Testing, Comparing, and Investing with Zero-Beta (Factor-Neutral) Portfolios

Rong Wang *

August 2025

Abstract

Zero-beta portfolios, also known as factor-neutral portfolios, are explicitly constructed to have no exposure to systematic risk within a factor model. Despite their foundational role in asset pricing theory, their empirical applications have been relatively limited. This paper investigates the role and implications of zero-beta portfolios in empirical asset pricing research. First, I develop a unified framework for model testing and comparison based on the maximum Sharpe ratio of zero-beta portfolios, applicable to a broad class of factor models. While all models are formally mispecified, machine learning-based models offer clear advantages as the dimensionality of the factor structure grows. Second, I introduce an optimal zero-beta investment strategy that exploits model mispricing, delivering robust out-of-sample performance and outperforming most established strategies even after accounting for transaction costs.

1. Introduction

The landscape of empirical asset pricing has long been characterized by the search for factor models that can accurately explain the cross-section of expected asset returns. This effort has led to the development of a vast "factor zoo", with hundreds of proposed risk factors and a growing array of sophisticated econometric techniques designed to model their influence (Cochrane, 2011; Harvey et al., 2016; Karolyi and Van Nieuwerburgh, 2020). Recent advances in machine learning have further accelerated this trend, introducing powerful conditional models that allow risk exposures (betas) to vary over time as complex, non-linear functions of a high-dimensional set of asset characteristics. These machine learning methods have demonstrated superior explanatory power compared to traditional unconditional models.

Despite these methodological advances, a fundamental challenge remains unresolved: the lack of unified statistical frameworks for formally testing and comparing increasingly complex models. Standard evaluation practices typically rely on metrics such as total or predictive

*Rong Wang: Ph.D. candidate at the Fuqua School of Business, Duke University (rw196@duke.edu).

R^2, which, while informative, are only point estimates of goodness-of-fit and do not allow for formal hypothesis testing or rigorous model comparison under sampling uncertainty. This gap is particularly acute in the context of modern machine learning models, where many traditional procedures—such as the Gibbons, Ross, and Shanken (1989) (GRS) test—are not directly applicable to models with time-varying parameters.

This paper addresses this critical gap by developing a novel, universally applicable framework for model testing and comparison based on the properties of zero-beta portfolios. Zero-beta portfolios—also known as factor-neutral portfolios—are constructed to have no exposure to any systematic risk factors in a given factor model. The core intuition is straightforward yet powerful: if a factor model is correctly specified, any zero-investment portfolio[1] that is fully hedged against all proposed risk factors should, by construction, have an expected return of zero. Persistent deviations from zero therefore reveal unpriced risks or mispricing, indicating model misspecification. I formalize this insight into a comprehensive methodology for both testing and comparing asset pricing models.

The first major contribution of this paper is to introduce a unified framework for model evaluation centered on the maximum Sharpe ratio attainable by zero-beta portfolios. For any given factor model, the maximum Sharpe ratio of such portfolios provides a direct and intuitive measure of model misspecification. A low maximum Sharpe ratio indicates that the model leaves little scope for generating risk-adjusted returns independent of its factors, consistent with a better-specified model. In contrast, a high maximum Sharpe ratio implies that the model fails to capture important sources of predictable returns.

This framework yields two practical applications. First, I introduce a zero-beta portfolio test based on the null hypothesis that all zero-beta portfolios have zero expected returns. To test this hypothesis, I employ a standard multivariate Hotelling's T^2 statistic, complemented by p-values derived from asymptotic theory, finite-sample distributions, and a specialized bootstrap procedure to ensure robust inference. The results show that the zero-beta portfolio test strongly rejects all models considered—including standard Fama–French style models with pre-specified factors, Arbitrage Pricing Theory (APT)-based models with statistical factors, and advanced machine learning models such as Instrumented Principal Component Analysis (IPCA) of Kelly et al. (2019) and Autoencoder neural networks (AE) of Gu et al. (2021). This provides formal evidence that even the most sophisticated machine learning models remain misspecified. Second, recognizing that all models are prone to misspecification, I develop a model comparison framework that ranks models by their

[1] Zero-investment, zero-beta portfolios are particularly useful for testing and comparison, since the unknown zero-beta rate cancels out. For simplicity, I use the term zero-beta portfolios to refer specifically to these zero-investment versions throughout the paper.

maximum Sharpe ratios of zero-beta portfolios. This approach reveals a clear hierarchy that evolves with model complexity. In-sample, while increasing the number factors to traditional Fama–French models yields little improvement over the CAPM, extensions of PCA, IPCA, and especially AE models produce substantially better specifications. The six-factor AE model in particular statistically dominates all other models in-sample, underscoring the advantages of flexible, nonlinear specifications when the factor structure is rich. This hierarchy largely persists in a more rigorous out-of-sample setting that incorporates transaction costs and leverage constraints. The initial in-sample advantage of simple one-factor models such as the CAPM vanishes, with the one-factor AE emerging as the most robust specification in sparse factor environments. Among multi-factor specifications, the six-factor AE again performs strongly, significantly outperforming all alternatives except for the six-factor IPCA. These results confirm that machine learning approaches, and the AE model in particular, offer clear advantages as the dimensionality of the factor structure grows.

The second major contribution of this paper is to examine the investment implications of model mispricing through an optimal zero-beta investment strategy. If factor models are misspecified, the resulting pricing errors may represent exploitable investment opportunities. My strategy is designed to capture these opportunities by constructing portfolios explicitly neutral to systematic factor risks. Using a standard mean–variance optimization framework, I derive the optimal portfolio weights under the constraints of zero investment and zero beta with respect to the model's factors. This approach seeks to combine the pricing error components of assets in a mean–variance optimal manner while fully hedging systematic risk exposures. Recognizing the importance of real-world frictions, I conduct a rigorous out-of-sample evaluation of the strategy, incorporating transaction costs calculated at the individual-stock level.

The optimal zero-beta investment strategy delivers robust and economically significant out-of-sample performance. Across all single-factor model specifications, the strategies achieve annualized net-of-cost (before-cost) Sharpe ratios of 1.15–1.17 (1.48–1.49), more than doubling the market portfolio's Sharpe ratio of 0.52 over the same period from January 1990 to December 2024. In the six-factor case, Sharpe ratios decline to 0.77–1.07 (1.12–1.38), reflecting a trade-off between model specification and investment profitability: simpler, more misspecified one-factor models often generate the highest-performing strategies, as their larger pricing errors are more readily exploitable. This contrasts sharply with standard factor-based investing. While factor mean-variance portfolios constructed from machine learning models exhibit spectacular gross Sharpe ratios exceeding 3.0, and the Sharpe ratio rises with the number of factors, their performance is entirely offset by the high transaction costs required for implementation, rendering them economically unviable. Net-of-cost Sharpe ratios and mean returns for factor tangency portfolios in IPCA and AE models are strongly negative. Volatility-managed multi-factor mean-variance portfolios, as proposed by Moreira and Muir (2017) and DeMiguel et al. (2024), improve performance before costs, but net-of-cost results deteriorate dramatically. Taken together, this evidence highlights that practical frictions often prevent investors from implementing extreme strategies suggested by asset pricing models. In contrast, the proposed zero-beta strategy proves far more robust to these real-world constraints, offering economically meaningful investment opportunities.

The practical robustness and distinctive properties of the optimal zero-beta strategy are further highlighted by a series of additional tests. Kelly et al. (2019) and Kim et al. (2021) construct "arbitrage portfolios"[2] using IPCA and PPCA (projected PCA of Fan et al., 2016), respectively. While these portfolios perform well before costs, their profitability is largely eliminated once transaction costs are accounted for. Time-series regressions against a suite of common risk factors confirms that the optimal zero-beta strategy's returns are not driven by exposures to known factors, as it generates significant time-series alphas. An examination of the portfolio composition further shows that, unlike the well-known Betting-Against-Beta (BAB) strategy, this approach does not systematically tilt toward low-beta or low-idiosyncratic-volatility stocks. Transaction costs, while elevated during market crises, remain manageable, and the portfolio weights are well-diversified, avoiding extreme concentrations in individual stocks.

[Mechanism: why do zero-beta strategies outperform?]

This paper contributes to the literature on testing and comparing asset pricing models. Broadly, model tests fall into two categories. The first focuses on the significance of individual factors, asking whether a proposed factor earns a risk premium beyond what existing models explain. Methods in this category include the conventional two-step procedure of Fama and MacBeth (1973), misspecification-robust standard errors in (Kan et al., 2013), and the three-pass procedure of (Giglio and Xiu, 2021), among others. The second category evaluates overall model specification, assessing how well the model explains cross-sectional returns. This approach typically employs classical time-series mean-variance efficiency tests such as the Gibbons et al. (1989) (GRS) test with a risk-free asset, as well as its counterparts for cases without a risk-free asset (Gibbons, 1982; Jobson and Korkie, 1982; Kandel, 1984, 1986; Shanken, 1986). Although cross-sectional regression R^2 is commonly used as a goodness-of-fit measure, sampling uncertainty complicates the interpretation of these point estimates. To address this issue, Kan et al. (2013) generalize the pricing errors test of Shanken (1985), providing statistical inference for R^2, while Bryzgalova et al. (2023) apply a Bayesian approach to construct confidence intervals for R^2. Against this backdrop, the zero-beta port-

[2]Here, arbitrage refers to portfolio constructions that hedge out exposure to systematic risk factors.

folio framework developed in this paper offers a unified approach for both model testing and comparison. By focusing on the maximum Sharpe ratios of zero-beta portfolios—specialized combinations of pricing errors—the method effectively tests aggregated mispricing, generalizing the work of Shanken (1985) and Kan et al. (2013). Its primary strength lies in its broad applicability: it accommodates unconditional or conditional machine-learning-based models, models with known or latent factors, and settings with or without an observable risk-free rate.

On the model comparison side, the widely cited framework of Barillas and Shanken (2017); Barillas et al. (2020) emphasizes the Sharpe ratios of factor portfolios. This paper shows that comparing the maximum attainable Sharpe ratio of zero-beta portfolios is fundamentally equivalent to comparing the Sharpe ratios of factor portfolios. A lower maximum Sharpe ratio of zero-beta portfolios implies a smaller distance between the factors and the efficiency frontier, which in turn indicates a higher Sharpe ratio of the factors. While preserving this economic intuition, the use of zero-beta portfolios is broadly applicable: it accommodates unconditional or conditional models, observable or unobservable risk-free rates, and factors that are traded, non-traded, or purely statistical. As long as zero-beta portfolios can be constructed, model comparison can be performed. In contrast, when the risk-free rate is unobservable, computing Sharpe ratios for factors becomes ambiguous, whereas the Sharpe ratios of zero-investment portfolios do not rely on the risk-free rate. Related refinements in model comparison consider out-of-sample evaluation (Fama and French, 2018; Kan et al., 2024), short-selling constraints (Fama and French, 2015), and transaction costs (Detzel et al., 2023); these considerations are naturally incorporated in the zero-beta portfolio framework. In summary, this generality makes zero-beta portfolios a powerful and practical common ground (as advocated by Karolyi and Van Nieuwerburgh, 2020) for evaluating the full spectrum of modern asset pricing models, including the increasingly complex models that dominate contemporary research.

The proposed investment strategy is conceptually related to market-neutral and statistical arbitrage approaches. These strategies typically follow a two-step process: first, an "alpha generation" step identifies potentially mispriced assets by modeling their returns, either through purely statistical methods or factor models; second, a "hedging" step constructs a portfolio by going long on underpriced assets and short on overpriced ones while neutralizing exposure to systematic risk factors. Equity market-neutral strategies are widely employed by hedge funds (Fung and Hsieh, 2001), and other common statistical arbitrage approaches include pairs trading (Gatev et al., 2006; Avellaneda and Lee, 2010) and the betting-against-beta strategy (Frazzini and Pedersen, 2014). Kelly et al. (2019) and Kim et al. (2021) construct "arbitrage portfolios" to exploit the mispricing from factor models.

More recently, Guijarro-Ordonez et al. (2021) apply deep learning (a convolutional transformer) to extract complex time-series trading signals from the residuals of a multi-factor model, which are then mapped to an optimal trading policy using a flexible neural network designed to maximize the Sharpe ratio. My approach aligns with this principle by using the factor model's pricing errors as the alpha signal and mean-variance optimization to construct a factor-neutral portfolio. Rather than introducing new techniques for alpha generation or hedging, the key contribution of this paper is to rigorously apply this established framework to compare the out-of-sample investment performance of different underlying factor models, with particular attention to the role of transaction costs.

The remainder of the paper is organized as follows. Section 2 develops the conceptual framework for using zero-beta portfolios in model testing and comparison, and details the associated statistical procedures. Section 3 outlines the methodology for the optimal zero-beta investment strategy and the computation of transaction costs. Section 4 presents the empirical results, covering the data, candidate models, and findings from model tests, model comparisons, and investment backtests. Section 5 discusses potential economic mechanisms underlying the performance of zero-beta strategies. Section 6 concludes.

2. Testing and Comparing Models using Zero-Beta Portfolios

This section explores the implications of zero-beta portfolios on testing and comparing a broad class of factor models. Section 2.1 introduces a standard mean-variance framework and demonstrates that the mean returns and maximum Sharpe ratio of zero-beta portfolios provide convenient and intuitive metrics for model testing and comparison, respectively. Section 2.2 and 2.3 elaborate on the new methods and statistical procedures. In Section 2.4 and 2.5, I connect the new methods to the existing literature, highlighting their strengths.

2.1. Conceptual Framework

A factor model posits that asset return $r_{i,t+1}$ follows a K-factor structure and the expected return in excess of the zero-beta rate is determined by risk loadings and factor risk premia:

$$r_{i,t+1} = \alpha_{i,t} + \boldsymbol{\beta}'_{i,t}\mathbf{f}_{t+1} + \varepsilon_{i,t+1} \tag{1}$$

$$\mathbb{E}_t[r_{i,t+1}] - r_{z,t} = a_{i,t} + \boldsymbol{\beta}'_{i,t}\boldsymbol{\lambda}_t \tag{2}$$

Equation (1) is the statistical assumption of realized asset returns where $\alpha_{i,t}$ is the intercept term, $\boldsymbol{\beta}_{i,t}$ is the $K \times 1$ vector of risk loadings, and $\mathbf{f}_{t+1}$ is the $K \times 1$ vector of factors.

In the theoretical expected returns model of equation (2), $a_{i,t}$ represents the pricing errors which should be zero if the factor model is perfect, $\boldsymbol{\lambda}_t$ is the $K \times 1$ vector of factor risk premia, and $r_{z,t}$ denotes the zero-beta rate—expected return for not taking any (systematic) risk. This is a general factor model framework incorporating a wide range of model specifications. The conventional Fama-French type of models assume constant betas and pre-specify known risk factors; The Arbitrage Pricing Theory (APT) models retain the constant betas but rely on principal component analysis (PCA) to extract statistical factors. With the recent development of machine learning techniques, new conditional models are able to formulate betas as functions of asset characteristics.

For any type of factor models, I focus on zero-beta portfolios that are explicitly constructed to avoid any exposure to risk factors. Suppose there are N assets and the vector form of equation (2) is $\mathbb{E}_t[\mathbf{r}_{t+1}] - r_{z,t}\boldsymbol{\iota}_N = \mathbf{a}_t + \boldsymbol{\beta}_t \lambda_t$ where $\mathbf{r}_{t+1}$ is the $N \times 1$ vector of returns, $\boldsymbol{\iota}_N$ is a $N \times 1$ vector of ones, $\mathbf{a}_t$ is the $N \times 1$ vector of pricing errors, and $\boldsymbol{\beta}_t$ is the $N \times K$ matrix of risk loadings. Consider a zero-investment zero-beta portfolio with a $N \times 1$ vector of weights $\boldsymbol{\omega}_z$. By definition, $\boldsymbol{\omega}_z' \boldsymbol{\beta}_t = \mathbf{0}_K$ and $\boldsymbol{\omega}_z' \boldsymbol{\iota}_N = 0$. Therefore, the zero-beta portfolio should have an expected return: $\mathbb{E}_t[\mathbf{r}_{t+1}^{(z)}] = \boldsymbol{\omega}_z' \mathbb{E}_t[\mathbf{r}_{t+1}] = r_{z,t} \boldsymbol{\omega}_z' \boldsymbol{\iota}_N + \boldsymbol{\omega}_z' \mathbf{a}_t + \boldsymbol{\omega}_z' \boldsymbol{\beta}_t \lambda_t = \boldsymbol{\omega}_z' \mathbf{a}_t$. A correctly specified factor model with zero pricing errors ($a_{i,t} = 0$) should produce zero-beta portfolios that have zero expected returns. This is a simple idea but it is applicable to all types of factor models no matter how different they formulate the risk loading or the risk premia.

The following proposition incorporate the above arguments and link zero-beta portfolios to model misspecification. (The complete proof is provided in Appendix C.1.)

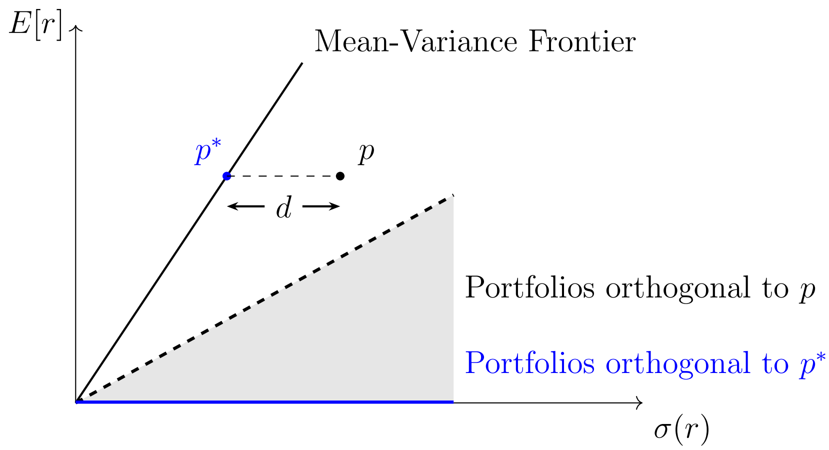
Proposition 1. *There exists an infinite number of zero-investment, zero-beta portfolios defined within a factor model.*

(i) If the factor model is correctly specified, all the zero-beta portfolios have zero expected returns.

(ii) If the factor model is misspecified, the Sharpe ratios of zero-beta portfolios are bounded above. The zero-beta frontier—the set of zero-beta portfolios that minimize variance for a given level of mean return—forms a straight line in the mean–standard deviation space. Furthermore, the slope of this frontier, which corresponds to the maximum Sharpe ratio attainable by zero-beta portfolios, can be interpreted as a measure of model misspecification.

To understand the above proposition, consider the standard textbook mean-variance framework. The mean-variance frontier for zero-investment portfolios is a straight line pass-

Figure. 1. The Zero-Investment Mean-Variance Frontier and Zero-Beta Portfolios



Notes: This figure shows the zero-investment mean-variance frontier and the zero-beta (orthogonal) sets for two assets, p^* and p, in the mean-standard deviation space. p^* is on the mean-variance frontier. p has the same mean as p^*. The blue horizontal solid line on the x-axis represents the orthogonal set with respect to p^*. The shaded area represents the orthogonal set with respect to p. The black dashed line represents the zero-beta frontier with respect to p.

ing through the origin[3]. According to asset pricing theory, a multivariate beta equation model corresponds to a mean-variance efficient portfolio on the frontier, denoted by p^* and illustrated in Figure 1. Proposition 1 *(i)* indicates that the zero-beta portfolios (also known as orthogonal portfolios, originally introduced by Roll, 1980) with respect to p^* lie on the blue horizontal solid line, characterized by zero expected returns. In contrast, a misspecified factor model corresponds to an inefficient portfolio p, also depicted in Figure 1. Proposition 1 *(ii)* indicates that the zero-beta portfolios with respect to p fall within the shaded area. The black dashed line is referred to as the zero-beta frontier—the set of portfolios that have minimum variance for a given level of mean return and zero beta with respect to p. The Sharpe ratios of these zero-beta portfolios are therefore bounded above by the slope of the zero-beta frontier. Interestingly, Appendix C.1 proves that this slope is positively related to d, the distance between p and the efficient portfolio p^* with the same expected return. This relationship allows the slope of the zero-beta frontier to serve as a measure of the inefficiency of p. In other words, the maximum Sharpe ratio of zero-beta portfolios associated with a given factor model can serve as a proxy for the degree of model misspecification.

A major advantage of this analysis is that it applies not only to unconditional factor models but also to conditional models, where the mean-variance frontier and zero-beta frontier are time-varying. Importantly, the zero-beta frontier remains a straight line in the

[3]Conventional mean-variance analysis typically focuses on unit-investment portfolios, where portfolio weights sum to 1. In that case, the mean-variance frontier takes the familiar parabolic shape.

mean–standard deviation space when viewed ex post. The following proposition formalizes this result.

Proposition 2. *For both unconditional and conditional factor models, if the same target mean return is used in constructing the conditional maximum Sharpe ratio zero-beta portfolio in each period, the ex-post full-sample Sharpe ratios of these portfolios are identical.*

The complete proof is provided in Appendix C.2. The intuition is straightforward. For each period, the zero-beta portfolios on the zero-beta frontier are all perfectly correlated with each other. Their weights are the same up to a constant. The only difference between them is the leverage used to achieve a different target mean return. This lockstep relationship holds true for every single period. Therefore, when I finally compute the Sharpe ratio over the full sample, I am taking the mean and standard deviation of these perfectly scaled time series, where the scaling constant for each portfolio cancels out, resulting in an identical Sharpe ratio for all of them.

This proposition illustrates the generality of working with zero-beta portfolios. Although the conditional mean–variance relations may vary over time, the ex-post zero-beta frontier remains linear when constructed over the full sample. This property enables a unified framework for model testing and comparison, as developed in the following sections.

Comments on the bounded Sharpe ratio of zero-beta portfolios: The statement of bounded Sharpe ratio of zero-beta portfolios is not inconsistent with the idea that arbitrage opportunities can have unbounded Sharpe ratios. The Sharpe ratios are bounded because they are constrained by the model structure. When I am constructing zero-beta portfolios within the factor model by imposing constraints (e.g., zero exposure to estimated betas and zero-investments), I only use information allowed by the model rather than discover true arbitrage opportunities. The maximum Sharpe ratio of such zero-beta portfolios reflects how misspecified the model is instead of an actual arbitrage opportunity. In summary, while true arbitrage opportunities (in the market) may have infinite or unbounded Sharpe ratios, model-implied zero-beta portfolios can only reflect misspecification within the model, and therefore have bounded Sharpe ratios. Therefore, I use the maximum Sharpe ratio among model-based zero-beta portfolios as a diagnostic tool — not a literal arbitrage detector.

2.2. The Zero-Beta Portfolio Test

Proposition 1 offers a unified framework for testing and comparing different factor models. After estimating the factor model betas, we can construct any number of zero-beta portfolios using the null space of betas. Let $r^{(z)}_{j,t+1}$ denote the realized return of the j-th zero-beta portfolio, where $j = 1, 2, \ldots, N_z$. If the factor model is correctly specified, then all zero-beta portfolios should have zero expected returns. This leads to a *zero-beta portfolio test* with the null hypothesis:

$$\mathcal{H}_0 : \mathbb{E}[r^{(z)}_{j,t+1}] = 0 \quad \text{for all } j \tag{3}$$

To test this hypothesis, I use the standard multivariate approach based on Hotelling's T^2 statistic:

$$Q = T \cdot \bar{\mathbf{r}}^{(z)\prime}\hat{\mathbf{V}}^{-1}\bar{\mathbf{r}}^{(z)} \tag{4}$$

where $\bar{\mathbf{r}}^{(z)}$ is the $N_z \times 1$ vector of time-series mean returns of the zero-beta portfolios, and $\hat{\mathbf{V}}$ is the sample covariance matrix. Inference is conducted using three p-values, following standard practice in the financial econometrics literature.

First, under asymptotic theory, the Hotelling's T^2 statistic converges in distribution to a chi-squared distribution with N_z degrees of freedom:

$$Q \xrightarrow{d} \chi^2(N_z) \tag{5}$$

from which p-values can be computed.

However, due to typically limited sample sizes in asset pricing applications, asymptotic approximations may not be reliable. The literature has long recognized that relying solely on asymptotic tests can lead to substantial Type I errors (Jobson and Korkie, 1982; Stambaugh, 1982; Shanken, 1985, among others). Under standard assumptions such as multivariate normality, an exact finite-sample distribution can be derived. Specifically, the transformed statistic follows an F-distribution:

$$F = \frac{T - N_z + 1}{N_z} \cdot \bar{\mathbf{r}}^{(z)\prime}\hat{\mathbf{V}}^{-1}\bar{\mathbf{r}}^{(z)} \sim F(N_z, T - N_z + 1) \tag{6}$$

and corresponding p-values can be computed from this distribution.

Finally, I also implement a bootstrap procedure that avoids strong distributional assumptions. Ideally, one would impose the null hypothesis and resample the raw asset returns, re-estimating the factor model and repeating the test in each bootstrap sample. However, because the null pertains to the means of zero-beta portfolios—which are non-unique combinations of the raw asset returns—it is not feasible to reconstruct the raw returns period by period based solely on moment conditions.[4]

[4]Bootstrapping is more straightforward when testing structural model parameters, where the null can be imposed directly on the parameter of interest. For example, Kelly et al. (2019) test for zero alphas in their IPCA model by setting alphas to zero, resampling returns, and re-estimating the model in each bootstrap iteration.

Given this constraint, I adopt a narrower form of bootstrapping that conditions on estimated betas. Specifically, I estimate the factor model once and treat the beta estimates as fixed, assuming minimal estimation uncertainty. In each bootstrap iteration, I randomly select a subset of zero-beta portfolios from a large pool generated from the model. I then impose the null by demeaning the returns of the selected zero-beta portfolios and compute residuals. Each residual is multiplied by a Student's t random variable with five degrees of freedom and unit variance. I use these bootstrapped returns to compute the Q-statistic, repeating the procedure across bootstrap samples. The empirical p-value is then computed as the fraction of bootstrapped Q-statistics that exceed the Q-statistic from the actual data.

2.3. Model Comparison using the Maximum Sharp ratio of Zero-Beta Portfolios

According to Proposition 1, the maximum Sharpe ratio of zero-beta portfolios can be interpreted as a measure of model misspecification. Therefore, comparing models reduces to comparing the maximum Sharpe ratios of the zero-beta portfolios generated by different models. To begin, I construct the zero-beta frontier for a given factor model analytically. Denote the mean asset returns by $\boldsymbol{\mu}$ and the variance-covariance matrix by $\boldsymbol{\Sigma}$. I compute the zero-beta portfolio weights $\boldsymbol{\omega}$ ($N \times 1$) vector by minimizing portfolio variance, $\boldsymbol{\omega}'\boldsymbol{\Sigma}\boldsymbol{\omega}$, subject to three constraints: (i) zero-investment constraint $\boldsymbol{\omega}'\boldsymbol{\iota} = 0$; (ii) zero-beta constraint $\boldsymbol{\omega}'\boldsymbol{\beta} = \mathbf{0}_K$; (iii) target mean return $\boldsymbol{\omega}'\boldsymbol{\mu} = r$ where r is a scaler. This problem has an analytical solution (Proof is in Appendix C.3):

$$\boldsymbol{\omega}(r) = \boldsymbol{\Sigma}^{-1} \begin{bmatrix} \boldsymbol{\iota} & \boldsymbol{\beta} & \boldsymbol{\mu} \end{bmatrix} \left(\begin{bmatrix} \boldsymbol{\iota}' \\ \boldsymbol{\beta}' \\ \boldsymbol{\mu}' \end{bmatrix} \boldsymbol{\Sigma}^{-1} \begin{bmatrix} \boldsymbol{\iota} & \boldsymbol{\beta} & \boldsymbol{\mu} \end{bmatrix} \right)^{-} \begin{bmatrix} 0 \\ \mathbf{0}_K \\ r \end{bmatrix} \tag{7}$$

Note that I omit the time subscript t for notational simplicity; this does not imply the portfolio weights are time-invariant. In conditional models with time-varying betas, I solve the same optimization problem in each time period and thus obtain time-varying zero-beta portfolio weights. I write $\boldsymbol{\omega}(r)$ to emphasize that the zero-beta portfolio weights depend on target mean return r. To trace out the zero-beta frontier, I generate 100 target mean returns evenly spaced between 0% and 10% (monthly values) and calculate the corresponding portfolio weights. In this way, the empirical zero-beta frontier associated with each factor model can be easily visualized in a mean-standard deviation graph. Recall that proposition 1 and 2 show that all zero-beta portfolios on the ex-post zero-beta frontier share the same Sharpe ratio, and this common value represents the maximum Sharpe ratio attainable by any zero-beta portfolio. Let $MSR_A^{(z)}$ denote the maximum Sharpe ratio of zero-beta portfolios for model A, and $MSR_B^{(z)}$ the maximum Sharpe ratio of zero-beta portfolios for model B.

