11)$$

where $p = (c_1, \dots, c_F)$ is a path that is defined as a vector of choices, c_f , from Figure 2. Therefore, each path corresponds to one portfolio based on estimator e and portfolio type p . Note, that we include transaction costs as a construction choice in this case because it is calculated before the performance measure and its treatment as a construction choice grants insights into how its inclusion/neglection affects portfolio performance.

We calculate the differences in performance measures over the entire sample for each of the matched

paths in $S_f^{e,p}$. We also create an overall estimate across all matched paths, S_f^p , for each portfolio type over all risk estimators. In this case, we also include the risk estimator as a decision fork for which we calculate the MAD. The mean absolute difference for each fork f , portfolio type p , and estimator e can be formalized as

$$MAD_f^p = \frac{1}{|S_f^{e,p}|} \sum_{(i,j) \in S_f^{e,p}} |m_i^e - m_j^e|, \quad (12)$$

for a selected performance measure m . With the MAD we can quantify the impact of each construction choice on a selected performance measure. We report MADs across risk estimators and portfolio types. MAD results for Sharpe ratios are presented in [Table 1](#) where decision forks are sorted from the largest overall impact (top) to smallest overall impact (bottom) across all risk estimators. Further results for mean return, standard deviation, and maximum drawdown are presented in [Tables A.2–A.4](#).

From [Table 1](#), we identify the portfolio construction decisions that induce the largest variation in Sharpe ratios when all other decisions at the respective fork are held constant. Generally, the ranking among risk estimators and (fully-invested) portfolio types is relatively consistent.

The most influential decision fork is the weighting scheme, followed by the consideration of transaction costs and the choice of the risk estimator. Across all risk estimators (Overall), the impact of equally-weighting versus value-weighting on the (MAD) SR ranges from 0.15 (100-0) to 0.20 (130-30). In comparison, transaction cost considerations impact the (MAD) SR from 0.07 (100-0) to 0.18 (100-100), and the risk estimator choice impacts the (MAD) SR from 0.07 (100-0) to 0.10 (130-30).⁹ The smallest impact generally arises from the price exclusion fork. Additionally, the length of the lookback window, rebalancing frequency, number of portfolio buckets, and size restrictions induce considerable and quantitatively similar variations in Sharpe ratios.

Consistent with the nonstandard error analysis in [Figure 5](#), we find that portfolio construction choices most significantly affect ewma and beta portfolios, while idiosyncratic volatility portfolios (ff3 and ff5) are influenced the least. Finally, we observe larger (aggregated) MAD numbers for dollar-neutral portfolios, followed by fully-invested limited long-short portfolios and long-only portfolios. This provides further evidence that nonstandard errors and thus the impact of decision forks

⁹Note that for the “Risk Estimator” fork only the Overall score can be computed.

Table 1: Mean absolute differences at decision forks with respect to Sharpe ratio

This table presents the average mean absolute Sharpe ratio differences p.a., as defined in Equation 12, for each decision fork. For each fork, we compare Sharpe ratio differences between matched decision paths that vary only in the specific decision under consideration. These mean absolute differences are calculated individually for each risk estimator and averaged across all matched paths, as well as jointly for all risk estimators. The columns report averages for all risk estimators combined (Overall) and for individual risk estimators separately. The forks (rows) are ordered by their overall impact in descending order. The analysis is carried out separately for long-only (100-0), limited-short (130-30), and long-short (100-100) portfolios. Note that the impact of the risk estimator can only be calculated for the joint analysis across all estimators.

Portfolio	Fork	Overall	Hist.	EWMA	FF3	FF5	Beta
100-0	Weighting	0.15	0.17	0.17	0.14	0.13	0.15
	Trans. Costs	0.07	0.04	0.07	0.03	0.03	0.20
	Risk Estimator	0.07					
	Rebalancing	0.03	0.02	0.03	0.01	0.01	0.07
	Lookback	0.02	0.02	0.02	0.02	0.02	0.05
	Drop Size	0.02	0.01	0.01	0.01	0.01	0.07
	Port. Buckets	0.02	0.02	0.02	0.01	0.01	0.05
	Drop Price	0.01	0.00	0.00	0.00	0.00	0.02
130-30	Weighting	0.20	0.21	0.23	0.15	0.15	0.28
	Trans. Costs	0.16	0.11	0.18	0.09	0.09	0.31
	Risk Estimator	0.10					
	Drop Size	0.06	0.05	0.06	0.05	0.05	0.10
	Lookback	0.06	0.05	0.06	0.05	0.05	0.07
	Rebalancing	0.06	0.04	0.07	0.04	0.04	0.10
	Port. Buckets	0.05	0.03	0.04	0.03	0.03	0.10
	Drop Price	0.02	0.02	0.03	0.02	0.02	0.03
100-100	Trans. Costs	0.18	0.17	0.21	0.18	0.19	0.16
	Weighting	0.17	0.18	0.19	0.18	0.18	0.10
	Drop Size	0.11	0.12	0.12	0.13	0.13	0.06
	Port. Buckets	0.10	0.11	0.11	0.12	0.11	0.06
	Risk Estimator	0.10					
	Lookback	0.09	0.10	0.09	0.11	0.11	0.05
	Rebalancing	0.09	0.09	0.11	0.09	0.09	0.07
	Drop Price	0.05	0.06	0.06	0.06	0.06	0.02

become more pronounced as the degrees of freedom in portfolio construction increase.

Tables A.2–A.4 show a similar relationship between the impact of portfolio construction choices and other performance measures. For instance, the ‘weighting’ decision fork’s MAD of annualized mean returns, as well as for standard deviation and maximum drawdown, consistently exceeds 1% p.a. across all estimators and portfolio types. A few notable cases emerge: As expected, transaction cost considerations have less influence on risk metrics such as standard deviation and maximum drawdown. However, other portfolio construction choices, such as the number of portfolio

buckets and size filters, are more relevant for risk metrics compared to returns. For the long-only portfolios and in terms of out-of-sample risk metrics, the choice of the risk estimator is even the most influential portfolio construction decision.

Based on the size differences of MADs among different decision variables, we conclude that our inclusion of the risk estimator computation (choice of estimation methodology and lookback window length) and transaction costs consideration are together with the weighting the most influential decisions. As the impact of the weighting is generally known and usually controlled for in the literature, we infer that our newly added decision forks are (practically) more relevant than the construction choices already covered by Hasler (2023), Walter et al. (2023), and Soebhag et al. (2024). Given the different order of decision forks across the performance metrics we investigate, portfolio managers should carefully assess their choices to optimize the performance metric(s) most aligned with their objectives. Focusing solely on average returns is insufficient to capture the complete picture.

3.3 Transaction Costs

A topic often overlooked in the academic literature on (factor) portfolio selection and nonstandard error analysis is the careful consideration of transaction costs. As shown in Table 1, the impact of including transaction costs in portfolio evaluation is as significant as the decision to invest using an equally-weighted or value-weighted approach. Therefore, an important contribution of our paper is the examination of the robustness of low-risk investing under realistic and practically relevant transaction cost considerations. To achieve this, we do not assume fixed transaction costs across stocks and time. Instead, we approximate transaction costs using half spreads, following the methodology of Novy-Marx and Velikov (2016) that is detailed in Appendix B.

To highlight the impact of transaction costs, we present the first quartile (Q1), median (Q2), and third quartile (Q3) of cumulative returns over time for 648 long-only portfolios, across different risk estimators and transaction cost considerations in Figure 6. The cumulative returns are plotted on a logarithmic scale for improved readability, and we compare the performance dispersion with our market benchmark.

From Figure 6, we observe that the impact of transaction costs is substantial, affecting not only Sharpe ratios but also cumulative returns. Similar findings are reported in Table A.2 for MADs of average returns. Additionally, we find that the impact of transaction costs varies across risk

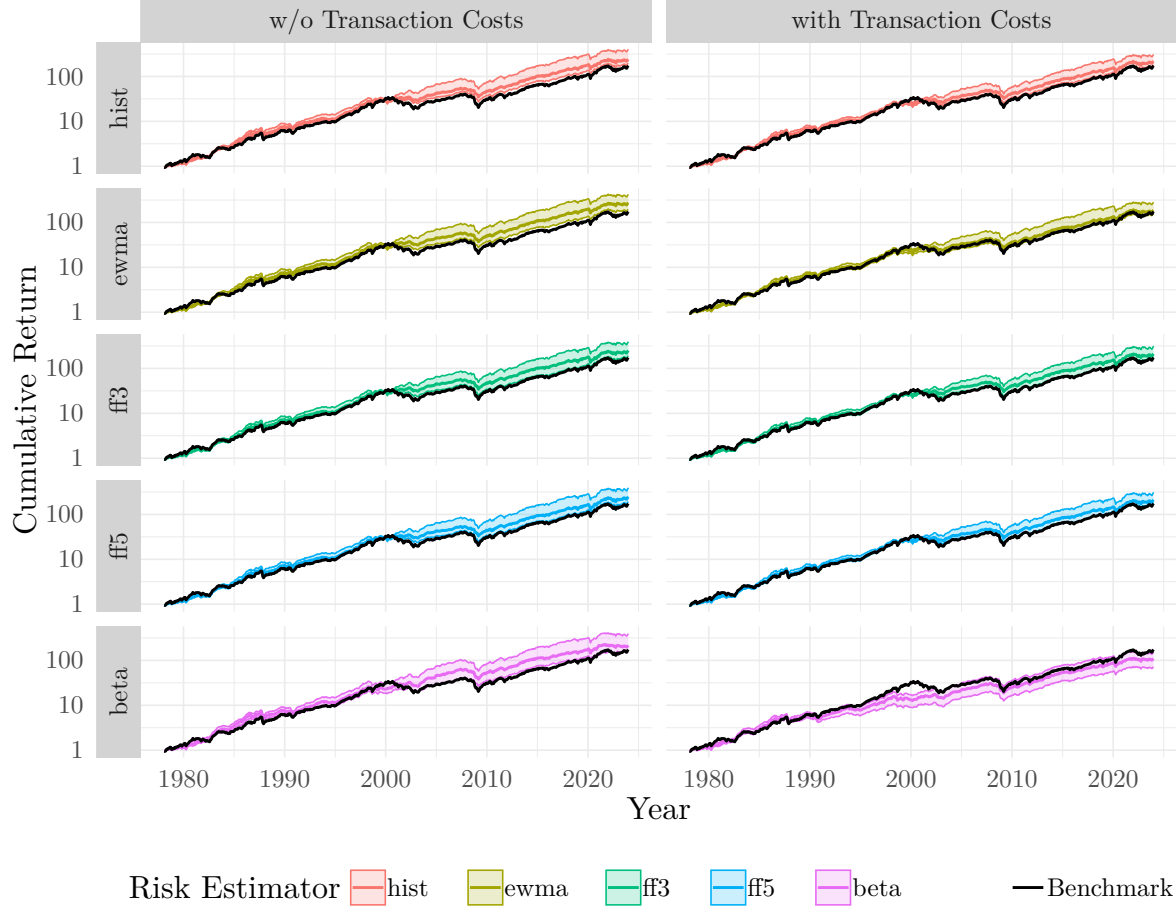


Figure 6: Interquartile range (Q1 and Q3) of cumulative returns for long-only portfolios with and without transaction costs vs. benchmark

The plot shows the interquartile range of cumulative returns for long-only portfolios, both without (left panels) and with (right panels) transaction costs. Each color represents a different risk estimator, illustrating the range across 648 portfolios. The black line represents the benchmark, a value-weighted market portfolio. The median cumulative return is highlighted in darker shade. The returns are calculated based on 11,598 daily out-of-sample returns including transaction costs from January 1978 to December 2023 and presented on a log-10 y-axis.

estimators and, as shown in Table A.2, also across portfolio types. The panels on the left in Figure 6 illustrate that without considering transaction costs, all risk estimators exhibit comparable performance (in terms of dispersion). Notably, for most of the analyzed period and for all risk estimators, approximately 75% of the portfolios outperform the benchmark. However, once transaction costs are deducted, the net outperformance decreases, and the impact of transaction costs varies significantly across risk estimators. Specifically, we find that the ewma and, particularly, the beta portfolios experience the largest negative impact from transaction costs due to higher turnover; see also the top panel of Table A.2. The panels on the right in Figure 6 reveal that,

even with transaction costs, most of the total and idiosyncratic volatility portfolios continue to outperform the benchmark, albeit with smaller margins and reduced dispersion. In contrast, only around 50% of the ewma portfolios and a minority of the beta portfolios manage to outperform the benchmark once transaction costs are considered. In conclusion, the inclusion of transaction costs has a profound influence on determining which portfolios perform best.

To closer examine what choices impact transaction costs, we focus on the two components that determine the total transaction costs of our portfolios, turnover and stock-level costs. We calculate the total transaction costs of portfolio p at time t as

$$TC_{p,t} = \sum_{i=1}^{N_{p,t}} to_{i,p,t} c_{i,t} , \quad (13)$$

where $c_{i,t}$ is the stock-level transaction cost, $N_{p,t}$ is the number of stocks at rebalancing, and $to_{i,p,t}$ is the turnover of stock i that is calculated as the absolute difference in weights before and after rebalancing, $to_{i,p,t} = |w_{i,p,t} - \tilde{w}_{i,p,t-}|$. More details on our transaction costs calculation can be found in [Appendix B](#).

We first examine what portfolios have high turnover to see what characteristics lead to higher transaction costs due to more extensive reallocations of portfolios—the first part of the right-hand side in (13). In [Figure 7](#) we plot the median yearly turnover for portfolios split by rebalancing frequencies and risk estimators as these are arguably the most influential choices on turnover. While the overall median turnover across all portfolios is 272%, the median turnover varies significantly by rebalancing frequency: 502% for weekly rebalancing, 297% for monthly rebalancing, 172% for quarterly rebalancing, and 87% for yearly rebalancing. For each rebalancing frequency, a consistent pattern emerges: Total and idiosyncratic volatility portfolios exhibit greater robustness over time, resulting in lower turnover compared to ewma and beta portfolios. This difference is a primary reason why net-of-transaction-cost results are less favorable for ewma and beta portfolios. The higher turnover for ewma portfolios arises because these portfolios place greater weight on recent observations to improve volatility estimation. However, this approach has the drawback of being more erratic and susceptible to influence from a few observations, depending on the chosen decay factor, λ . Similarly, the turnover for beta portfolios is elevated due to the instability of Ordinary Least Squares (OLS) regression coefficient estimates.

