Journal of Financial Economics 162 (2024) 103949

Contents lists available at ScienceDirect

Journal of Financial Economics

journal homepage: www.elsevier.com/locate/finec

Comparing factor models with price-impact costs[☆]

Sicong Li [a], Victor DeMiguel [b,*], Alberto Martín-Utrera [c]

[a] The Chinese University of Hong Kong, Hong Kong
[b] London Business School, United Kingdom
[c] Iowa State University, United States of America

ARTICLE INFO

Dataset link: Comparing Factor Models with Price-Impact Costs (Original data)

Keywords:
Trading costs
Mean–variance utility
Statistical test

ABSTRACT

We propose a formal statistical test to compare asset-pricing models in the presence of price impact. In contrast to the case without trading costs, we show that in the presence of price-impact costs different models may be best at spanning the investment opportunities of different investors depending on their absolute risk aversion. Empirically, we find that the five-factor model of Hou et al. (2021), the six-factor model of Fama and French (2018) with cash-based operating profitability, and a high-dimensional model are best at spanning the investment opportunities of investors with high, medium, and low absolute risk aversion, respectively.

1. Introduction

Prominent asset-pricing models include factors constructed using time-varying firm characteristics such as profitability and momentum. The investment opportunity set implied by these models requires investors to execute sizable trades whenever the conditioning information in firm characteristics changes. For the large institutional investors that manage most of the capital in financial markets, the price impact of these trades affects their optimal portfolio choices, and thus, it also affects the overall achievable investment opportunity set. We propose a formal statistical test to compare asset-pricing factor models in the presence of price impact. In contrast to the cases without transaction costs and with proportional costs, we show that in the presence of price-impact costs different models may be best at spanning the investment opportunities of different investors depending on their absolute risk aversion. Empirically, we find that the five-factor model of Hou et al. (2021), the six-factor model of Fama and French (2018) with cash-based operating profitability, and the high-dimensional model of DeMiguel et al. (2020) are best at spanning the investment opportunities of investors with high, medium, and low absolute risk aversion, respectively.

A popular approach to compare asset-pricing models is the GRS test of Gibbons et al. (1989), which evaluates the ability of the factors in a model to span the investment opportunity set generated by certain test assets. The GRS statistic is a quadratic form of the time-series intercept (alpha) obtained from the regression of the test-asset returns on the factor returns. Gibbons et al. (1989) show that this quadratic form measures the squared Sharpe ratio improvement that an investor can achieve by having access to the test assets, in addition to the factors in the model. Moreover, Barillas and Shanken (2017) show that, under the tenet that a good model should span not only the test assets but also the factors in other models, test assets are irrelevant and it suffices to compare models in terms of the squared Sharpe ratio generated by their factors.

Detzel et al. (2023), however, point out that one has to account for *trading costs* when comparing factor models. In particular, they explain that the framework underpinning these models, the arbitrage pricing theory (APT) of Ross (1976), relies on the assumption that investment opportunities that deliver abnormal returns attract arbitrage capital until such opportunities vanish. However, arbitrageurs allocate capital only to investment opportunities that are profitable after trading costs, and thus, Detzel et al. (2023) propose comparing factor models in terms of their squared Sharpe ratio of returns net of *proportional* transaction costs, which measures the ability of the factor model to span the *achievable* investment opportunity set.

Proportional transaction costs capture the trading costs of retail investors, but price-impact costs are more relevant for the large institutional investors who manage most of the capital in financial markets.

☆ Nikolai Roussanov was the editor for this article. We are grateful to the editor and an outstanding referee who helped us greatly improve our empirical contribution. We are also grateful for comments from Pedro Barroso, Svetlana Bryzgalova, Andrew Chen, Andrew Detzel, Gavin Feng, Jinyu He, Raymond Kan, Nicos Savva, Stephen Schaefer, Raman Uppal, Mihail Velikov, Guofu Zhou, Paolo Zaffaroni and seminar participants at 4th Frontiers of Factor Investing Conference at Lancaster University, AFA Ph.D. poster session, City University of Hong Kong, FMA Annual Meeting, INFORMS Annual Meeting, International Young Finance Scholars' Conference, Maastricht University, Tilburg University, and Trans-Atlantic Doctoral Conference at London Business School.

* Corresponding author.
E-mail addresses: sicongli@cuhk.edu.hk (S. Li), avmiguel@london.edu (V. DeMiguel), amutrera@iastate.edu (A. Martín-Utrera).

https://doi.org/10.1016/j.jfineco.2024.103949
Received 8 November 2023; Received in revised form 16 September 2024; Accepted 18 September 2024
Available online 21 September 2024

For instance, Gârleanu and Pedersen (2022) show that institutional investors held around 50% of the US equity market in 2017, and Edelen et al. (2007) show that price-impact costs represent 65% of the total trading costs of mutual funds, whereas proportional (bid–ask spread) costs represent only 17%.