The null hypothesis for model comparison becomes:

$$\mathcal{H}_0 : MSR_A^{(z)} - MSR_B^{(z)} \geq 0, \quad \mathcal{H}_1 : MSR_A^{(z)} - MSR_B^{(z)} < 0 \tag{8}$$

This is a one-sided test of whether model A's maximum Sharpe ratio of zero-beta portfolios is significantly lower than that of model B. Not rejection of the null implies that model A is not significantly better than model B. Conversely, rejecting the null supports the conclusion that model A is less misspecified than model B.

Statistical inferences are based on the following specialized bootstrapping procedure. First, I compute the observed difference between the maximum Sharpe ratios of model A versus model B, $\Delta_{obs} = MSR_A^{(z)} - MSR_B^{(z)}$. According to the previous propositions, since all zero-beta portfolios on the zero-beta frontier share the same ex-post Sharpe ratios over the full sample, I can choose a random target mean return r to compute the maximum Sharpe ratio based on equation (7). Next, I generate bootstrap samples for both portfolio returns by resampling the original return series in blocks of random length, following the stationary block bootstrap method of Politis and Romano (1994). For each bootstrap sample, I calculate the Sharpe ratios and record the differences $\Delta_{boot} = MSR_{A,boot}^{(z)} - MSR_{B,boot}^{(z)}$. The empirical p-value is then computed as the proportion of bootstrap differences that are less than or equal to the observed difference.

2.4. Connecting to the Model Testing Literature

There is an extensive literature on asset pricing model testing. I begin by briefly reviewing the existing approaches, and then discuss the advantages of the zero-beta portfolio test introduced in the previous sections.

Asset pricing model tests can broadly be divided into two main categories. First, *significance tests of risk factors* focus on whether a proposed new factor is a legitimate source of risk that commands a risk premium in the market. The goal is to see if the factor adds explanatory power beyond what existing models can already explain. To name a few, popular approaches include the Fama and MacBeth (1973) two-pass procedure, model misspecification-robust standard errors of Kan et al. (2013), GRS-FAR confidence bounds proposed by Kleibergen and Zhan (2020), the three-pass procedure of Giglio and Xiu (2021), and the Bayesian confidence bounds developed by Bryzgalova et al. (2023), among others.

Second, *model specification tests* evaluate how well an entire asset pricing model explains the returns of test assets. The most common type of specification tests, often referred to as *time-series tests* or *mean-variance efficiency tests*, focus on the time-series intercepts based on the equivalence between beta pricing and mean–variance efficiency. To elaborate, recall that equations (1) and (2) jointly define a factor model. Taking (conditional) expectations on both sides of equation (1), we obtain:

$$\mathbb{E}_t[r_{i,t+1}] = \alpha_{i,t} + \beta_{i,t}' \mathbb{E}_t[f_{t+1}] \tag{9}$$

If the factor $f_{k,t+1}$ is a return, then the model implies that the factor expected return equals to its risk premium plus the zero-beta rate: $\mathbb{E}_t[f_{t+1}] = \lambda_{k,t} + r_{z,t}$. If the factor $f_{k,t+1}$ is an excess return, then the model implies that the factor expected return simply equals to its risk premium: $\mathbb{E}_t[f_{t+1}] = \lambda_{k,t}$. Assume for simplicity that there is only one return factor (level factor) and the other factors are excess returns (long-short factors). Then equation (2) and (9) combined implies the following constraint for the time-series intercept α_{it} (Proof is provided in Appendix C.4):

$$\alpha_{i,t} = (1 - \beta_{1,i,t}) r_{z,t} + a_{i,t} \tag{10}$$

When Treasury yields are used as a proxy for $r_{z,t}$ and excess returns are analyzed, equation (10) reduces to $\alpha_{i,t} = a_{i,t}$, and the null hypothesis becomes $\mathcal{H}_0 : \alpha_{i,t} = 0$. The most widely used method in this setting is the Gibbons, Ross, and Shanken (1989) (GRS) test. When working with total returns and acknowledging that $r_{z,t}$ is unobservable, a standard approach is to estimate the zero-beta rate and test the model based on equation (10), where the null hypothesis becomes $\mathcal{H}_0 : \alpha_{i,t} = (1 - \beta_{1,i,t}) r_{z,t}$. For example, Gibbons (1982), Jobson and Korkie (1982), Kandel (1984, 1986), and Shanken (1986) estimate $r_{z,t}$ via maximum likelihood estimation (MLE) and apply likelihood ratio tests (LRT). These likelihood ratio tests serve as counterparts to the GRS test in the absence of a risk-free rate. All of these tests have a clear interpretation within the mean–variance framework—they are equivalent to testing whether the factor tangency portfolio is mean–variance efficient. It is also possible to bypass the estimation of the zero-beta rate. For example, Chou (2000) rearrange the null hypothesis to $\mathcal{H}_0 : \frac{\alpha_{i,t}}{1-\beta_{1,i,t}} = \frac{\alpha_{j,t}}{1-\beta_{1,j,t}}$ for all i, j and apply either Wald tests or GMM tests.

Another group of model specification tests focuses directly on the cross-sectional beta pricing equation (2) and investigates whether the model adequately explains the cross-section of expected returns. Shanken (1985) proposes a cross-sectional regression test (CSRT). It first calculates the quadratic aggregate pricing errors $Q_e = \hat{e}' \hat{V}(e)^{-1} \hat{e}$ where $\hat{e}$ is the error term in the cross-sectional regression and $\hat{V}(e)$ is a consistent estimator of the asymptotic variance of the sample pricing errors. Q_e is a Hotelling's T^2 statistic that follows an asymptotic χ^2 distribution and a transformation of Q_e follows an exact finite-sample F distribution. Kan et al. (2013) generalizes the CSRT and derives an asymptotic distribution of the cross-sectional R^2. Specifically, the test statistic $T(1 - \hat{R}^2)$ converges to a weighted sum of chi-squared distributions. Using a Bayesian framework, Bryzgalova et al. (2023) produces a full posterior probability distribution for the R^2.

The zero-beta portfolio test proposed in this paper is fundamentally a generalization of the cross-sectional pricing error tests of Shanken (1985) and Kan et al. (2013) discussed in the previous paragraph. It explicitly constructs zero-beta portfolios, which are essentially specialized weighted averages of cross-sectional pricing errors. As a result, testing whether the mean returns of zero-beta portfolios are zero is equivalent to testing whether the average pricing errors are zero. The primary advantage of the zero-beta portfolio test lies in its broad applicability to a wide range of factor models. Time-series mean-variance efficiency tests, by contrast, critically depend on the constraint of time-series intercepts and model parameter, as in equation (10). The derivation of equation (10) in turn relies on a well-defined relationship between factor expected returns and factor risk premia, which holds naturally for traded factors (or mimicking portfolios for macroeconomic factors). However, this relationship becomes ambiguous for latent statistical factors that lack a portfolio interpretation. As a result, time-series tests may not be applicable to many recent machine-learning-based factor models. The zero-beta portfolio test remains applicable regardless of whether the model is unconditional or conditional, whether the risk-free rate is observable or not, and whether the factors are traded, non-traded, or purely statistical. As long as zero-beta portfolios can be constructed, the test can be implemented.

2.5. Connecting to the Model Comparison Literature

For decades, the standard approach to comparing asset pricing models has focused on their ability to explain the returns of a set of test assets, often using metrics such as the average absolute alpha or the GRS F-statistic. However, this approach has been shown to be problematic, as model rankings can depend heavily on the specific choice of test assets. A pivotal shift in methodology is proposed by Barillas and Shanken (2017), who argue that a proper comparison should assess a model's ability to price all available assets, including the factors from competing models. Under this framework, the choice of test assets becomes mathematically irrelevant for model comparison. As a result, the exercise reduces to comparing the maximum squared Sharpe ratio $SR^2(f)$ that each model's factors can generate. The paper's title, *"Which Alpha?"*, highlights that the relevant alpha for model comparison is not the test-asset alpha but the "excluded-factor alpha"—the pricing error of one model's factors with respect to another's. Building directly on their earlier work, Barillas et al. (2020) develop asymptotically valid tests to compare the squared Sharpe ratios of both nested and, crucially, non-nested models. For nested models, the standard GRS F-test on the excluded-factor alphas is sufficient. For non-nested models, they derive the asymptotic distribution of the difference in sample squared Sharpe ratios, allowing for a formal statistical test of which model is superior[5].

With the factor maximum Sharpe ratio framework becoming the dominant approach in model comparison, subsequent research has addressed several important practical challenges. First, in-sample (ex-post) Sharpe ratios are known to be upward biased due to estimation risk. Fama and French (2018) and Kan et al. (2024) compare models based on the out-of-sample Sharpe ratios of their factors. Kan et al. (2024) further show that models with more factors are subject to greater estimation risk, which leads to a larger deterioration in out-of-sample performance. They propose that a more complex model must not only have a higher population Sharpe ratio, but one that exceeds a "break-even" Sharpe ratio—a higher threshold required to offset its greater estimation risk relative to a simpler benchmark. Second, real-world frictions may materially impact and reverse the conclusions of model comparisons. Fama and French (2015) show that the factor tangency portfolios that generate high maximum Sharpe ratios are not practical for real-world investors since they require extremely high levels of short selling. The paper re-calculates the maximum Sharpe ratios under a no-short-selling constraint, which better reflects the opportunity set of a long-only investor. Detzel et al. (2023) compare the Sharpe ratios accounting for transaction costs of factors.

Comparing models based on the maximum Sharpe ratio of zero-beta portfolios, as introduced in this paper, is fundamentally equivalent to comparing the maximum Sharpe ratio of the model factors. According to Proposition 1, the slope of the zero-beta frontier is positively related to the distance between the highest-Sharpe-ratio factor portfolio and the efficient portfolio with the same expected return. Therefore, a lower maximum Sharpe ratio of zero-beta portfolios implies a smaller distance between the factors and the efficiency frontier, which in turn indicates a higher Sharpe ratio of the factors. In addition, utilizing zero-beta portfolios offers several advantages. First, the method is applicable regardless of whether the model is unconditional or conditional, whether the risk-free rate is observable or not, and whether the factors are traded, non-traded, or purely statistical. As long as zero-beta portfolios can be constructed, model comparison can be performed. In contrast, if the risk-free rate is unobservable, it becomes ambiguous how to calculate Sharpe ratios for the factors, whereas the Sharpe ratios of zero-investment portfolios do not rely on the risk-free rate. Second, as suggested by Fama and French (2018) and Kan et al. (2024), Sharpe ratios

[5]Another related metric of model misspecification, introduced by Hansen and Jagannathan (1997), is the HJ-distance, which measures the distance between a model's stochastic discount factor (SDF) and the closest valid SDF that prices all assets correctly. Kan and Robotti (2009) use the HJ-distance for model comparison, and Barillas et al. (2020) show that comparing models using a modified HJ-distance is equivalent to comparing their maximum squared Sharpe ratios.

of zero-beta portfolios can be compared both in-sample and out-of-sample. Third, following the ideas of Fama and French (2015) and Detzel et al. (2023), leverage and transaction costs can also be incorporated when calculating Sharpe ratios. Regarding the short-selling constraint, although no such restriction is imposed on zero-beta portfolios in this paper, the portfolio weights are generally small, mitigating the leverage concerns highlighted by Fama and French (2018). Unlike Barillas and Shanken (2017), my approach is not independent of the choice of test assets, as the zero-beta portfolios are constructed directly from the selected test assets.

3. Investing in Zero-Beta Portfolios

Zero-beta portfolios naturally offer investment benefits such as diversification and risk reduction. This section introduces an intuitive optimal zero-beta (factor-neutral) investment strategy, designed to eliminate—or at least mitigate—exposure to systematic factor risks. Section 3.1 outlines the investment methodology, which adopts a mean–variance optimization framework, and discusses practical market-neutral strategies commonly employed by hedge funds and mutual funds. Section 3.2 details how transaction costs are computed in the context of this strategy.

3.1. Methodology

The conditional zero-beta investment strategy is constructed by solving a constrained mean-variance optimization problem on a monthly basis and applying the optimal portfolio weights to the out-of-sample returns:

$$\begin{aligned} &\max_{\boldsymbol{\omega}} \boldsymbol{\omega}'\boldsymbol{\mu} - \frac{\gamma}{2}\boldsymbol{\omega}'\boldsymbol{\Sigma}\boldsymbol{\omega} \\ &s.t. \quad \boldsymbol{\omega}'\boldsymbol{\Pi} = \mathbf{0}_{1+K} \end{aligned} \tag{11}$$

where $\boldsymbol{\mu}$ is the asset mean returns, $\boldsymbol{\Sigma}$ is the variance-covariance matrix, γ is the risk aversion coefficient, and $\boldsymbol{\Pi} = [\boldsymbol{\iota},\ \boldsymbol{\beta}]$ captures the zero-investment constraint and the zero-beta constraint. Using the Lagrangian Multiplier method, we can analytically derive the optimal zero-beta portfolio weights:

$$\boldsymbol{\omega}_z^* = \frac{1}{\gamma}\boldsymbol{\Sigma}^{-1}\left[\mathbf{I} - \boldsymbol{\Pi}\left(\boldsymbol{\Pi}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Pi}\right)\boldsymbol{\Pi}'\boldsymbol{\Sigma}^{-1}\right]\boldsymbol{\mu} \tag{12}$$

Recall that the standard unconstrained optimal portfolio is $\boldsymbol{\omega}^* = \frac{1}{\gamma}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$. In comparison, the solution (12) can be viewed as the optimal portfolio based on the transformed returns $\tilde{\boldsymbol{\mu}} \equiv \mathbf{P}_{\Pi}\boldsymbol{\mu}$, where $\mathbf{P}_{\Pi} \equiv \left[\mathbf{I} - \boldsymbol{\Pi}\left(\boldsymbol{\Pi}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Pi}\right)\boldsymbol{\Pi}'\boldsymbol{\Sigma}^{-1}\right]$ is a generalized projection matrix with weights $\boldsymbol{\Sigma}^{-1}$. This matrix projects the mean return vector onto the subspace orthogonal to $\boldsymbol{\Pi}$. $\tilde{\boldsymbol{\mu}}$ can then be interpreted as an alpha (pricing error) component defined by the given factor model. Therefore, $\boldsymbol{\omega}_z^*$ represents the optimal combination of alphas while simultaneously achieving beta neutrality. In this sense, the zero-beta strategy can also be viewed as an alpha strategy—extracting alphas and constructing optimal long–short portfolios accordingly.

Factor neutrality is not a new concept. Maintaining low or minimal correlation between hedge fund returns and market returns has long been considered a desirable feature of hedge funds (Fung and Hsieh, 2001). In particular, the equity market-neutral strategy (also known as the beta-neutral strategy) is popular among hedge funds and mutual funds. These strategies are designed to eliminate systematic market risk and deliver returns that are independent of broad market movements.[6] According to BarclayHedge, the total assets under management for Equity Market-Neutral hedge funds was \$76.6 billion as of the first quarter of 2025.[7] In practice, portfolio construction is considerably more complex. Typically, an Equity Market-Neutral fund (EMN) uses proprietary quantitative models to go long undervalued stocks and short overvalued stocks (alpha generation) and then manages the aggregate portfolio to maintain a market beta of approximately zero (the zero-beta constraint). Frazzini and Pedersen (2014)'s betting-against-beta (BAB) strategy follows the same logic. According to the prospectus of AQR's Equity Market Neutral Fund, "The Adviser employs a model which aggregates many measures, or signals, that are used to determine a stock's relative attractiveness, utilizing a wide variety of traditional and non-traditional, public and proprietary data sources...Applying these signal categories, the Adviser takes long or short positions in sectors, industries and companies that it believes are attractive or unattractive (to achieve market neutrality)."

The notion of market neutrality has been extended to multi-factor neutrality through statistical arbitrage. Building on the idea of pairs trading, Avellaneda and Lee (2010) propose a factor model-based statistical arbitrage strategy. They decompose a stock's return into a "systematic" component (driven by multiple risk factors) and an "idiosyncratic" component (the residual). The residual is modeled as an Ornstein–Uhlenbeck process, reflecting the key assumption that it is mean-reverting. A trading signal is then triggered by the deviation of the current residual from its estimated long-term mean (analogous to an alpha).

[6] Quote from the prospectus of AQR's Equity Market Neutral Fund: "The Fund seeks to provide investors with returns from the potential gains from its long and short equity positions. The Fund is designed to be market- or beta-neutral, which means that the Fund seeks to achieve returns that are not closely correlated with the returns of the equity markets in which the Fund invests. Accordingly, the Adviser, on average, intends to target a portfolio beta of zero to equity markets in which the Fund invests over a normal business cycle. Achieving zero portfolio beta would result in returns with no correlation to the returns of equity markets in which the Fund invests over a normal business cycle."

[7] See some additional analysis of Equity Market Neutral hedge funds in Appendix B.1.

After this alpha-generation step, they go long the underpriced stock and simultaneously short a basket of risk factors in proportion to the stock's estimated betas. By construction, the combined position has net-zero exposure to the systematic risk factors. More recently, Guijarro-Ordonez et al. (2021) employ deep learning (a convolutional transformer) to extract complex time-series trading signals from the residual portfolios of a multi-factor model. These signals are then mapped to an optimal trading policy using another flexible neural network designed to explicitly maximize the Sharpe ratio.

All beta-neutral strategies, including equation (12), follow the same general principle of portfolio construction. The process begins with alpha extraction—identifying underpriced and overpriced stocks—followed by establishing long and short positions to enforce the zero-beta constraint. Different approaches vary in both the method of alpha generation and the design of the long/short positions. Specifically, the zero-beta strategy in equation (12) relies on factor models for alpha extraction and mean–variance optimization for portfolio construction. Rather than proposing new techniques for each step, this paper focuses on: (i) Comparing the out-of-sample investment performance across different factor models, serving as a tool for factor model comparison; (i) evaluating the strategies while accounting for transaction costs; (iii) providing a comprehensive comparison between factor-neutral investing versus factor-based investing; and (iv) constructing strategies based on characteristics-sorted portfolios versus individual stocks.

3.2. Modeling Transaction Costs

I follow DeMiguel et al. (2024) to compute the transaction costs associated with the zero-beta portfolios. These portfolios are constructed based on characteristics-sorted portfolio betas. Suppose an optimal zero-beta portfolio is constructed from P characteristics-sorted portfolios, with a $N \times P$ matrix of weights $\mathbf{x}_{t+1}$, where the p-th column represents portfolio p's weights on N individual stocks at time $t+1$. The zero-beta portfolio itself is a combination of these sorted portfolios, with a $P \times 1$ vector of weights $\boldsymbol{\omega}_{t+1}$. Let $\boldsymbol{\iota}_N$ denote an $N \times 1$ vector of ones, and $\mathbf{r}_t$ the $N \times 1$ vector of individual returns. In this setup, the $N \times 1$ turnover vector of individual stocks required to rebalance the zero-beta portfolio is:

$$TO_{t+1} = (\mathbf{x}_{i,t+1} \cdot \boldsymbol{\omega}_{t+1}) - (\mathbf{x}_t \cdot \boldsymbol{\omega}_t) \circ (\boldsymbol{\iota} + \mathbf{r}_t) \tag{13}$$

and the aggregate transaction cost is:

$$TO_{t+1} = \left\| \Lambda_{t+1} \circ TO_{t+1} \right\|_1 \tag{14}$$

where $\circ$ is the Hadamard or component-wise product, $\|\cdot\| = \sum_{i=1}^{N} |\cdot|$ denotes the 1-norm, and

Λ_{t+1} is a $N \times 1$ vector of transaction-cost parameters. An important insight from DeMiguel et al. (2024) is that netting trades across multiple factors—a form of trading diversification—can lead to substantial cost savings in conditional volatility-managed factor portfolios. Applying the same idea here, in equation (14), I first net the rebalancing trades across P characteristics-sorted portfolios and then charge the transaction costs at the individual-stock level. This approach captures the savings from offsetting trades across sorted-portfolios while accurately reflecting the costs incurred when adjusting positions in the underlying individual stocks.

The individual transaction-cost parameter, $\Lambda_{i,t+1}$, is measured using the average low-frequency (LF) effective bid–ask spreads described in Chen and Velikov (2023). They provide both high-frequency (HF) measures, derived from intraday trade and quote data, and low-frequency (LF) measures, based only on daily price and volume data. Since HF measures are available only from 1983 onward, I use the average of four LF measures (Hasbrouck, 2009; Corwin and Schultz, 2012; Kyle and Obizhaeva, 2016; and Abdi and Ranaldo, 2017), which are available across my full sample. Chen and Velikov (2023) finds that LF measures tend to be biased upward compared to HF measures in the modern era of electronic trading (post-2005). Moreover, Frazzini et al. (2018) argues that actual trading costs may be substantially lower than suggested by previous studies. Consequently, the transaction costs in this analysis may be overestimated, implying that the investment performance reported in Section 4 could be understated. Figure B.14 shows the time variation of the mean, median, 5th percentile, and 95th percentile of individual transaction costs from January 1960 to December 2024.

4. Empirical Results

4.1. Data

I obtain monthly individual stock returns and characteristics from the Global Factor Data website organized by Jensen, Kelly, and Pedersen (2023) (JKP).[8] My sample spans January 1960 to December 2024, covering 780 months (65 years). In total, there are 3,658,843 stock-month observations for 28,828 unique stocks, averaging 4,691 stocks per month. For each characteristic, I fill missing values with the cross-sectional median by 2-digit SIC industry each month. After this step, I retain 136 characteristics with complete coverage across the full sample (see Appendix A.1 for the full list). All characteristics are lagged one month. JKP update characteristics using the most recent accounting data four months after the fiscal period ends, ensuring that lagged characteristics are in the public information set and

[8] I thank the authors for making the data easily accessible (a WRDS account with access to CRSP and Compustat is required).

avoiding look-ahead bias.

In addition to individual stocks, I also use characteristic-sorted portfolios in the empirical analysis. Following Jensen et al. (2023), I construct portfolios and factors for each characteristic and retain the two corner portfolios (top and bottom terciles), since much of the relevant information resides in the extremes (Lettau and Pelger, 2020). This yields a total of $136 \times 2 = 272$ univariate-sorted portfolios. The timing of my portfolio formation differs a bit from standard practice: while Fama–French form portfolios annually in June and JKP form them monthly, I construct portfolios each December, aligned with the rolling out-of-sample periods in my following analysis. During this procedure, I store portfolio and factor weights on individual stocks to compute transaction costs for each investment strategy.

In conditional factor models (described in Section 4.2), lagged characteristics also serve as determinants of model parameters. Following Gu et al. (2020), Gu et al. (2021), and others, I cross-sectionally rank-normalize all characteristics into the $(-1, 1)$ interval each month.[9]

As described in Section 3.2, transaction costs for individual stocks are measured using average low-frequency effective spreads.[10]

Although this paper studies factor models without a risk-free rate, I still require a funding cost of capital to calculate Sharpe ratios for comparability with standard investment strategies. Rather than the Treasury bill rate, which is distorted by convenience yields, I use short-term unsecured borrowing rates, which more closely reflect actual institutional funding costs. Specifically, I use: (i) the 30-Day AA Financial Commercial Paper rate (FRED: CPF1M) from January 1997 to December 2024, (ii) the 1-Month Commercial Paper Rate (FRED: CP1M) from April 1971 to December 1996, and (iii) the 1-Month Finance Paper Placed Directly Average Offering Rate (FRED: H0RIFSPPFM01NM) from January 1960 to March 1971.

Finally, I collect European-style S&P 500 index options from OptionMetrics, covering January 1996 to August 2023 (332 months). I apply standard cleaning procedures: dropping options with bid prices below \$0.5 or where bids exceed asks; excluding contracts with less than 7 days to maturity; retaining only options expiring on the third Friday (or following Saturday) of each month; adjusting AM-settled options to expire the previous day; and excluding in-the-money contracts. The final sample includes 346,740 option contracts across 332 month-end dates.

[9] Stock characteristics often display high skewness and kurtosis. This rank transformation reduces sensitivity to outliers.

[10] I thank professor Robert Novy-Marx and Mihail Velikov for providing computation codes on the Assaying Anomalies website.

4.2. Candidate Factor Models

I study a wide range of factor models in this paper. For each model, I consider two asset universes: individual stocks and 272 characteristic-sorted portfolios (see Section 4.1). I evaluate specifications with 1, 3, 5, and 6 factors. A general representation of a factor model is given by

$$r_{i,t+1} = \alpha_{i,t}(z_{i,t}) + \boldsymbol{\beta}(z_{i,t})'\mathbf{f}_{t+1} + \varepsilon_{i,t+1} \tag{15}$$

where $\alpha_{i,t}(z_{i,t})$ and $\boldsymbol{\beta}(z_{i,t})$ denote the intercept and risk loadings, potentially functions of the 136 stock characteristics.

First, I consider *unconditional linear models with pre-specified factors*, the most widely used class of models. These models assume that a small set of observable, economically motivated factors explain stock returns, with constant intercepts and loadings: $\alpha_{i,t}(z_{i,t}) = \alpha_i$ and $\boldsymbol{\beta}(z_{i,t}) = \boldsymbol{\beta}$. Prominent examples include Fama and French (1993), Carhart (1997), Hou et al. (2015), Stambaugh and Yuan (2017), and Fama and French (2018). In the 1-factor case, I include only the market factor. In the 3-factor case, I include market, size, and value factors. In the 5-factor case, I include market, size, value, profitability, and investment factors. In the 6-factor case, I add the momentum factor to the 5-factor specification.[11] For individual stocks, I require at least 60 months of observations to include them in the time-series regressions used for estimating betas. I refer to these models collectively as "FF".

Second, I turn to *unconditional linear models with PCA factors*, rooted in the Arbitrage Pricing Theory (APT) (Ross, 1976; Huberman, 1982; Chamberlain and Rothschild, 1983; Ingersoll Jr, 1984; Connor and Korajczyk, 1986, among others). Unlike the CAPM or the Intertemporal CAPM (ICAPM), which derive from equilibrium models with explicit preferences and market assumptions, APT is a reduced-form framework.[12] It assumes a factor structure in which returns decompose into systematic and idiosyncratic components. With sufficiently many assets, idiosyncratic risk diversifies away, and the absence of arbitrage opportunities yields an approximate linear beta-pricing relation. As in the "FF" case, loadings are static: $\alpha_{i,t}(z_{i,t}) = \alpha_i$ and $\boldsymbol{\beta}(z_{i,t}) = \boldsymbol{\beta}$. The APT naturally motivates the use of principal component analysis (PCA) to extract statistical factors. Cooper et al. (2021) demonstrate that such statistically constructed factors outperform most of the traditional "FF"-style multi-factor models, in both economic and statistical terms. Following this insight, I extract

[11] The market factor is the weighted average of all stocks. The size, value, profitability, investment, and momentum factor are constructed using the characteristics "market_equity", "be_me", "ope_me", "at_gr1", and "ret_12_1".

[12] Extensions that embed APT in an equilibrium setting include Connor (1984) and Connor and Korajczyk (1988).

1, 3, 5, and 6 factors using the standard PCA. For individual stocks, I require at least 60 months of observations to include them in the PCA used for factor extraction and beta estimation.[13] I refer to these specifications collectively as "PCA".

Third, I examine *conditional linear models with latent factors*, which allow risk loadings to vary with stock characteristics. Specifically, I implement the instrumented principal component analysis (IPCA) models of Kelly et al. (2019) with 1, 3, 5, and 6 factors. In this framework, the intercepts and loadings are modeled as linear functions of observable characteristics: $\alpha_{i,t}(z_{i,t}) = \Gamma'_{\alpha} z_{i,t}$ and $\boldsymbol{\beta}(z_{i,t}) = \Gamma'_{\beta} z_{i,t}$. Unlike FF models, which rely on pre-specified factors, or PCA models, which assume static loadings, IPCA jointly estimates latent factors and their time-varying exposures using an alternating least squares (ALS) algorithm. I collectively refer to these models as "IPCA".[14]

Finally, I consider *conditional non-linear models with latent factors*, which leverage machine learning methods to capture richer relationships between characteristics and risk exposures. While IPCA models imposes linearity, neural networks can approximate complex nonlinear mappings. I use the conditional autoencoder model of Gu et al. (2021). Autoencoders are neural networks designed for unsupervised dimension reduction, which can be viewed as nonlinear analogues of PCA. They aim to learn a compressed, low-dimensional representation of input data by training the network to reconstruct their own inputs as accurately as possible. A standard latent factor model can be interpreted as a simple autoencoder, while conditional autoencoders extend this by incorporating observable characteristics. The architecture consists of two networks: a multi-layer beta network capturing non-linear mappings from characteristics to loadings, and a single-layer factor network generating latent factors as linear combinations of portfolios. The two are then combined as in equation (15). My implementation follows Gu et al. (2021) but adds an intercept term in the beta network, allowing $\alpha_{i,t}$ to vary flexibly with characteristics, and uses the 272 characteristic-sorted portfolios as the input layer to the factor network. Estimation relies on stochastic gradient descent (SGD), with learning rate tuning, LASSO (l_1) penalization, and early stopping for regularization.[15]

[13] Unlike IPCA and AE, which do not impose filters on individual stocks, both FF and PCA require a minimum number of observations for regressions and for PCA. As a result, Tables 1 and B.1 do not compare models rigorously on exactly the same universe of individual stocks. This limitation does not affect the new model comparison method, since the comparison is based on zero-beta portfolios regardless of their composition.