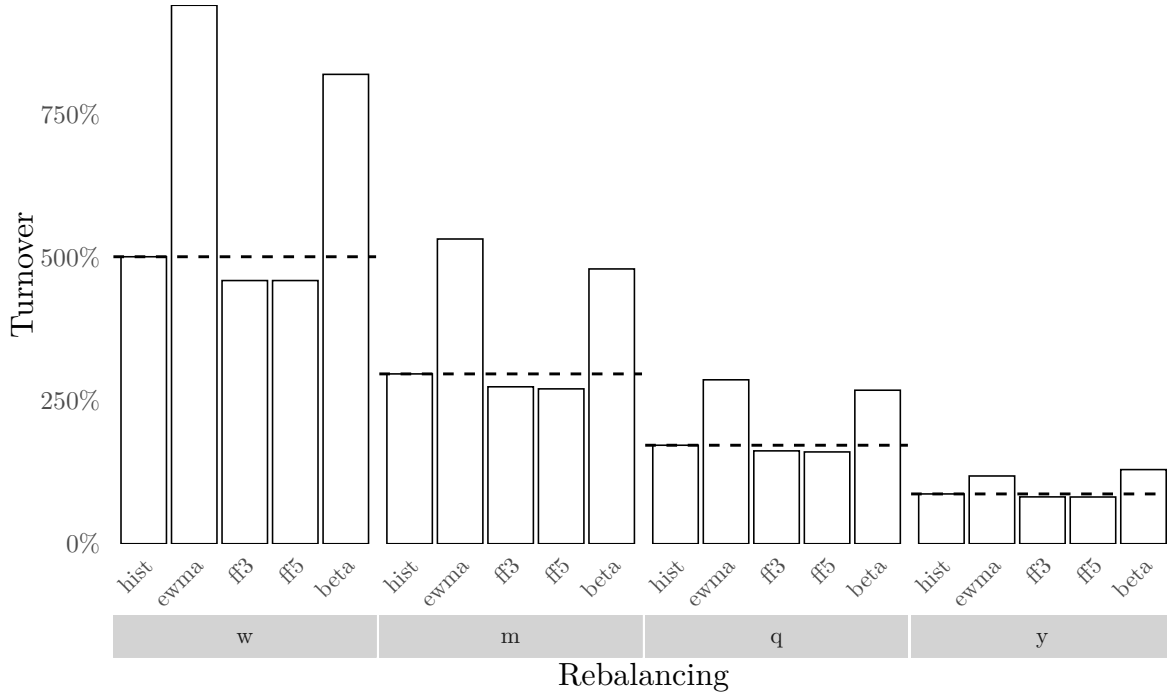


Figure 7: Median yearly turnover for portfolios with varying rebalancing frequency and risk estimator

The plot illustrates median yearly turnover for portfolios across construction choices. Turnover is measured for portfolios grouped by the same risk estimator and rebalancing frequency: weekly (w), monthly (m), quarterly (q), and yearly (y). The median yearly turnover across all portfolios is 272%. The turnover is calculated based on 11,598 daily out-of-sample returns from January 1978 to December 2023.

Next, we investigate what characteristics lead to higher transaction costs due to putting more weight on stocks that are more costly to trade—the second part of the right-hand side in (13). In Figure 8, we analyze the median transaction costs (per unit of turnover) and split our results by portfolio type, portfolio weighting and size filters. We perform this “size focused” split because, like many other anomalies, low-risk is criticized to short expensive small stocks that lead to high transaction costs that are usually neglected (Baker et al., 2011; Novy-Marx & Velikov, 2022). We find that the overall median transaction cost across all portfolios is 0.45%, it varies by portfolio type: 0.37% for long-only portfolios, 0.46% for limited long-short portfolios, and 0.52% for dollar-neutral portfolios. This shows that as criticized by literature, the short leg of low-volatility portfolios is indeed more costly to trade. As expected, we observe lower transaction costs for value-weighted portfolios compared to equally-weighted ones, as well as for portfolios that exclude small stocks. Both the weighting methodology and the size filter influence exposure to expensive small stocks, directly impacting the costs per unit of turnover for the respective portfolios. Notably, however,

only equally weighted portfolios without a size filter exhibit exceptionally high trading costs, while other specifications show comparatively small variation in costs. These findings highlight why "Drop Size" is one of the influential portfolio construction choices discussed in [Section 3.2](#) and should not be underestimated.

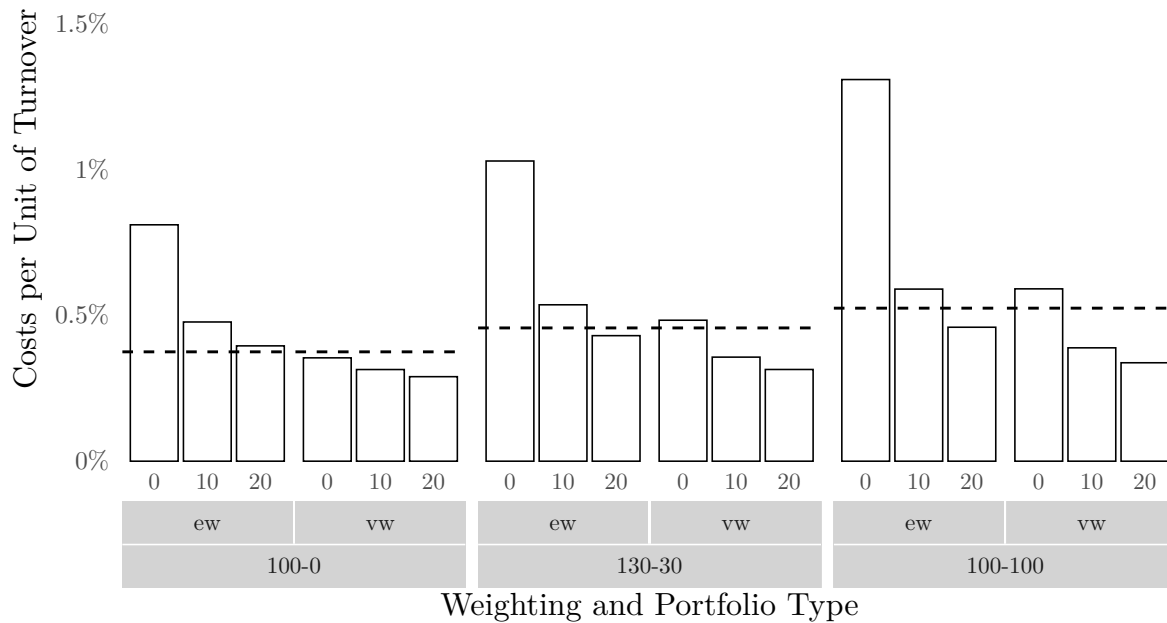


Figure 8: Median transaction costs per unit of turnover for portfolios with varying portfolio type, weighting scheme, and size filters

The plot illustrates median costs per unit of turnover for portfolios across construction choices, categorized by three decision variables: size exclusion, weighting scheme, and portfolio type. Size exclusion refers to the exclusion of stocks that are smaller than the 0, 10, or 20 percentile of stocks with respect to market equity. Weighting scheme indicates whether portfolios are constructed using value-weighting (vw) or equal-weighting (ew). Considered portfolio types are long-only (100-0), limited-short (130-30), and long-short (100-100) portfolios. The median costs per turnover across all portfolios within a portfolio type is indicated by the dashed black lines. The median transaction costs per unit of turnover across all portfolios is 0.45%. The costs are calculated based on 11,598 daily out-of-sample returns from January 1978 to December 2023.

Finally, we examine the impact of transaction costs on conclusions drawn on what characteristics lead to superior portfolio performance. We do this by selecting the top 100 winner and loser portfolios in terms of average Sharpe ratio (for fully-invested long-only and limited-short portfolios) and count how often each decision variable in each decision fork is represented in the winner and loser portfolios. We repeat the process twice, once with and once without transaction costs and plot the results in [Figure 9](#).

The focus of this analysis lies on how transaction costs change the conclusions about preferences

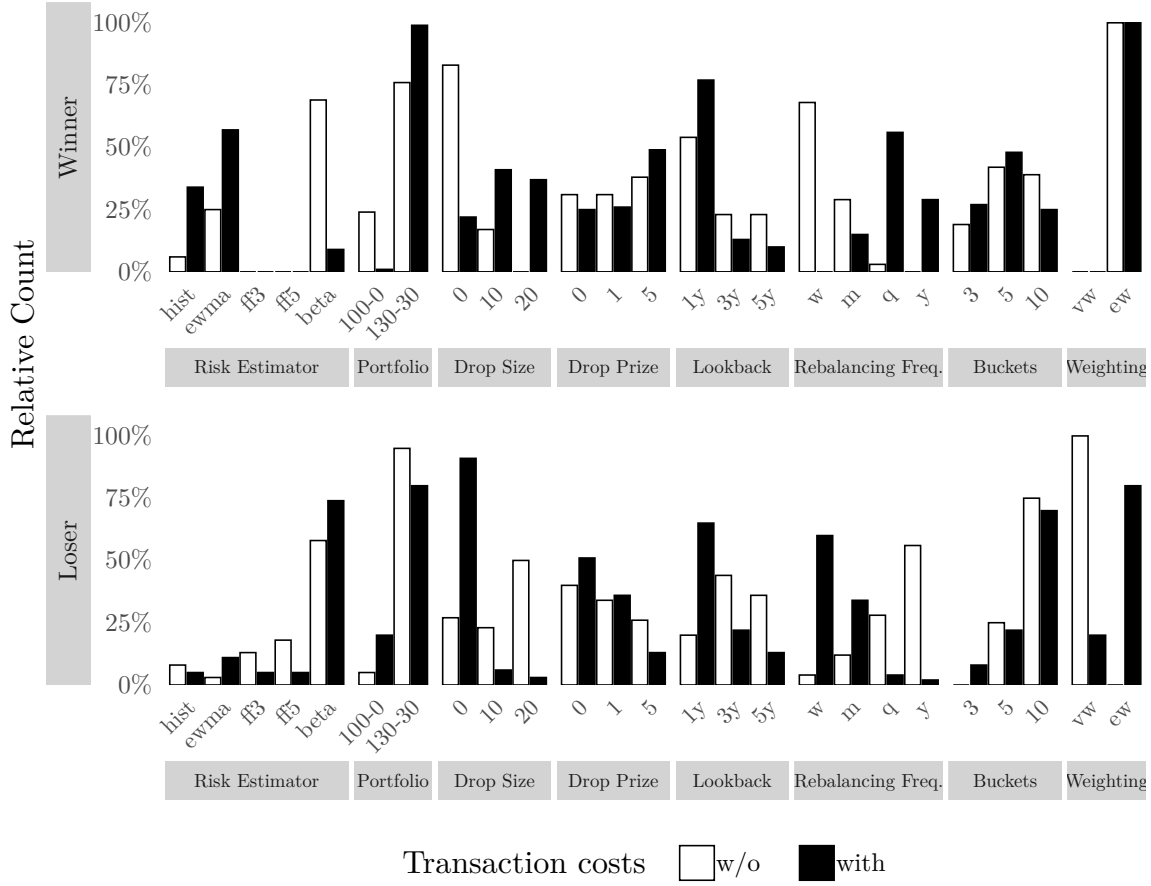


Figure 9: Relative appearance of decision choices in top 100 and bottom 100 long-only and limited-short portfolios by gross and net Sharpe ratio

The figure illustrates the relative occurrence of decision variables in the top 100 winner portfolios (upper panel) and bottom 100 loser portfolios (lower panel) based on Sharpe ratios. Portfolios, including both long-only and limited-short strategies, are ranked by Sharpe ratio, and the relative appearance of decision variables is calculated for each decision fork. The plot distinguishes between portfolios evaluated with transaction costs (black bars) and without transaction costs (white bars). The Sharpe ratios are calculated based on 11,598 daily out-of-sample returns from January 1978 to December 2023.

for decision variables. A more extensive discussion on what characteristics are commonly shared by winner and loser portfolios can be found in [Section 3.5](#).

We observe the following influences of transaction costs on the choices in our decision forks:

- **Risk Estimator:** We find that without transaction costs most winner portfolios are constructed using beta as the risk estimator. However, due to its high turnover, the majority of winner portfolios with transaction costs employ the hist or ewma estimators, which are more cost-efficient and thus yield better net performance. Among loser portfolios, the beta

estimator is predominantly used, regardless of whether transaction costs are accounted for.

- **Portfolio Type:** Although limited-short portfolios are more expensive than long-only portfolios, they appear more frequently among winners and less frequently among losers when transaction costs are included. This indicates that the performance gains from shorting outweigh the relatively high costs.
- **Size Exclusion:** Without considering transaction costs, most winner portfolios do not apply a size filter. When transaction costs are factored in, however, approximately 75% of winner portfolios exclude at least the smallest 10% of stocks due to their high trading costs. The opposite trend can be observed for loser portfolios.
- **Price Exclusion:** Transaction costs appear to have no impact on the relative prevalence of price filters in winner and loser portfolios.
- **Lookback Window:** For winner portfolios, the distribution of lookback periods remains relatively consistent across both gross and net cases. In contrast, for loser portfolios, the relative share of portfolios with shorter lookback periods increases when transaction costs are considered. This is intuitive, as shorter lookback periods enable more responsive risk estimators, which increase turnover and, consequently, transaction costs that diminish performance.
- **Rebalancing Frequency:** Without transaction costs most winning portfolios are rebalanced weekly. As shown before, these portfolios have the highest turnover which is why most winning portfolios are rebalanced quarterly and yearly when transaction costs are considered. In contrast, loser portfolios are predominantly rebalanced yearly without transaction costs but shift to weekly rebalancing when costs are included.
- **Portfolio Buckets:** We do not find a strong impact of transaction costs on the choice of portfolio buckets.
- **Weighting Scheme:** Our results indicate that all winning portfolios are equal-weighted, regardless of whether transaction costs are accounted for. In contrast, for losing portfolios, all are value-weighted when transaction costs are excluded, while approximately 80% shift to equal-weighting when costs are considered. This aligns with our earlier analysis, as equal-weighted portfolios incur higher costs than value-weighted ones.

In unreported results, we also analyze the impact of transaction costs on the top 100 winner

and loser portfolios with respect to other performance measures. Generally, risk metrics such as standard deviation and maximum drawdown are less affected by transaction costs.

From our analysis, we conclude that portfolio construction choices significantly impact the turnover and average stock-level costs of low-risk portfolios that ultimately determine total transaction costs. We argue that for practical applications, transaction costs must be properly accounted for, as they can significantly influence inferences about what portfolio construction choices lead to positive performance. Our findings show that transaction costs most notably influence the selection of the optimal risk estimator, the use of size filters, and the preferred rebalancing frequency.

3.4 Evaluation of Implementable Portfolios

In this section, we focus on the portfolios that are arguably the most relevant from a practical investor perspective: fully invested portfolios without (100-0) and with limited short selling (130-30). In [Table 2](#), we present the number of portfolios that outperform the benchmark across various performance measures.