Despite the importance of price impact for the large investors that dominate financial markets, price impact would not affect the achievable investment opportunity set if large investors did not have to trade or they had to execute only small trades. However, prominent asset-pricing models include factors constructed using firm characteristics such as profitability and momentum that vary substantially over time. These characteristics encapsulate conditioning information that investors can optimally exploit when choosing their portfolios (Cochrane, 2009, p. 134). As a result, the investment opportunity set implied by these factor models requires investors to execute *sizable* trades at regular intervals—whenever the conditioning information in firm characteristics varies. For large investors, the price impact of these sizable trades affects their optimal portfolio choices, and thus, price impact affects the overall achievable investment opportunity set, which includes the optimal portfolio of every investor. In this manuscript, we propose a methodological framework to compare factor models in terms of their ability to span the achievable investment opportunity set in the presence of price impact.

Our contribution to the literature is fourfold. Our first contribution is to propose comparing factor models in terms of the *mean–variance utility net of price-impact costs* generated by their factors, and to show that different models may be better at spanning the investment opportunities of investors with different absolute risk aversion. In particular, we prove that the achievable efficient frontier in the presence of price impact is strictly concave, and thus, the squared Sharpe ratio criterion is no longer sufficient to compare factor models because each efficient portfolio has a different Sharpe ratio of returns net of price-impact costs. Moreover, the objective of investors is not to maximize Sharpe ratio, but rather their utility of returns net of price-impact costs, which is therefore the economically meaningful criterion to compare the ability of different factor models to span the achievable investment opportunity set. We show that our proposed criterion is equivalent to the squared Sharpe ratio in the cases without costs and with proportional costs. In addition, we generalize the result of Gibbons et al. (1989) to show that the increase in the mean–variance utility net of price-impact costs of an investor when she has access to a set of test assets in addition to the factors in a model is a quadratic form of the alpha (net of price impact). Finally, we also generalize the result of Barillas and Shanken (2017) to show that test assets are irrelevant for model comparison *also* in the presence of price impact under the tenet that a good model should span not only the test assets but also the factors in other models.

The intuition behind our first contribution is illustrated in Fig. 1, which depicts the achievable efficient frontier (black solid line) as well as the efficient frontiers spanned by the factors in models A (red dotted line) and B (blue dashed line) in the presence of price impact. Each portfolio in the achievable frontier maximizes the mean–variance utility net of price-impact costs of investors with a particular absolute risk aversion, which can be defined as the ratio of the investor's relative risk aversion to her endowment (Gârleanu and Pedersen, 2013). Intuitively, larger investors have lower absolute risk aversion, and thus, they are willing to take on larger investment positions to maximize their net mean return at the expense of higher return variance. The figure depicts the indifference curves of an investor with low absolute risk aversion (brown dash-dotted lines) and an investor with high absolute risk aversion (green dash-dotted lines). Each investor's optimal portfolio is at the tangent between the investor's indifference curve and the efficient frontier. The figure shows that model B spans better the investment opportunities of the low-absolute-risk-aversion investor and model A those of the high-absolute-risk-aversion investor. This is because the low-absolute-risk-aversion investor is willing to take on larger investment positions that incur higher price-impact costs. Because the price-impact costs from exploiting the factors in model B are much lower than those from exploiting the factors in model A, model B is better at spanning the achievable investment opportunities of the low-absolute-risk-aversion investor. On the other hand, model A is better at spanning the investment opportunities of the high-absolute-risk-aversion investor because she takes smaller investment positions that incur lower price-impact costs, and the factors of model A offer a better risk–return tradeoff when price-impact costs are lower.

Our second contribution is to develop statistical tests to compare factor models in terms of mean–variance utility net of price-impact costs. For *pairwise* model comparisons, we derive two asymptotic distributions that allow us to compare two factor models for the cases when they are nested or non-nested by generalizing the pairwise tests of Kan and Robotti (2009) and Barillas et al. (2020) to the case with price impact. We also develop a closed-form expression for the variance of the asymptotic distribution and use it to show that it is easier to reject the null hypothesis that the mean–variance utilities net of price-impact costs of two models are equal not only when the mean–variance portfolio returns of the two models are positively correlated as shown by Barillas et al. (2020) for the case without trading costs, but also when the mean–variance portfolio return of each model is highly correlated with the rebalancing trades of the portfolio of the other model, and when the rebalancing trades of the two portfolios are highly correlated. For *multiple* model comparisons, we use the approach of Barillas et al. (2020) to test the null hypothesis that a benchmark model has a mean–variance utility net of price-impact costs at least as high as that of any other model in a set of alternative models.