[14] Another strand of conditional linear factor models emphasizes time-varying risk premia in addition to time-varying loadings, pioneered by Ferson and Harvey (1991), who attribute much of cross-sectional return predictability to variations in risk premia than by variations in betas. Gagliardini et al. (2016) further develop econometric methods for large panels of individual stocks, modeling both risk premia and risk loadings as parametric functions of macro instruments and stock characteristics.

[15] Other machine learning approaches include Feng et al. (2024), who use feed-forward networks to map characteristics into deep characteristics that generate latent deep factors, and Chen et al. (2024), who incorporate no-arbitrage directly into the loss function via a generative adversarial network (GAN) framework.

I refer to these models collectively as "AE".[16]

For the conditional autoencoder models, I split the full sample into training, validation, and testing sets. The initial training period is 1960–1977 (18 years), the validation period is 1978–1989 (12 years), and the testing period is 1990–1991 (1 year). Following the literature (e.g., Gu et al., 2020), I refit the models annually. At each refit, the training sample expands by one year, while the validation sample is rolled forward with a fixed length, always including the most recent 12 years. This setup yields an out-of-sample period from 1990 to 2024, totaling 35 years. Since non–deep learning models typically do not require hyperparameter tuning, I combine the training and validation samples for estimation and use the same 1-year testing window for out-of-sample evaluation.

Kelly et al. (2019) and Gu et al. (2021) propose to evaluate factor models using total R^2 and predictive R^2. The total R^2 measures the model explanatory power of test assets using contemporaneous factor realizations:

$$R^2_{\text{total}} = 1 - \frac{\sum_{i,t \in OOS} \left(r_{i,t} - \hat{\boldsymbol{\beta}}_i' \hat{\boldsymbol{f}}_t \right)^2}{\sum_{i,t \in OOS} r_{i,t}^2}. \tag{16}$$

The predictive R^2 measures the model explanatory power of test assets using the factor risk premia, calculated as the prevailing sample average of factors up to the last month:

$$R^2_{\text{pred}} = 1 - \frac{\sum_{i,t \in OOS} \left(r_{i,t} - \hat{\boldsymbol{\beta}}_i' \hat{\boldsymbol{\lambda}}_{t-1} \right)^2}{\sum_{i,t \in OOS} r_{i,t}^2}. \tag{17}$$

In summary, the total R^2 measures how well the realized factor returns explain realized asset returns, whereas the predictive R^2 evaluates how well a model's conditional expected returns explain realized asset returns. Both metrics are reported in Table 1.[17] Consistent with Gu et al. (2021), I find that IPCA and AE models outperform the standard FF and PCA benchmarks in terms of both total and predictive R^2, for both individual stocks and characteristic-sorted portfolios. As expected, predictive R^2 values for individual stocks are negative under the FF and PCA models. Comparing IPCA and AE, I find that IPCA achieves slightly higher total R^2, whereas AE delivers higher predictive R^2, particularly for

Their architecture pairs an SDF network that constructs the pricing kernel with a conditional network that selects assets and moments yielding the largest mispricings, iterating until arbitrage opportunities are eliminated.

[16] I use two hidden layers in the beta network, with 32 and 16 neurons, respectively. The empirical results are robust to the choice of network depth. For robustness, I also consider architectures with a single hidden layer and with three hidden layers in Appendix B.3.

[17] The in-sample versions of these statistics are reported in Table B.1.

Table 1: Out-of-Sample Model Performance

Models	Test Assets	Metrics	# Factors			
			1	3	5	6
FF	Individual Stocks	Total R^2	6.5	7.4	2.7	0.3
		Pred R^2	-0.23	-0.21	-0.24	-0.25
	Portfolios	Total R^2	86.6	92.1	93.6	93.9
		Pred R^2	2.89	2.88	2.89	2.89
PCA	Individual Stocks	Total R^2	7.4	6.9	7.1	7.2
		Pred R^2	-1.09	-1.06	-1.07	-1.07
	Portfolios	Total R^2	92.3	96.2	97.1	97.5
		Pred R^2	2.88	2.88	2.88	2.88
IPCA	Individual Stocks	Total R^2	11.0	13.3	13.9	14.1
		Pred R^2	0.54	0.55	0.54	0.52
	Portfolios	Total R^2	73.6	93.0	93.8	94.4
		Pred R^2	3.15	3.22	3.23	3.27
AE	Individual Stocks	Total R^2	10.2	11.3	11.3	11.3
		Pred R^2	0.66	0.78	0.71	0.81
	Portfolios	Total R^2	79.5	91.7	93.3	92.5
		Pred R^2	3.23	3.02	3.27	3.13

Notes: This table reports the out-of-sample total $R2$ and predictive R^2 in percentages (%) for FF, PCA, ICA, and AE with 1, 3, 5, and 6 factors.

individual stocks. [18]

As noted earlier, a single R^2 value is not sufficient for accurately comparing models due to sampling uncertainty. The next two sections present formal model testing and comparison based on the new methods introduced in this paper.

Table 2: Model Testing Results

Models	Metrics	# Factors 1	3	5	6
FF	Asymptotic p-values	0.00	0.00	0.00	0.00
	Finite-sample p-values	0.00	0.00	0.00	0.00
	Bootstrapped p-values	0.00	0.00	0.00	0.00
PCA	Asymptotic p-values	0.00	0.00	0.00	0.00
	Finite-sample p-values	0.00	0.00	0.00	0.00
	Bootstrapped p-values	0.00	0.00	0.00	0.00
IPCA	Asymptotic p-values	0.00	0.00	0.00	0.00
	Finite-sample p-values	0.00	0.00	0.00	0.00
	Bootstrapped p-values	0.00	0.00	0.00	0.00
AE	Asymptotic p-values	0.00	0.00	0.00	0.00
	Finite-sample p-values	0.00	0.00	0.00	0.00
	Bootstrapped p-values	0.00	0.00	0.00	0.00

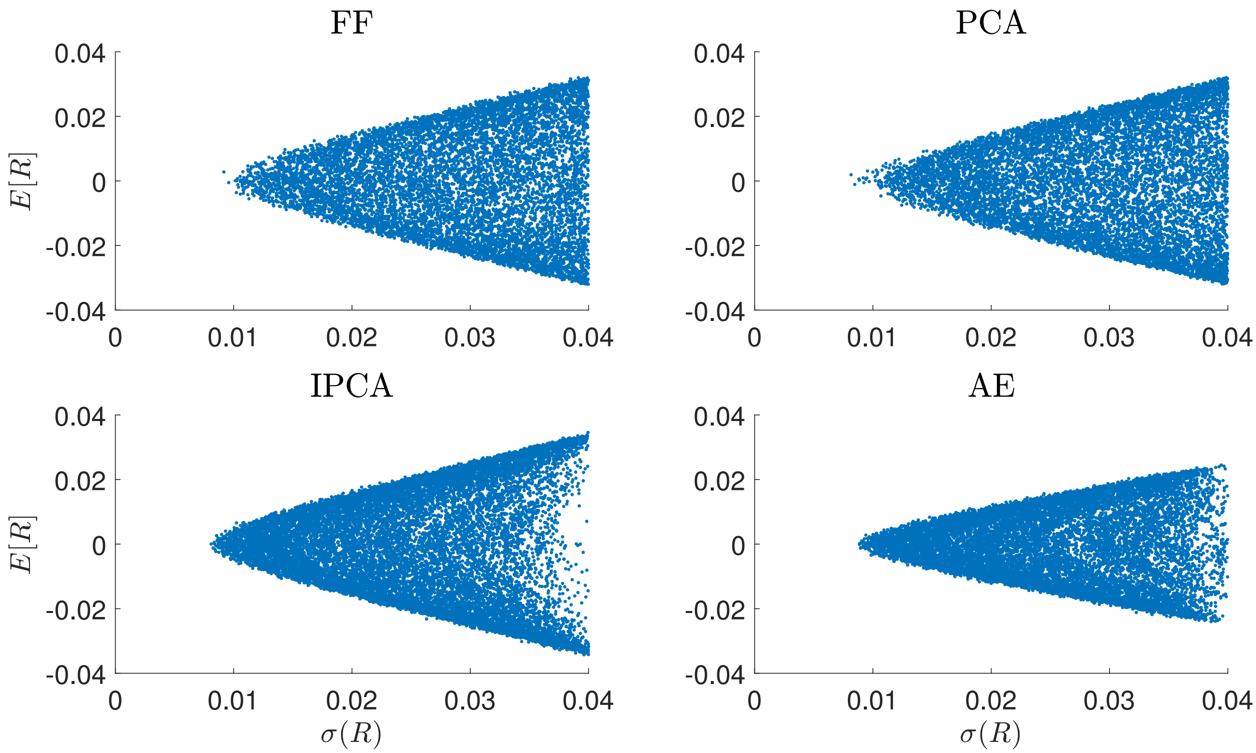
Notes: This table reports the p-values of the zero-beta portfolio test, computed using three approaches: the asymptotic χ^2 distribution, the finite-sample F distribution, and a bootstrap procedure. Results are shown for FF, PCA, IPCA, and AE models with 1, 3, 5, and 6 factors.

4.3. Model Testing

I use the 272 characteristic-sorted portfolios as base assets for model testing and comparison. The choice of test assets has long been debated in the literature. Lewellen et al. (2010) critique the conventional practice of testing models on a limited set of characteristic-sorted portfolios (e.g., the 25 Fama-French portfolios formed on size and book-to-market). Because these portfolios exhibit a strong factor structure, they fail to provide a sufficiently diverse set of return patterns to rigorously challenge new models. Similarly, Daniel et al. (2011) and Ang et al. (2020) show that sorted portfolios inherently compress the dispersion of betas, thereby reducing valuable information and weakening the power of model tests.

[18]Table 1 replicates Tables 1 and 2 of Gu et al. (2021), but the magnitudes of predictive R^2 for characteristic-sorted portfolios are considerably larger in my results. For IPCA and AE, this difference likely stems from portfolio construction. I form 272 extreme-tercile characteristic-sorted portfolios, whereas Gu et al. (2021) uses managed portfolios constructed from the characteristics matrix, $x_t = (Z'_{t-1}Z_{t-1})Z_{t-1}r_t$. For FF and PCA, the discrepancy primarily reflects my inclusion of an intercept in model estimation. Since both FF and PCA are estimated via OLS, including an intercept implies that predicted returns reduce to historical averages. Consequently, the predictive R^2 for FF and PCA essentially capture the predictive R^2 associated with expanding mean returns.

Figure. 2. Zero-Investment, Zero-Beta Portfolios



Notes: This figure shows the zero-investment, zero-beta portfolio returns in the mean-standard deviation space for the FF, PCA, IPCA, and AE models with six factors. Visible deviations from the horizontal zero line indicate model misspecification, since correctly specified models should yield portfolio mean returns around zero.

More recently, Giglio et al. (2025) propose a methodology called Supervised Principal Component Analysis (SPCA) to systematically select test assets. In this paper, I address the test asset problem by testing and comparing models using zero-beta portfolios. For the more advanced IPCA and AE models, portfolio betas are computed in a "bottom-up" manner as the weighted average of individual stock betas. Thus, constructing zero-beta portfolios from the portfolio universe is effectively equivalent to constructing them from the individual stock universe. As a result, model testing and comparison can be viewed as being conducted simultaneously on both sorted portfolios and individual stocks.[19] For computational tractability, I use the sorted portfolios when solving for zero-beta portfolio weights, thereby avoiding the inversion of an extremely large individual-stock covariance matrix.

This section implements the zero-beta portfolio tests described in Section 2.2. The null hypothesis in (3) states that the zero-investment, zero-beta portfolios have zero expected

[19]This argument applies only to the IPCA and AE models, not to the FF and PCA models, where beta estimation is performed separately for sorted portfolios and individual stocks.

returns. I compute the Hotelling's T^2 statistic in (4) and obtain p-values using three approaches: the asymptotic χ^2 distribution, the finite-sample F distribution, and a bootstrap procedure. Table 2 reports the resulting p-values. Across all specifications, the null is strongly rejected. This indicates that the zero-investment, zero-beta portfolios constructed from these models exhibit nonzero mean returns, providing evidence of misspecification in all of the factor models examined in this paper.

An additional advantage of the zero-beta portfolio test is that the zero-beta portfolios can be easily visualized. Figure 2 shows the zero-investment, zero-beta portfolio returns in the mean-standard deviation space for the FF, PCA, IPCA, and AE models with six factors. Visible deviations from the horizontal zero line indicate model misspecification, since correctly specified models should yield portfolio mean returns around zero.[20] Visualizing these portfolios offers preliminary insights for model comparison: the closer the portfolios are to the horizontal zero line, the flatter the zero-beta frontier is, the more accurately the model is specified. A detailed discussion of model comparison follows in the next section.

4.4. Model Comparison

As discussed in Section 2.3, the maximum Sharpe ratio attainable by zero-investment, zero-beta portfolios can serve as an indicator of model misspecification. Recall from Proposition 2 that the zero-beta frontier remains linear for both in-sample and out-of-sample comparisons. Consequently, the maximum Sharpe ratio is easy to visualize, as it corresponds to the slope of the zero-beta frontier in mean–standard deviation space. Figures B.8 and B.9 plot the in-sample and out-of-sample zero-beta frontiers for the FF, PCA, IPCA, and AE models with 1, 3, 5, and 6 factors. A flatter zero-beta frontier indicates a smaller distance between the factors and the efficiency frontier, signaling a better-specified model. Visual inspection suggests that adding more factors in the FF model does not improve specification relative to the 1-factor case (CAPM), whereas additional factors in PCA and IPCA appear beneficial. Notably, the 5-factor and 6-factor AE models may dominate the smaller-factor versions. Formal analysis follows.

Focusing on within-model class comparisons, Table 3 reports in-sample and out-of-sample annualized maximum Sharpe ratios attainable by zero-investment, zero-beta portfolios for the FF, PCA, IPCA, and AE models with 1, 3, 5, and 6 factors. Out-of-sample calculations incorporate transaction costs. Within each model class, the 1-factor case serves as the benchmark; an asterisk (*) indicates that the Sharpe ratio is significantly lower than the benchmark at the 5% level, based on p-values computed using the stationary block-bootstrap

[20]Appendix B.4 shows similar plots for the FF, PCA, IPCA, and AE models with 1, 3, and 5 factors.

Table 3: Within-Model Class Comparison Results

# Factors	1	3	5	6
Panel A: In-Sample Comparison				
FF	2.97	2.96	2.91	2.89
PCA	2.97	2.96	**2.87***	**2.87***
IPCA	3.45	3.34	**3.12***	**3.01***
AE2	3.45	3.45	**2.70***	**2.18***
Panel B: Out-of-Sample Comparison				
FF	1.17	1.17	1.08	1.07
PCA	1.17	1.17	**1.07***	**0.99***
IPCA	1.15	1.10	**0.92***	**0.88***
AE2	1.05	**0.92***	**0.86***	**0.72***

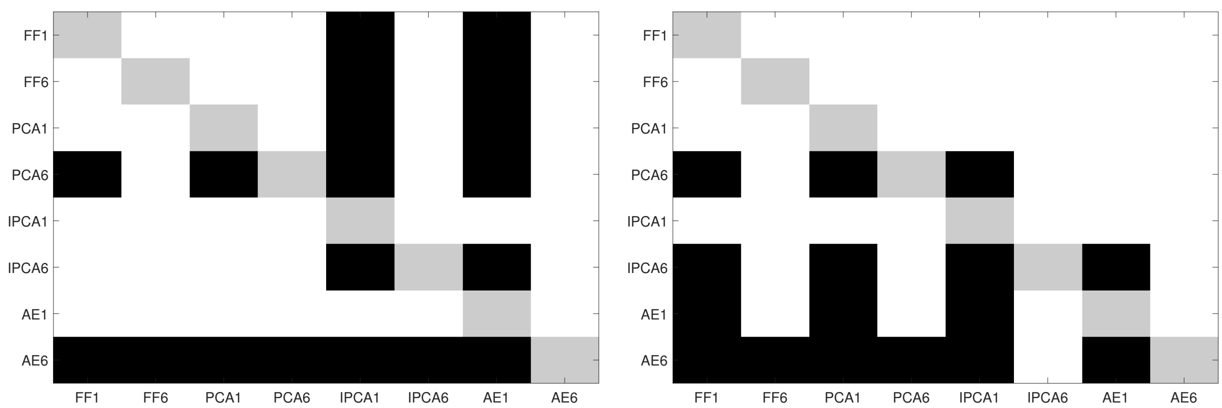
Notes: This table reports in-sample and out-of-sample annualized maximum Sharpe ratios attainable by zero-investment, zero-beta portfolios for the FF, PCA, IPCA, and AE models with 1, 3, 5, and 6 factors. Out-of-sample calculations incorporate transaction costs. Within each model class, the 1-factor case serves as the benchmark; an asterisk (*) indicates that the Sharpe ratio is significantly lower than the benchmark at the 5% level, based on p-values computed using the stationary block-bootstrap method of Politis and Romano (1994).

method of Politis and Romano (1994). A significantly lower Sharpe ratio indicates a better-specified model.

In-sample results show that increasing the number of factors in the FF model does not significantly improve specification relative to the 1-factor case (CAPM). For PCA, IPCA, and AE, the 3-factor specifications do not improve upon the 1-factor case, but the 5- and 6-factor models clearly dominate the smaller-factor versions. Out-of-sample, attainable Sharpe ratios fall substantially—consistent with Fama and French (2018), Kan et al. (2024), and others—yet the conclusions mirror those of the in-sample analysis.

Next, I turn to cross-model class comparisons. This generality is a key advantage of the zero-beta portfolio approach. The Sharpe ratios are reported in Table 3, and Figure 3 visualizes the significance of model improvements using a heat map. I compare eight models: the 1-factor and 6-factor versions of FF, PCA, IPCA, and AE. In the plots, models on the horizontal axis serve as the benchmark; a black block indicates that the model on the vertical axis delivers a significantly lower attainable maximum Sharpe ratio, hence a better-specified model. Focusing first on Panel (a), the CAPM (FF1) and the 1-factor PCA significantly improve upon both IPCA1 and AE1, suggesting that simple factor structures outperform these machine-learning models in the 1-factor case. However, when additional factors are

Figure. 3. Cross-Model Class Comparison Results



(a) In-Sample (b) Out-of-Sample

Notes: Panel (a) shows the significance of in-sample model comparison. Panel (b) shows the significance of out-of-sample model comparison. I compare eight models: the 1-factor and 6-factor versions of FF, PCA, IPCA, and AE. In the plots, models on the horizontal axis serve as the benchmark; a black block indicates that the model on the vertical axis delivers a significantly lower attainable maximum Sharpe ratio.

introduced, the relative ranking changes. The 6-factor version of AE displays substantial gains: AE6 dominates all other models. This highlights the benefit of richer factor structures, particularly when using more flexible machine-learning approaches. Turning to Panel (b), which is arguably more informative for model comparisons from an investment perspective. The initial advantage of FF1 and PCA1 disappears; they no longer dominate the 1-factor machine-learning models. Instead, AE1 outperforms all the other 1-factor specifications, showing robustness of the machine learning approach in sparse factor environments. Among the 6-factor models, AE6 again performs strongly, significantly improving upon most alternatives, though it does not dominate IPCA6. This suggests that both IPCA and AE provide strong and competitive specifications when many factors are used, with AE delivering the broadest improvements. Overall, the cross-model comparison suggests that while traditional factor models remain competitive in simpler 1-factor settings, machine-learning approaches become clearly advantageous as the factor dimension grows. In particular, AE stands out as the most robust model, delivering substantial improvements both in-sample and out-of-sample.

Table 4: Investment Performance of Zero-Beta Portfolios

Models	# Factors	Sharpe Ratio		Mean (%)		STD (%)		MDD (%)		Hit Rate (%)	
		Gross	Net	Gross	Net	Gross	Net	Gross	Net	Gross	Net
FF	1	1.49	1.17	25.4	20.1	17.0	17.1	37.0	44.8	67.6	65.0
	3	1.49	1.17	25.7	20.3	17.2	17.3	33.6	43.0	67.9	65.2
	5	1.39	1.08	24.2	18.8	17.4	17.4	32.8	41.5	66.9	63.3
	6	1.38	1.07	24.4	18.8	17.7	17.7	32.7	40.2	66.7	62.6
PCA	1	1.49	1.17	25.3	20.0	17.0	17.1	37.1	44.7	67.6	64.8
	3	1.49	1.17	25.5	20.2	17.1	17.2	36.9	44.6	66.9	64.3
	5	1.39	1.07	24.3	18.8	17.4	17.6	36.5	45.5	66.4	62.1
	6	1.32	0.99	23.5	17.8	17.8	17.9	37.9	46.8	64.0	60.5
IPCA	1	1.48	1.15	24.7	19.5	16.7	16.9	41.1	49.3	68.1	65.0
	3	1.42	1.10	23.7	18.4	16.6	16.8	43.2	50.9	67.6	63.8
	5	1.25	0.92	19.8	14.6	15.8	15.9	41.4	49.3	66.7	63.3
	6	1.23	0.88	18.7	13.5	15.1	15.3	39.2	47.2	66.0	62.4
AE	1	1.48	1.16	24.9	19.6	16.8	16.9	40.8	49.4	67.9	65.0
	3	1.34	1.02	22.3	17.1	16.7	16.7	38.3	46.6	65.5	63.8
	5	1.36	0.98	19.9	14.6	14.7	14.9	38.9	45.3	66.4	62.4
	6	1.12	0.77	17.8	12.4	15.9	16.1	36.4	43.7	65.7	63.1

Notes: The table reports the performance of optimal zero-beta portfolios for the FF, PCA, IPCA, and AE models with 1, 3, 5, and 6 factors. Reported statistics include annualized Sharpe ratios, mean returns (%), standard deviations (%), maximum drawdowns (%), and hit rates (%). Zero-beta portfolio weights are rescaled each month to target a 15% annualized volatility, based on historical estimates of betas and the covariance matrix. Performance is shown both before transaction costs ("gross") and after transaction costs ("net").

4.5. Zero-Beta (Factor-Neutral) Investment

Out-of-sample Sharpe ratios reported in Table 3 not only facilitate model comparisons but also provide direct evaluations of investment opportunities. To sharpen this perspective, this section further examines the performance of zero-beta portfolios as actual investment strategies. The zero-beta portfolios are constructed on a purely out-of-sample basis by solving the constrained optimization problem (11) using the beta estimates through t, and tracking the post-formation $t+1$ return[21]. These portfolios are constructed based on characteristics-sorted portfolio betas.[22] Table 4 reports backtesting metrics for the FF, PCA, IPCA, and AE models with 1, 3, 5, and 6 factors, including annualized Sharpe ratios, annualized mean returns (%), annualized standard deviations (%), maximum drawdowns (%)[23], and hit rates (%)[24]. Although the models are re-estimated annually, portfolio positions are rebalanced monthly based on the most recent estimates of betas. The conditional optimal zero-beta portfolio weights are computed each month using equation (12), and then rescaled to target a 15% annualized volatility using historical estimates of betas and the covariance matrix. Investment performance is reported both before transaction costs ("gross") and after transaction costs ("net").

During the out-of-sample period from January 1990 to December 2024, the zero-beta portfolios deliver consistently strong investment performance, even after accounting for transaction costs. The Sharpe ratios across all specifications are remarkably high relative to the market portfolio benchmark, whose annualized Sharpe ratio is 0.53 before costs and 0.52 after costs.[25] By comparison, the zero-beta strategies achieve annualized net Sharpe ratios between 0.72 and 1.17, with many gross Sharpe ratios exceeding 1.3. A first observation is that the FF and PCA models stand out for their stability and persistence of high Sharpe ratios. For both models, the one- and three-factor portfolios achieve a gross Sharpe ratio of 1.49 (1.17 net), more than doubling the performance of the market. As dimensionality increases, performance modestly declines: the six-factor versions still deliver net Sharpe ratios around 1.07–1.08, which remain very strong in absolute terms. In contrast, IPCA and AE display a sharper deterioration in Sharpe ratios as the number of factors rises. For IPCA, the one-factor specification attains a net Sharpe ratio of 1.15, competitive with FF and PCA, but performance falls to 0.88 with six factors. AE follows a similar trajectory, starting at 1.16 net with one factor and declining to 0.77 with six factors. These declines are consistent with the findings in Section 4.4: as model specification improves—particularly through machine-learning methods—the profitability of zero-beta portfolios diminishes. The strong performance of one-factor strategies across all models is particularly striking. Net Sharpe ratios of 1.17 for FF and PCA, 1.15 for IPCA, and 1.16 for AE indicate that even sim-

[21] Instead of relying on the analytical solution in equation (12), I implement a numerical procedure that imposes the additional constraint that zero-beta portfolio weights cannot exceed one. The main results are robust to the exclusion of these constraints.

[22] I do not use individual stock betas to construct zero-beta portfolios, since inverting the individual-stock covariance matrix is notoriously unstable and beta estimates are considerably noisier than those obtained from characteristic-sorted portfolios.

[23] Maximum drawdown (MDD): largest peak-to-trough decline in investment values over a specific period.

[24] Hit rate: percentage of months in which the portfolio generates a positive return.

[25] Transaction costs associated with trading the market portfolio are minimal.

ple specifications—though far from fully capturing true risk structure—translate into highly profitable investment strategies. This suggests that from an investment perspective, a simple market-neutral strategy may suffice, while extending to multi-factor, beta-neutral portfolios offers limited incremental benefits in real-world trading.

Turning to returns and volatility, the mean annualized net returns range between 12% and 20% depending on the model and factor count. This compares favorably to typical equity premium estimates. Standard deviations are tightly centered around 15%–18% by construction, reflecting the monthly volatility-scaling rule. Importantly, volatility is well-controlled across models and specifications, which means that differences in Sharpe ratios are primarily driven by variations in mean returns rather than excess volatility. The risk management properties of these portfolios are more nuanced. Maximum drawdowns (MDD) are typically around 40–50% net of costs, comparable in magnitude to that of the market portfolio (51.9%). This suggests that, although the strategies deliver superior risk-adjusted returns on average, they remain exposed to occasional severe losses. In other words, high Sharpe ratios coexist with drawdowns of a scale similar to traditional equity markets (further discussed in Section **??**). Importantly, however, these drawdowns are smaller than those of the standard BAB and momentum strategy, whose maximum drawdown reached 74.7% and 75.3%, respectively, over the same period. Finally, the hit rate (the fraction of months with positive returns) provides additional insight into return consistency. Across models, hit rates range from roughly 60% to 70%, closely matching the market's 64.5%. This suggests that zero-beta strategies succeed not by winning more often than the market, but by delivering larger average payoffs per winning trade, consistent with their elevated Sharpe ratios.

Overall, the evidence indicates that zero-beta portfolios represent attractive investment opportunities, particularly in parsimonious one-factor implementations. Despite lacking systematic factor risk exposures by construction, they achieve high and persistent risk-adjusted returns, maintain volatility near targeted levels, and generate hit rates on par with the market. Particularly noteworthy is that one-factor models based on FF and PCA specifications deliver net Sharpe ratios near 1.2—more than double the market's Sharpe ratio—underscoring their remarkable efficiency as practical investment strategies.

Table 5: Investment Performance of Fator Mean-Variance Portfolios

Models	# Factors	Sharpe Ratio		Mean (%)		STD (%)		MDD (%)		Hit Rate (%)	
		Gross	Net	Gross	Net	Gross	Net	Gross	Net	Gross	Net
FF	1	0.53	0.51	8.0	7.8	15.1	15.1	54.3	55.1	62.9	62.9
	3	0.46	0.38	8.1	6.8	17.9	17.9	57.2	59.2	56.9	56.7
	5	0.64	0.50	17.2	13.6	26.9	27.1	79.6	82.5	61.7	60.5
	6	0.63	0.51	15.7	12.7	24.9	25.1	73.3	76.4	64.0	62.6
PCA	1	0.46	0.44	7.1	6.8	15.3	15.4	51.1	51.6	60.0	60.0
	3	0.58	0.52	10.2	9.3	17.7	17.8	71.9	73.0	66.0	66.0
	5	0.59	0.49	12.2	10.3	20.8	20.9	63.5	65.8	60.7	59.8
	6	0.93	0.79	15.3	13.1	16.4	16.7	47.6	51.4	65.5	65.0
IPCA	1	0.42	-4.76	6.5	--	15.5	++	44.6	++	57.6	0.0
	3	1.82	-3.48	35.2	--	19.3	++	58.4	++	74.3	0.0
	5	3.99	-4.66	85.7	--	21.5	++	19.6	++	91.0	0.0
	6	3.88	-4.69	87.8	--	22.6	++	21.0	++	91.2	0.0
AE	1	0.52	-5.11	8.4	--	16.1	++	42.6	++	59.2	0.0
	3	1.73	-4.11	34.8	--	20.1	++	59.1	++	73.1	0.0
	5	3.85	-4.89	85.7	--	22.2	++	20.7	++	89.7	0.0
	6	3.68	-4.97	84.3	--	22.9	++	21.2	++	90.2	0.0

Notes: The table reports the performance of factor mean-variance portfolios for the FF, PCA, IPCA, and AE models with 1, 3, 5, and 6 factors. Reported statistics include annualized Sharpe ratios, mean returns (%), standard deviations (%), maximum drawdowns (%), and hit rates (%). I scale the mean-variance portfolio weights each month to target a 15% annualized volatility using historical estimates of mean returns and the covariance matrix. Performance is shown both before transaction costs ("gross") and after transaction costs ("net"). "++" indicates a number above 100 (%), and "--" indicates a number below -100 (%).