Table 2: Share of portfolios that outperform their benchmark with and without transaction costs
The table reports the number of portfolios within each portfolio type that outperform their respective benchmarks based on six performance metrics: cumulative return, average return, alpha, standard deviation, Sharpe ratio, and maximum drawdown. The results are further divided by whether transaction costs are included in the evaluation. For long-only (100-0) and limited-short (130-30) portfolios, the benchmark is a value-weighted market portfolio. For long-short portfolios (100-100), the benchmark is zero. Due to this difference in benchmarks, not all performance metrics can be evaluated for long-short portfolios. The statistics are calculated based on 11,598 daily out-of-sample returns from January 1978 to December 2023.

Portfolio	Trans. Costs	Cum.	Ret.	Alpha	Std.	SR	m. DD
100-0	w/o	80%	66%	45%	100%	99%	97%
100-0	with	63%	44%	20%	100%	94%	96%
130-30	w/o	78%	69%	47%	100%	94%	84%
130-30	with	45%	31%	14%	100%	72%	72%
100-100	w/o	11%	48%	9%			
100-100	with	0%	13%	1%			

In general, we observe that the percentage of portfolios outperforming the benchmark is higher for long-only portfolios across performance measures and transaction cost considerations. This is primarily due to the reduced impact of nonstandard errors, as discussed in [Section 3.2](#). Specifically, around 80% of the gross cumulative portfolio returns are higher for long-only portfolios compared to the value-weighted benchmark, while 63% of net cumulative returns outperform. However, these

percentages are noticeably smaller when comparing annualized arithmetic mean returns: 66% of long-only portfolios outperform the benchmark in terms of gross arithmetic mean returns, but only 44% outperform in terms of net arithmetic mean returns.

This difference can be attributed to compounding effects, where higher (lower) volatility leads to lower (higher) geometric average returns. This phenomenon, which disproportionately affects the most volatile stocks and benchmarks, is further discussed van Vliet et al. (2011). Arithmetic averages ignore compounding effects, which are particularly important when comparing portfolios with significantly different volatility characteristics, as is evident in our empirical analysis. Arguably, and even though less common in the financial industry and not yet sufficiently emphasized in the literature, geometric (compounded) average returns are more relevant to long-term investors than arithmetic (simple) averages.

The next column of Table 2 shows the percentage of portfolios with positive alpha from the daily Fama and French (2015) five-factor regression (6). For gross returns, less than half of the portfolios exhibit positive alpha, and for net returns, fewer than a fifth do. Novy-Marx (2015) argues that the low-beta and low-volatility anomalies can be explained by a three-factor model augmented with a profitability factor. Similarly, Fama and French (2016) find that their five-factor model, which adds profitability and investment factors to the original three-factor model, explains returns on low-risk-sorted portfolios. Nevertheless, Blitz and Vidojevic (2017) and related work contend that the low-risk anomaly is not fully explained by the five-factor model: *“But as long as the data indicates that portfolios with higher risk do not generate higher returns, it is premature to conclude that the low-risk anomaly has been resolved.”*

The last three columns of Table 2 present risk(-adjusted performance) metrics. As we focus on low-risk portfolios it comes as no surprise that risk metrics improve relative to the benchmark. Nevertheless, it is impressive that the standard deviation of all portfolios is always smaller than benchmark and that also Sharpe ratios as well as maximum drawdowns improve almost for all portfolio constructions. We find that around 99% (97%) of the gross SRs (m. DD) are larger for the long-only portfolios compared to benchmark, respectively 94% (97%) for net SRs (m. DD).

Note that almost half of the gross and 13% of the net dollar-neutral portfolios exhibit positive returns, thereby outperforming the benchmark. However, as shown in Figure 3, only 7% of all dollar-neutral portfolios achieve a positive Sharpe ratio. This indicates that the deduction of the risk-free rate has a substantial impact on flipping the sign of Sharpe ratios, particularly because

many mean returns are close to zero. Another notable observation is that the number of dollar-neutral portfolios with cumulative returns greater than one is significantly lower than the number of portfolios with mean returns greater than zero. This results in a reversal of the ordering between geometric (compounded) and arithmetic (simple) average returns, in contrast to fully-invested low-risk portfolios. This is attributable to the fact that among all our gross and net long-short portfolios roughly 48% experience a drawdown of more than 90% throughout our sample period.

The relatively weak performance of our long-short portfolios can largely be attributed to two factors. First, as evidenced by the significant differences between gross and net results, transaction costs have a substantial impact on performance, a finding also documented by Detzel et al. (2023). Second, our methodology for calculating returns on long-short portfolios differs from that in existing literature. While previous studies make the simplifying assumption that the long and short legs at each observation t to maintain equal dollar values, we allow the total dollar value of the long and short legs to diverge. This approach means that adverse movements (i.e., the appreciation of the short leg) accumulate more quickly and can lead to portfolio defaults—a scenario that is not possible under the simplifying assumption in current literature. Further details on our return calculation method for long-short portfolios can be found in [Appendix B](#).

For the remainder of our paper, we restrict our attention to the implementable fully-invested 100-0 and 130-30 portfolios net of transaction costs. This decision is motivated by the limited practical relevance of dollar-neutral portfolios and the need to focus on portfolio constructions and evaluations that are both relevant and implementable. For these portfolios, the average Sharpe ratio across all 6,480 strategies is 0.57, which exceeds the benchmark Sharpe ratio of 0.5. In [Figure 10](#), we illustrate the distribution of Sharpe ratios for the implementable portfolios net of transaction costs, categorized by various construction choices.

The (on average) results with respect to SR can be summarized as follows:

- **Risk Estimator:** The total volatility estimator performs the best, followed by ewma and idiosyncratic volatility estimators. Portfolios based on beta sorting perform the worst by far and also have the largest dispersion.
- **Portfolio Type:** Interestingly, long-only (100-0) portfolios outperform limited-short (130-30) portfolios and exhibit lower dispersion. In general, shorting markedly increases tail risk.
- **Size Exclusion:** The inclusion of size filters boosts performance.

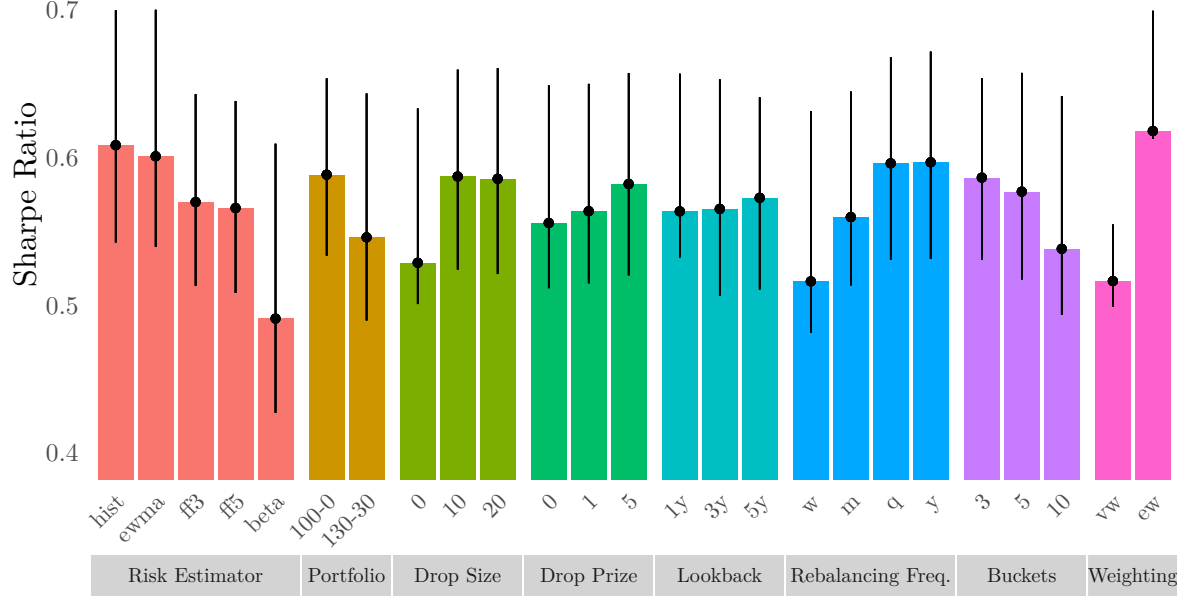


Figure 10: Average Sharpe ratios for long-only and limited-short portfolios by decision variable including transaction costs

The plot displays the average annualized Sharpe ratios of long-only and limited-short portfolios after accounting for transaction costs. Each bar represents portfolios sharing a specific decision variable within a given decision fork, differentiated by colors. The black lines show the inter quartile range of Sharpe ratios. Sharpe ratios are calculated based on 11,598 daily out-of-sample returns from January 1978 to December 2023.

- **Price Exclusion:** The impact of price filters is less pronounced, but higher price filters are preferred.
- **Lookback Window:** There is no obvious pattern regarding lookback window length. However, a shorter lookback of one year provides a more attractive overall distribution by reducing left tail risk (increasing Q1).
- **Rebalancing Frequency:** Less frequent rebalancing (quarterly or annual) outperforms weekly and monthly updates.
- **Portfolio Buckets:** A smaller number of buckets, resulting in a larger number of stocks per portfolio, outperforms.
- **Weighing Scheme:** Equally-weighted portfolios substantially outperform value-weighted portfolios, though they also exhibit markedly larger dispersion.

To further understand which portfolio construction choices are most suitable for benchmarked

managers, we plot the characteristics of the implementable portfolios net of transaction costs with a positive information ratio (IR) in [Figure 11](#). This subset comprises 2,434 out of 6,480 (38%) benchmark-outperforming portfolios, and the relative distribution of construction choices closely mirrors the Sharpe ratio evaluation above. The only exceptions are the lookback window, where 1-year portfolios are more prevalent, and the size filter, which does not increase the likelihood of a positive IR.

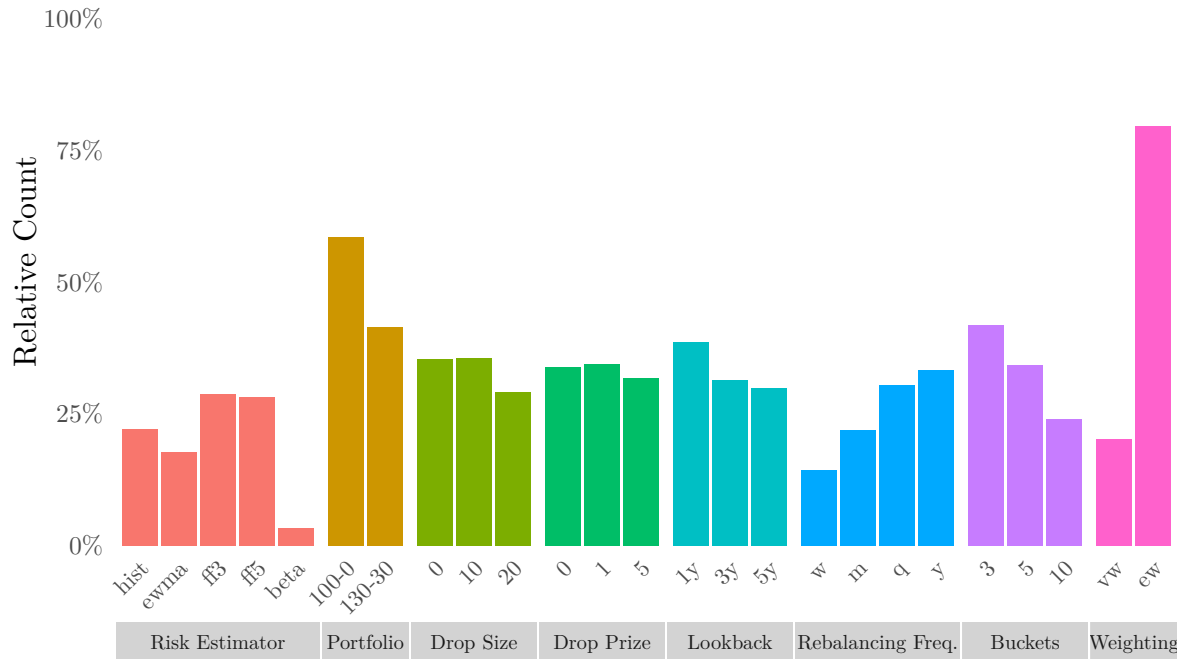


Figure 11: Relative appearance of decision variables in long-only and limited-short portfolios with positive information ratio after transaction costs

The figure illustrates the relative frequency of decision variables in long-only and limited-short portfolios that achieve a positive information ratio after accounting for transaction costs. The relative frequency is calculated for each decision fork, distinguished by different colors. The benchmark for calculating the information ratio is a value-weighted market portfolio. Information ratios are calculated based on 11,598 daily out-of-sample returns from January 1978 to December 2023.

3.5 Winner Loser Evaluation

It is natural to ask which implementable portfolio construction choices perform the best and worst with respect to various performance measures, net of transaction costs. To address this question, we report the relative counts of decision variables of the 100 winner (W) and 100 loser (L) portfolios in [Table 3](#). We highlight results in (light) green if they are more than (two) three standard deviations away from an equal distribution of variables in a positive direction. For winning portfolios, this

indicates that the variable is overrepresented, while for loser portfolios, it suggests the variable is underrepresented. Conversely, we apply (light) red coloring for the opposite scenario.

This differentiation of results allows for two evaluations of each performance metric. First, we can assess which decisions improve or worsen performance. If a decision variable is overrepresented in winner portfolios and underrepresented in loser portfolios (both green), it indicates an improvement in performance for that metric, with the opposite implying a worsening (red coloring). Second, we can evaluate how decisions affect the dispersion of results. If a variable is overrepresented in both winner portfolios (green) and loser portfolios (red), it increases dispersion. Conversely, red coloring in the winner column and green in the loser column suggests a reduction in dispersion.

The results can be summarized as follows:

- **Risk Estimator:** Portfolios based on idiosyncratic volatility improve mean returns and information ratios, while those based on beta result in lower average returns. Interestingly, beta-sorted portfolios reduce volatility, whereas idiosyncratic volatility-sorted portfolios increase it. Total volatility estimators (hist and ewma) perform best in terms of Sharpe ratios, whereas the beta estimator performs the worst. Additionally, beta-sorted portfolios show wider dispersion in maximum drawdowns.
- **Portfolio Type:** With the exception of maximum drawdowns, where limited-short portfolios result in worse performance, they increase dispersion across all other performance metrics. This further emphasizes the heightened impact of nonstandard errors as the degrees of freedom increase. Also, limited-short portfolios lead to lower information ratios.
- **Price Exclusion:** Price filters appear to have a positive impact on results. With the exception of information ratio, they increase the likelihood of being among winning portfolios and reducing the likelihood of being among losing portfolios across all performance metrics.
- **Size Exclusion:** There is no consistent finding for size filters. Such filters significantly reduce the dispersion of results, for average returns, alpha, and volatility. However, they are clearly beneficial to improve Sharpe ratios and avoid maximum drawdowns.
- **Lookback Window:** There is no clear winner regarding the lookback window length. Generally, shorter lookback windows lead to greater dispersion in outcomes, while longer windows tend to reduce dispersion.

Table 3: Relative appearance of portfolio construction choices in top 100 winner (W) and loser (L) portfolios for long-only and limited-short portfolios with transaction costs

The table presents the relative occurrence of decision variables in the top 100 winner (W) and bottom 100 loser (L) portfolios, ranked according to various performance metrics. These portfolios, encompassing both long-only and limited-short strategies, are evaluated net of transaction costs. The relative appearance of decision variables is calculated for each decision fork across the following performance metrics: average return, alpha, volatility, Sharpe ratio, information ratio, and maximum drawdown. Results are highlighted in (light) green if they are more than (two) three standard deviations away from an uniform distribution of variables in a positive direction. For winning portfolios, this indicates that the variable is overrepresented, while for loser portfolios, it suggests the variable is underrepresented. Conversely, we apply (light) red coloring for the opposite scenario. The performance metrics are calculated based on 11,598 daily out-of-sample returns from January 1978 to December 2023.

Variable	Return		Alpha		Volatility		SR		IR		max. DD	
	W	L	W	L	W	L	W	L	W	L	W	L
Risk Estimator												
hist	0%	5%	12%	13%	0%	0%	34%	5%	1%	9%	3%	7%
ewma	21%	11%	33%	24%	7%	0%	57%	11%	6%	13%	3%	17%
ff3	42%	3%	0%	19%	0%	50%	0%	5%	48%	9%	1%	6%
ff5	37%	4%	0%	19%	0%	46%	0%	5%	45%	9%	4%	6%
beta	0%	77%	55%	25%	93%	4%	9%	74%	0%	60%	89%	64%
Portfolio Type												
100-0	36%	22%	24%	7%	15%	2%	1%	20%	99%	31%	55%	17%
130-30	64%	78%	76%	93%	85%	98%	99%	80%	1%	69%	45%	83%
Size Exclusion												
0	43%	91%	64%	92%	81%	78%	22%	91%	34%	96%	47%	95%
10	47%	6%	25%	5%	10%	10%	41%	6%	42%	4%	26%	5%
20	10%	3%	11%	3%	9%	12%	37%	3%	24%	0%	27%	0%
Prize Exclusion												
0	36%	50%	34%	56%	12%	55%	25%	51%	44%	54%	34%	56%
1	36%	36%	29%	40%	28%	33%	26%	36%	38%	38%	32%	37%
5	28%	14%	37%	4%	60%	12%	49%	13%	18%	8%	34%	7%
Lookback Window												
1y	73%	60%	42%	51%	40%	22%	77%	65%	40%	53%	53%	60%
3y	18%	27%	26%	27%	32%	39%	13%	22%	48%	31%	28%	26%
5y	9%	13%	32%	22%	28%	39%	10%	13%	12%	16%	19%	14%
Rebalancing Frequency												
w	18%	60%	1%	83%	39%	13%	0%	60%	33%	64%	8%	59%
m	3%	35%	3%	17%	26%	13%	15%	34%	1%	33%	39%	25%
q	40%	4%	38%	0%	20%	18%	56%	4%	29%	3%	42%	6%
y	39%	1%	58%	0%	15%	56%	29%	2%	37%	0%	11%	10%
Portfolio Buckets												
3	72%	9%	12%	21%	26%	2%	27%	8%	100%	16%	36%	7%
5	9%	22%	26%	27%	44%	8%	48%	22%	0%	29%	27%	23%
10	19%	69%	62%	52%	30%	90%	25%	70%	0%	55%	37%	70%
Weighting Scheme												
vw	18%	17%	0%	10%	0%	86%	0%	20%	0%	0%	100%	14%
ew	82%	83%	100%	90%	100%	14%	100%	80%	100%	100%	0%	86%

- **Rebalancing Frequency:** Less frequent rebalancing schedules (quarterly and annual) are better for returns, alpha, Sharpe ratio, and information ratio, but worse for reducing volatility.

Monthly and quarterly rebalancing perform best in terms of maximum drawdown.

- **Portfolio Buckets:** Larger portfolios (e.g., terciles) are least frequent in the loser bucket and often among the most frequent in the winner bucket. Thus larger portfolios generally lead to improved performance.
- **Weighting Scheme:** As observed previously, the results for equally-weighted versus value-weighted portfolios are mixed. Equal-weighted portfolios increase the dispersion of results with regard to average returns, alpha, Sharpe ratio, and information ratio. However, for volatility, equally-weighted portfolios are the clear winners, whereas for maximum drawdown, value-weighted portfolios perform best.

Note that the results discussed above regarding the construction choices for winner and loser portfolios are robust over time and during recessions. In [Appendix E](#), we present and discuss results for the above winner loser analysis to check its robustness over time and during recession periods. The results of our time-series analyses align with the overall results presented here, reinforcing the robustness of our findings. Thus, we find that the long-term results are not significantly driven by single events, and the top (bottom) portfolio constructions consistently outperform (underperform) over time.

3.6 A Recommendation for Nonstandard Errors in Low-Risk Portfolios

As the final step in our analysis, we aim to recommend an attractive and robust construction methodology for low-risk portfolios. While we acknowledge that this ex-post approach is susceptible to a look-ahead bias, it nonetheless provides valuable insights into the potential upper limit of low-risk portfolio performance.

Based on our findings, we conclude that the performance of low-risk premia is highly sensitive to the construction choices made during the portfolio-building process. To enhance the robustness of our evaluation, we constructed an aggregated low-risk portfolio by selecting the most favorable variations at each decision point in the construction process. The return time series of this aggregated portfolio is computed as the equally-weighted net return time series of all portfolios deemed optimal according to our evaluation criteria.

For our aggregated portfolio, we adopt the following construction choices:

1. **Risk Estimator:** We select total and idiosyncratic volatility risk estimators, as they achieve

the highest Sharpe ratios while maintaining robust mid-range performance during crises.

2. **Portfolio Type:** We focus on long-only and limited-short strategies, given that most dollar-neutral long-short strategies yield negative Sharpe ratios. This approach also aligns with our emphasis on implementable portfolios for benchmarked managers.
3. **Size Exclusion:** Size filters at the 10% and 20% levels are applied, as portfolios with these filters deliver the highest Sharpe ratios and demonstrate robust performance during crises. However, this choice reduces the likelihood of extreme winners.
4. **Price Exclusion:** We adopt a stringent price filter, excluding stocks priced below 5 USD. This filter consistently outperforms less restrictive alternatives across all summary statistics, during crises, and over time.
5. **Lookback Window:** All available lookback periods are incorporated due to inconclusive results. While a 1-year lookback excels in crisis performance and information ratios, it produces more extreme summary statistics. Conversely, a 5-year lookback performs best over longer periods, with comparable Sharpe and information ratios across time.
6. **Rebalancing Frequency:** Quarterly and yearly rebalancing frequencies are selected for their superior Sharpe and information ratios, consistent performance across summary statistics, and robustness during crises and over time.
7. **Portfolio Buckets:** Portfolios are divided into three buckets, as this configuration achieves the highest Sharpe and information ratios, strong summary statistics, and solid long-term performance. However, it tends to produce more extreme outcomes during crises.
8. **Weighting Scheme:** Equal weighting is chosen for its superior Sharpe and information ratios, despite resulting in more extreme outcomes across summary statistics, during crises, and over time.

This analysis results in a total of 72 portfolios, forming the foundation of our aggregated portfolio. An illustration of our selected subset of variables is outlined in [Figure A.2](#). To ensure that the evaluation reflects realistic investment scenarios, we limit our focus to portfolios that account for transaction costs.

The cumulative returns of the aggregated portfolio (red), its underlying portfolios (gray), and the benchmark (black) are illustrated in [Figure 12](#). The aggregated portfolio significantly outperforms

the benchmark over the sample period, with much of the outperformance driven by the dot-com bubble in the early 2000s. Notably, while low-risk strategies underperformed during the buildup of the bubble, they recovered strongly thereafter. Additionally, the relatively low dispersion among the underlying portfolios suggests that nonstandard errors for this subset are minimal.

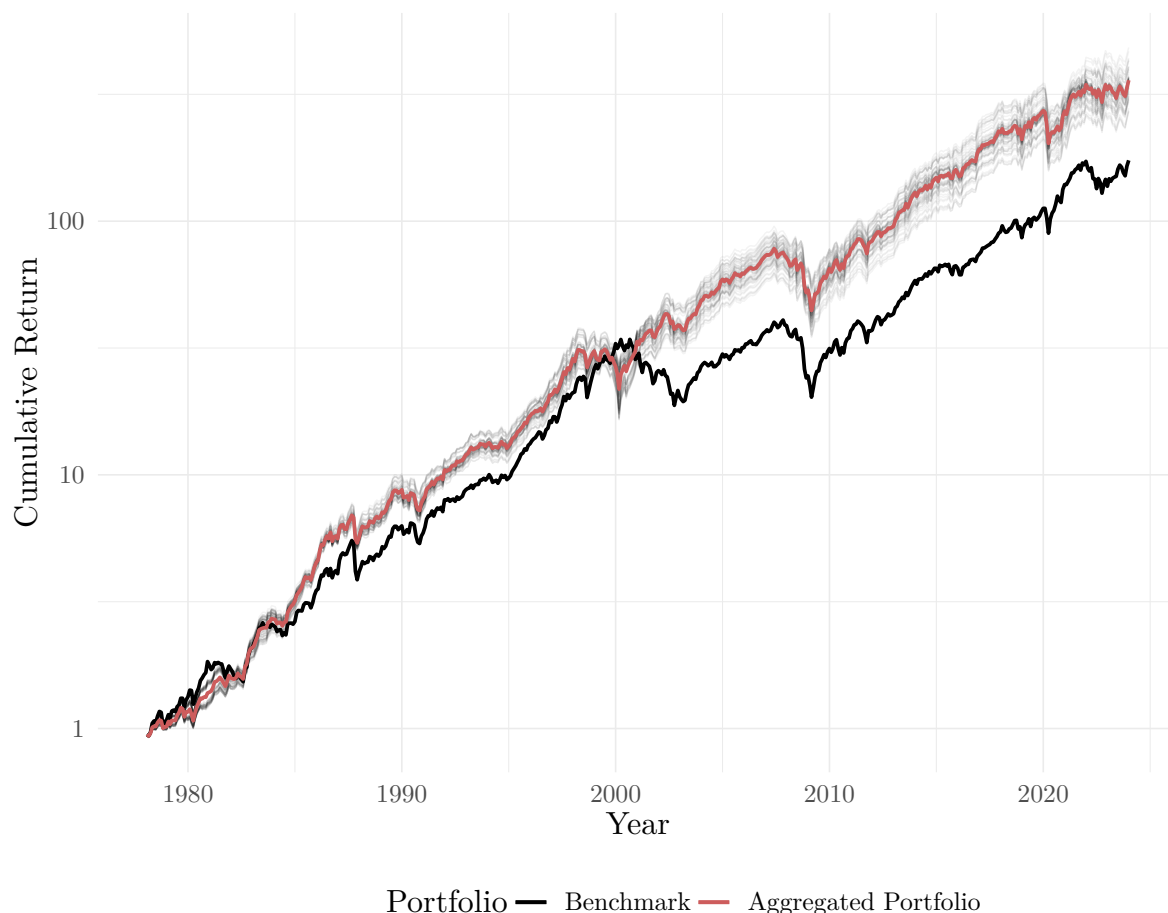


Figure 12: Cumulative returns of the benchmark and the aggregated portfolio with its underlying strategies

The figure illustrates cumulative returns for an aggregated portfolio (red), derived as the average of 72 individual portfolios (gray), alongside a value-weighted market benchmark (black). The 72 portfolios represent those with optimal decision variables identified in our analysis. These are all portfolios that are built with total or idiosyncratic volatility risk estimators, long-only or limited-short portfolio types, 10% or 20% size exclusion, a price filter of \$5, all lookback periods, quarterly or yearly rebalancing, three portfolio buckets, and equal weighting scheme. The cumulative returns are calculated based on 11,598 daily out-of-sample returns from January 1978 to December 2023 and presented on a log-10 y-axis.

The outperformance of the aggregated strategy is further examined in [Table 4](#), which presents selected performance metrics. The results indicate that the aggregated low-risk portfolio consistently outperforms the benchmark across all metrics. It achieves a higher average return and a lower stan-

dard deviation, resulting in a superior Sharpe ratio and a positive information ratio. Additionally, the aggregated portfolio exhibits a smaller maximum drawdown compared to the benchmark, underscoring its resilience during market downturns. We further evaluate whether the outperformance of the aggregated portfolio in terms of lower standard deviation and higher Sharpe ratio compared to the benchmark is statistically significant and economically meaningful. The **bold** numbers for standard deviation (Std.) and Sharpe ratio (SR) indicate that these differences are statistically significant at the 1% significance level for Std. and at the 5% significance level for SR.¹⁰

Table 4: Summary statistics of aggregated portfolio vs. benchmark

The table holds portfolio summary statistics for an aggregated portfolio, derived as the average of 72 individual portfolios and for a value-weighted market benchmark. The 72 portfolios represent those with optimal decision variables identified in our analysis. These are all portfolios that are built with total or idiosyncratic volatility risk estimators, long-only or limited-short portfolio types, 10% or 20% size exclusion, a price filter of \$5, all lookback periods, quarterly or yearly rebalancing, three portfolio buckets, and equal weighting scheme. The presented performance metrics are cumulative return, average return, average standard deviation, Sharpe ratio, information ratio, and maximum drawdown. The standard deviation and Sharpe ratio of the aggregated portfolio are printed in bold as they are statistically significantly different from those of the benchmark at the 1% and 5% levels, respectively. The performance metrics are calculated based on 11,598 daily out-of-sample returns from January 1978 to December 2023.

Portfolio	Cum.	Ret.	Std.	SR	IR	m. DD
Aggregated	358.43	14%	14%	0.68	0.13	49%
Benchmark	172.66	13%	17%	0.50		55%

Overall, this section demonstrates that, when managed effectively, low-risk strategies can achieve substantial outperformance relative to the market. Furthermore, it highlights that nonstandard errors have a small impact on well-managed portfolios that are diversified across the most attractive and robust portfolio construction choices.

4 Conclusion

This study provides a comprehensive examination of the impact of portfolio construction choices on the performance of low-risk portfolios. Our analysis highlights that the methodological decisions during the portfolio construction process significantly influence performance metrics and are crucial for optimizing investment outcomes.

¹⁰To test the outperformance of the aggregated portfolio over the benchmark a two-sided p -value for the null hypothesis of equal standard deviations, respectively equal Sharpe ratios, is obtained by the prewhitened HAC_{PW} method described in Ledoit and Wolf (2011, Section 3.1), respectively Ledoit and Wolf (2008, Section 3.1). As the out-of-sample size is very large at 11,598, there is no need to use the computationally more involved bootstrap method described in Ledoit and Wolf (2011, Section 3.2), respectively in Ledoit and Wolf (2008, Section 3.2), which is preferred for small sample sizes.