4.5.1. Factor-Neutral Investing vs Factor-Based Investing

While the literature on factor investing has documented the strong performance of factor-based strategies, it has not comprehensively compared these approaches with factor-neutral investing in a consistent framework. This section addresses this gap by evaluating the out-of-sample performance of factor mean-variance portfolios (FMV), thereby complementing the results in the previous section. Table 5 reports the investment performance metrics across different factor models and numbers of factors.

Before accounting for transaction costs, the performance of factor-based portfolios is striking, with machine learning models clearly dominating the conventional FF and PCA benchmarks. For example, portfolios with five or six IPCA and AE factors achieve Sharpe ratios above 3.0 and annualized gross returns exceeding 80%, highlighting the large potential gains from exploiting systematic factor exposures. These results are consistent with Kelly et al. (2019) and Gu et al. (2021). Moreover, maximum drawdowns for these portfolios are as low as 20% and hit rates approach 100%, indicating highly favorable risk–return trade-offs in gross terms.

However, these extraordinary results do not survive once transaction costs are incorporated. Net Sharpe ratios and net returns fall substantially for all models, and in the case of IPCA and AE, net performance turns highly negative, reflecting the extreme trading turnover required to implement such strategies. This high turnover not only erodes mean returns but also inflates standard deviations and maximum drawdowns, making these strategies economically unviable. The contrast highlights the importance of accounting for practical frictions: while machine learning methods can generate striking gross performance, their real-world implementability is severely limited. By comparison, factor mean-variance portfolios based on FF and PCA deliver modest but positive net Sharpe ratios, outperforming machine learning models once costs are considered. Nonetheless, their net Sharpe ratios remain lower than those of zero-beta investing in Table 4, although maximum drawdowns and hit rates are broadly similar.

One way to enhance multi-factor portfolio performance is through volatility management. Prior work has shown that scaling down factor exposure during high-volatility periods improves Sharpe ratios. Moreira and Muir (2017) (MM) manage the FMV portfolio as a whole based on portfolio volatility, while DeMiguel et al. (2024) (DMU) manage each factor individually based on its factor volatility and then optimally combine them with their managed counterparts. Both approaches significantly improve multi-factor portfolio performance. I replicate their results in Tables B.2 and B.3, confirming that MM and DMU portfolios generate higher Sharpe ratios for FF and PCA factors, both before and after transaction costs.[26] By contrast, while IPCA and AE improve mean-variance efficiency in theory, they fail as implementable trading strategies due to extreme transaction costs.

In summary, although the literature has highlighted the remarkable factor mean-variance efficiency via machine learning methods, with reported Sharpe ratios easily exceeding 3.0, these results are not economically meaningful once transaction costs are accounted for. In

[26]For FF models, the improvement of MM and DMU is clearer when expanding the factor set beyond the six standard factors to include a BAB factor and two mispricing factors from Stambaugh and Yuan (2017). In unreported results, the net-of-cost (before-cost) Sharpe ratios are 0.74 (0.96) for FMV, 0.82 (1.05) for MM, and 0.83 (1.03) for DMU portfolios.

practice, factor-neutral investing through the zero-beta strategy delivers stronger and more robust investment performance, both before and after costs, than factor-based mean-variance portfolios.

4.5.2. Portfolio Construction: Individual Stocks vs Sorted Portfolios

The optimal zero-beta strategies are constructed using characteristic-sorted portfolio betas rather than individual stock betas, since mean–variance optimization requires the asset covariance matrix as input, and the individual-stock covariance matrix is typically less reliable. The literature has proposed alternative approaches to constructing beta-neutral portfolios that hedge risks associated with stock characteristics while exploiting the "mispricing" component of individual stock returns. These are referred to as "arbitrage" portfolios.[27] Table 6 reports the performance of the "arbitrage" portfolios of Kelly et al. (2019) (KPS) and Kim et al. (2021) (KKN) with 1, 3, 5, and 6 factors. Both approaches aim to extract the mispricing component (alphas) of individual stocks, orthogonal to the risks associated with stock characteristics (betas). Portfolio weights are then set proportional to the estimated mispricing signals, implying that the strategy goes long stocks with high predicted alphas and short stocks with low or negative predicted alphas.

Table 6 highlights that while these "arbitrage" portfolios may perform impressively before transaction costs, their performance deteriorates sharply once costs are accounted for. The KPS portfolios are particularly striking: gross Sharpe ratios approach 3.0, annualized volatility remains around 10% despite targeting 15% in-sample, maximum drawdowns are minimal, and hit rates exceed 80%. Yet the high trading turnover wipes out these gains—net Sharpe ratios turn highly negative, and hit rates fall below 30%. The KKN portfolios perform more modestly, with gross Sharpe ratios around 1, but transaction costs again reduce them to near zero. These findings suggest that zero-beta strategies are more appropriately constructed using characteristic-sorted portfolios rather than individual stocks once trading frictions are incorporated.

A further observation is that the performance of the "arbitrage" portfolios does not materially change with the number of factors. This suggests that their profitability may not be driven by beta-neutrality, since removing additional systematic risk exposures does not alter performance. This echoes the concern that individual-stock beta estimates are relatively noisy, making them a weaker basis for portfolio construction.

[27] Both the zero-beta strategies and the "arbitrage" portfolios in Kelly et al. (2019) and Kim et al. (2021) are not arbitrage in the classical, risk-free sense. Rather, they are forms of statistical arbitrage that use quantitative models to identify potential mispricings and construct portfolios hedged against known sources of systematic risk. They still bear risk, as true risk-free arbitrage may not exist in practice due to frictions and limits to arbitrage (Shleifer and Vishny, 1997).

Table 6: Investment Performance of "Arbitrage" Portfolios

Methods	# Factors	Sharpe Ratio		Mean (%)		STD (%)		MDD (%)		Hit Rate (%)	
		Gross	Net	Gross	Net	Gross	Net	Gross	Net	Gross	Net
KPS	1	2.60	-1.97	31.5	-22.3	12.1	11.3	10.4	100.0	82.1	21.9
	3	2.96	-1.38	31.9	-14.4	10.8	10.4	9.8	99.5	84.3	29.3
	5	2.76	-1.84	28.8	-21.2	10.4	11.5	12.1	100.0	83.1	28.8
	6	2.78	-1.77	27.8	-19.4	10.0	11.0	10.4	99.9	81.7	28.3
KKN	1	1.01	0.05	19.8	1.0	19.6	18.7	63.7	87.3	64.8	52.4
	3	1.01	0.04	19.7	0.8	19.6	18.7	64.1	88.2	64.5	52.1
	5	1.01	0.05	19.8	0.9	19.6	18.7	63.5	87.8	64.8	51.9
	6	1.01	0.04	19.7	0.8	19.6	18.7	63.5	87.9	64.5	51.9

Notes: The table reports the performance of the "arbitrage" portfolios of Kelly et al. (2019) and Kim et al. (2021) with 1, 3, 5, and 6 factors. Reported statistics include annualized Sharpe ratios, mean returns (%), standard deviations (%), maximum drawdowns (%), and hit rates (%). I scale the mean-variance portfolio weights each month to target a 15% annualized volatility using historical estimates of mean returns and the covariance matrix. Performance is shown both before transaction costs ("gross") and after transaction costs ("net"). "++" indicates a number above 100 (%), and "--" indicates a number below -100 (%).

4.5.3. Risk-adjusted Returns

To further assess the performance of zero-beta portfolios, Table 7 reports their risk-adjusted returns (alphas, in percent) along with risk loadings from time-series regressions on common risk factors: the Fama–French six factors (MKT, SMB, HML, RMW, CMA, UMD), an idiosyncratic risk factor[28], and the Pastor–Stambaugh liquidity factor. The table presents portfolio net returns for models with one and six factors (similar results for gross returns and models with three or five factors are reported in Appendix B.8). Alphas are highly positive and statistically significant across all models. Risk exposures to most factors are negligible, and although there is some exposure to the value and momentum factors, adjusted R^2 remains low, with a maximum of 0.16. These results reinforce the interpretation that the portfolios' strong performance is not driven by traditional factor risk exposures.

[28] IVOL is measured as the standard deviation of daily CAPM residuals over the past year. Results are robust to using Fama–French three-factor residuals or a one-month lookback window.

Table 7: Regression of Zero-Beta Portfolio Net Returns on Benchmark Risk Factors

	FF		PCA		IPCA		AE	
# Factors	1	6	1	6	1	6	1	6
Alpha (%)	1.29***	1.46***	1.29***	1.26***	1.27***	0.88***	1.29***	0.86***
	(4.25)	(4.32)	(4.25)	(3.79)	(4.33)	(3.45)	(4.33)	(3.30)
MKT	0.13	0.06	0.12	0.10	0.08	-0.03	0.07	0.02
	(1.59)	(0.64)	(1.48)	(1.15)	(1.11)	(-0.39)	(0.98)	(0.24)
SMB	0.06	-0.07	0.09	-0.04	0.08	-0.07	0.08	-0.02
	(0.34)	(-0.38)	(0.46)	(-0.21)	(0.44)	(-0.53)	(0.47)	(-0.12)
HML	0.32***	0.11	0.31***	0.37***	0.25**	0.27***	0.25**	0.26*
	(2.54)	(0.80)	(2.55)	(2.52)	(2.06)	(2.55)	(2.05)	(1.77)
RMW	0.26	0.06	0.26	0.08	0.36*	0.41***	0.34*	0.18
	(1.31)	(0.28)	(1.34)	(0.35)	(1.88)	(2.69)	(1.73)	(0.98)
CMA	0.13	-0.06	0.14	-0.23	0.06	0.18	0.07	-0.12
	(0.64)	(-0.29)	(0.69)	(-1.02)	(0.34)	(1.22)	(0.37)	(-0.63)
UMD	0.28**	0.20	0.28**	0.25*	0.25**	0.22**	0.25*	0.31***
	(2.20)	(1.53)	(2.20)	(1.81)	(2.03)	(1.97)	(1.94)	(2.73)
IVOL	-0.05	-0.14	-0.05	-0.02	-0.00	-0.09	-0.01	-0.13
	(-0.35)	(-0.79)	(-0.34)	(-0.10)	(-0.01)	(-0.74)	(-0.05)	(-0.84)
LIQ	-0.06	-0.07	-0.06	-0.03	-0.06	-0.09	-0.06	-0.11
	(-0.88)	(-0.92)	(-0.82)	(-0.39)	(-0.91)	(-1.59)	(-0.87)	(-1.63)
Adj.R^2	0.11	0.04	0.12	0.06	0.12	0.16	0.12	0.09

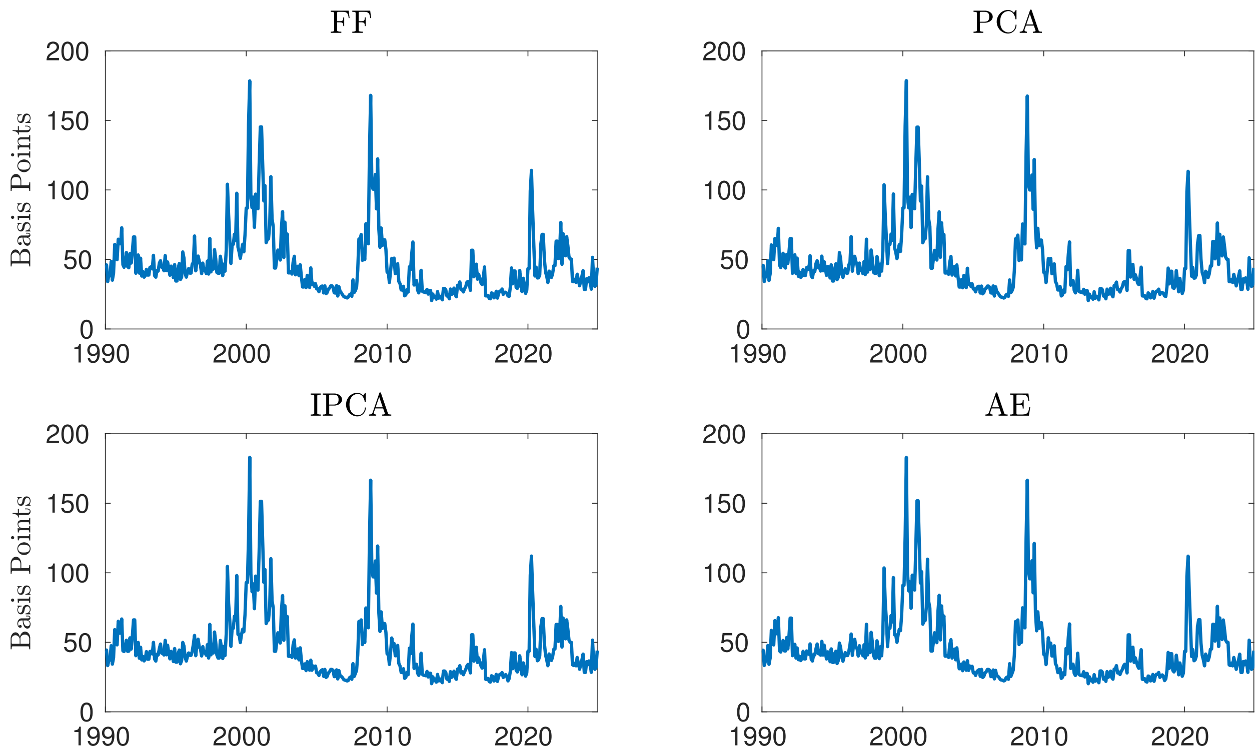
Notes: This table presents alphas (%), factor loadings, and adjusted R^2 from regressions of zero-beta portfolio returns (net-of-costs) on the Fama–French six factors, an idiosyncratic volatility factor, and a liquidity factor. The models include FF, PCA, IPCA, and AE with one and six factors. Newey-West t-statistics are shown in parentheses. Significance levels: *** $p < .01$, ** $p < .05$, * $p < .1$.

4.5.4. Portfolio Transaction Costs and Weights

I further investigate the transaction costs incurred by trading the optimal zero-beta portfolios, which are plotted in Figure 4 for FF, PCA, IPCA, and AE models with a single factor.[29] Since each model is re-estimated annually at the end of December, transaction costs exhibit pronounced spikes in January, reflecting adjustments in model parameter esti-

[29] Appendix B.11 reports zero-beta portfolio transaction costs for 3, 5, and 6 factors, which exhibit nearly identical magnitudes and time-series patterns.

Figure. 4. Zero-Beta Portfolio Transaction Costs (Single Factor)



Notes: This figure shows the transaction costs of the optimal zero-beta portfolios from January 1990 to December 2024, measured using the average low-frequency effective spreads in Chen and Velikov (2023). The models considered are FF, PCA, IPCA, and AE with a single factor.

mates. In computing the backtesting metrics in the previous sections, I maintain the realistic assumption that transaction costs are incurred when they actually occur. However, for visualization purposes, to avoid showing these sharp spikes, I distribute the January transaction costs evenly across all months of the same year, effectively assuming that the January costs are amortized over the year. Importantly, this assumption of cost amortization does not materially affect the results reported in the previous sections.

A clear pattern emerges across all models. Transaction costs of zero-beta portfolios are about 60–70 basis point on average across all models, with pronounced spikes exceeding 150 basis points during major market crises, notably the dot-com bust (2000), the Global Financial Crisis (2008), and the COVID-19 pandemic (2020). These peaks may be driven by two main factors. First, crises are periods of extreme volatility and strained liquidity, which widen bid–ask spreads on individual stocks and make trading inherently more expensive. Figure B.14 illustrates this by showing the time variation in the mean, median, 5th percentile, and 95th percentile of individual transaction costs from January 1960 to December 2024, with visible peaks around 2000, 2008, and 2020. Second, during crises, the risk characteristics of assets change rapidly, forcing the models to rebalance their portfolios more frequently and aggressively to preserve the zero-beta hedge. This leads to higher turnover and further increases in costs. Figures B.18 – B.21 plot the median absolute weight changes of optimal zero-beta portfolios across 272 characteristic-sorted portfolios from January 1990 to December 2024, showing that rebalancing activity does intensify in crises, although the magnitudes remain modest. Hence, the spikes in transaction costs are primarily driven by liquidity stress and widening bid–ask spreads rather than weight adjustments alone. Importantly, these cost spikes are not inconsistent with the beta neutrality. The purpose of a zero-beta portfolio is to neutralize systematic risk exposure, not to minimize implementation costs. Higher trading costs during crises thus represent the unavoidable price of maintaining the hedge precisely when markets are most turbulent and the hedge is most valuable.

A legitimate concern with zero-investment portfolios is that they may allocate unrealistically large weights to individual stocks. Figures B.22 - B.23 address this concern by plotting the maximum, minimum, 1st percentile, and 99th percentile of zero-beta portfolio weights on individual stocks from January 1990 to December 2024. The results show that the largest absolute weight assigned to any individual stock does not exceed 5%, indicating that the portfolios are well diversified and free from extreme concentration.

4.5.5. Market Betas and Idiosyncratic Risks of Long vs Short Positions

Another well-known beta-neutral strategy is the BAB portfolio of Frazzini and Pedersen (2014), which takes long positions in low-beta stocks and short positions in high-beta stocks. The strategy builds on the "beta anomaly" documented in the literature.[30] This naturally raises the question: do the proposed optimal zero-beta portfolios resemble BAB in their positioning? In other words, do they systematically go long low-beta stocks and short high-beta stocks? Figure B.24 - B.27 plot the weighted average market betas of the long and short sides of the optimal zero-beta portfolios. The evidence shows otherwise: long and short positions have similar weighted average betas, with no clear tilt toward low- or high-beta stocks. By contrast, Figure B.28 illustrates the BAB case, which clearly features low-beta

[30] There are a number of theories explaining the beta anomaly. Several theories have been proposed to explain the beta anomaly. For example, Black (1972), Frazzini and Pedersen (2014), and Asness et al. (2020) argue that leverage-constrained investors, unable to borrow to lever up market exposure, instead tilt toward high-beta stocks. This excess demand inflates the prices of high-beta stocks, leading to their underperformance. Bali et al. (2017) attribute the anomaly to behavioral biases. High-beta stocks often have lottery-like features (e.g., high skewness, high maximum daily returns), which attracts investors who are willing to accept lower average returns for the small chance of a large payoff. Liu et al. (2018) argue that the anomaly reflects the correlation between beta and idiosyncratic volatility (IVOL): high-beta stocks tend to have high IVOL, which deters arbitrage and makes them more likely to be overpriced. Schneider et al. (2020) find that low-beta stocks carry higher coskewness risk—underperforming in extreme downturns—so their higher returns may reflect rational compensation for higher-order risks rather than mispricing.

longs and high-beta shorts.

Liu et al. (2018) argue that the beta anomaly is largely driven by the strong positive cross-sectional correlation between beta and idiosyncratic volatility (IVOL). Figure B.33 confirms that the BAB strategy mainly takes long positions in low-IVOL stocks and short positions in high-IVOL stocks. To explore whether the optimal zero-beta portfolios display a similar pattern, I examine the weighted average IVOL of their long and short sides. Figures B.29–B.32 show that long and short positions have comparable average IVOL, with only a mild tilt toward low-IVOL in longs and high-IVOL in shorts.

4.5.6. Other Benchmark Strategies

I also consider prominent investment strategies proposed in the literature. Following Gu et al. (2020) and Gu et al. (2021), I construct long–short portfolios based on model-predicted returns. For each model, stocks are sorted into deciles according to their out-of-sample predicted returns. The strategy goes long the top-decile (highest expected return) stocks and short the bottom-decile (lowest expected return) stocks, with equal weights applied within each leg. Table B.4 reports the performance of these prediction-sorted portfolios. Since unconditional models are poor at out-of-sample return prediction (see Table 1), FF and PCA yield non-positive Sharpe ratios. In contrast, machine learning methods substantially improve predictability and, consequently, portfolio performance. However, their out-of-sample net-of-cost Sharpe ratios remain below those of zero-beta strategies.

I also examine conventional model-free investment strategies. Table B.5 reports results for the market portfolio (MKT), the mean–variance portfolio (MVP), the global minimum-variance portfolio (GMV), the equal-weighted portfolio (EW), the risk-parity portfolio (RP), and the Kan et al. (2024) optimal portfolio (KWZ). Consistent with the literature, MVP underperforms relative to simple rules such as GMV, EW, and RP. Among these model-free strategies, the KWZ rule achieves a before-cost Sharpe ratio of 1.31 and a net-of-cost Sharpe ratio of 1.02, offering attractive investment opportunities.

4.5.7. Robustness

Before Covid, shrinkage covariance matrix estimation, weights constraints, additional sets of stock characteristics,

5. Economic Mechanism and Implications

Why do the proposed optimal zero-beta portfolios outperform most prominent benchmark strategies? Since Section 4.5.5 shows that they do not primarily go long in low-beta/low-

IVOL stocks or short in high-beta/high-IVOL stocks, the existing beta-anomaly explanations are unlikely to account for their outperformance.

5.1. Mispricing

5.2. Limits to arbitrage

5.3. Behavioral biases

6. Conclusion

References

Abdi, F., Ranaldo, A., 2017. A simple estimation of bid-ask spreads from daily close, high, and low prices. The Review of Financial Studies 30, 4437–4480.

Ang, A., Liu, J., Schwarz, K., 2020. Using stocks or portfolios in tests of factor models. Journal of Financial and Quantitative Analysis 55, 709–750.

Asness, C., Frazzini, A., Gormsen, N. J., Pedersen, L. H., 2020. Betting against correlation: Testing theories of the low-risk effect. Journal of Financial Economics 135, 629–652.

Avellaneda, M., Lee, J.-H., 2010. Statistical arbitrage in the us equities market. Quantitative Finance 10, 761–782.

Bali, T. G., Brown, S. J., Murray, S., Tang, Y., 2017. A lottery-demand-based explanation of the beta anomaly. Journal of Financial and Quantitative Analysis 52, 2369–2397.

Barillas, F., Kan, R., Robotti, C., Shanken, J., 2020. Model comparison with sharpe ratios. Journal of Financial and Quantitative Analysis 55, 1840–1874.

Barillas, F., Shanken, J., 2017. Which alpha? The Review of Financial Studies 30, 1316–1338.

Black, F., 1972. Capital market equilibrium with restricted borrowing. The Journal of business 45, 444–455.

Bryzgalova, S., Huang, J., Julliard, C., 2023. Bayesian solutions for the factor zoo: We just ran two quadrillion models. The Journal of Finance 78, 487–557.

Carhart, M. M., 1997. On persistence in mutual fund performance. The Journal of finance 52, 57–82.

Chamberlain, G., Rothschild, M., 1983. Arbitrage, factor structure, and mean-variance analysis on large asset markets.

Chen, A. Y., Velikov, M., 2023. Zeroing in on the expected returns of anomalies. Journal of Financial and Quantitative Analysis 58, 968–1004.

Chen, L., Pelger, M., Zhu, J., 2024. Deep learning in asset pricing. Management Science 70, 714–750.

Chou, P.-H., 2000. Alternative tests of the zero-beta capm. Journal of Financial Research 23, 469–493.

Cochrane, J. H., 2011. Presidential address: Discount rates. The Journal of finance 66, 1047–1108.

Connor, G., 1984. A unified beta pricing theory. Journal of Economic Theory 34, 13–31.

Connor, G., Korajczyk, R. A., 1986. Performance measurement with the arbitrage pricing theory: A new framework for analysis. Journal of financial economics 15, 373–394.

Connor, G., Korajczyk, R. A., 1988. Risk and return in an equilibrium apt: Application of a new test methodology. Journal of financial economics 21, 255–289.

Cooper, I., Ma, L., Maio, P., Philip, D., 2021. Multifactor models and their consistency with the apt. The Review of Asset Pricing Studies 11, 402–444.

Corwin, S. A., Schultz, P., 2012. A simple way to estimate bid-ask spreads from daily high and low prices. The journal of finance 67, 719–760.

Daniel, K., Titman, S., et al., 2011. Testing factor-model explanations of market anomalies. Critical Finance Review 1, 103–139.

DeMiguel, V., Martín-utrera, A., Uppal, R., 2024. A multifactor perspective on volatility-managed portfolios. The Journal of Finance 79, 3859–3891.

Detzel, A., Novy-Marx, R., Velikov, M., 2023. Model comparison with transaction costs. The Journal of Finance 78, 1743–1775.

Fama, E. F., French, K. R., 1993. Common risk factors in the returns on stocks and bonds. Journal of financial economics 33, 3–56.

Fama, E. F., French, K. R., 2015. Incremental variables and the investment opportunity set. Journal of Financial Economics 117, 470–488.

Fama, E. F., French, K. R., 2018. Choosing factors. Journal of financial economics 128, 234–252.

Fama, E. F., MacBeth, J. D., 1973. Risk, return, and equilibrium: Empirical tests. Journal of political economy 81, 607–636.

Fan, J., Liao, Y., Wang, W., 2016. Projected principal component analysis in factor models. Annals of statistics 44, 219.

Feng, G., He, J., Polson, N. G., Xu, J., 2024. Deep learning in characteristics-sorted factor models. Journal of Financial and Quantitative Analysis 59, 3001–3036.

Ferson, W. E., Harvey, C. R., 1991. The variation of economic risk premiums. Journal of political economy 99, 385–415.

Frazzini, A., Israel, R., Moskowitz, T. J., 2018. Trading costs, vol. 3229719. SSRN.

Frazzini, A., Pedersen, L. H., 2014. Betting against beta. Journal of financial economics 111, 1–25.

Fung, W., Hsieh, D. A., 2001. The risk in hedge fund strategies: Theory and evidence from trend followers. The review of financial studies 14, 313–341.

Gagliardini, P., Ossola, E., Scaillet, O., 2016. Time-varying risk premium in large cross-sectional equity data sets. Econometrica 84, 985–1046.

Gatev, E., Goetzmann, W. N., Rouwenhorst, K. G., 2006. Pairs trading: Performance of a relative-value arbitrage rule. The review of financial studies 19, 797–827.

Gibbons, M. R., 1982. Multivariate tests of financial models: A new approach. Journal of financial economics 10, 3–27.

Gibbons, M. R., Ross, S. A., Shanken, J., 1989. A test of the efficiency of a given portfolio. Econometrica: Journal of the Econometric Society pp. 1121–1152.

Giglio, S., Xiu, D., 2021. Asset pricing with omitted factors. Journal of Political Economy 129, 1947–1990.

Giglio, S., Xiu, D., Zhang, D., 2025. Test assets and weak factors. The Journal of Finance 80, 259–319.

Gu, S., Kelly, B., Xiu, D., 2020. Empirical asset pricing via machine learning. The Review of Financial Studies 33, 2223–2273.

Gu, S., Kelly, B., Xiu, D., 2021. Autoencoder asset pricing models. Journal of Econometrics 222, 429–450.

Guijarro-Ordonez, J., Pelger, M., Zanotti, G., 2021. Deep learning statistical arbitrage. arXiv preprint arXiv:2106.04028 .

Hansen, L. P., Jagannathan, R., 1997. Assessing specification errors in stochastic discount factor models. The Journal of Finance 52, 557–590.

Harvey, C. R., Liu, Y., Zhu, H., 2016. … and the cross-section of expected returns. The Review of Financial Studies 29, 5–68.

Hasbrouck, J., 2009. Trading costs and returns for us equities: Estimating effective costs from daily data. The Journal of Finance 64, 1445–1477.

Hou, K., Xue, C., Zhang, L., 2015. Digesting anomalies: An investment approach. The Review of Financial Studies 28, 650–705.

Huberman, G., 1982. A simple approach to arbitrage pricing theory. Journal of Economic Theory 28, 183–191.

Ingersoll Jr, J. E., 1984. Some results in the theory of arbitrage pricing. The Journal of Finance 39, 1021–1039.

Jensen, T. I., Kelly, B., Pedersen, L. H., 2023. Is there a replication crisis in finance? The Journal of Finance 78, 2465–2518.

Jobson, J. D., Korkie, B., 1982. Potential performance and tests of portfolio efficiency. Journal of Financial Economics 10, 433–466.

Kan, R., Robotti, C., 2009. Model comparison using the hansen-jagannathan distance. The Review of Financial Studies 22, 3449–3490.

Kan, R., Robotti, C., Shanken, J., 2013. Pricing model performance and the two-pass cross-sectional regression methodology. The Journal of Finance 68, 2617–2649.

Kan, R., Wang, X., Zheng, X., 2024. In-sample and out-of-sample sharpe ratios of multi-factor asset pricing models. Journal of Financial Economics 155, 103837.

Kan, R., Wang, X., Zhou, G., 2022. Optimal portfolio choice with estimation risk: No risk-free asset case. Management Science 68, 2047–2068.

Kandel, S., 1984. The likelihood ratio test statistic of mean-variance efficiency without a riskless asset. Journal of Financial Economics 13, 575–592.