This paper makes several significant contributions to the literature. First, it systematically evaluates the influence of risk estimators on portfolio performance, demonstrating that the choice of estimation methodology—specifically, the risk estimator choice together with its lookback window length—accounts for a big share of performance variability. Second, the study investigates the role of transaction costs and performance evaluation metrics, shedding light on the trade-offs between return enhancement and cost efficiency—areas that have been largely neglected in the current literature Hasler (2023), Walter et al. (2023), and Soebhag et al. (2024). We demonstrate that these enhancements are essential, as both significantly shape preferences and inferences regarding portfolio construction choices. Third, our work contributes by focusing on realistic, implementable portfolios that address practical challenges such as transaction costs and leverage constraints. By bridging the gap between theory and practice, it offers valuable insights for both academics and practitioners.

Overall, this study highlights the dispersion in decision choices for sorting variable computation and portfolio construction. It underscores the risk of nonstandard errors while demonstrating how construction choices can be optimized to maximize the benefits of the low-risk anomaly. Our results indicate that optimal low-risk portfolios should also be constructed in a low-risk manner. This means that choices that arguable reduce risk, such as the exclusion of small and penny stocks, avoidance of extreme portfolio buckets, long rebalancing cycles, and little short selling lead to consistent performance over time that is not driven by single events.

Our findings provide valuable guidance for practitioners seeking to implement low-risk strategies and highlight opportunities for further research. Future studies should extend the impact analysis of practical low-risk portfolio construction choices by exploring more efficient and optimized portfolios, such as the global minimum variance (GMV), equal risk contribution (ERC), or hierarchical risk parity (HRP) portfolio. Whereas most sorting portfolios rely solely on stock-specific risk measures, efficient portfolios also incorporate cross-sectional information through the covariance matrix. Since large-dimensional portfolio optimization and covariance matrix estimation are more prone to errors compared to simple univariate risk sorting, the choice of portfolio optimization and covariance matrix estimators introduces even greater variability in the results. However, the need for optimization routines significantly increases computational requirements, making the inclusion of such portfolios challenging (if not impossible) when evaluating nonstandard errors.

References

- Abdi, F., & Ranaldo, A. (2017). [A simple estimation of bid-Ask spreads from daily close, high, and low prices](#). *Review of Financial Studies*, 30(12), 4437–4480.
- Ang, A., Hodrick, R. J., Xing, Y., & Zhang, X. (2006). [The Cross-Section of Volatility and Expected Returns](#). *The Journal of Finance*, 61(1), 259–299.
- Ang, A., Hodrick, R. J., Xing, Y., & Zhang, X. (2009). [High idiosyncratic volatility and low returns: International and further U.S. evidence](#). *Journal of Financial Economics*, 91(1), 1–23.
- Asness, C. S., Frazzini, A., & Pedersen, L. H. (2014). [Low-risk investing without industry bets](#). *Financial Analysts Journal*, 70(4), 24–41.
- Baker, M., Bradley, B., & Wurgler, J. (2011). [Benchmarks as Limits to Arbitrage: Understanding the Low Volatility Anomaly](#). *Financial Analysts Journal*, 67(1), 40–54.
- Baker, N. L., & Haugen, R. A. (2012). [Low Risk Stocks Outperform within All Observable Markets of the World](#). *SSRN Electronic Journal*.
- Blitz, D., Howard, C., Huang, D., & Jansen, M. (2024). [Low-Risk Alpha Without Low Beta](#). *SSRN Electronic Journal*.
- Blitz, D., Pang, J., & van Vliet, P. (2013). [The volatility effect in emerging markets](#). *Emerging Markets Review*, 16(1), 31–45.
- Blitz, D., van Vliet, P., & Baltussen, G. (2023). [The Volatility Effect Revisited](#). *The Journal of Portfolio Management*, 46(2), 45–63.
- Blitz, D., & van Vliet, P. (2007). [The volatility effect](#). *The Journal of Portfolio Management*, 34(1), 102–113.
- Blitz, D., & Vidojevic, M. (2017). [The profitability of low-volatility](#). *Journal of Empirical Finance*, 43, 33–42.
- Bollerslev, T. (1986). [Generalized Autoregressive Conditional Heteroskedasticity](#). *Journal of Econometrics*, 31(3), 307–327.
- Bryzgalova, S., Lerner, S., Lettau, M., & Pelger, M. (2024). [Missing Financial Data](#). *The Review of Financial Studies*, 00, 1–80.
- Cochrane, J. H. (2011). [Presidential address: Discount rates](#). *The Journal of Finance*, 66(4), 1047–1108.
- De Nard, G. (2022). [Oops! I shrunk the sample covariance matrix again: Blockbuster meets shrinkage](#). *Journal of Financial Econometrics*, 20(4), 569–611.

- Detzel, A., Novy-Marx, R., & Velikov, M. (2023). [Model Comparison with Transaction Costs](#). *Journal of Finance*, 78(3), 1227–1831.
- Dimson, E., Marsh, P., & Staunton, M. (2017). [Factor-Based Investing: The Long-Term Evidence](#). *The Journal of Portfolio Management*, 43(5), 15–37.
- Engle, R. F. (2001). [GARCH 101: The use of ARCH/GARCH models in applied econometrics](#). *Journal of Economic Perspectives*, 15(4), 157–168.
- Fama, E. F., & French, K. R. (1993). [Common risk factors in the returns on stocks and bonds](#). *Journal of Financial Economics*, 33(1), 3–56.
- Fama, E. F., & French, K. R. (2015). [A five-factor asset pricing model](#). *Journal of Financial Economics*, 116(1), 1–22.
- Fama, E. F., & French, K. R. (2016). Dissecting Anomalies with a Five-Factor Model. *Review of Financial Studies*, 29(1), 69–103.
- Fama, E. F., & MacBeth, J. (1973). [Risk, Return and Equilibrium: Empirical Tests](#). *Journal of Political Economy*, 81(3), 697–637.
- Frazzini, A., & Pedersen, L. H. (2014). [Betting Against Beta](#). *Journal of Financial Economics*, 111(1), 1–25.
- Harvey, C. R. (2017). [Presidential address: The scientific outlook in financial economics](#). *The Journal of Finance*, 72, 1399–1440.
- Hasler, M. (2023). [Looking Under the Hood of Data-Mining](#). *SSRN Electronic Journal*.
- Hou, K., Xue, C., & Zhang, L. (2020). [Replicating Anomalies](#). *Review of Financial Studies*, 33(5), 2019–2133.
- Invesco. (2023). *2023 Invesco Global Systematic Investing Study*.
- Invesco. (2024). *2024 Invesco Global Systematic Investing Study*.
- Jagannathan, R., & Ma, T. (2003). [Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps](#). *Journal of Finance*, 54(4), 1651–1684.
- Jensen, T. I., Kelly, B. T., & Pedersen, L. H. (2023). [Is There a Replication Crisis in Finance?](#) *The Journal of Finance*, 78(5), 2465–2518.
- Ledoit, O., & Wolf, M. (2011). [Robust performance hypothesis testing with the variance](#). *Wilmott Magazine*, September, 86–89.
- Ledoit, O., & Wolf, M. (2004). [A well-conditioned estimator for large-dimensional covariance matrices](#). *Journal of Multivariate Analysis*, 88(2), 365–411.

- Ledoit, O., & Wolf, M. (2008). [Robust performance hypothesis testing with the Sharpe ratio](#). *Journal of Empirical Finance*, 15(5), 850–859.
- Menkveld, A. J., Dreber, A., Holzmeister, F., Huber, J., Johannesson, M., Kirchler, M., Neus, S., Razen, M., & Weitzel, U. (2024). [Non-Standard Errors *](#). *The Journal of Finance*, 79(3), 2339–2390.
- Mina, J., & Xiao, J. Y. (2001). Return to RiskMetrics: The Evolution of a Standard.
- Novy-Marx, R. (2015). [How Can a Q-Theoretic Model Price Momentum?](#) *NBER Working Paper Series*, 20985.
- Novy-Marx, R., & Velikov, M. (2016). [A Taxonomy of Anomalies and Their Trading Costs](#). *Review of Financial Studies*, 29(1), 104–147.
- Novy-Marx, R., & Velikov, M. (2022). [Betting against betting against beta](#). *Journal of Financial Economics*, 143(1), 80–106.
- RiskMetrics. (1996). *RiskMetricsTM—Technical Document*.
- Soebhag, A., Van Vliet, B., & Verwijmeren, P. (2024). [Non-Standard Errors in Asset Pricing: Mind Your Sorts](#). *Journal of Empirical Finance*, 78, 101517.
- Traut, J. (2023). [What we know about the low-risk anomaly: a literature review](#). *Financial Markets and Portfolio Management*, 37(3), 297–324.
- van Vliet, P., Blitz, D., & van der Grient, B. (2011). [Is the relation between volatility and expected stock returns positive, flat or negative?](#) *SSRN working paper of Robeco*.
- Walter, D., Weber, R., & Weiss, P. (2023). [Methodological uncertainty in portfolio sorts](#). *SSRN Electronic Journal*.
- Wiest, T. (2023). [Momentum: what do we know 30 years after Jegadeesh and Titman’s seminal paper?](#) *Financial Markets and Portfolio Management*, 37(1), 95–114.
- Zumbach, G. O. (2007). [The Riskmetrics 2006 Methodology](#). *SSRN Electronic Journal*.

Appendices

A Additional Figures and Tables

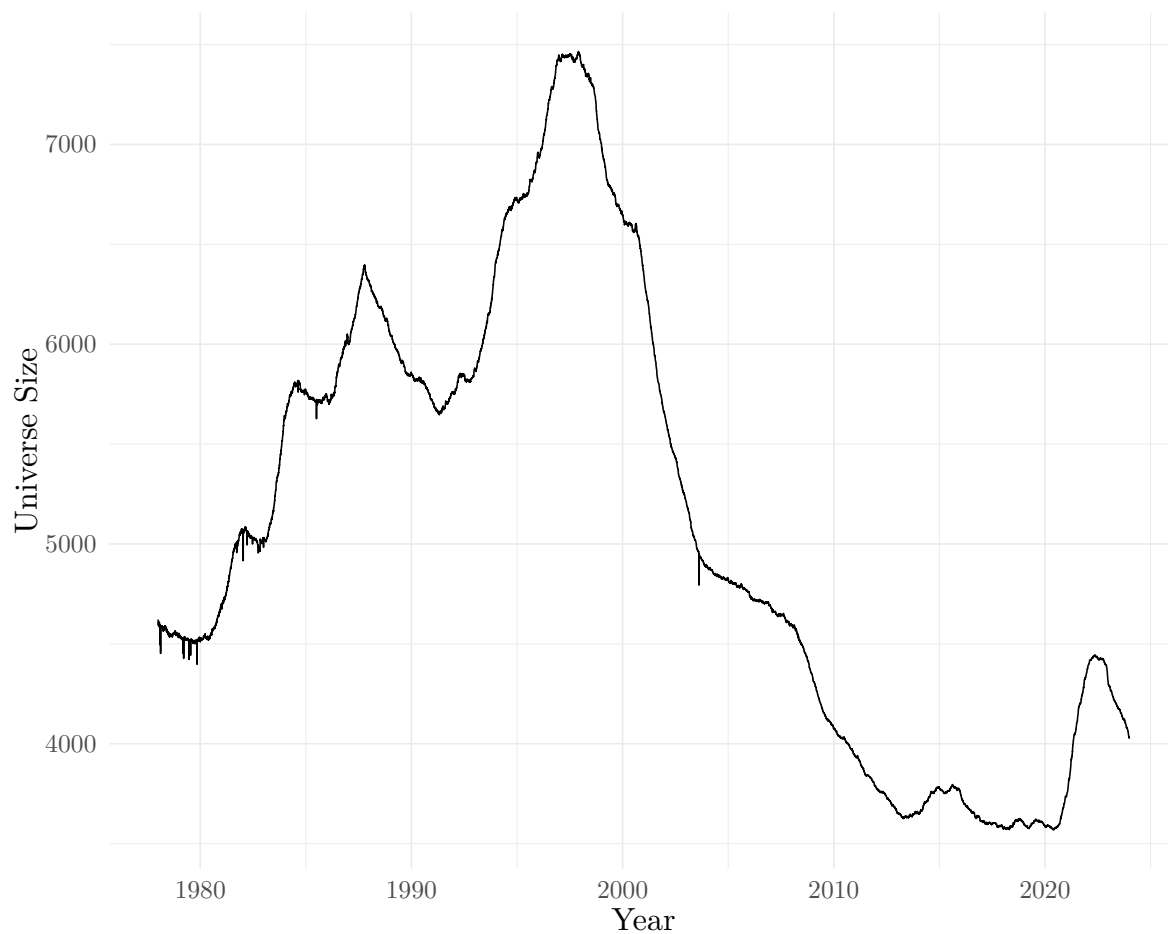


Figure A.1: Size of the eligible out-of-sample universe over time

The figure displays the total number of stocks that are available in our eligible universe for portfolio construction at each day between January 1978 and December 2023.