Kandel, S., 1986. The geometry of the maximum likelihood estimator of the zero-beta return. The Journal of Finance 41, 339–346.

Karolyi, G. A., Van Nieuwerburgh, S., 2020. New methods for the cross-section of returns. The Review of Financial Studies 33, 1879–1890.

Kelly, B. T., Pruitt, S., Su, Y., 2019. Characteristics are covariances: A unified model of risk and return. Journal of Financial Economics 134, 501–524.

Kim, S., Korajczyk, R. A., Neuhierl, A., 2021. Arbitrage portfolios. The Review of Financial Studies 34, 2813–2856.

Kleibergen, F., Zhan, Z., 2020. Robust inference for consumption-based asset pricing. The Journal of Finance 75, 507–550.

Kyle, A. S., Obizhaeva, A. A., 2016. Market microstructure invariance: Empirical hypotheses. Econometrica 84, 1345–1404.

Lettau, M., Pelger, M., 2020. Factors that fit the time series and cross-section of stock returns. The Review of Financial Studies 33, 2274–2325.

Lewellen, J., Nagel, S., Shanken, J., 2010. A skeptical appraisal of asset pricing tests. Journal of Financial economics 96, 175–194.

Liu, J., Stambaugh, R. F., Yuan, Y., 2018. Absolving beta of volatility's effects. Journal of Financial Economics 128, 1–15.

Moreira, A., Muir, T., 2017. Volatility-managed portfolios. The Journal of Finance 72, 1611–1644.

Politis, D. N., Romano, J. P., 1994. The stationary bootstrap. Journal of the American Statistical association 89, 1303–1313.

Roll, R., 1980. Orthogonal portfolios. Journal of Financial and Quantitative analysis 15, 1005–1023.

Ross, S. A., 1976. The arbitrage theory of capital asset pricing. In: *Journal of Economic Theory*, Elsevier, pp. 341–360.

Schneider, P., Wagner, C., Zechner, J., 2020. Low-risk anomalies? The Journal of Finance 75, 2673–2718.

Shanken, J., 1985. Multivariate tests of the zero-beta capm. Journal of financial economics 14, 327–348.

Shanken, J., 1986. Testing portfolio efficiency when the zero-beta rate is unknown: a note. The Journal of Finance 41, 269–276.

Shleifer, A., Vishny, R. W., 1997. The limits of arbitrage. The Journal of finance 52, 35–55.

Stambaugh, R. F., 1982. On the exclusion of assets from tests of the two-parameter model: A sensitivity analysis. Journal of financial economics 10, 237–268.

Stambaugh, R. F., Yuan, Y., 2017. Mispricing factors. The review of financial studies 30, 1270–1315.

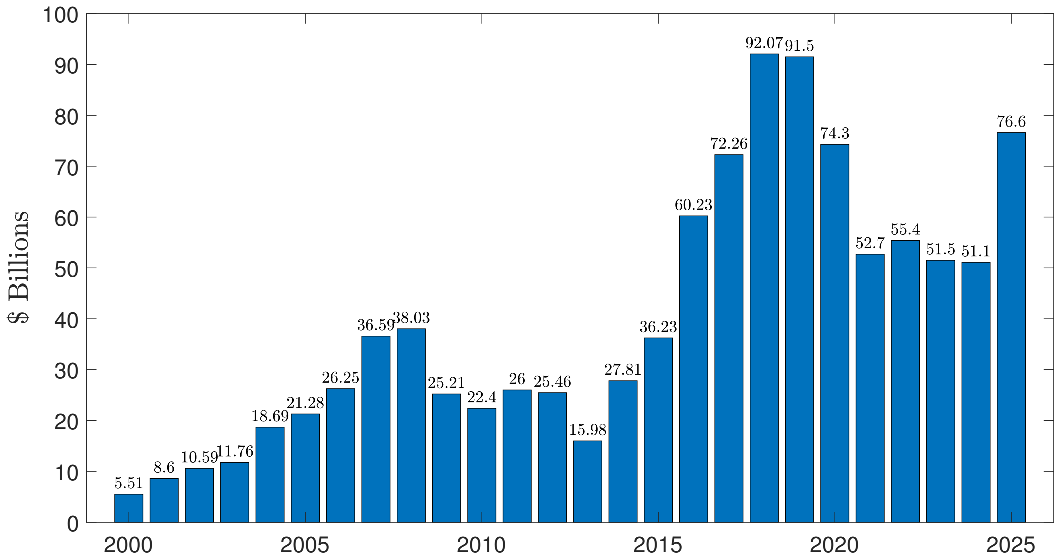
Appendix A. Data Construction

A.1. Stock Characteristics

Appendix B. Additional Empirical Results

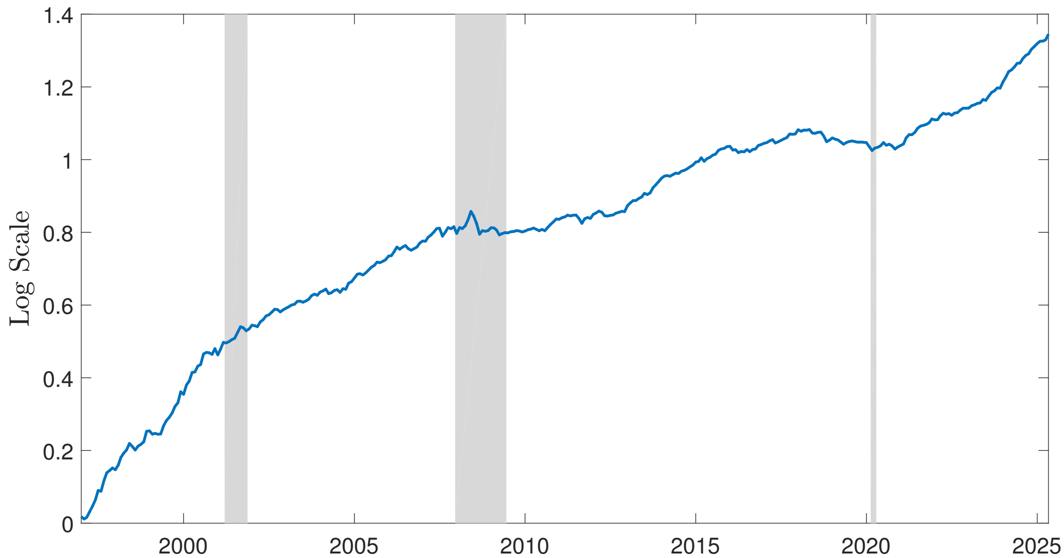
B.1. Equity Market Neutral Hedge Funds

Figure. B.1. Asset Under Management of Equity Market Neutral Hedge Funds



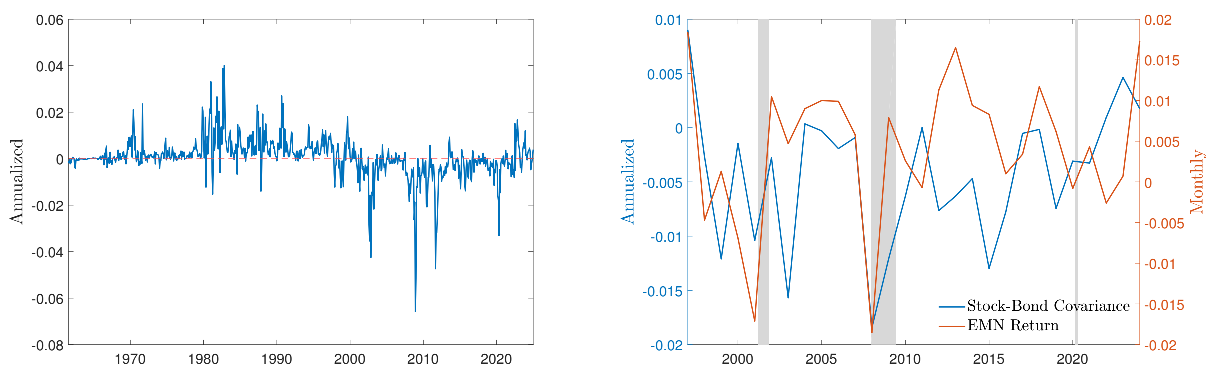
Notes: This figure shows the asset under management (AUM) of Equity Market Neutral Hedge Funds from the first quarter of 2000 to the first quarter of 2025. Source: BarclayHedge.

Figure. B.2. Barclay Equity Market Neutral Index



Notes: This figure shows the Barclay Equity Market Neutral Index, calculated and updated real-time as the cumulative equal-weighted average monthly (net) return of hedge funds in the Equity Market Neutral category. Source: BarclayHedge.

Figure. B.3. Stock-Bond Covariance and EMN Returns



(a) Monthly Stock-Bond Covariance (b) Annual Covariance and EMN Returns

Notes: Panel (a) shows the monthly stock-bond covariance between nominal 10-year constant maturity bond returns and S&P 500 returns using a 30 trading-day rolling window. Panel (b) shows the annual trend in this covariance together with Barclay EMN returns.

The correlation between the annual trend in stock–bond covariance and Barclay EMN returns is 0.40. A regression of monthly EMN returns on monthly stock–bond covariance (including a constant) yields a significant coefficient of 0.23 with a t-statistic of 4.76. Alternatively, Using percentage changes in AUM as a proxy for returns at the quarterly frequency produces a coefficient of 3.93 with a t-statistic of 4.06. These results indicate that the performance of equity market-neutral strategies is linked to the hedging role of bonds. When stock–bond covariance rises, bonds become less reliable as hedges for equities, making market-neutral strategies appear more attractive as alternative diversification tools. The traditional 60/40 allocation rule gained popularity partly because stock–bond covariance had been negative since 2000. However, this pattern has recently reversed, with covariance turning positive after Covid. This shift suggests that market-neutral strategies may attract greater investor interest going forward.

B.2. In-Sample Factor Model Performance

Table B.1: In-Sample Model Performance

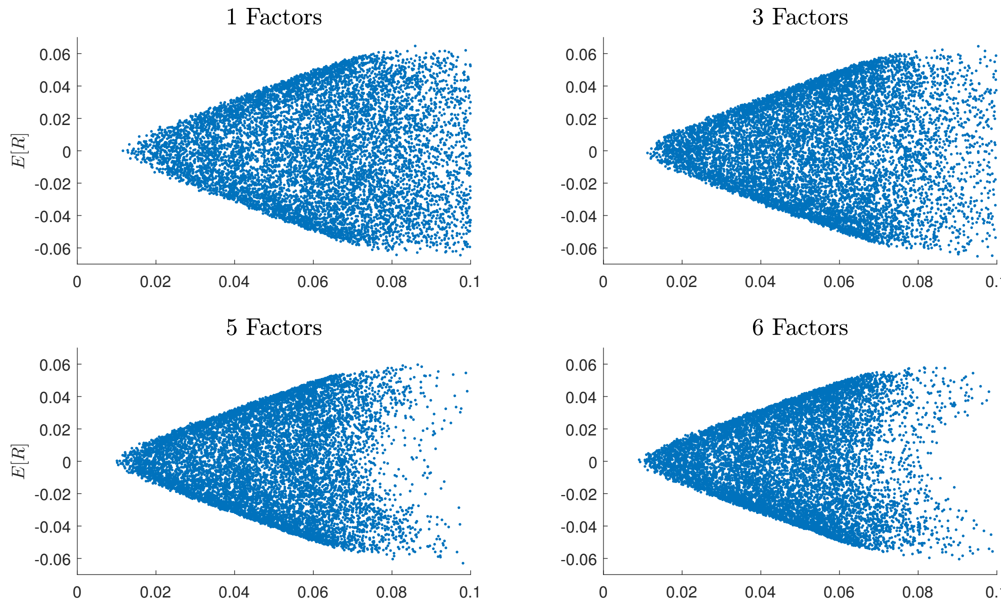
Models	Test Assets	Metrics	# Factors			
			1	3	5	6
FF	Individual Stocks	Total R^2	10.9	17.0	18.7	19.5
		Pred R^2	0.93	0.90	0.89	0.87
	Portfolios	Total R^2	89.4	94.1	95.2	95.4
		Pred R^2	3.42	3.42	3.42	3.42
PCA	Individual Stocks	Total R^2	4.5	10.3	13.1	14.3
		Pred R^2	0.01	0.28	0.34	0.34
	Portfolios	Total R^2	94.1	97.4	98.1	98.3
		Pred R^2	3.42	3.42	3.42	3.42
IPCA	Individual Stocks	Total R^2	12.8	15.1	15.8	16.0
		Pred R^2	0.81	0.79	0.78	0.77
	Portfolios	Total R^2	78.7	94.0	95.0	95.5
		Pred R^2	3.01	3.10	3.18	3.17
AE	Individual Stocks	Total R^2	12.4	13.3	13.5	13.5
		Pred R^2	1.12	1.02	1.09	1.16
	Portfolios	Total R^2	84.1	92.9	92.9	93.3
		Pred R^2	3.52	3.50	3.51	3.37

Notes: This table reports the in-sample total $R2$ and predictive R^2 in percentages (%) for FF, PCA, ICA, and AE with 1, 3, 5, and 6 factors.

B.3. *Autoencoders with Different Number of Hidden Layers*

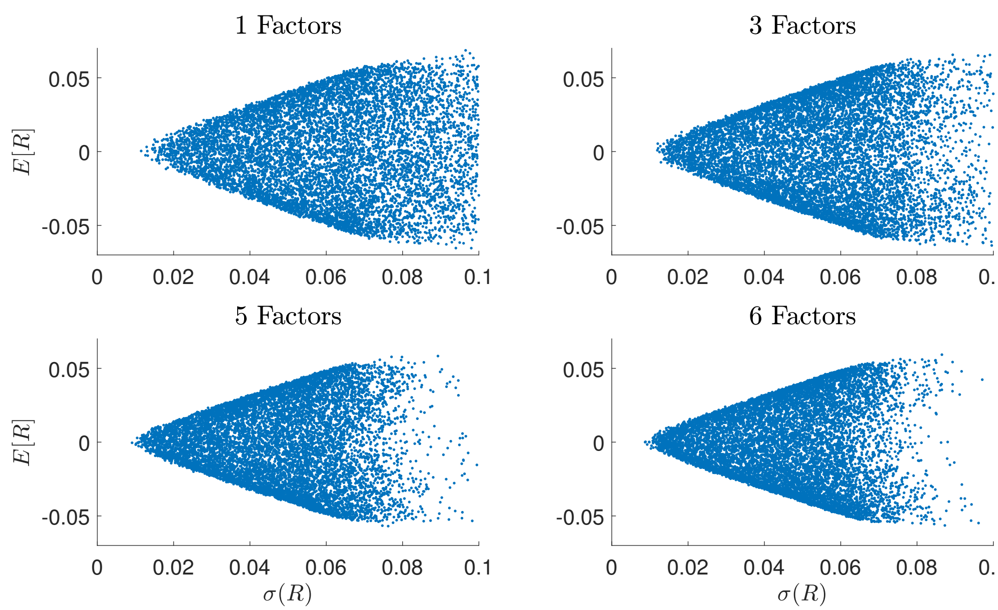
B.4. *Visualization of Model Testing and Comparison*

Figure. B.4. Zero-Investment, Zero-Beta Portfolios: FF Models



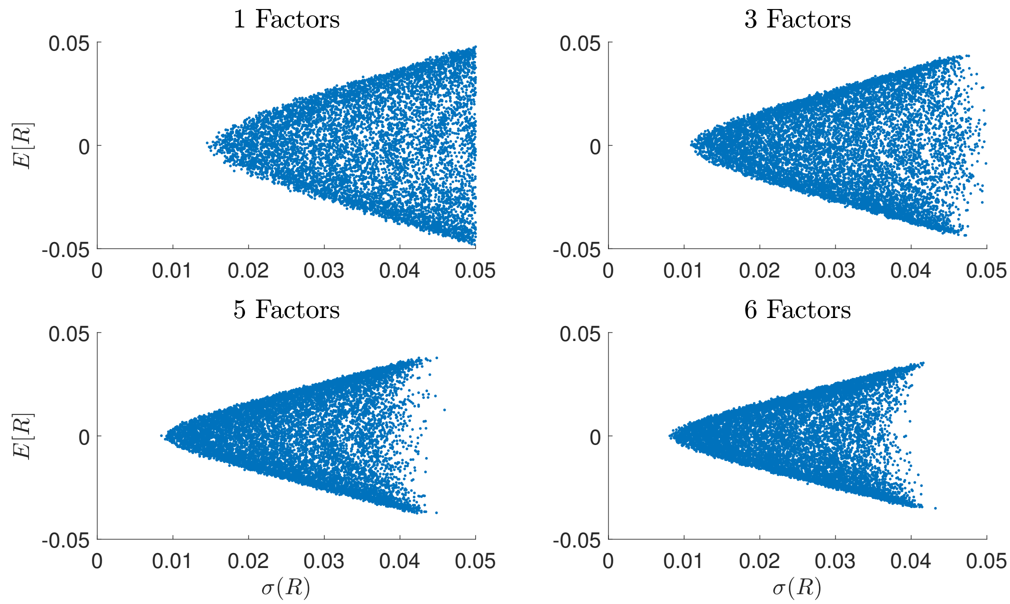
Notes: This figure shows the zero-investment, zero-beta portfolios for FF models.

Figure. B.5. Zero-Investment, Zero-Beta Portfolios: PCA Models



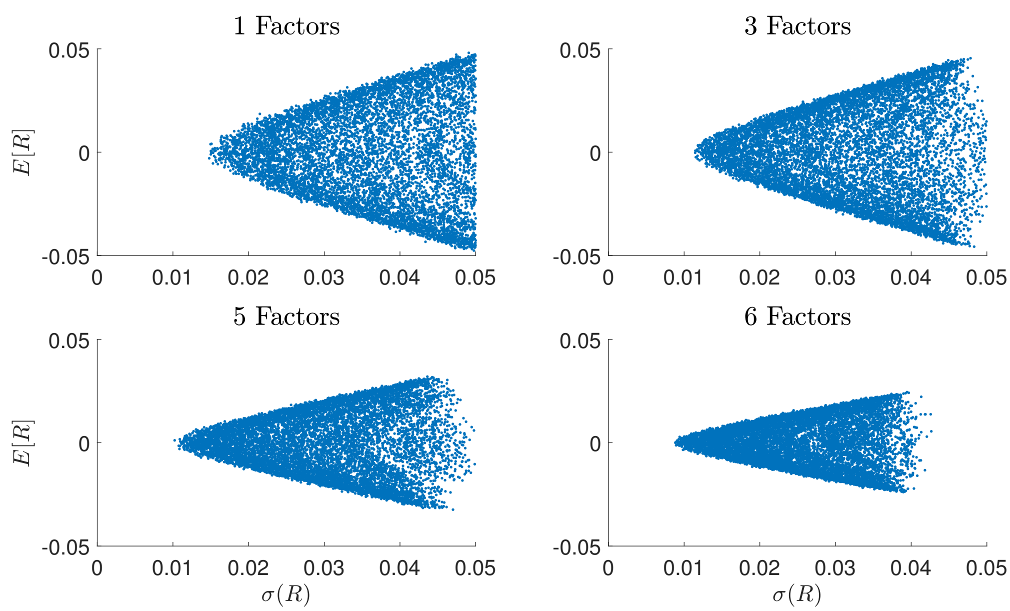
Notes: This figure shows the zero-investment, zero-beta portfolios for PCA models.

Figure. B.6. Zero-Investment, Zero-Beta Portfolios: IPCA Models



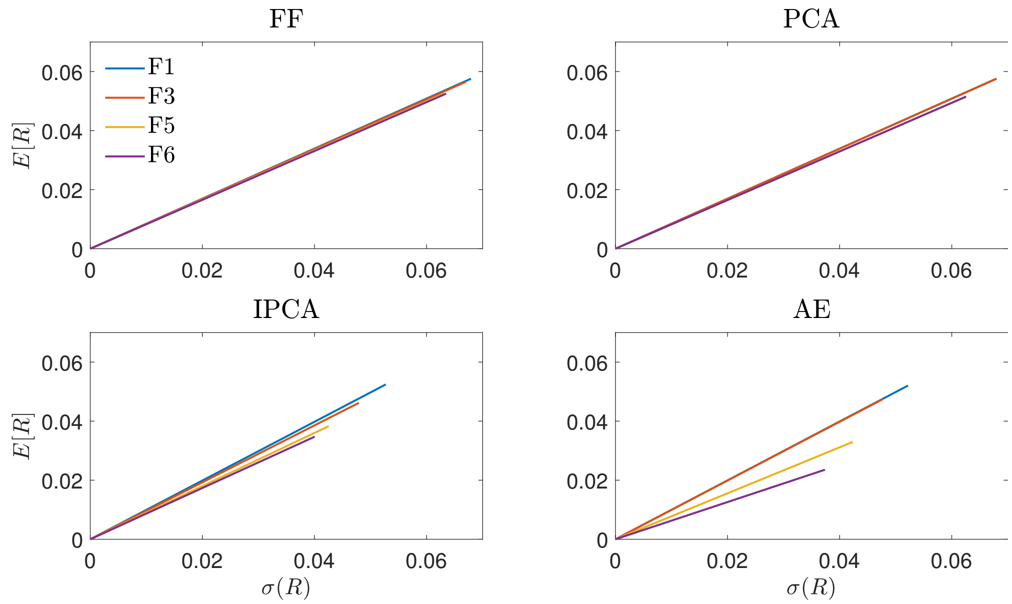
Notes: This figure shows the zero-investment, zero-beta portfolios for IPCA models.

Figure. B.7. Zero-Investment, Zero-Beta Portfolios: AE Models



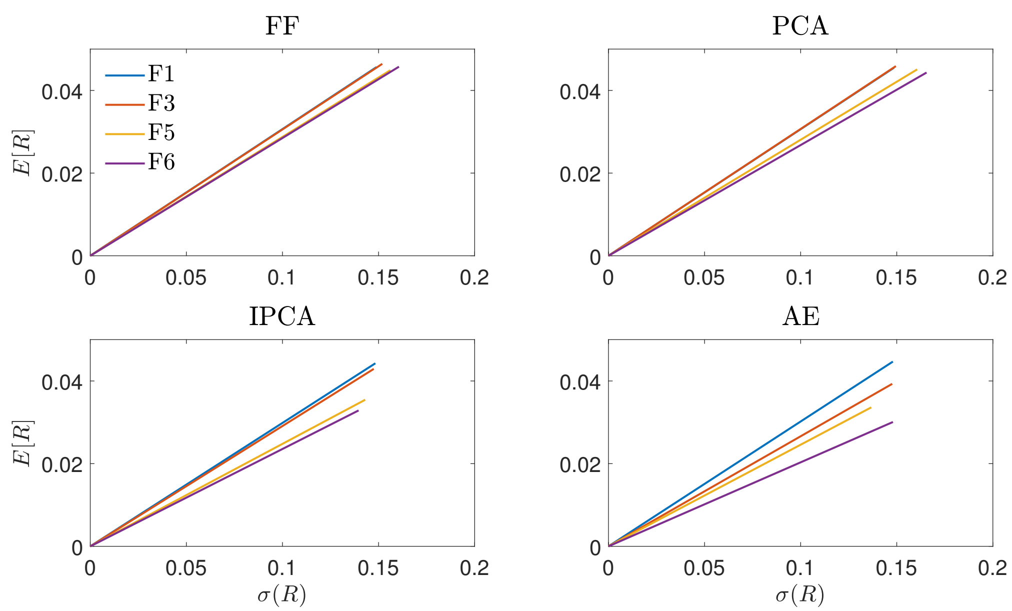
Notes: This figure shows the zero-investment, zero-beta portfolios for AE models.

Figure. B.8. Zero-Investment, Zero-Beta Frontiers



Notes: This figure shows the in-sample zero-investment, zero-beta frontiers for the FF, PCA, IPCA, and AE models with 1, 3, 5, and 6 factors. The slope of each zero-beta frontier corresponds to the in-sample maximum Sharpe ratio attainable by zero-beta portfolios.

Figure. B.9. Zero-Investment, Zero-Beta Frontiers



Notes: This figure shows the out-of-sample zero-investment, zero-beta frontiers for the FF, PCA, IPCA, and AE models with 1, 3, 5, and 6 factors. The slope of each zero-beta frontier corresponds to the out-of-sample maximum Sharpe ratio attainable by zero-beta portfolios.

B.5. *Multi-Factor Volatility-Managed Portfolios*

Table B.2: Investment Performance of Moreira-Muir Volatility-Managed Portfolios

Models	# Factors	Sharpe Ratio		Mean (%)		STD (%)		MDD (%)		Hit Rate (%)	
		Gross	Net	Gross	Net	Gross	Net	Gross	Net	Gross	Net
FF	1	0.53	0.51	8.0	7.8	15.1	15.1	54.3	55.1	62.9	62.9
	3	0.46	0.38	8.1	6.8	17.9	17.9	57.2	59.2	56.9	56.7
	5	0.64	0.50	17.2	13.6	26.9	27.1	79.6	82.5	61.7	60.5
	6	0.63	0.51	15.7	12.7	24.9	25.1	73.3	76.4	64.0	62.6
PCA	1	0.46	0.44	7.1	6.8	15.3	15.4	51.1	51.6	60.0	60.0
	3	0.58	0.52	10.2	9.3	17.7	17.8	71.9	73.0	66.0	66.0
	5	0.59	0.49	12.2	10.3	20.8	20.9	63.5	65.8	60.7	59.8
	6	0.93	0.79	15.3	13.1	16.4	16.7	47.6	51.4	65.5	65.0
IPCA	1	0.42	-4.76	6.5	--	15.5	++	44.6	++	57.6	0.0
	3	1.82	-3.48	35.2	--	19.3	++	58.4	++	74.3	0.0
	5	3.99	-4.66	85.7	--	21.5	++	19.6	++	91.0	0.0
	6	3.88	-4.69	87.8	--	22.6	++	21.0	++	91.2	0.0
AE	1	0.42	-4.76	6.5	--	15.5	++	44.6	++	57.6	0.0
	3	1.82	-3.48	35.2	--	19.3	++	58.4	++	74.3	0.0
	5	3.99	-4.66	85.7	--	21.5	++	19.6	++	91.0	0.0
	6	3.88	-4.69	87.8	--	22.6	++	21.0	++	91.2	0.0

Notes: The table reports the performance of Moreira and Muir (2017) volatility-managed portfolios for the FF, PCA, IPCA, and AE models with 1, 3, 5, and 6 factors. Reported statistics include annualized Sharpe ratios, mean returns (%), standard deviations (%), maximum drawdowns (%), and hit rates (%). I scale the mean-variance portfolio weights each month to target a 15% annualized volatility using historical estimates of mean returns and the covariance matrix. Performance is shown both before transaction costs ("gross") and after transaction costs ("net"). "++" indicates a number above 100 (%), and "--" indicates a number below -100 (%).

Table B.3: Investment Performance of DMU Volatility-Managed Portfolios

Models	# Factors	Sharpe Ratio		Mean (%)		STD (%)		MDD (%)		Hit Rate (%)	
		Gross	Net	Gross	Net	Gross	Net	Gross	Net	Gross	Net
FF	1	0.53	0.51	8.0	7.8	15.1	15.1	54.3	55.1	62.9	62.9
	3	0.46	0.38	8.1	6.8	17.9	17.9	57.2	59.2	56.9	56.7
	5	0.64	0.50	17.2	13.6	26.9	27.1	79.6	82.5	61.7	60.5
	6	0.63	0.51	15.7	12.7	24.9	25.1	73.3	76.4	64.0	62.6
PCA	1	0.46	0.44	7.1	6.8	15.3	15.4	51.1	51.6	60.0	60.0
	3	0.58	0.52	10.2	9.3	17.7	17.8	71.9	73.0	66.0	66.0
	5	0.59	0.49	12.2	10.3	20.8	20.9	63.5	65.8	60.7	59.8
	6	0.93	0.79	15.3	13.1	16.4	16.7	47.6	51.4	65.5	65.0
IPCA	1	0.42	-4.76	6.5	--	15.5	++	44.6	++	57.6	0.0
	3	1.82	-3.48	35.2	--	19.3	++	58.4	++	74.3	0.0
	5	3.99	-4.66	85.7	--	21.5	++	19.6	++	91.0	0.0
	6	3.88	-4.69	87.8	--	22.6	++	21.0	++	91.2	0.0
AE	1	0.42	-4.76	6.5	--	15.5	++	44.6	++	57.6	0.0
	3	1.82	-3.48	35.2	--	19.3	++	58.4	++	74.3	0.0
	5	3.99	-4.66	85.7	--	21.5	++	19.6	++	91.0	0.0
	6	3.88	-4.69	87.8	--	22.6	++	21.0	++	91.2	0.0

Notes: The table reports the performance of DeMiguel et al. (2024) volatility-managed portfolios for the FF, PCA, IPCA, and AE models with 1, 3, 5, and 6 factors. Reported statistics include annualized Sharpe ratios, mean returns (%), standard deviations (%), maximum drawdowns (%), and hit rates (%). I scale the mean-variance portfolio weights each month to target a 15% annualized volatility using historical estimates of mean returns and the covariance matrix. Performance is shown both before transaction costs ("gross") and after transaction costs ("net"). "++" indicates a number above 100 (%), and "--" indicates a number below -100 (%).