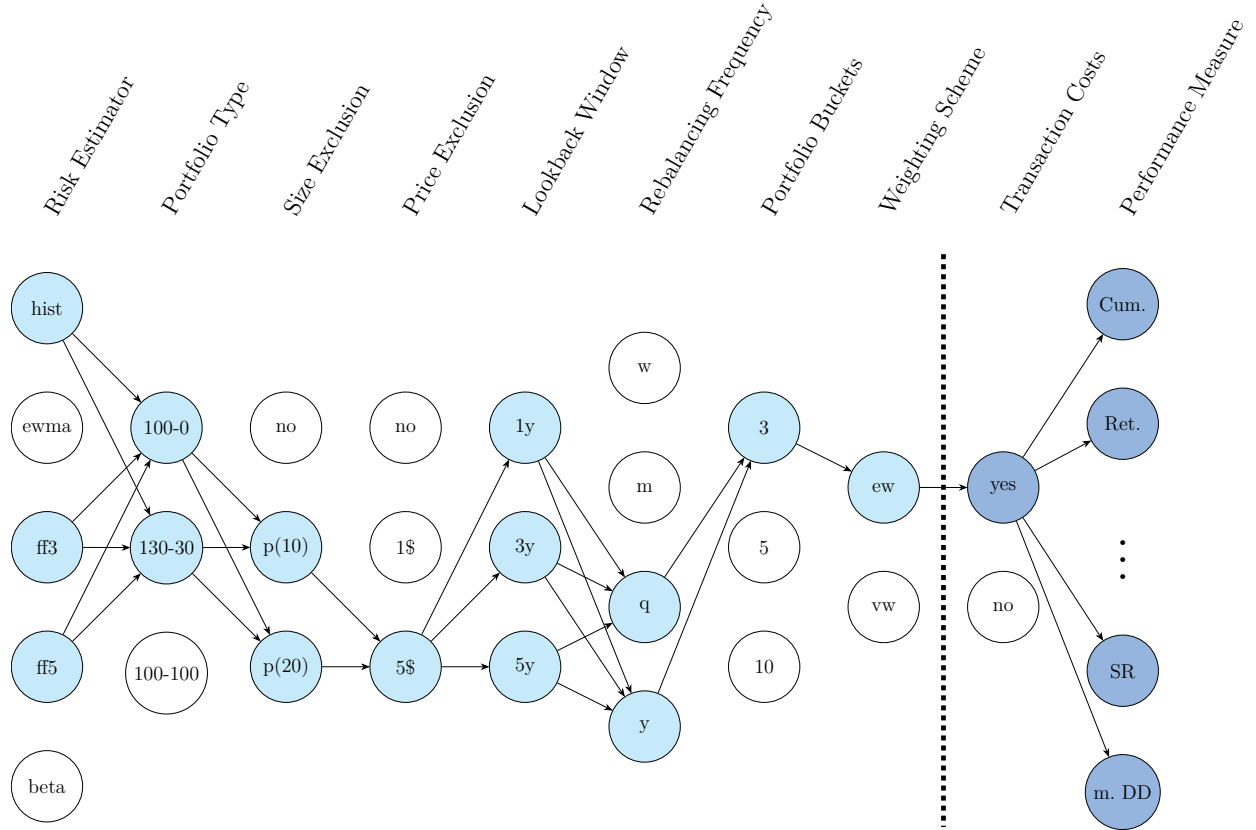


Figure A.2: Decision tree of selected portfolio construction choices for aggregated portfolio

The figure illustrates our selected portfolio construction choices for an aggregated portfolio across all decision paths. The selected choices per decision forks are as follows: risk estimation using historical volatility or idiosyncratic volatility with respect to the Fama and French 3-factor or 5-factor model, long-only (100-0) or limited-short (130-30) portfolios, exclusion of small stocks that are smaller than 10% or 20% of stocks regarding market equity, exclusion of stocks with prices below \$5, estimation of risk using the past one, three, or five years of return data, sorting stocks into three portfolio buckets, and equally weighting stocks. With our evaluation we focus on metrics after the deduction of transaction costs. In total our aggregated portfolio includes 72 portfolios.

	Top 5					Bottom 5					Top 5					Bottom 5				
	Top	2	3	4	5	5	4	3	2	Bot.	Top	2	3	4	5	5	4	3	2	Bot.
	100-0 without transaction costs										100-0 with transaction costs									
SR	1.75	1.74	1.72	1.62	1.59	0.45	0.45	0.45	0.39	0.34	0.78	0.75	0.75	0.75	0.75	−0.17	−0.42	−0.51	−0.54	−0.55
Estimator	beta	beta	beta	beta	beta	beta	beta	beta	beta	beta	beta	ewma	hist	hist	ewma	beta	beta	beta	beta	beta
Lookback	1y	3y	5y	5y	1y	5y	3y	5y	1y	1y	5y	5y	3y	3y	5y	1y	1y	1y	1y	1y
Weighting	ew	ew	ew	ew	ew	vw	vw	vw	vw	vw	ew	ew	ew	ew	ew	ew	ew	ew	ew	ew
Rebalancing	w	w	w	w	w	y	y	y	y	y	y	w	w	w	q	m	m	w	w	m
Port. Buckets	10	10	10	10	5	10	10	10	10	10	10	10	10	10	10	5	10	10	10	10
Size	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no
Price	no	no	no	1\$	1\$	no	no	no	1\$	no	5\$	5\$	1\$	1\$	1\$	no	1\$	1\$	no	no
	130-30 without transaction costs										130-30 with transaction costs									
SR	2.07	2.02	1.99	1.97	1.96	0.29	0.28	0.28	0.18	0.17	0.88	0.86	0.86	0.86	0.85	−0.63	−0.69	−0.81	−0.97	−1.00
Estimator	beta	beta	beta	beta	beta	beta	beta	beta	beta	beta	ewma	ewma	ewma	ewma	ewma	ewma	ewma	ewma	ewma	ewma
Lookback	1y	3y	5y	5y	3y	5y	3y	3y	1y	1y	5y	1y	1y	1y	1y	1y	1y	1y	1y	1y
Weighting	ew	ew	ew	ew	ew	vw	vw	vw	vw	vw	ew	ew	ew	ew	ew	ew	ew	ew	ew	ew
Rebalancing	w	w	w	w	w	y	y	y	y	y	q	q	y	q	y	w	m	m	w	w
Port. Buckets	5	5	5	10	10	10	10	10	10	10	5	10	10	5	5	10	10	10	10	10
Size	no	no	no	no	no	no	no	no	no	no	no	no	no	$p(10)$	no	no	no	no	no	no
Price	no	no	no	no	no	1\$	no	1\$	1\$	no	5\$	5\$	5\$	5\$	5\$	no	1\$	no	1\$	no
	100-100 without transaction costs										100-100 with transaction costs									
SR	0.39	0.34	0.32	0.30	0.30	−0.73	−0.73	−0.74	−0.77	−0.78	0.22	0.22	0.20	0.18	0.18	−1.84	−1.85	−1.89	−1.89	−2.03
Estimator	ewma	ewma	ff3	ff5	ewma	ff3	ff5	ff3	ff5	ewma	hist	ewma	ewma	ewma	ewma	ff3	ff5	ff3	ff5	ewma
Lookback	1y	1y	1y	1y	1y	5y	5y	5y	5y	5y	3y	5y	1y	1y	1y	3y	3y	1y	1y	1y
Weighting	ew	vw	vw	vw	ew	ew	ew	ew	ew	ew	ew	ew	vw	vw	ew	ew	ew	ew	ew	ew
Rebalancing	m	m	w	w	m	w	w	w	w	w	y	y	q	y	y	w	w	w	w	w
Port. Buckets	5	5	5	10	10	5	5	3	3	10	5	10	10	5	5	10	10	10	10	10
Size	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no
Price	1\$	no	1\$	no	5\$	no	no	no	no	no	no	no	no	1\$	1\$	no	no	no	no	no

Table A.1: Characteristics of top and bottom portfolios

This table shows the construction choices of the top five and bottom five portfolios in term of average Sharpe ratio. The analysis is conducted separately for each portfolio type: long-only (100-0) on top, limited-short (130-30) in the middle, and long-short (100-100) on the bottom as well as without transaction costs (left) and with transaction costs (right). The Sharpe ratio is calculated based on 11,598 daily out-of-sample returns from January 1978 to December 2023.

Table A.2: Mean absolute differences at decision forks with respect to mean returns

This table presents the average mean absolute return differences p.a., as defined in Equation 12, for each decision fork. For each fork, we compare average return differences between matched decision paths that vary only in the specific decision under consideration. These mean absolute differences are calculated individually for each risk estimator and averaged across all matched paths, as well as jointly for all risk estimators. The columns report averages for all risk estimators combined (Overall) and for individual risk estimators separately. The forks (rows) are ordered by their overall impact in descending order. The analysis is carried out separately for long-only (100-0), limited-short (130-30), and long-short (100-100) portfolios. Note that the impact of the risk estimator can only be calculated for the joint analysis across all estimators.

Portfolio	Fork	Overall	Hist.	EWMA	FF3	FF5	Beta
100-0	Weighting	1.39%	1.29%	1.26%	1.23%	1.23%	1.93%
	Risk Estimator	0.90%					
	Trans. Costs	0.87%	0.47%	0.83%	0.38%	0.37%	2.32%
	Port. Buckets	0.47%	0.51%	0.54%	0.31%	0.36%	0.64%
	Rebalancing	0.31%	0.18%	0.29%	0.16%	0.15%	0.79%
	Lookback	0.30%	0.18%	0.27%	0.19%	0.21%	0.64%
	Drop Size	0.29%	0.19%	0.20%	0.14%	0.15%	0.79%
	Drop Price	0.10%	0.06%	0.06%	0.05%	0.05%	0.27%
130-30	Trans. Costs	1.95%	1.46%	2.17%	1.39%	1.38%	3.36%
	Weighting	1.69%	1.50%	1.56%	1.46%	1.46%	2.46%
	Risk Estimator	1.33%					
	Drop Size	0.75%	0.68%	0.75%	0.63%	0.63%	1.05%
	Rebalancing	0.68%	0.51%	0.83%	0.50%	0.50%	1.06%
	Lookback	0.67%	0.55%	0.70%	0.63%	0.65%	0.82%
	Port. Buckets	0.55%	0.42%	0.52%	0.45%	0.45%	0.91%
	Drop Price	0.27%	0.25%	0.28%	0.24%	0.24%	0.36%
100-100	Weighting	5.23%	6.39%	5.67%	5.86%	5.51%	2.75%
	Trans. Costs	3.95%	3.70%	4.86%	3.73%	3.74%	3.70%
	Drop Size	3.90%	4.96%	4.20%	4.45%	4.22%	1.63%
	Port. Buckets	3.38%	3.93%	3.74%	3.70%	3.32%	2.21%
	Risk Estimator	3.34%					
	Rebalancing	3.15%	3.64%	3.77%	3.14%	2.88%	2.33%
	Lookback	2.79%	3.57%	2.98%	2.92%	2.74%	1.76%
	Drop Price	2.47%	3.06%	2.76%	3.02%	2.76%	0.74%

Table A.3: Mean absolute differences at decision forks with respect to standard deviation

This table presents the average mean absolute standard deviation differences p.a., as defined in Equation 12, for each decision fork. For each fork, we compare differences in standard deviation between matched decision paths that vary only in the specific decision under consideration. These mean absolute differences are calculated individually for each risk estimator and averaged across all matched paths, as well as jointly for all risk estimators. The columns report averages for all risk estimators combined (Overall) and for individual risk estimators separately. The forks (rows) are ordered by their overall impact in descending order. The analysis is carried out separately for long-only (100-0), limited-short (130-30), and long-short (100-100) portfolios. Note that the impact of the risk estimator can only be calculated for the joint analysis across all estimators.

Portfolio	Fork	Overall	Hist.	EWMA	FF3	FF5	Beta
100-0	Risk Estimator	1.71%					
	Weighting	1.30%	1.54%	1.58%	1.40%	1.38%	0.58%
	Port. Buckets	0.73%	1.03%	1.04%	0.44%	0.44%	0.72%
	Drop Size	0.23%	0.22%	0.21%	0.11%	0.11%	0.52%
	Lookback	0.18%	0.25%	0.20%	0.15%	0.15%	0.15%
	Rebalancing	0.15%	0.16%	0.27%	0.07%	0.07%	0.19%
	Drop Price	0.08%	0.09%	0.09%	0.05%	0.05%	0.14%
	Trans. Costs	0.01%	0.00%	0.01%	0.00%	0.00%	0.06%
130-30	Weighting	1.92%	2.22%	2.51%	1.61%	1.54%	1.69%
	Risk Estimator	1.83%					
	Drop Size	0.47%	0.45%	0.44%	0.45%	0.46%	0.54%
	Port. Buckets	0.44%	0.32%	0.33%	0.33%	0.30%	0.92%
	Rebalancing	0.25%	0.28%	0.38%	0.17%	0.17%	0.26%
	Lookback	0.24%	0.28%	0.23%	0.23%	0.23%	0.21%
	Drop Price	0.23%	0.24%	0.25%	0.21%	0.21%	0.23%
	Trans. Costs	0.05%	0.02%	0.04%	0.02%	0.02%	0.12%
100-100	Port. Buckets	10.22%	11.64%	11.55%	10.15%	9.11%	8.65%
	Weighting	8.95%	12.14%	10.45%	9.82%	8.70%	3.63%
	Risk Estimator	6.71%					
	Drop Size	6.17%	7.94%	7.79%	6.93%	6.25%	1.94%
	Rebalancing	5.32%	6.66%	5.92%	5.75%	5.25%	3.03%
	Lookback	5.31%	8.33%	5.17%	4.88%	4.93%	3.25%
	Drop Price	5.23%	6.69%	6.35%	6.52%	5.35%	1.23%
	Trans. Costs	0.13%	0.11%	0.20%	0.12%	0.12%	0.10%

Table A.4: Mean absolute differences at decision forks with respect to maximum drawdown

This table presents the average mean absolute maximum drawdown differences, as defined in Equation 12, for each decision fork. For each fork, we compare differences in maximum drawdowns between matched decision paths that vary only in the specific decision under consideration. These mean absolute differences are calculated individually for each risk estimator and averaged across all matched paths, as well as jointly for all risk estimators. The columns report averages for all risk estimators combined (Overall) and for individual risk estimators separately. The forks (rows) are ordered by their overall impact in descending order. The analysis is carried out separately for long-only (100-0), limited-short (130-30), and long-short (100-100) portfolios. Note that the impact of the risk estimator can only be calculated for the joint analysis across all estimators.

Portfolio	Fork	Overall	Hist.	EWMA	FF3	FF5	Beta
100-0	Risk Estimator	4.84%					
	Weighting	3.89%	1.86%	2.21%	1.30%	1.27%	12.82%
	Port. Buckets	3.10%	3.63%	3.01%	2.85%	2.83%	3.19%
	Drop Size	2.13%	1.56%	1.47%	1.15%	1.15%	5.30%
	Lookback	1.28%	1.22%	0.73%	1.43%	1.32%	1.68%
	Rebalancing	1.20%	1.01%	1.76%	0.73%	0.66%	1.81%
	Drop Price	0.59%	0.42%	0.42%	0.36%	0.36%	1.36%
	Trans. Costs	0.45%	0.22%	0.40%	0.14%	0.14%	1.36%
130-30	Weighting	6.62%	4.96%	5.97%	6.17%	6.65%	9.36%
	Port. Buckets	5.30%	5.35%	6.83%	3.72%	3.53%	7.08%
	Risk Estimator	4.72%					
	Drop Size	3.56%	3.14%	3.33%	2.68%	2.59%	6.06%
	Lookback	3.18%	2.85%	2.74%	3.21%	3.13%	3.95%
	Trans. Costs	3.17%	2.23%	4.11%	1.76%	1.71%	6.05%
	Rebalancing	2.90%	2.64%	3.13%	2.48%	2.44%	3.82%
	Drop Price	1.61%	1.60%	1.84%	1.47%	1.45%	1.67%
100-100	Trans. Costs	5.12%	3.90%	4.40%	4.15%	4.19%	8.97%
	Port. Buckets	4.50%	3.54%	4.27%	3.77%	3.86%	7.06%
	Weighting	4.32%	3.88%	3.92%	4.20%	4.29%	5.31%
	Risk Estimator	3.44%					
	Rebalancing	3.18%	2.84%	3.86%	2.66%	2.59%	3.96%
	Drop Size	2.65%	2.48%	2.56%	2.79%	2.92%	2.51%
	Lookback	2.09%	1.72%	1.92%	2.21%	2.36%	2.27%
	Drop Price	1.47%	1.33%	1.57%	1.57%	1.61%	1.26%

B Transaction Costs Calculation

We calculate turnover and transaction costs following Novy-Marx and Velikov (2016). The turnover at time t of stock i in portfolio p is denoted as $to_{i,p,t}$. The total transaction costs of portfolio p at time t is denoted as $TC_{p,t}$. Both can be expressed as

$$to_{i,p,t} = |w_{i,p,t} - \tilde{w}_{i,p,t-}| , \quad (B.1)$$

$$TC_{p,t} = \sum_{i=1}^{N_{p,t}} to_{i,p,t} c_{i,p,t} , \quad (B.2)$$

where $c_{i,t}$ represents the one-way transaction cost for stock i at time t , $N_{p,t}$ is the total number of stocks at time t in p , and $w_{i,p,t}$ is the portfolio weight of stock i after rebalancing at time t , while $\tilde{w}_{i,p,t-}$ denotes the weight of stock i before rebalancing.