Multi-factor portfolios with vs without an intercept.

B.6. Prediction-Sorted Portfolios

Table B.4: Investment Performance of Prediction-Sorted Portfolios

Models	# Factors	Sharpe Ratio		Mean (%)		STD (%)		MDD (%)		Hit Rate (%)	
		Gross	Net	Gross	Net	Gross	Net	Gross	Net	Gross	Net
FF	1	-0.54	-0.80	-9.3	-14.4	17.2	18.0	98.6	99.8	47.1	43.8
	3	-0.46	-0.75	-7.4	-12.6	16.1	16.8	97.6	99.6	45.2	41.9
	5	-0.47	-0.76	-7.4	-12.6	15.9	16.6	97.6	99.6	45.7	41.7
	6	-0.50	-0.81	-7.7	-12.9	15.3	16.0	97.7	99.6	44.8	41.2
PCA	1	-0.62	-0.86	-12.8	-18.9	20.7	21.9	99.7	100.0	48.3	43.3
	3	-0.62	-0.86	-12.8	-18.9	20.7	21.9	99.7	100.0	48.3	43.3
	5	-0.62	-0.86	-12.8	-18.9	20.7	21.9	99.7	100.0	48.3	43.3
	6	-0.62	-0.86	-12.8	-18.9	20.7	21.9	99.7	100.0	48.3	43.3
IPCA	1	0.97	0.70	18.2	12.3	18.9	17.5	44.8	52.5	64.3	59.5
	3	1.08	0.81	19.6	13.4	18.1	16.7	39.8	48.3	64.3	61.2
	5	1.04	0.76	18.5	12.4	17.8	16.3	34.6	43.8	65.0	59.5
	6	0.96	0.68	17.4	11.4	18.1	16.7	43.3	51.3	64.5	60.7
AE	1	1.30	0.73	15.8	8.4	12.2	11.4	25.0	31.9	66.4	60.7
	3	1.06	0.47	13.1	5.5	12.4	11.6	33.2	43.1	66.7	60.2
	5	1.16	0.60	15.0	7.4	13.0	12.3	29.1	36.9	64.8	57.4
	6	1.13	0.57	14.2	6.5	12.5	11.4	20.4	24.0	63.8	57.6

Notes: The table reports the performance of prediction-sorted portfolios for the FF, PCA, IPCA, and AE models with 1, 3, 5, and 6 factors. Reported statistics include annualized Sharpe ratios, mean returns (%), standard deviations (%), maximum drawdowns (%), and hit rates (%). Performance is shown both before transaction costs ("gross") and after transaction costs ("net").

B.7. Model-Free Portfolios

Table B.5: Investment Performance of Zero-Beta Portfolios

Portfolio Rules	Sharpe Ratio		Mean (%)		STD (%)		MDD (%)		Hit Rate (%)	
	Gross	Net	Gross	Net	Gross	Net	Gross	Net	Gross	Net
MKT	0.53	0.52	11.1	11.0	15.5	15.5	51.6	51.9	64.5	64.5
MVP	0.79	0.39	9.6	4.9	12.1	12.5	36.0	40.3	61.0	56.7
GMV	0.69	0.49	7.5	5.5	10.9	11.3	26.8	30.3	61.2	58.8
EW	0.48	0.46	8.5	8.1	17.7	17.7	56.7	57.2	60.0	60.0
RP	0.50	0.48	8.6	8.3	17.2	17.2	56.3	56.7	61.0	60.2
KWZ	1.31	1.02	27.2	21.7	20.8	21.2	49.0	53.6	68.6	65.0

Notes: The table reports the performance of conventional model-free portfolios, including the market portfolio (MKT), mean-variance portfolio (MVP), global minimum-variance portfolio (GMV), equal-weighted portfolio (EW), risk-parity portfolio (RP), Kan et al. (2022) optimal portfolio (KWZ). I scale the MVP and KWZ portfolio weights each month to target a 15% annualized volatility using historical estimates of mean returns and the covariance matrix. Investment performance is reported both before transaction costs ("gross") and after transaction costs ("net").

B.8. Regression of Zero-Beta Portfolio Returns on Benchmark Risk Factors

Table B.6: Regression of Zero-Beta Portfolio Gross Returns on Benchmark Risk Factors

	FF		PCA		IPCA		AE	
# Factors	1	6	1	6	1	6	1	6
Alpha (%)	1.79***	1.98***	1.78***	1.78***	1.76***	1.37***	1.79***	1.36***
	(5.78)	(5.66)	(5.79)	(5.29)	(5.95)	(5.38)	(5.90)	(5.40)
MKT	0.09	0.02	0.08	0.06	0.04	-0.07	0.04	-0.02
	(1.17)	(0.18)	(1.04)	(0.73)	(0.62)	(-1.01)	(0.50)	(-0.26)
SMB	0.05	-0.10	0.07	-0.06	0.06	-0.09	0.07	-0.04
	(0.24)	(-0.49)	(0.35)	(-0.30)	(0.33)	(-0.63)	(0.36)	(-0.20)
HML	0.30***	0.10	0.30***	0.35***	0.24*	0.25***	0.23*	0.24*
	(2.40)	(0.71)	(2.41)	(2.43)	(1.90)	(2.42)	(1.88)	(1.73)
RMW	0.26	0.06	0.26	0.08	0.35*	0.41***	0.34*	0.18
	(1.32)	(0.28)	(1.36)	(0.36)	(1.92)	(2.78)	(1.77)	(1.00)
CMA	0.18	-0.02	0.18	-0.18	0.11	0.23	0.12	-0.07
	(0.85)	(-0.09)	(0.90)	(-0.78)	(0.58)	(1.47)	(0.60)	(-0.35)
UMD	0.25**	0.17	0.25**	0.22	0.22*	0.19*	0.22*	0.28***
	(2.03)	(1.29)	(2.03)	(1.63)	(1.86)	(1.77)	(1.77)	(2.60)
IVOL	-0.09	-0.17	-0.08	-0.05	-0.03	-0.13	-0.04	-0.16
	(-0.57)	(-1.03)	(-0.57)	(-0.31)	(-0.22)	(-1.00)	(-0.26)	(-1.06)
LIQ	-0.07	-0.08	-0.07	-0.04	-0.08	-0.10*	-0.07	-0.12*
	(-1.03)	(-1.09)	(-0.97)	(-0.56)	(-1.06)	(-1.82)	(-1.03)	(-1.78)
Adj.R^2	0.10	0.03	0.10	0.05	0.11	0.16	0.10	0.08

Notes: This table presents alphas (%), factor loadings, and adjusted R^2 from regressions of zero-beta portfolio returns (before-costs) on the Fama–French six factors, an idiosyncratic volatility factor, and a liquidity factor. The models include FF, PCA, IPCA, and AE with one and six factors. Newey-West t-statistics are shown in parentheses. Significance levels: *** $p < .01$, ** $p < .05$, * $p < .1$.

Table B.7: Regression of Zero-Beta Portfolio Net Returns on Benchmark Risk Factors

	FF		PCA		IPCA		AE	
# Factors	3	5	3	5	3	5	3	5
Alpha (%)	1.29***	1.34***	1.27***	1.31***	1.18***	0.96***	1.21***	0.96***
	(4.22)	(4.08)	(4.25)	(4.27)	(4.41)	(3.77)	(4.00)	(4.28)
MKT	0.15*	0.10	0.16**	0.13	0.08	-0.02	0.00	0.08
	(1.83)	(1.18)	(1.96)	(1.50)	(1.08)	(-0.32)	(0.04)	(1.22)
SMB	-0.10	-0.14	0.03	0.02	0.13	-0.04	0.05	0.02
	(-0.53)	(-0.76)	(0.14)	(0.11)	(0.86)	(-0.23)	(0.29)	(0.14)
HML	0.25*	0.26*	0.33***	0.33***	0.36***	0.27***	0.19	0.20*
	(1.95)	(1.88)	(2.67)	(2.40)	(3.47)	(2.60)	(1.40)	(1.82)
RMW	0.28	0.07	0.28	0.11	0.39**	0.46***	0.16	0.18
	(1.40)	(0.30)	(1.45)	(0.50)	(2.28)	(2.85)	(0.81)	(1.10)
CMA	0.12	-0.11	0.09	-0.19	0.08	0.25	0.14	-0.07
	(0.56)	(-0.49)	(0.47)	(-0.87)	(0.47)	(1.62)	(0.64)	(-0.37)
UMD	0.31***	0.36***	0.27**	0.23*	0.15	0.16	0.30***	0.30***
	(2.39)	(2.66)	(2.15)	(1.84)	(1.38)	(1.44)	(2.54)	(3.45)
IVOL	-0.05	-0.16	-0.01	-0.01	0.06	-0.08	-0.07	-0.03
	(-0.34)	(-0.97)	(-0.10)	(-0.06)	(0.41)	(-0.62)	(-0.48)	(-0.25)
LIQ	-0.07	-0.06	-0.05	-0.07	-0.05	-0.07	-0.08	-0.12**
	(-0.98)	(-0.80)	(-0.75)	(-0.92)	(-0.69)	(-1.17)	(-1.13)	(-1.98)
Adj. R^2	0.12	0.10	0.12	0.06	0.19	0.16	0.10	0.11

Notes: This table presents alphas (%), factor loadings, and adjusted R^2 from regressions of zero-beta portfolio returns (net-of-costs) on the Fama–French six factors, an idiosyncratic volatility factor, and a liquidity factor. The models include FF, PCA, IPCA, and AE with three and five factors. Newey-West t-statistics are shown in parentheses. Significance levels: *** $p < .01$, ** $p < .05$, * $p < .1$.

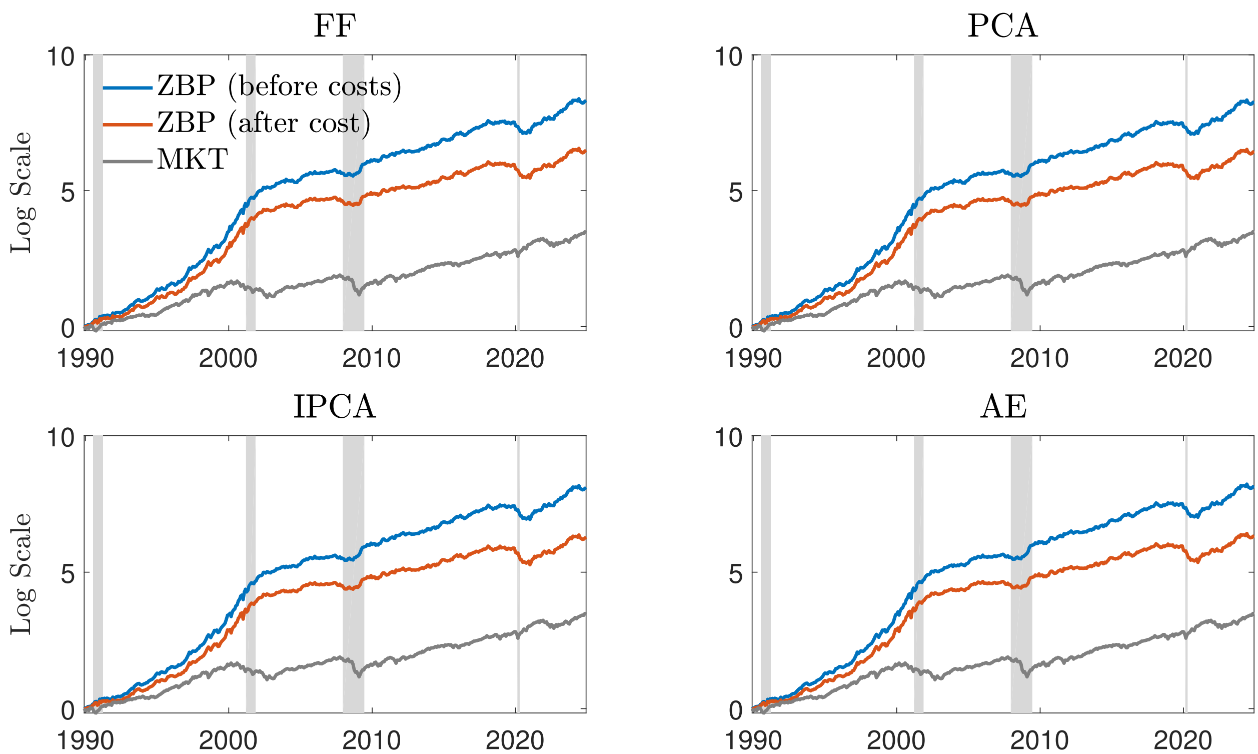
Table B.8: Regression of Zero-Beta Portfolio Gross Returns on Benchmark Risk Factors

	FF		PCA		IPCA		AE	
# Factors	3	5	3	5	3	5	3	5
Alpha (%)	1.80***	1.84***	1.77***	1.83***	1.67***	1.45***	1.70***	1.47***
	(5.77)	(5.47)	(5.84)	(5.83)	(6.21)	(5.70)	(5.53)	(6.61)
MKT	0.11	0.06	0.12	0.09	0.04	-0.06	-0.04	0.04
	(1.43)	(0.76)	(1.55)	(1.07)	(0.57)	(-0.92)	(-0.47)	(0.66)
SMB	-0.12	-0.16	0.01	0.00	0.11	-0.05	0.03	0.00
	(-0.61)	(-0.86)	(0.04)	(0.01)	(0.72)	(-0.33)	(0.18)	(0.00)
HML	0.24*	0.25*	0.31***	0.32**	0.35***	0.25***	0.17	0.19*
	(1.82)	(1.74)	(2.52)	(2.30)	(3.16)	(2.42)	(1.25)	(1.67)
RMW	0.28	0.07	0.28	0.11	0.39***	0.45***	0.16	0.18
	(1.41)	(0.30)	(1.47)	(0.50)	(2.38)	(2.96)	(0.82)	(1.13)
CMA	0.16	-0.06	0.14	-0.14	0.13	0.30*	0.19	-0.02
	(0.76)	(-0.28)	(0.69)	(-0.63)	(0.73)	(1.87)	(0.84)	(-0.09)
UMD	0.28**	0.32***	0.24**	0.20	0.12	0.13	0.27***	0.27***
	(2.24)	(2.47)	(1.98)	(1.63)	(1.14)	(1.21)	(2.38)	(3.29)
IVOL	-0.08	-0.20	-0.05	-0.04	0.02	-0.12	-0.10	-0.06
	(-0.57)	(-1.19)	(-0.33)	(-0.27)	(0.17)	(-0.85)	(-0.70)	(-0.53)
LIQ	-0.08	-0.07	-0.07	-0.08	-0.06	-0.09	-0.09	-0.13**
	(-1.13)	(-0.96)	(-0.90)	(-1.07)	(-0.86)	(-1.35)	(-1.29)	(-2.14)
Adj.R^2	0.10	0.08	0.11	0.05	0.18	0.16	0.09	0.09

Notes: This table presents alphas (%), factor loadings, and adjusted R^2 from regressions of zero-beta portfolio returns (before-costs) on the Fama–French six factors, an idiosyncratic volatility factor, and a liquidity factor. The models include FF, PCA, IPCA, and AE with three and five factors. Newey-West t-statistics are shown in parentheses. Significance levels: *** $p < .01$, ** $p < .05$, * $p < .1$.

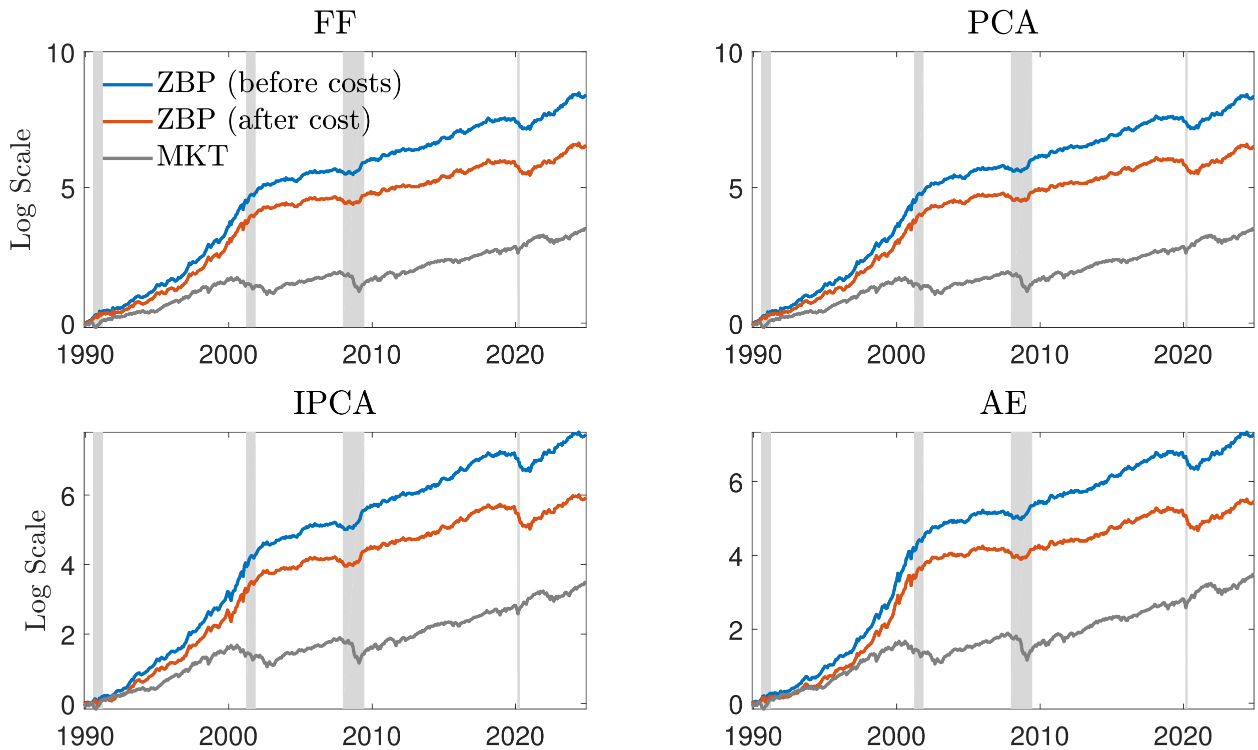
B.9. Price Path of Zero-Beta Portfolios

Figure. B.10. Price Path of Zero-Beta Portfolios (Single Factor)



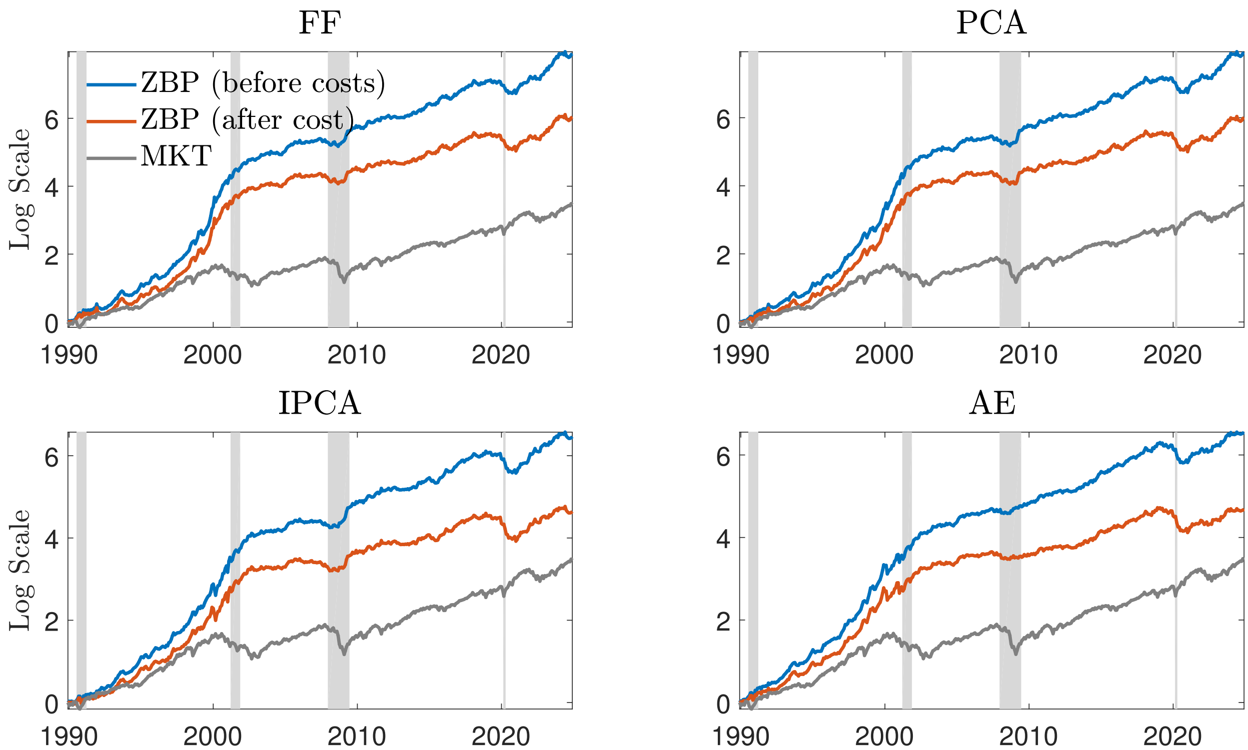
Notes: This figure shows the logarithmic price paths (cumulative returns) of zero-beta portfolios before and after transaction costs, alongside the market portfolio. Gray-shaded areas denote NBER recessions. The models considered are FF, PCA, IPCA, and AE with a single factors.

Figure. B.11. Price Path of Zero-Beta Portfolios (Three Factors)



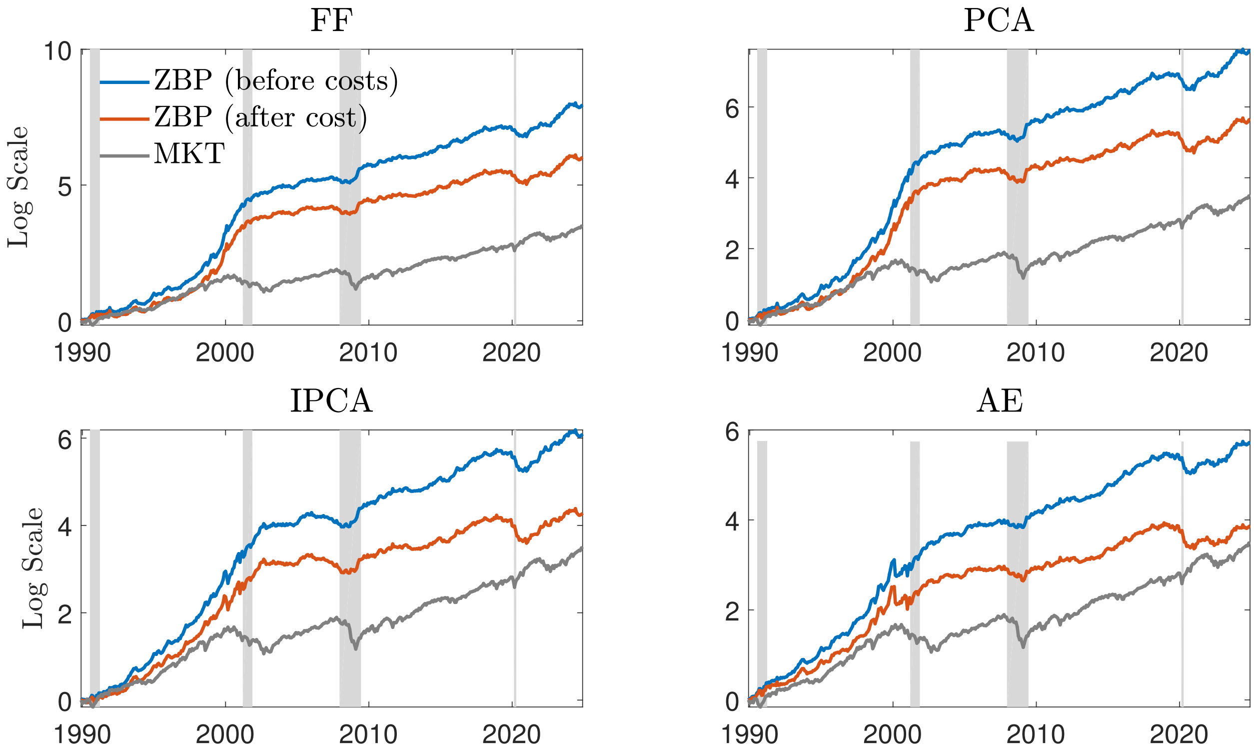
Notes: This figure shows the logarithmic price paths (cumulative returns) of zero-beta portfolios before and after transaction costs, alongside the market portfolio. Gray-shaded areas denote NBER recessions. The models considered are FF, PCA, IPCA, and AE with three factors.

Figure. B.12. Price Path of Zero-Beta Portfolios (Five Factors)



Notes: This figure shows the logarithmic price paths (cumulative returns) of zero-beta portfolios before and after transaction costs, alongside the market portfolio. Gray-shaded areas denote NBER recessions. The models considered are FF, PCA, IPCA, and AE with five factors.

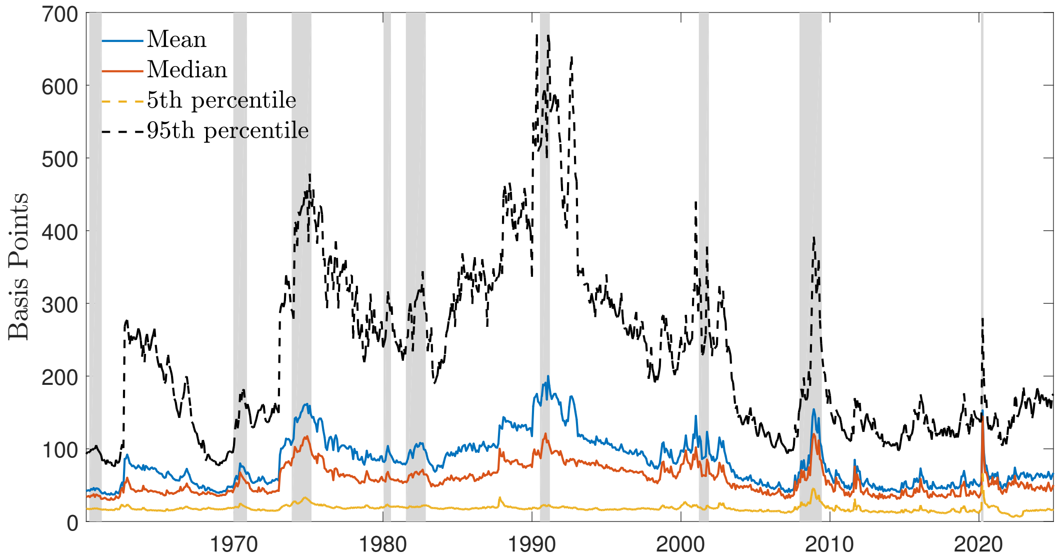
Figure. B.13. Price Path of Zero-Beta Portfolios (Six Factors)



Notes: This figure shows the logarithmic price paths (cumulative returns) of zero-beta portfolios before and after transaction costs, alongside the market portfolio. Gray-shaded areas denote NBER recessions. The models considered are FF, PCA, IPCA, and AE with six factors.

B.10. Individual-Stock Level Transaction Costs

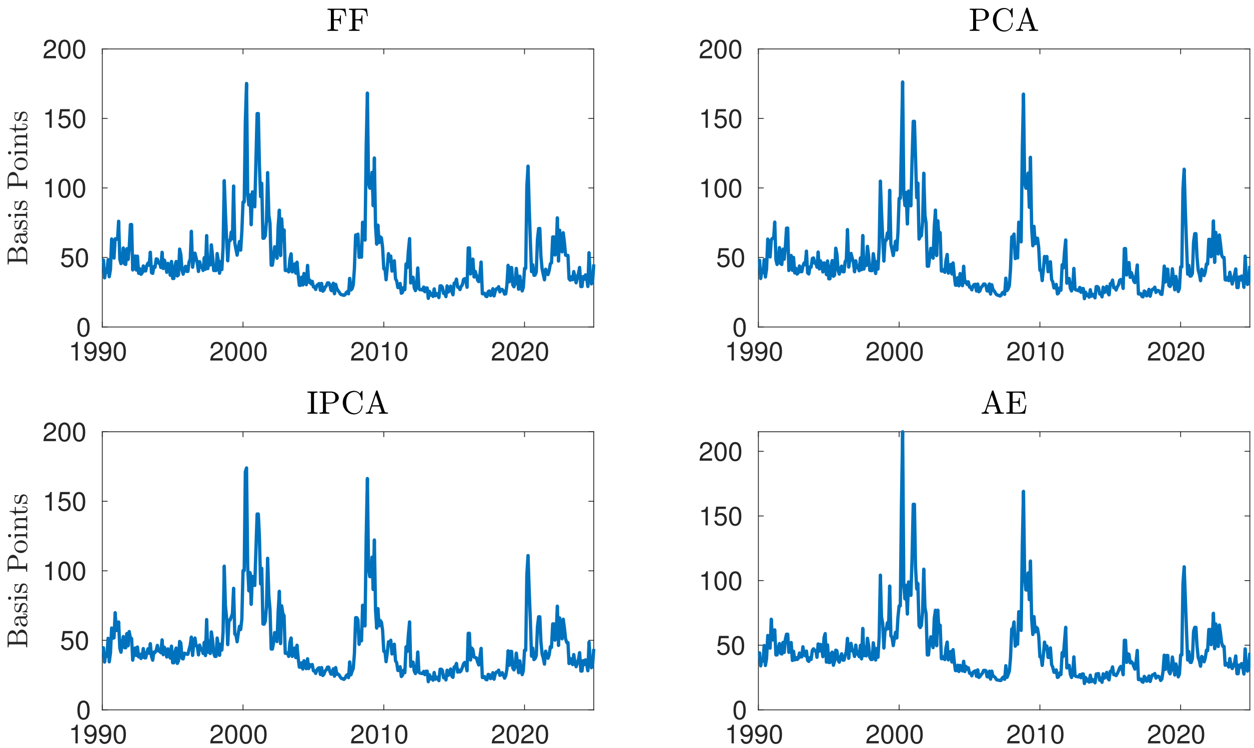
Figure. B.14. Individual-Stock Level Transaction Costs



Notes: This figure shows the time variation of the mean, median, 5th percentile, and 95th percentile of individual transaction costs from Jan 1960 to Dec 2024, measured using the average low-frequency effective spreads described in Chen and Velikov (2023).

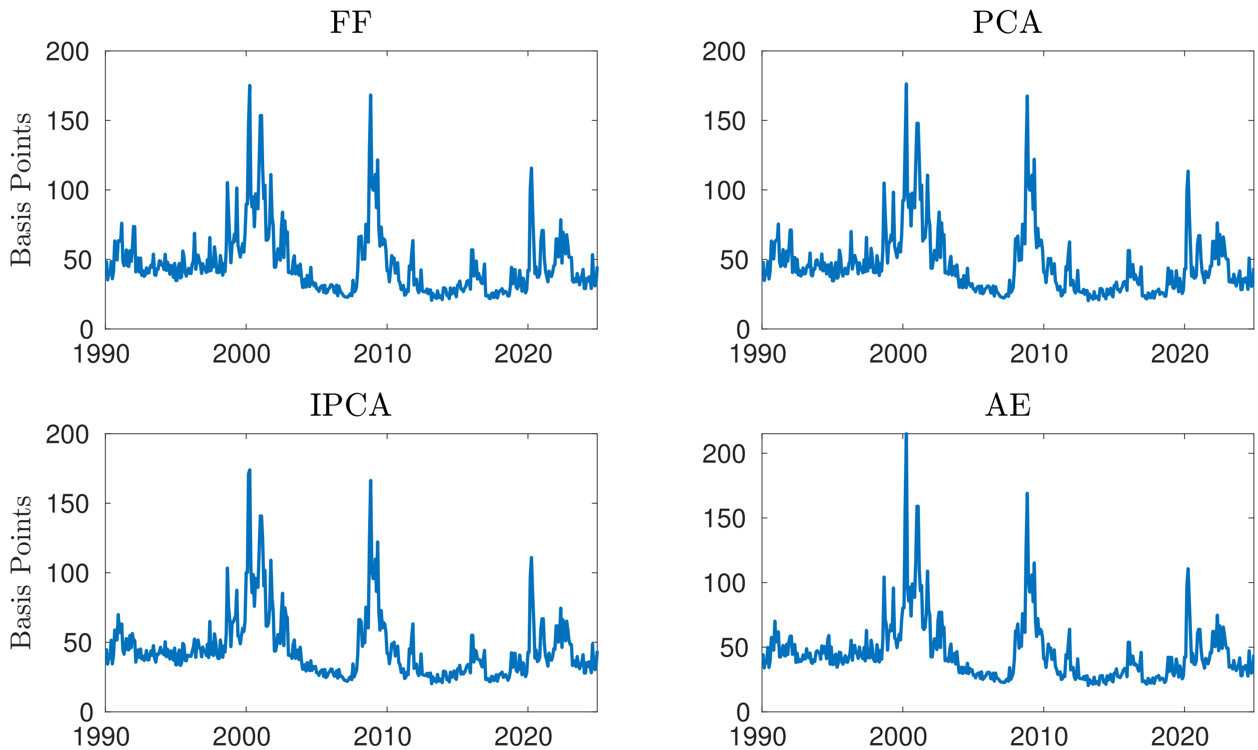
B.11. *Zero-Beta Portfolio Transaction Costs*

Figure. B.15. Zero-Beta Portfolio Transaction Costs (Three Factors)



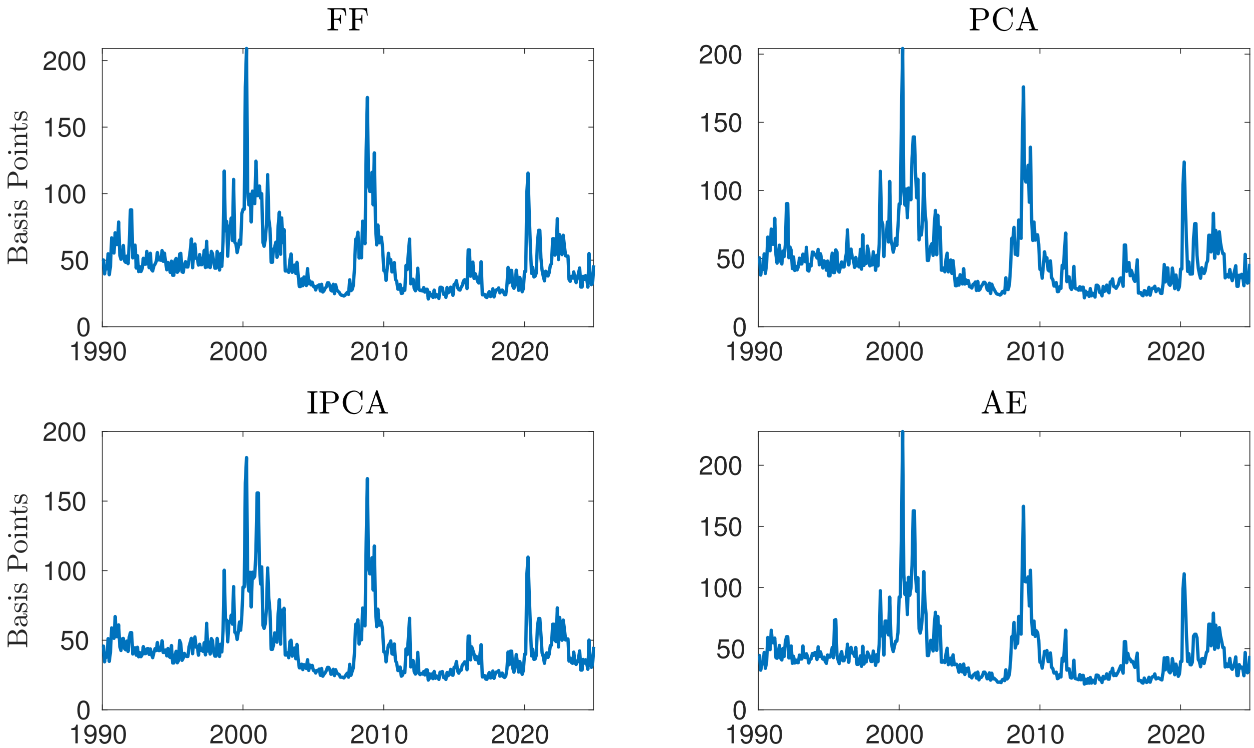
Notes: This figure shows the transaction costs of the optimal zero-beta portfolios from January 1990 to December 2024, measured using the average low-frequency effective spreads in Chen and Velikov (2023). The models considered are FF, PCA, IPCA, and AE with three factors.

Figure. B.16. Zero-Beta Portfolio Transaction Costs (Five Factors)



Notes: This figure shows the transaction costs of the optimal zero-beta portfolios from January 1990 to December 2024, measured using the average low-frequency effective spreads in Chen and Velikov (2023). The models considered are FF, PCA, IPCA, and AE with five factors.

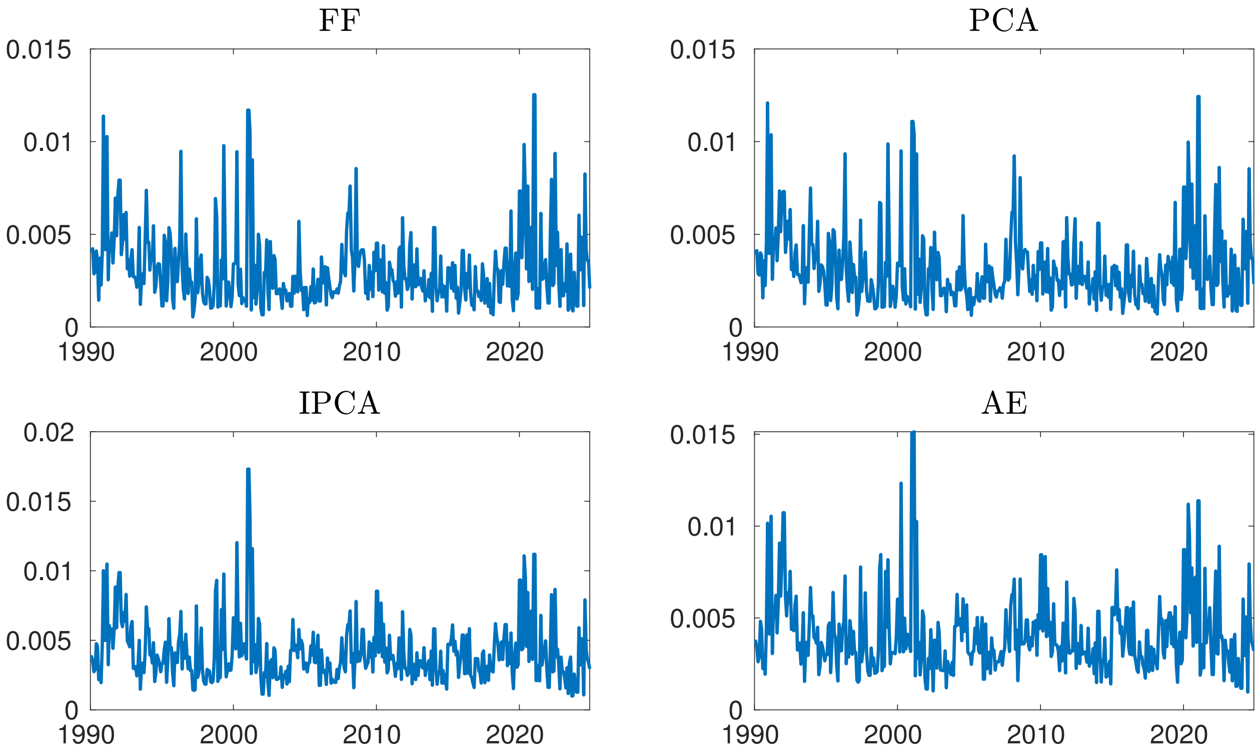
Figure. B.17. Zero-Beta Portfolio Transaction Costs (Six Factors)



Notes: This figure shows the transaction costs of the optimal zero-beta portfolios from January 1990 to December 2024, measured using the average low-frequency effective spreads in Chen and Velikov (2023). The models considered are FF, PCA, IPCA, and AE with six factors.

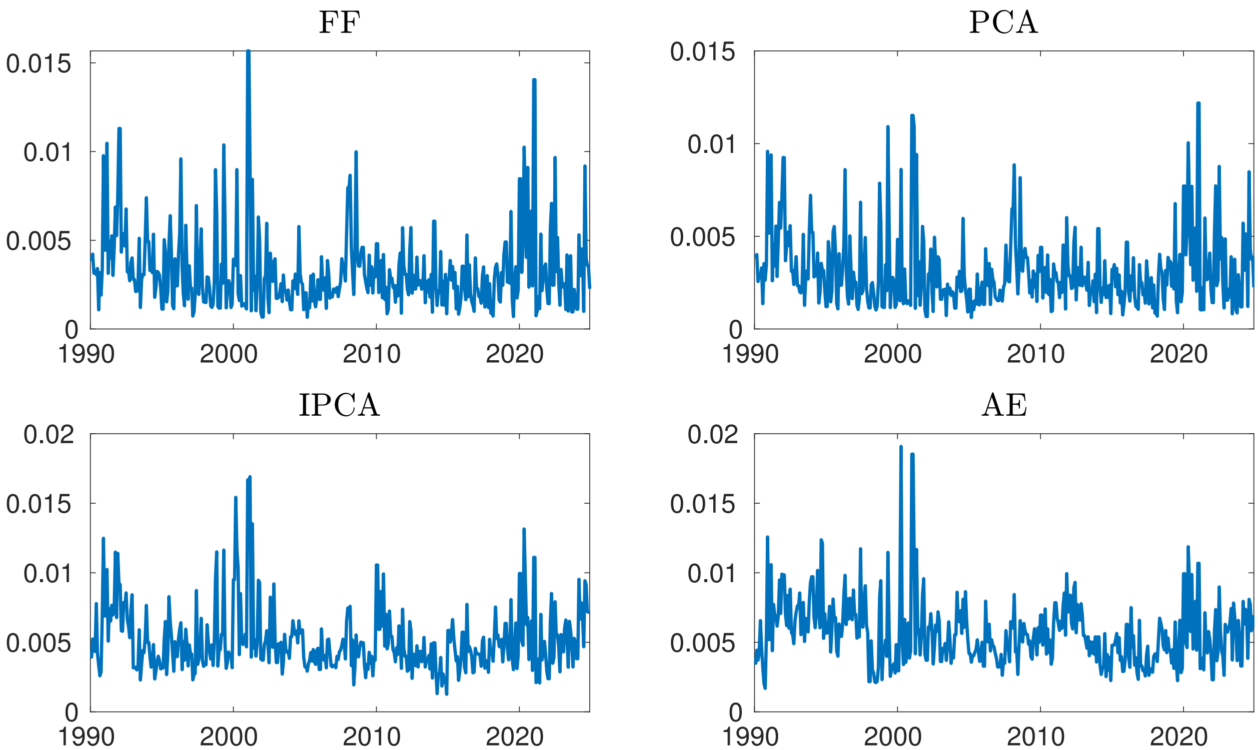
B.12. Median Absolute Zero-Beta Portfolio Weight Changes

Figure. B.18. Median Absolute Zero-Beta Portfolio Weight Changes (Single Factors)



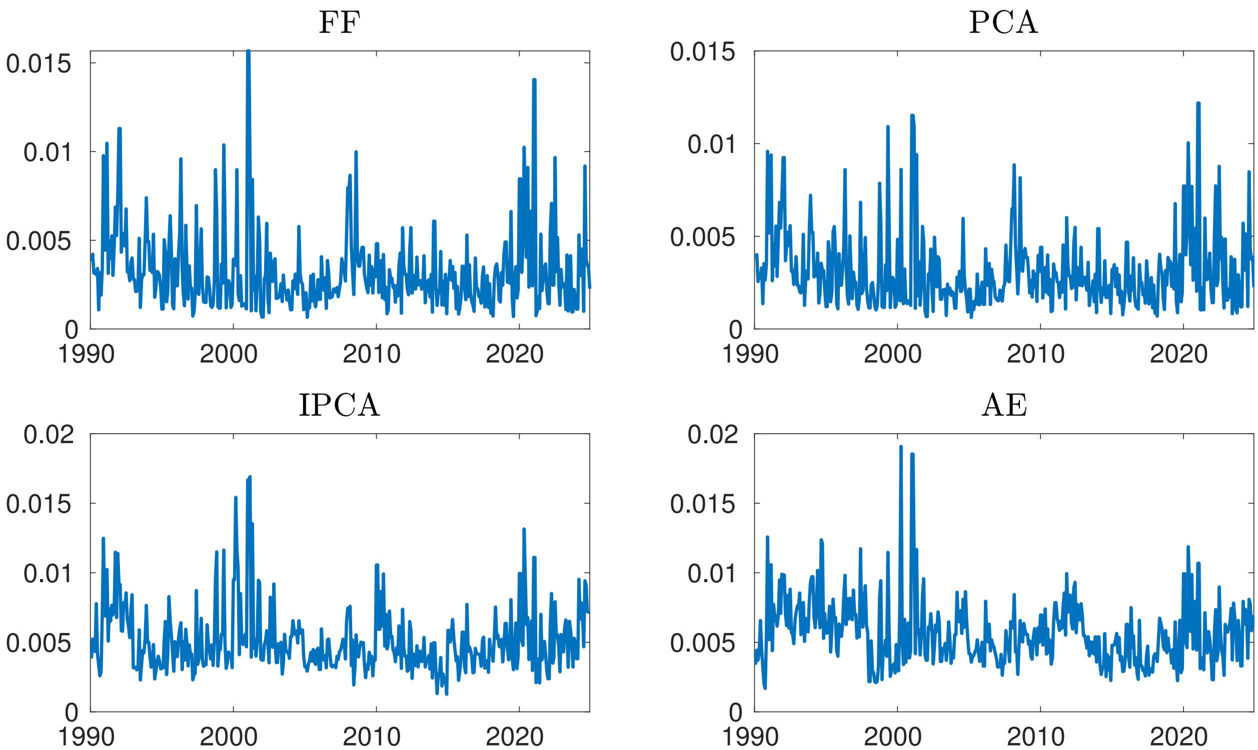
Notes: This figure shows the median absolute weight changes of optimal zero-beta portfolios across the 272 characteristic-sorted portfolios from January 1990 to December 2024. The models considered are FF, PCA, IPCA, and AE with a single factors.

Figure. B.19. Median Absolute Zero-Beta Portfolio Weight Changes (Three Factors)



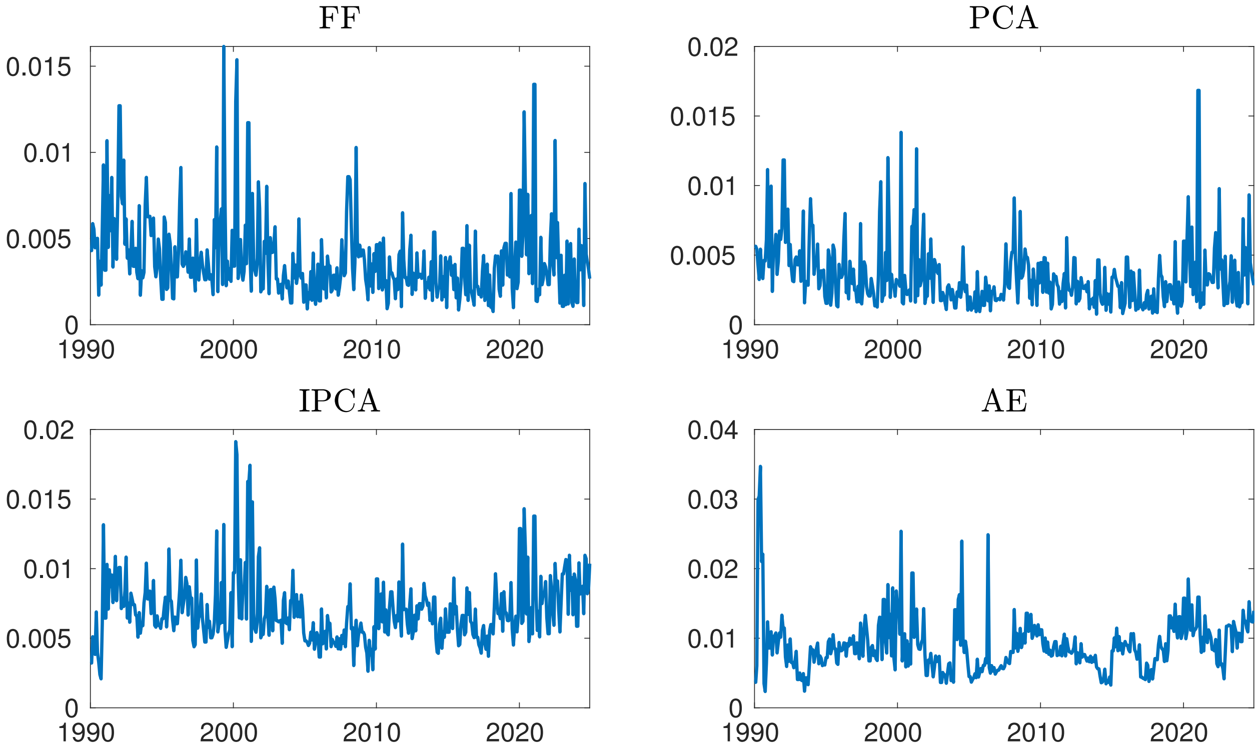
Notes: This figure shows the median absolute weight changes of optimal zero-beta portfolios across the 272 characteristic-sorted portfolios from January 1990 to December 2024. The models considered are FF, PCA, IPCA, and AE with three factors.

Figure. B.20. Median Absolute Zero-Beta Portfolio Weight Changes (Five Factors)



Notes: This figure shows the median absolute weight changes of optimal zero-beta portfolios across the 272 characteristic-sorted portfolios from January 1990 to December 2024. The models considered are FF, PCA, IPCA, and AE with five factors.

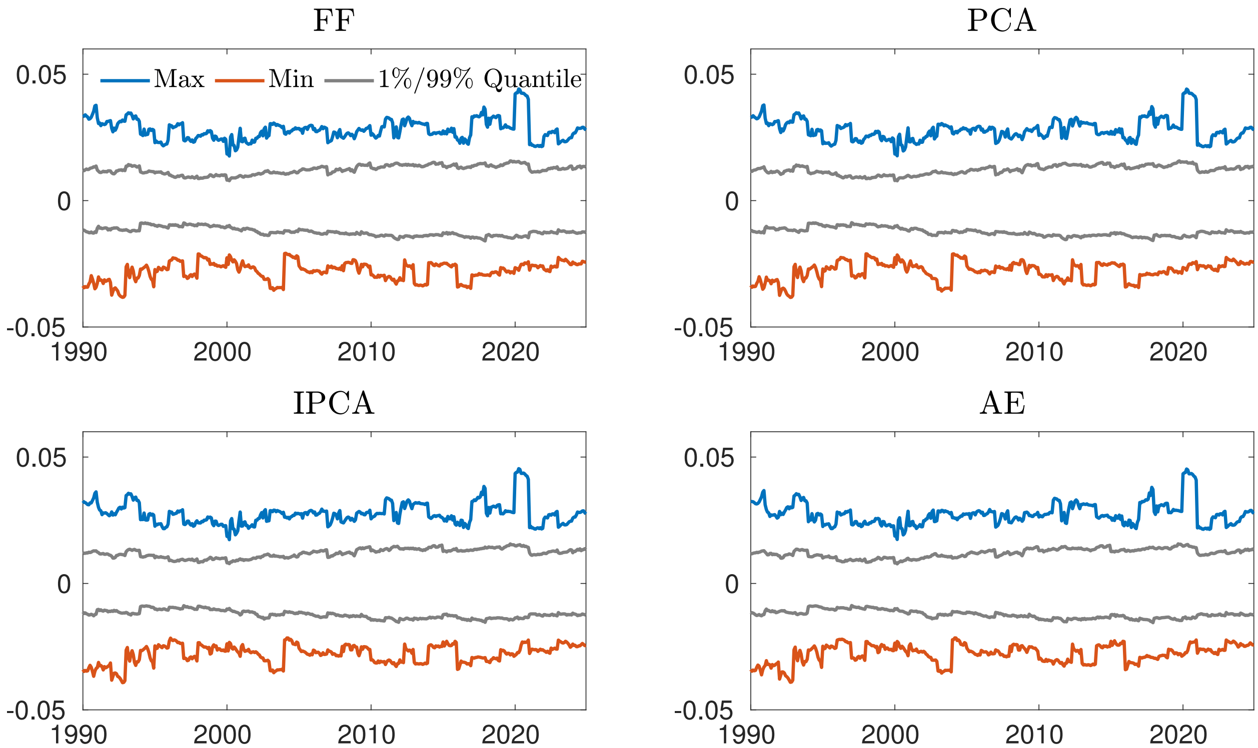
Figure. B.21. Median Absolute Zero-Beta Portfolio Weight Changes (Six Factors)



Notes: This figure shows the median absolute weight changes of optimal zero-beta portfolios across the 272 characteristic-sorted portfolios from January 1990 to December 2024. The models considered are FF, PCA, IPCA, and AE with six factors.

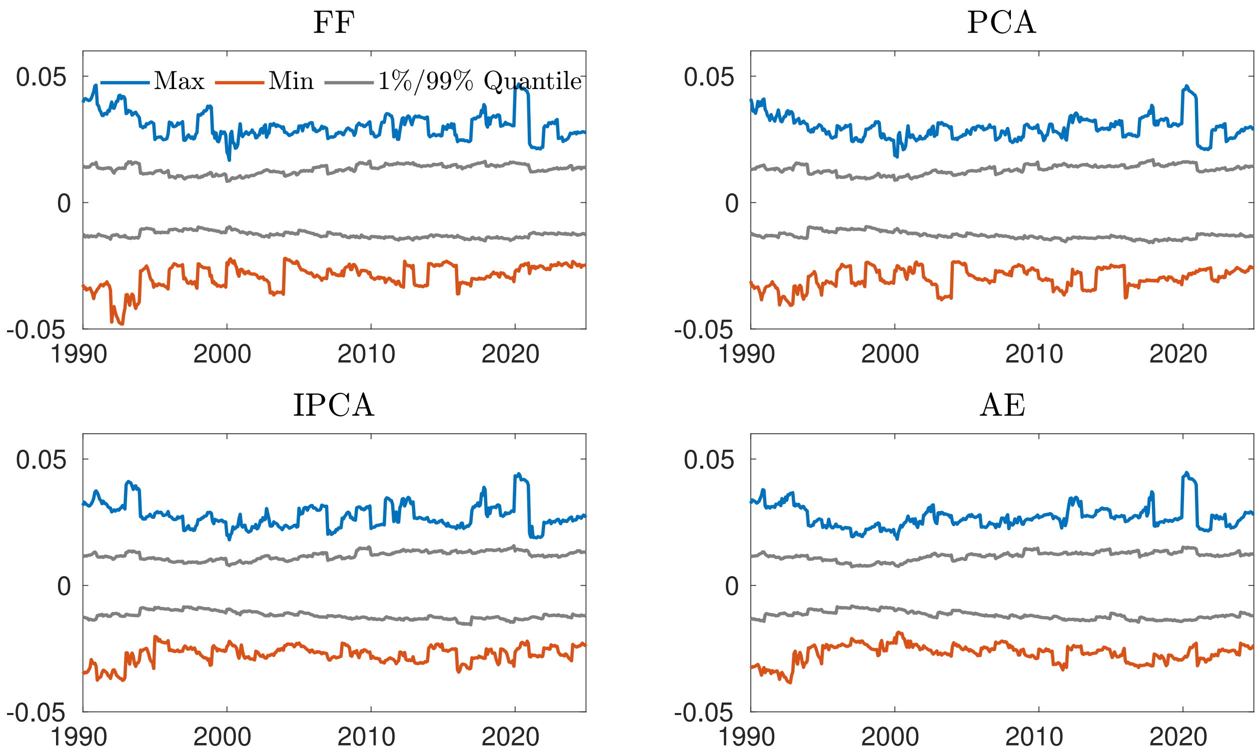
B.13. Zero-Beta Portfolio Weights on Individual Stocks

Figure. B.22. Zero-Beta Portfolio Weights on Individual Stocks (Single Factors)



Notes: This figure shows the maximum, minimum, 1% quantile, and 99% quantile of the zero-beta portfolio weights on individual stocks from January 1990 to December 2024. The models considered are FF, PCA, IPCA, and AE with a single factors.