The net-of-costs return of the portfolio p , denoted as r_p^{net} , is given by:

$$r_{p,t}^{net} = r_{p,t}^{gross} - TC_{p,t} , \quad (B.3)$$

where $r_{p,t}^{gross}$ is the return of p that ignores costs.

Following the literature on transaction costs, we calculate $c_{i,t}$ as half spreads. Our primary measure for spreads are quoted spread estimates following Chung and Zhang (2014). We discard days with non-positive close, bid, or ask prices. Additionally, we ensure that bid-ask spreads are non-negative for each observation. The relative bid-ask half-spreads, $s_{i,m,t}^Q$, are given by:

$$s_{i,m,t}^Q = \frac{A_{i,m,t} - B_{i,m,t}}{M_{i,m,t}} , \quad (B.4)$$

where $A_{i,m,t}$ and $B_{i,m,t}$ are the closing ask and bid prices quoted on day t of month m for stock i . We denote the prevailing end-of-day mid-quote as $M_{i,m,t} = \frac{A_{i,m,t} + B_{i,m,t}}{2}$. Following Chung and Zhang (2014), we then calculate $s_{i,m}^Q$ as the average of $s_{i,m,t}^Q$ estimates over month m , after discarding days with spreads exceeding 50% of the mid-quote. Furthermore, we treat $s_{i,m}^Q$ as missing if we have fewer than 12 valid observations for a given month.

We compute a second set of effective spread estimates, $s_{i,m}^{CHL}$, using the methodology proposed by Abdi and Rinaldo (2017). As recommended by them, we utilize the 2-day corrected version of their estimator. We discard observations with non-positive close, high, or low prices, and exclude stock-months with fewer than 12 valid observations. The proportional cost estimator is given by:

$$s_{i,m}^{CHL} = \frac{1}{T_m} \sum_{t=1}^{T_m} \sqrt{\max \{ (p_{i,m,t} - \eta_{i,m,t})(p_{i,m,t} - \eta_{i,m,t+1}), 0 \}} , \quad (\text{B.5})$$

where $p_{i,m,t}$ and $\eta_{i,m,t}$ represent the log of the closing price and the log mid-range, respectively, with $\eta_{i,m,t} = \frac{\log(H_{i,m,t}) + \log(L_{i,m,t})}{2}$ on day t . If the leading mid-range $\eta_{i,m,t+1}$ is missing, we substitute it with the prevailing log midpoint, as suggested by Abdi and Rinaldo (2017). Furthermore, if after the prevailing imputation the price, high, or low are either missing or zero, we see a day with zero volume, or the high equals the low, we treat the respective day as missing.

For all stocks, we set $s_{i,m}$ equal to $s_{i,m}^Q$ if available, and use the CHL estimator $s_{i,m}^{CHL}$ otherwise. For all stocks where $s_{i,m}$ is missing, we employ the non-parametric method proposed by Novy-Marx and Velikov (2016). We impute missing $s_{i,m}$ values with the respective $s_{j,m}$ of stock j , which minimizes the following distance:

$$\sqrt{(\text{rank}(ME)_{i,m} - \text{rank}(ME)_{j,m})^2 + (\text{rank}(FF3IVOL)_{i,m} - \text{rank}(FF3IVOL)_{j,m})^2} , \quad (\text{B.6})$$

where ME represents market equity, and $FF3IVOL$ refers to the idiosyncratic volatility derived from Fama and French (1993) 3-factor model regressions, estimated using three months of daily data. Because we observe some outliers in our spread estimators, we winsorize our $s_{i,m}^Q$ and $s_{i,m}^{CHL}$ estimators at the 99.5 percentile before we perform the imputation. Our final cost estimate is calculated as half spread as

$$c_{i,m} = \frac{s_{i,m}}{2} . \quad (\text{B.7})$$

C Cumulative Returns for Long-Short Portfolios

We define the cumulative return, C , of a long-short portfolio as

$$C = \prod_{t=1}^T (1 + r_{L,t} - r_{S,t}) - 1, \quad (\text{C.1})$$

where $r_{L,t}$ and $r_{S,t}$ represent the daily returns on the long and short portfolios, respectively. This expression captures the cumulative performance of a strategy that initially buys \$1 of the long portfolio and short-sells \$1 of the short portfolio. Motivated by regulation, the strategy posts cash collateral equal to 50% of the total equity positions in a non-interest-bearing margin account.

Alternatively, this can be viewed as the performance of a trader’s book when the margin account earns the risk-free rate, but the trader’s firm charges for the use of capital at the same rate. A common alternative approach, as discussed by Daniel and Moskowitz (2016), assumes that investors do not pay fees and thus earn the risk-free rate. However, we prefer our formulation because it more accurately reflects economic profitability and avoids inflating performance in high-inflation environments.

Furthermore, it is important to highlight that we allow portfolio weights to float between rebalancing dates relative to the total portfolio value, including the initial cash value. As a result, if the long and short portfolios develop differently, their total value no longer sums to zero. In contrast, most of the literature on low-risk investing evaluates long-short portfolios by simply subtracting the return of the short portfolio from that of the long portfolio (see Traut, 2023, for an overview). This approach effectively rebalances both portfolios to have the same value at each return observation—an assumption that is unrealistic for real-world implementations.¹¹

Our approach ensures that long-short portfolios remain directly comparable to their long-only and limited-short counterparts, as portfolio weights for all portfolios are allowed to float freely between rebalancing dates. However, this approach also implies that under highly unfavorable market conditions, losses can exceed the posted cash collateral ($C_t \leq -1$) of the long-short portfolio, potentially leading to a default. As illustrated in Figure C.1, this happens for roughly 7% of all

¹¹For example, if the long portfolio doubles in value while the short portfolio remains unchanged, the investor would naturally hold a net long position. Our methodology accounts for this dynamic, whereas a simple return subtraction would incorrectly assume equal portfolio values in the next period.

long-short portfolios of which most fail in the early 2000s, which was a very unfavorable period for low-risk strategies (evident from [Figure 12](#)). It is worth noting that we calculate the performance measures (like average return and volatility) for all defaulting portfolios for the time before default. For all other portfolios, we use the full sample period.

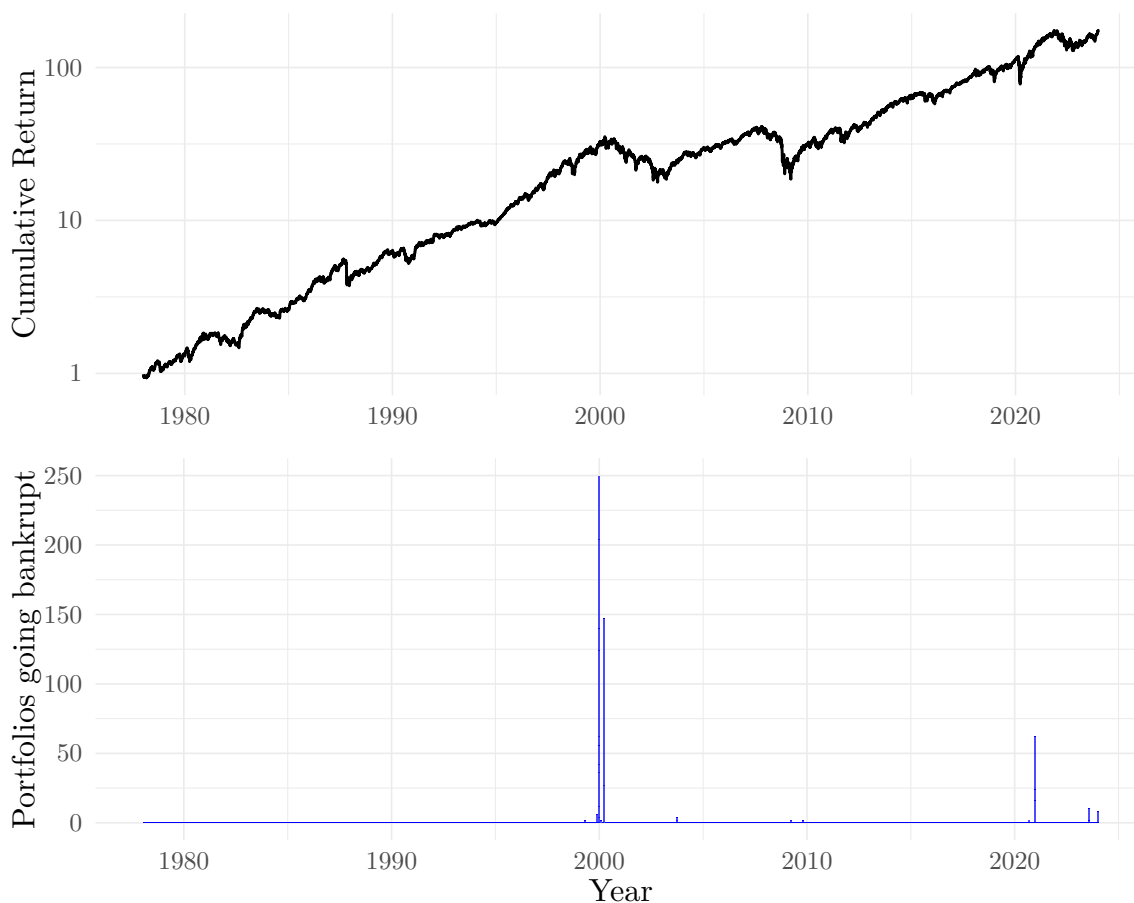


Figure C.1: Long-short strategies going bankrupt over time

The figure displays the cumulative return of our benchmark, a value-weighted market portfolio, in the upper plot, and the number of our long-short portfolios that default over time in the lower plot. A default is defined as a portfolio experiencing a cumulative return of -100% or lower. Both analysis are calculated based on 11,598 daily out-of-sample returns from January 1978 to December 2023.

[Figure C.2](#) presents the characteristic split of long-short portfolios that default in our analysis. It is clear that most of the defaulting portfolios are rebalanced yearly. This is expected, as the long and short legs of these portfolios remain unadjusted for a longer period, allowing for a greater accumulation of losses that eventually exceed the posted cash collateral. Additionally, it is evident that more risky portfolios, particularly those without a size or price filter and those in more extreme

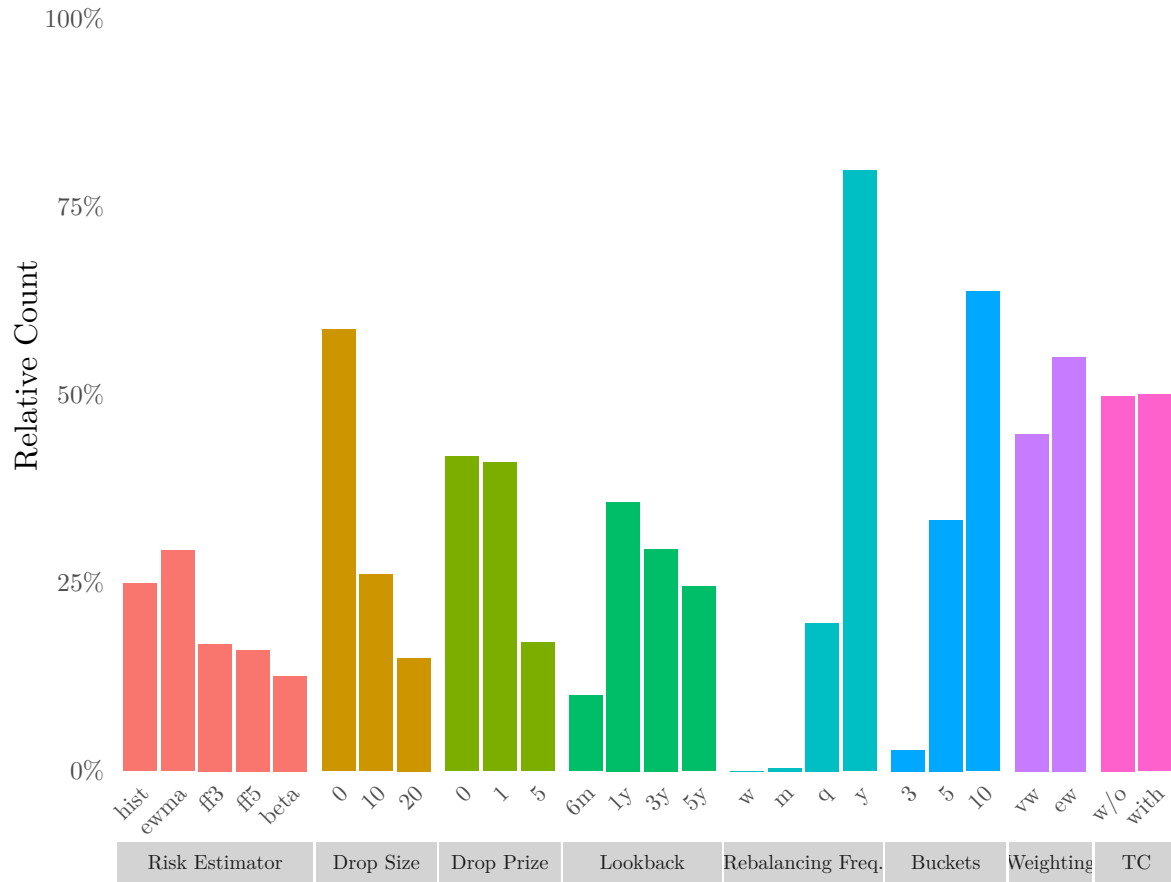


Figure C.2: Relative appearance of decision variables in defaulted long-short portfolios

The figure illustrates the relative occurrence of decision variables in long-short portfolios that default over our sample period. A default is defined as a portfolio experiencing a cumulative return of -100% or lower. The relative frequency is calculated for each decision fork, distinguished by different colors. Cumulative returns are calculated based on 11,598 daily out-of-sample returns from January 1978 to December 2023.

portfolio buckets, are more prone to default.

D EWMA RiskMetrics Specification

In this section we formalize how we match our selection of λ with the respective lookback periods and demonstrate why the recommendation of $\lambda = 0.94$ for daily data (RiskMetrics, 1996; Mina & Xiao, 2001; Zumbach, 2007) is suboptimal for our analysis.

Rearranging the definition of our *ewma* estimator in (1) it can be shown that the weights of return observation r_{t-k} can be calculated as

$$w_{t-k} = (1 - \lambda) \cdot \lambda^k . \quad (\text{D.1})$$

Following RiskMetrics (1996), the corresponding $\tilde{\lambda}$ for a given k and cumulative weight target, W , can be calculated as

$$\tilde{\lambda} = e^{\ln(1-W)/k} . \quad (\text{D.2})$$

We set $W = 99.9\%$ and calculate $\tilde{\lambda}$ for one, three and five years respectively. The corresponding $\tilde{\lambda}$ is roughly equal to 0.97, 0.99, and 0.995. For more intuition, these λ ensure that the weights within one, three, and five years sum to 99.9%.

Even for our shortest lookback period our λ is significantly larger than the recommended value of 0.94 for daily data (RiskMetrics, 1996; Mina & Xiao, 2001; Zumbach, 2007). This implies that the recommended *ewma* decay factor places substantially more emphasis on recent observations, allowing it to incorporate new information more quickly. To see whether a more responsive estimator yields economic gains, we compare average performance metrics for portfolios constructed with our most responsive $\lambda = 0.97$ to the recommended value of $\lambda = 0.94$. For each λ we evaluate 1,296 portfolios using all decision variables forks in our decision tree.

The results of the analysis are documented in Table D.1. It holds the relative average difference between the portfolios with $\lambda = 0.97$ and $\lambda = 0.94$. A positive value shows that the portfolios using $\lambda = 0.94$ perform better in the respective metrics and vice versa. Results are reported separately for the different portfolio types and with/without transaction costs.

Table D.1: Relative comparison of summary statistics for EWMA with $\lambda = 0.94$ and $\lambda = 0.97$

The table presents the relative performance of the EWMA specifications with $\lambda = 0.94$ and $\lambda = 0.97$ across various summary statistics (cumulative return, average return, alpha, standard deviation, Sharpe ratio, and maximum drawdown). Each specification generates 1,296 portfolios. The relative difference is assessed by comparing which specification performs better; a value below zero indicates that the $\lambda = 0.94$ specification underperforms relative to $\lambda = 0.97$, while a value above zero indicates the opposite. The analysis is carried out for a specification without transaction costs (upper row) and with transaction costs (lower row). The performance measures are based on 11,598 daily out-of-sample returns from January 1978 to December 2023.

Trans. Costs	Portfolio	Cum.	Ret.	Alpha	Std.	SR	m. DD
w/o	100-0	0.4%	0.0%	13.5%	0.3%	0.4%	-0.6%
	130-30	12.1%	1.2%	17.8%	1.0%	3.0%	3.9%
	100-100	65.9%	9.8%	10.4%	-10.0%	87.4%	2.1%
with	100-0	-14.5%	-5.4%	-78.0%	0.2%	-8.5%	-2.9%
	130-30	-9.1%	-9.5%	-61.3%	0.7%	-15.2%	-2.8%
	100-100	-0.7%	-49.8%	-18.9%	-10.1%	-19.2%	-0.1%

The findings reveal that portfolios with $\lambda = 0.94$ achieve greater total risk reduction compared to those with $\lambda = 0.97$. This outcome is intuitive, as a more responsive EWMA estimator improves risk forecasting accuracy. However, after accounting for transaction costs, $\lambda = 0.94$ portfolios significantly underperform across all other metrics, making them inferior to the $\lambda = 0.97$ specification. The agility of the $\lambda = 0.94$ estimator leads to substantially higher portfolio turnover, which in turn reduces average returns. To maintain comparability among portfolio settings and due to its inferior performance compared to the $\lambda = 0.97$ specification, we exclude $\lambda = 0.94$ portfolios from our main analysis.

E Robustness Over Time

Up to this point, we have only examined the performance of portfolios over the full time horizon. However, it is crucial to determine whether the top-performing portfolios achieve their success through consistent outperformance or benefit from a one-time event that skews results over the sample period. Furthermore, as low-risk portfolios are designed to prioritize risk reduction, evaluating their performance during periods of financial distress and heightened uncertainty is particularly important.

To go after these two concerns, we perform two analyses that are similar to the analysis in [Table 3](#). To see whether the performance of portfolios is consistent over time, we sort all portfolios based on their realized Sharpe ratios each month for each portfolio type separately and track whether

a portfolio is in the upper (“winner”) or lower (“loser”) quartile of portfolios. [Table E.1](#) holds the relative occurrence of construction of the top and bottom 100 winner (W) and loser (L) portfolios. On average, a winning (losing) portfolio occurred 43% (48%) of the time in the upper (lower) quartile of Sharpe ratios, indicating that certain construction choices result in consistent out/underperformance rather than being caused by a one-time event.

Comparing the results from this table to the Sharpe ratio column of [Table 3](#) we find that results are relatively similar. Consistent with the full-horizon findings, size and price filters enhance performance, particularly by mitigating downside losses. Longer lookback windows and rebalancing frequencies improve long-term performance, less extreme portfolio buckets bolster results, and equal weighting leads to more extreme portfolio outcomes.

Contrary to the full-horizon findings, the time-series analysis reveals more extreme outcomes for the choice of the low-risk estimator. Nearly all winning and losing portfolios in the long-only and limited-short portfolios are sorted on beta. While beta-sorted portfolios account for most losers in the full-horizon analysis in [Table 3](#), only a few beta portfolios emerge as winners. This suggests that beta-sorted portfolios exhibit cyclical behavior, outperforming in certain periods while underperforming in others. This is intuitive when considering beta-sorted portfolios as deleveraged market portfolios, which typically underperform during economic upswings and outperform during downturns. In contrast, other risk estimators lack this interpretation and are therefore less likely to exhibit such cyclical behavior.

To assess portfolio performance during periods of market turmoil, we construct [Table 3](#) for all NBER recession periods in our sample combined. As before, we analyze the top 100 winner (W) and loser (L) portfolios in [Table E.2](#) using various performance metrics. In this analysis, alpha and maximum drawdowns are omitted due to the need for uninterrupted time-series data for these two metrics. Instead, we count the number of times that our low-risk portfolios had a higher return than the benchmark, $r_{p,t} \geq r_{bm,t}$, to capture the consistency of relative outperformance or underperformance of our strategies.

As before, we observe that many construction choices, such as portfolio type, size and price filters, lookback windows, and weighting schemes, yield results during recession periods consistent with the full time-horizon analysis. However, during financial distress, more frequent rebalancing proves beneficial for adapting to market changes. In contrast, the performance of portfolio buckets is less conclusive compared to the clear advantage of using less extreme buckets in the full-horizon analysis.

Table E.1: Relative appearance of portfolio construction choices in top 100 winner (W) and loser (L) portfolios with transaction costs over time

The table shows the relative occurrence of decision variables in the top 100 winner (W) and bottom 100 loser (L) portfolios, ranked based on consistently high or low Sharpe ratios over time. Portfolios are evaluated net of transaction costs. Each month, portfolios are sorted into quartiles based on their Sharpe ratios. The portfolios that most frequently appear in the top quartile (Q4, “winners”) and the bottom quartile (Q1, “losers”) over the evaluation period are identified. From these, the 100 portfolios with the highest frequency in Q4 (winners) and Q1 (losers) are selected. On average, the winning portfolios are in Q4 43% of the time, while the losing portfolios are in Q1 48% of the time. For the top and bottom 100 portfolios, the relative occurrence of decision variables is calculated for each decision fork. The analysis is conducted separately for each portfolio type: long-only (100-0), limited-short (130-30), and long-short (100-100). Results are highlighted in (light) green if they are more than (two) three standard deviations away from an uniform distribution of variables in a positive direction. For winning portfolios, this indicates that the variable is overrepresented, while for loser portfolios, it suggests the variable is underrepresented. Conversely, we apply (light) red coloring for the opposite scenario. The performance metrics are calculated based on 11,598 daily out-of-sample returns between January 1978 to December 2023.

Variable	100-0		130-30		100-100	
	W	L	W	L	W	L
Risk Estimator						
hist	1%	0%	0%	3%	25%	21%
ewma	0%	0%	3%	7%	27%	23%
ff3	0%	0%	0%	3%	24%	25%
ff5	0%	1%	0%	3%	22%	26%
beta	99%	99%	97%	84%	2%	5%
Size Exclusion						
0	40%	88%	37%	85%	98%	100%
10	29%	8%	29%	8%	2%	0%
20	31%	4%	34%	7%	0%	0%
Prize Exclusion						
0	24%	43%	30%	51%	61%	39%
1	32%	39%	33%	32%	36%	38%
5	44%	18%	37%	17%	3%	23%
Lookback Window						
1y	17%	49%	39%	54%	48%	36%
3y	39%	32%	31%	29%	30%	31%
5y	44%	19%	30%	17%	22%	33%
Rebalancing Frequency						
w	9%	35%	3%	53%	0%	100%
m	3%	34%	1%	33%	16%	0%
q	42%	18%	41%	12%	46%	0%
y	46%	13%	55%	2%	38%	0%
Portfolio Buckets						
3	8%	15%	54%	12%	31%	37%
5	38%	18%	34%	20%	14%	34%
10	54%	67%	12%	68%	55%	29%
Weighting Scheme						
vw	6%	34%	0%	27%	59%	0%
ew	94%	66%	100%	73%	41%	100%

Table E.2: Relative appearance of portfolio construction choices in top 100 winner (W) and loser (L) portfolios for long-only and limited-short portfolios with transaction costs during recessions

The table presents the relative occurrence of decision variables in the top 100 winner (W) and bottom 100 loser (L) portfolios, ranked according to various performance metrics during periods of recessions. These portfolios, encompassing both long-only and limited-short strategies, are evaluated net of transaction costs. The relative appearance of decision variables is calculated for each decision fork across the following performance metrics: number of times that the portfolio had a higher return than the benchmark, $r_{p,t} \geq r_{bm,t}$, average return, alpha, volatility, Sharpe ratio, and information ratio. A value-weighted market portfolio serves as the benchmark. Results are highlighted in (light) green if they are more than (two) three standard deviations away from an uniform distribution of variables in a positive direction. For winning portfolios, this indicates that the variable is overrepresented, while for loser portfolios, it suggests the variable is underrepresented. Conversely, we apply (light) red coloring for the opposite scenario. The performance metrics are calculated based on 1,216 daily out-of-sample returns in recession periods between January 1978 to December 2023.

Variable	$r_{p,t} \geq r_{bm,t}$		Return		Volatility		SR		IR	
	W	L	W	L	W	L	W	L	W	L
Risk Estimator										
hist	11%	2%	20%	7%	0%	0%	21%	4%	41%	9%
ewma	43%	0%	59%	15%	10%	0%	58%	14%	4%	16%
ff3	5%	53%	13%	6%	0%	45%	10%	1%	30%	9%
ff5	9%	45%	2%	5%	0%	52%	2%	1%	25%	9%
beta	32%	0%	6%	67%	90%	3%	9%	80%	0%	57%
Portfolio Type										
100-0	87%	40%	0%	22%	9%	33%	0%	25%	93%	27%
130-30	13%	60%	100%	78%	91%	67%	100%	75%	7%	73%
Size Exclusion										
0	13%	60%	40%	97%	88%	89%	38%	96%	18%	99%
10	32%	28%	32%	3%	9%	11%	31%	4%	41%	1%
20	55%	12%	28%	0%	3%	0%	31%	0%	41%	0%
Prize Exclusion										
0	38%	40%	29%	49%	25%	74%	27%	36%	40%	53%
1	37%	33%	34%	37%	42%	26%	35%	40%	42%	33%
5	25%	27%	37%	14%	33%	0%	38%	24%	18%	14%
Lookback Window										
1y	44%	8%	51%	54%	42%	2%	47%	57%	3%	54%
3y	29%	26%	28%	29%	31%	35%	30%	27%	48%	28%
5y	27%	66%	21%	17%	27%	63%	23%	16%	49%	18%
Rebalancing Frequency										
w	16%	45%	0%	58%	38%	27%	0%	48%	67%	64%
m	34%	14%	50%	26%	18%	25%	49%	28%	7%	24%
q	39%	8%	21%	10%	20%	23%	17%	13%	6%	8%
y	11%	33%	29%	6%	24%	25%	34%	11%	20%	4%
Portfolio Buckets										
3	82%	49%	0%	9%	18%	60%	1%	8%	32%	17%
5	9%	43%	6%	30%	38%	18%	4%	28%	64%	30%
10	9%	8%	94%	61%	44%	22%	95%	64%	4%	53%
Weighting Scheme										
vw	25%	82%	90%	1%	0%	4%	85%	0%	8%	0%
ew	75%	18%	10%	99%	100%	96%	15%	100%	92%	100%

Among estimators, the ewma estimator performs better during recessions, which is intuitive given its high responsiveness to changes in the underlying data. The performance of other risk estimators remains relatively stable.

Overall, the time-series analysis confirms that the favorable construction choices for low-risk portfolios are not tied to one-off events. These choices consistently deliver outperformance over time and provide a valuable edge during periods of financial distress.

References

- Abdi, F., & Ranaldo, A. (2017). [A simple estimation of bid-Ask spreads from daily close, high, and low prices](#). *Review of Financial Studies*, 30(12), 4437–4480.
- Chung, K. H., & Zhang, H. (2014). [A simple approximation of intraday spreads using daily data](#). *Journal of Financial Markets*, 17(1), 94–120.
- Daniel, K., & Moskowitz, T. J. (2016). [Momentum crashes](#). *Journal of Financial Economics*, 122(2), 221–247.
- Fama, E. F., & French, K. R. (1993). [Common risk factors in the returns on stocks and bonds](#). *Journal of Financial Economics*, 33(1), 3–56.
- Mina, J., & Xiao, J. Y. (2001). Return to RiskMetrics: The Evolution of a Standard.
- Novy-Marx, R., & Velikov, M. (2016). [A Taxonomy of Anomalies and Their Trading Costs](#). *Review of Financial Studies*, 29(1), 104–147.
- RiskMetrics. (1996). *RiskMetrics™—Technical Document*.
- Traut, J. (2023). [What we know about the low-risk anomaly: a literature review](#). *Financial Markets and Portfolio Management*, 37(3), 297–324.
- Zumbach, G. O. (2007). [The Riskmetrics 2006 Methodology](#). *SSRN Electronic Journal*.