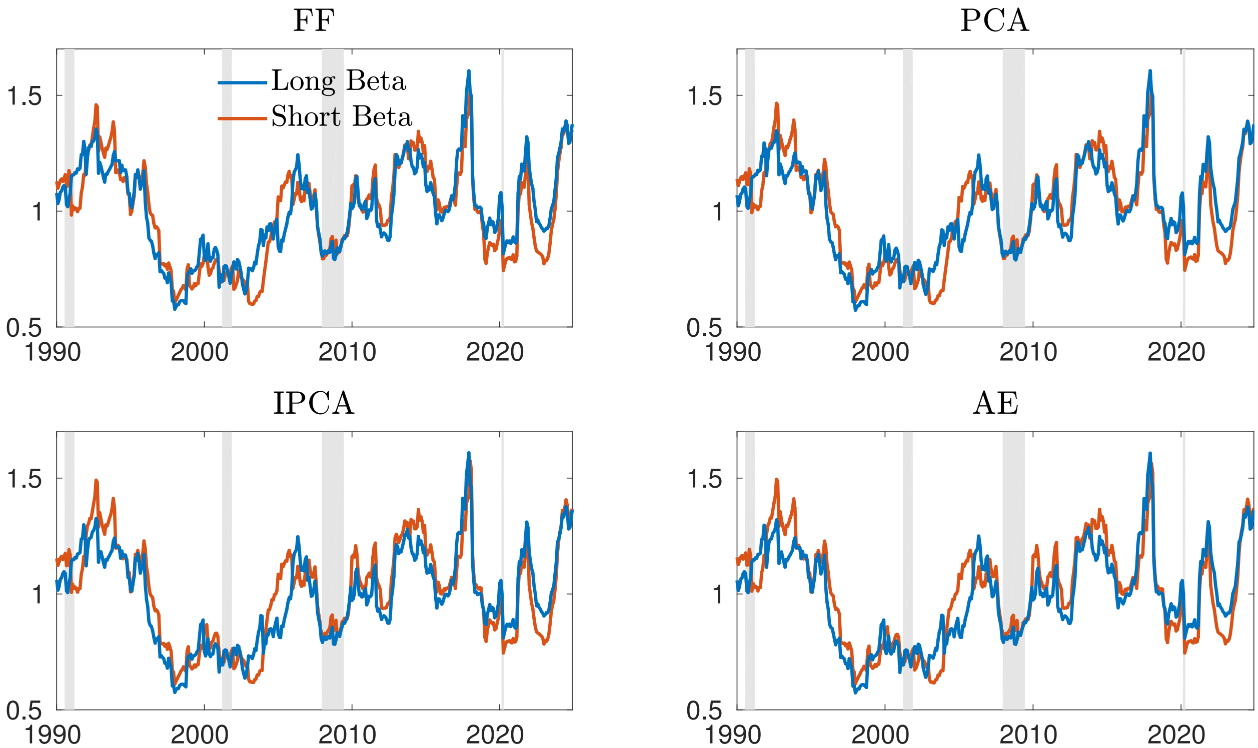
Figure. B.23. Zero-Beta Portfolio Weights on Individual Stocks (Six Factors)



Notes: This figure shows the maximum, minimum, 1% quantile, and 99% quantile of the zero-beta portfolio weights on individual stocks from January 1990 to December 2024. The models considered are FF, PCA, IPCA, and AE with six factors.

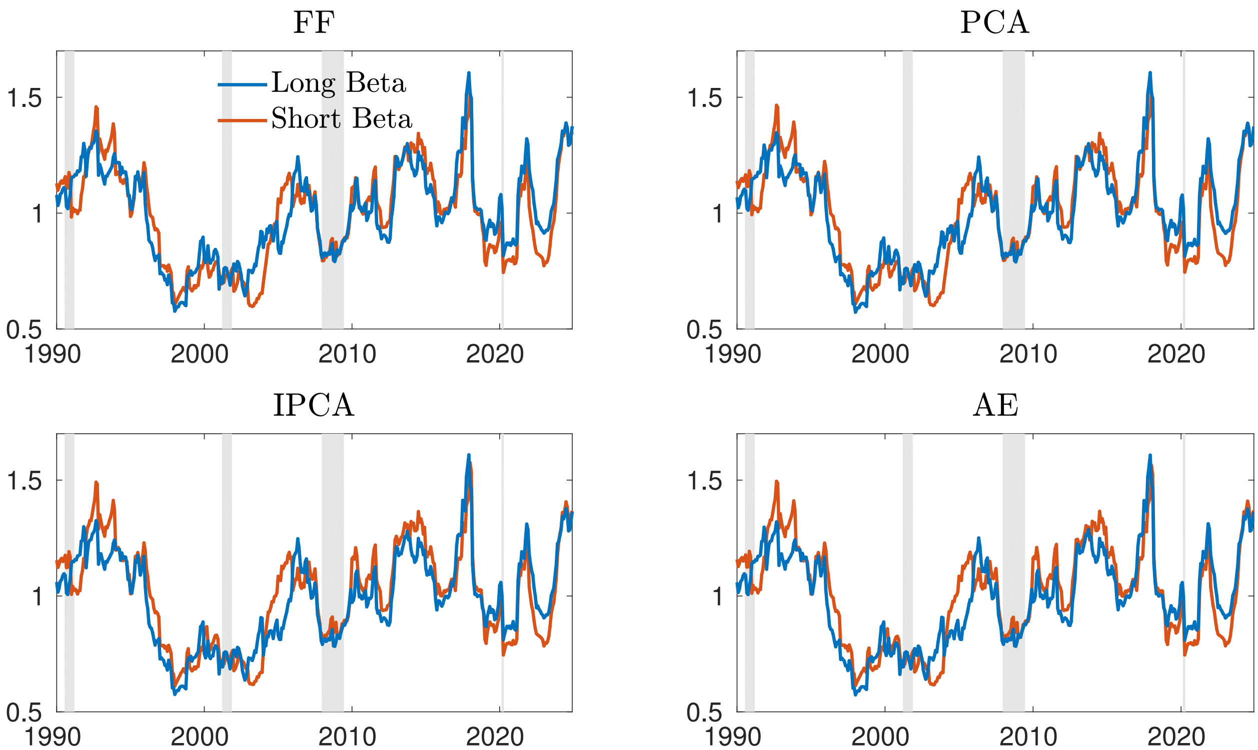
B.14. Market Betas of Long vs Short Positions

Figure. B.24. Market Betas of Long vs Short Positions (Single Factor)



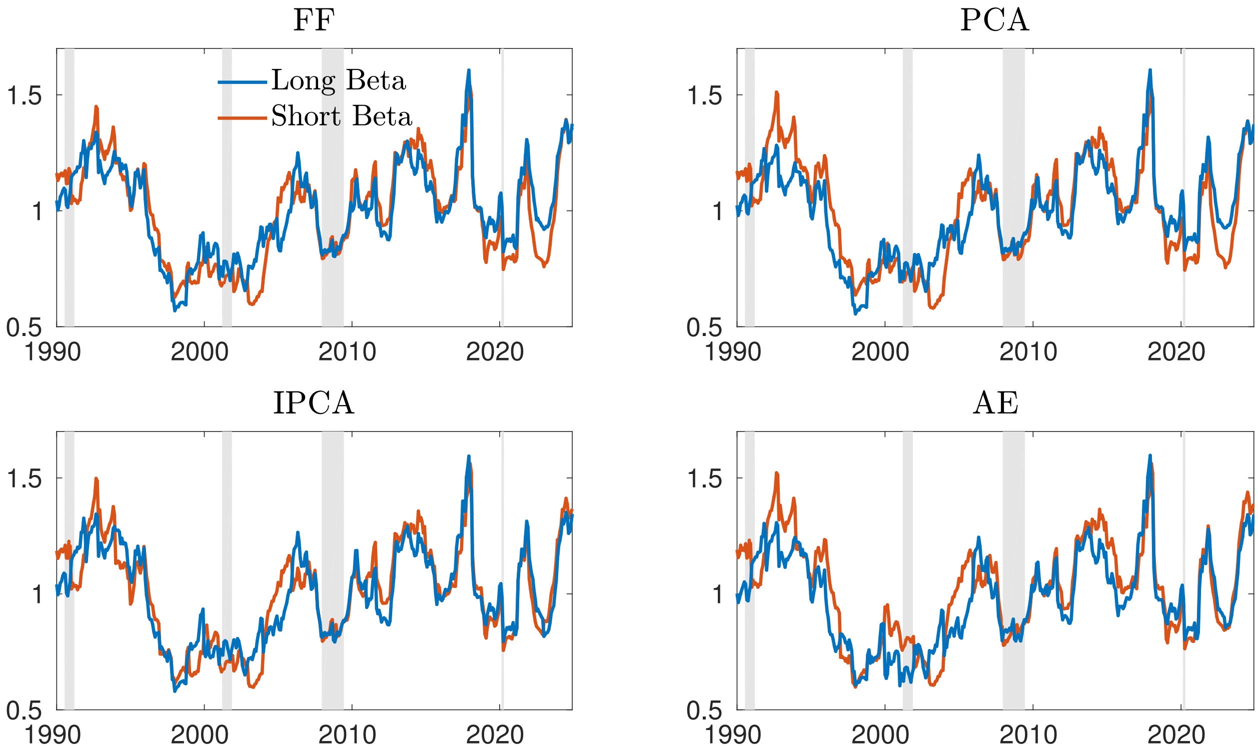
Notes: This figure shows the weighted average market betas of long and short positions for the optimal zero-beta portfolios. Gray-shaded areas denote NBER recessions. The models considered are FF, PCA, IPCA, and AE with a single factors.

Figure. B.25. Market Betas of Long vs Short Positions (Three Factor)



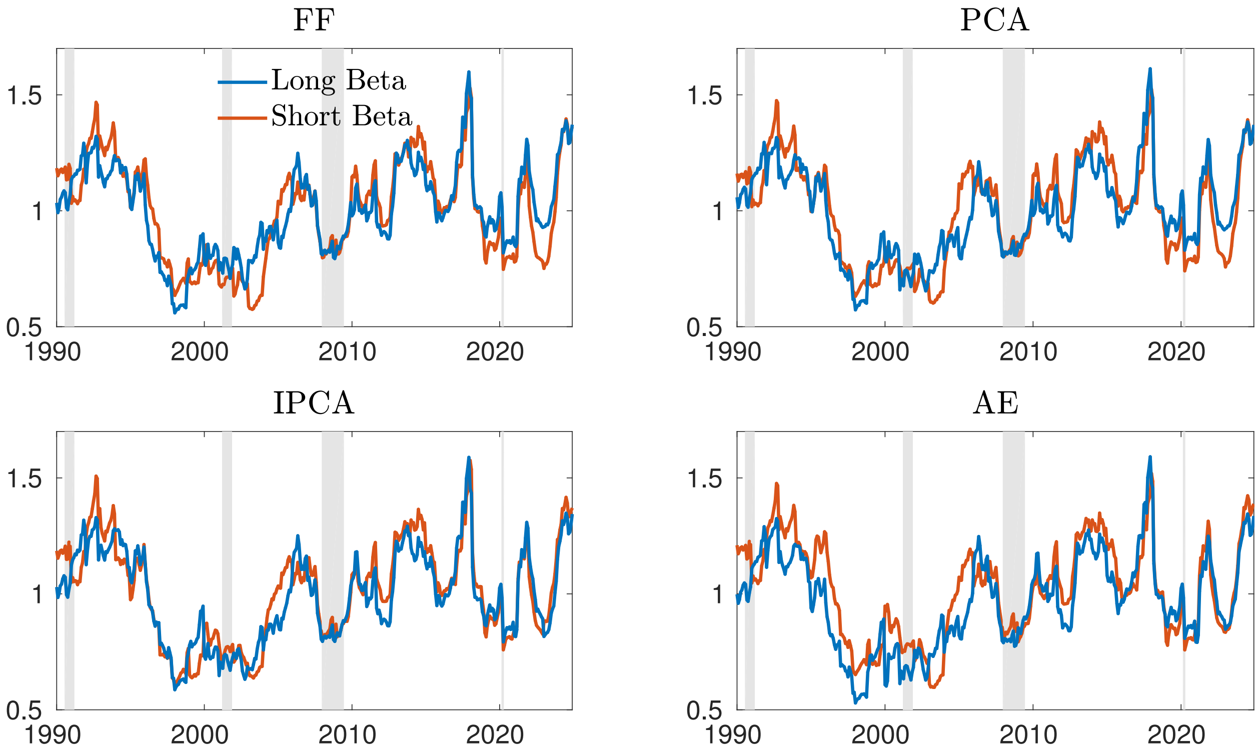
Notes: This figure shows the weighted average market betas of long and short positions for the optimal zero-beta portfolios. Gray-shaded areas denote NBER recessions. The models considered are FF, PCA, IPCA, and AE with three factors.

Figure. B.26. Market Betas of Long vs Short Positions (Five Factor)



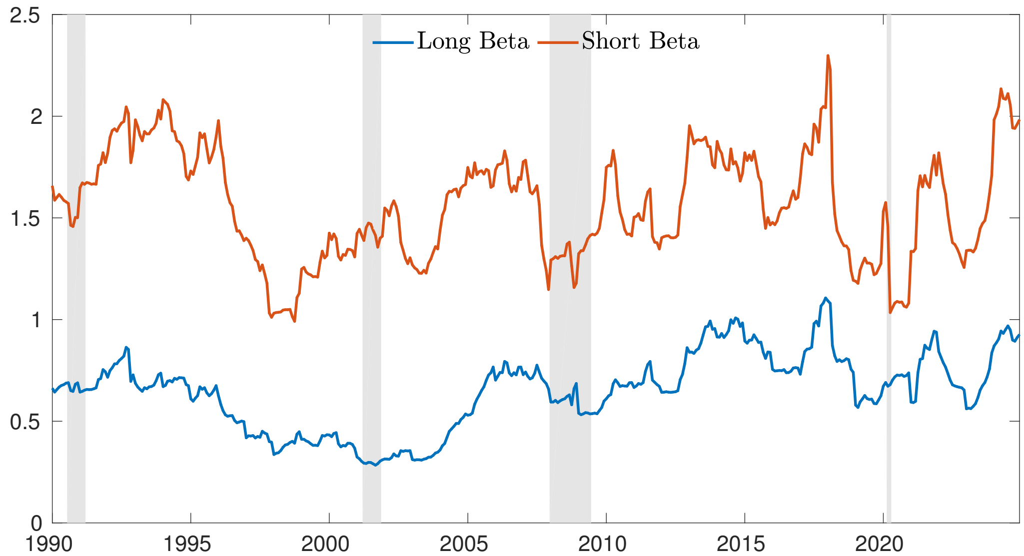
Notes: This figure shows the weighted average market betas of long and short positions for the optimal zero-beta portfolios. Gray-shaded areas denote NBER recessions. The models considered are FF, PCA, IPCA, and AE with three factors.

Figure. B.27. Market Betas of Long vs Short Positions (Six Factor)



Notes: This figure shows the weighted average market betas of long and short positions for the optimal zero-beta portfolios. Gray-shaded areas denote NBER recessions. The models considered are FF, PCA, IPCA, and AE with three factors.

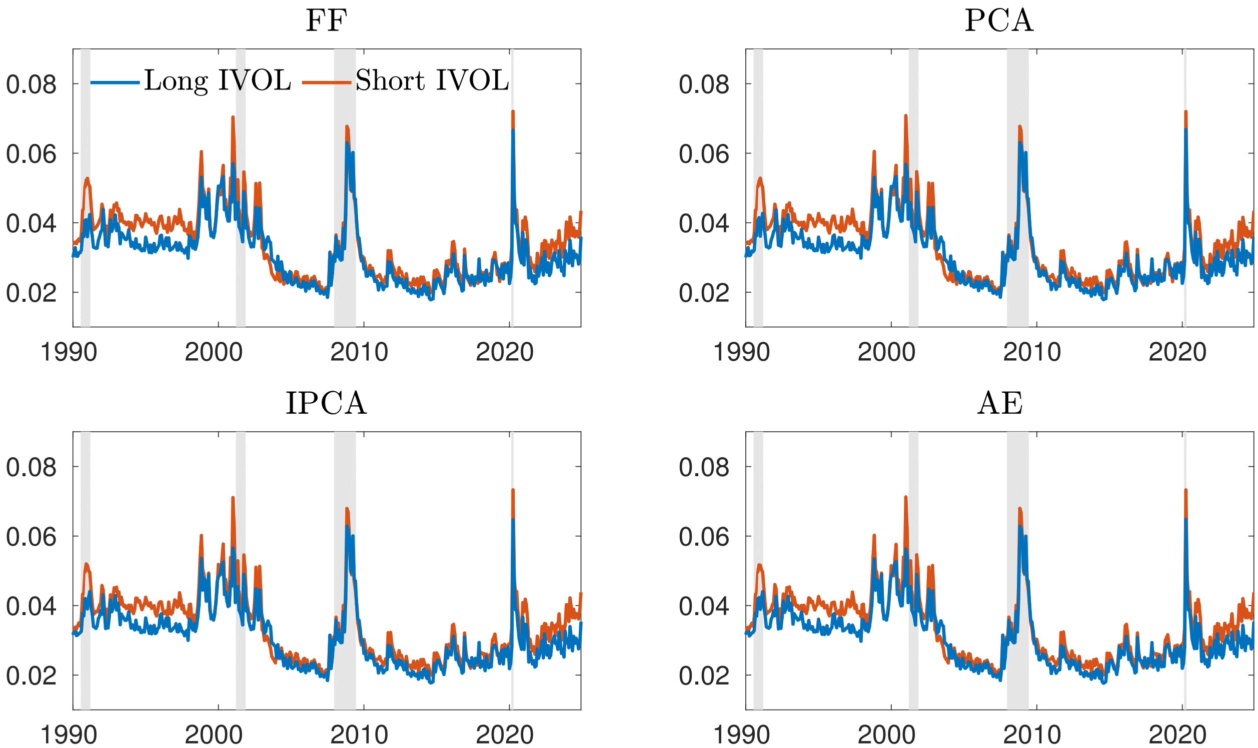
Figure. B.28. Market Betas of Long vs Short Positions (BAB)



Notes: This figure shows the weighted average market betas of long and short positions for the BAB factor portfolio. Gray-shaded areas denote NBER recessions.

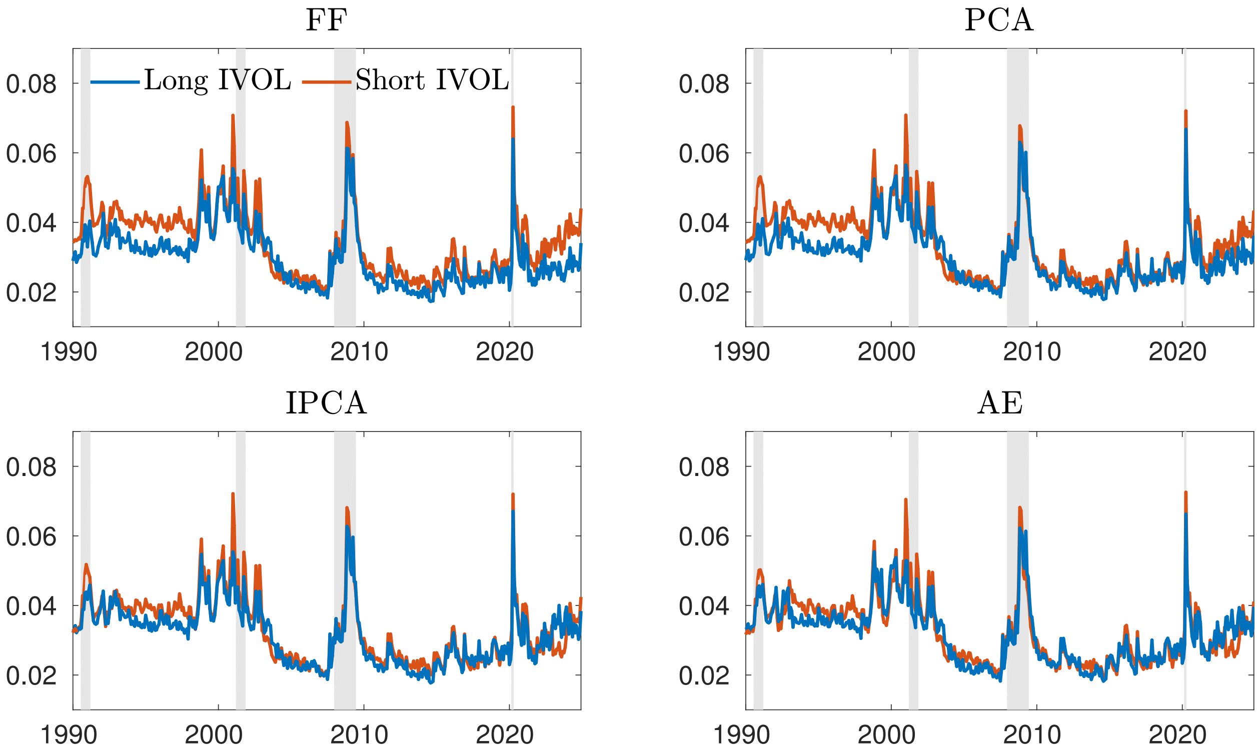
B.15. Idiosyncratic Risks of Long vs Short Positions

Figure. B.29. Idiosyncratic Risks of Long vs Short Positions (Single Factor)



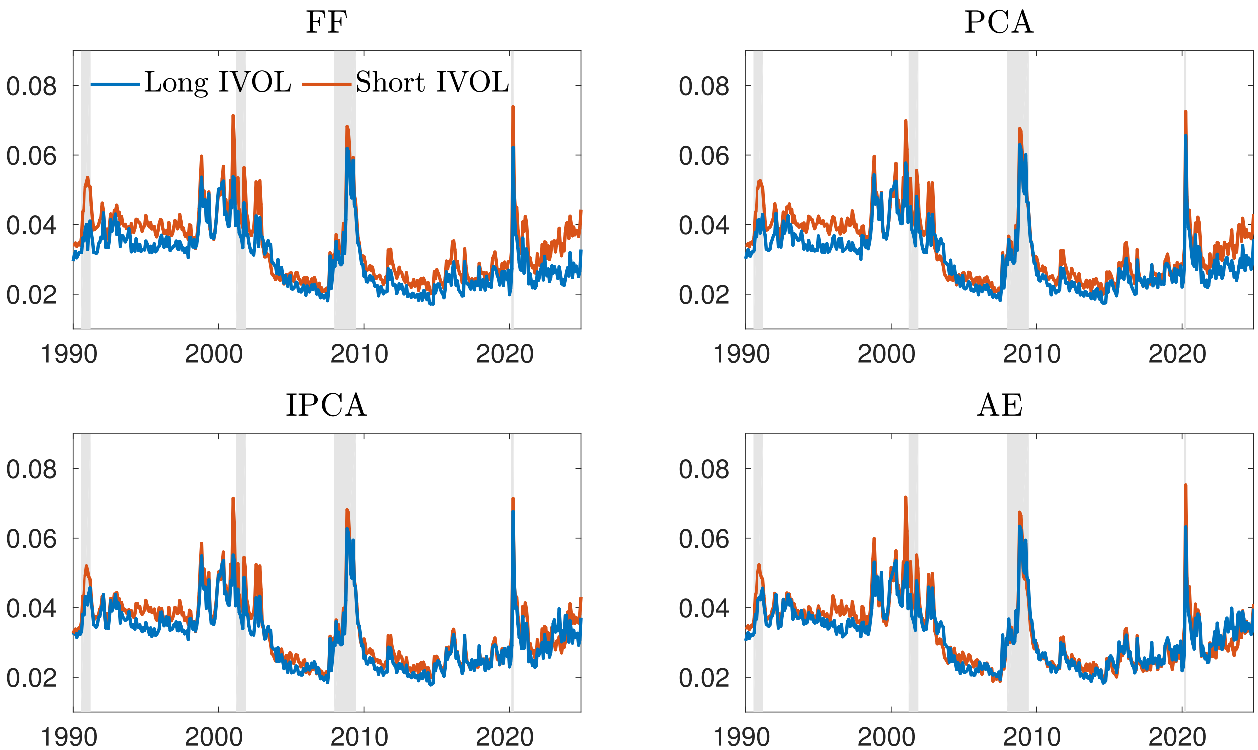
Notes: This figure shows the weighted average idiosyncratic risks of long and short positions for the optimal zero-beta portfolios. Gray-shaded areas denote NBER recessions. The models considered are FF, PCA, IPCA, and AE with a single factors.

Figure. B.30. Idiosyncratic Risks of Long vs Short Positions (Three Factor)



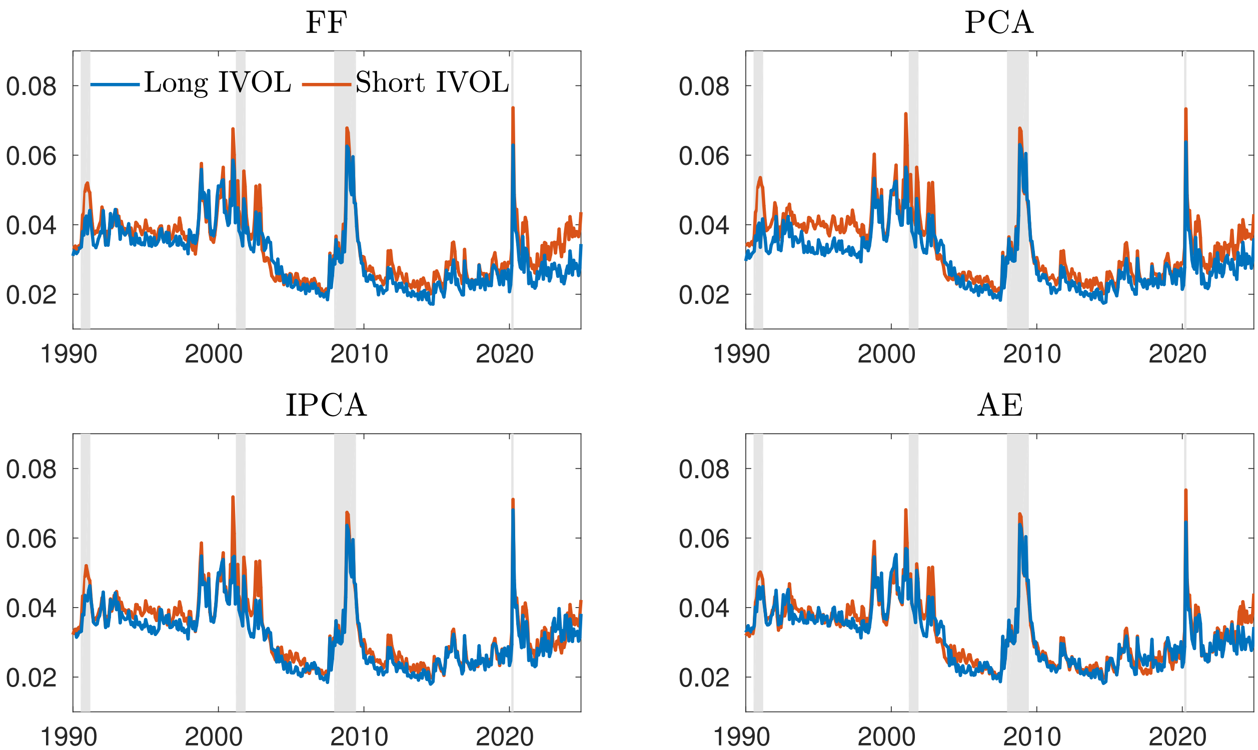
Notes: This figure shows the weighted average idiosyncratic risks of long and short positions for the optimal zero-beta portfolios. Gray-shaded areas denote NBER recessions. The models considered are FF, PCA, IPCA, and AE with three factors.

Figure. B.31. Idiosyncratic Risks of Long vs Short Positions (Five Factor)



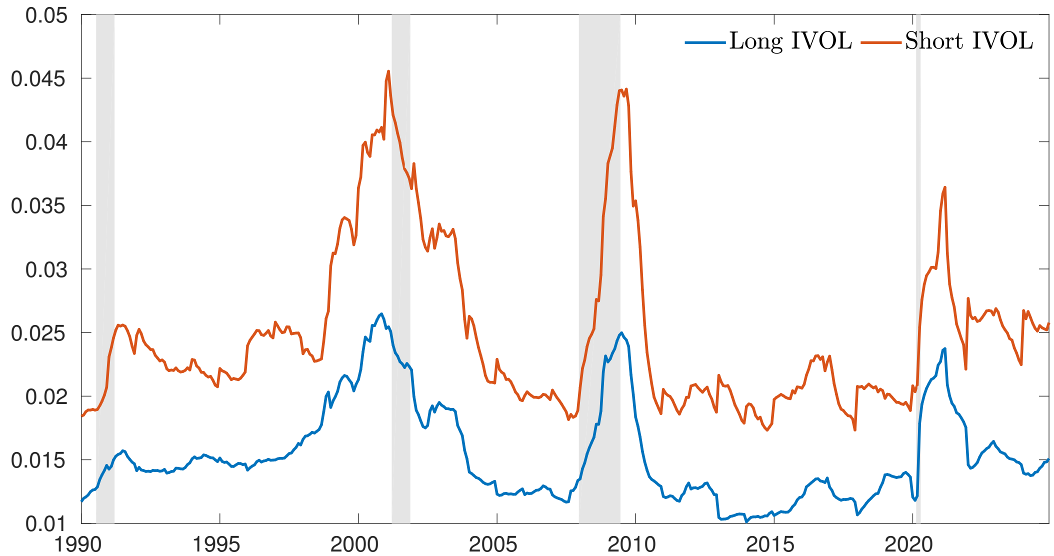
Notes: This figure shows the weighted average idiosyncratic risks of long and short positions for the optimal zero-beta portfolios. Gray-shaded areas denote NBER recessions. The models considered are FF, PCA, IPCA, and AE with three factors.

Figure. B.32. Idiosyncratic Risks of Long vs Short Positions (Six Factor)



Notes: This figure shows the weighted average idiosyncratic risks of long and short positions for the optimal zero-beta portfolios. Gray-shaded areas denote NBER recessions. The models considered are FF, PCA, IPCA, and AE with three factors.

Figure. B.33. Idiosyncratic Risks of Long vs Short Positions (BAB)



Notes: This figure shows the weighted average idiosyncratic risks of long and short positions for the BAB factor portfolio. Gray-shaded areas denote NBER recessions.

Appendix C. Proofs

C.1. Proof of Proposition 1

C.2. Proof of Proposition 2

C.3. Proof of Equation (7)

C.4. Proof of Equation (10)