Our third contribution is to use our statistical tests to compare the empirical performance of ten factor models in terms of the mean–variance utility net of price-impact costs generated by their factors. We consider nine *low-dimensional* factor models: the CAPM model of Sharpe (1964) and Lintner (1965), the four-factor model of Hou et al. (2015), HXZ4, the five-factor model of Hou et al. (2021), HMXZ5, the four-factor model of Fama and French (1993) and Carhart (1997), FFC4, the five-factor model of Fama and French (2015), FF5, the six-factor model of Fama and French (2018), FF6, and the six-factor model of Barillas and Shanken (2018), BS6. Note that Fama and French (2018) and Ball et al. (2016) show that using cash-based operating profitability instead of accrual-based operating profitability can improve model performance, and thus, following Detzel et al. (2023) we also consider versions of the five- and six-factor Fama–French models constructed using cash-based operating profitability, FF5c and FF6c. In addition, DeMiguel, Martin-Utrera, Nogales, and Uppal (2020) show that trading costs provide an economic rationale to consider high-dimensional factor models. In particular, they show that combining factors helps to reduce transaction costs because the trades required to rebalance different factor portfolios often cancel out, a phenomenon they term *trading diversification*. Moreover, they show that the benefits from trading diversification increase with the number of factors combined. For this reason, we consider a tenth factor model containing the 20 factors that DeMiguel et al. (2020) find statistically significant in the presence of price-impact costs, DMNU20.

We highlight two empirical findings. First, in the presence of price impact, model performance depends not only on the portfolio turnover required to trade the factors in the model, as pointed out by Detzel et al. (2023) for the case with proportional costs, but also on the liquidity of the stocks traded. In particular, we find that, compared to their Fama–French counterparts, the investment and profitability factors of Hou et al. (2015, 2021) not only involve higher portfolio turnover, but also require trading stocks with lower market capitalization, which are more illiquid and subject to higher price-impact costs. As a result, while in the absence of trading costs the five-factor model of Hou et al. (2021) outperforms the six-factor model of Fama and French (2018) with cash-based operating profitability, in the presence of price-impact costs the

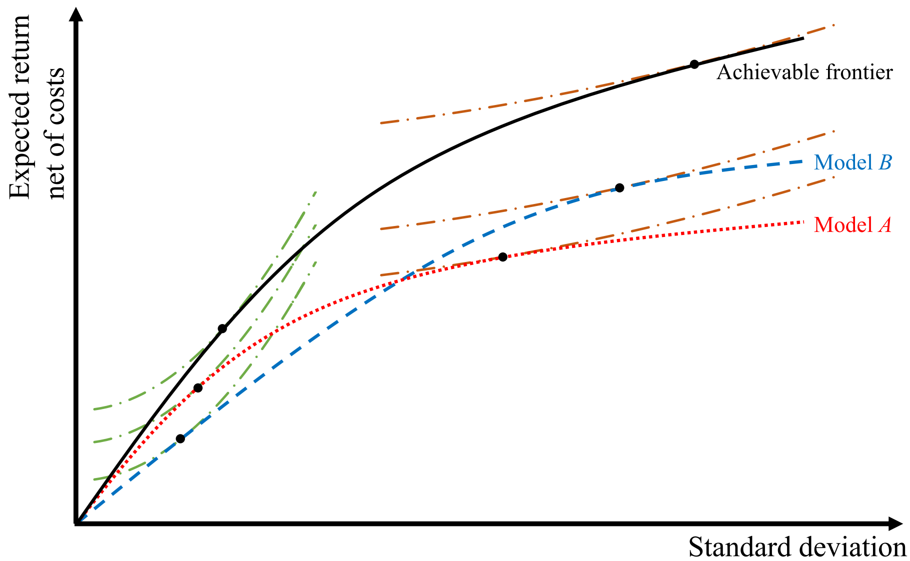

Fig. 1. Achievable efficient frontier and frontiers spanned by two factor models.
This figure illustrates the achievable efficient frontier in the presence of price impact (black solid line) as well as the efficient frontiers spanned by the factors in models *A* (red dotted line) and *B* (blue dashed line). The figure also depicts the indifference curves of an investor with low absolute risk aversion (brown dash-dotted lines) and an investor with high absolute risk aversion (green dash-dotted lines). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

six-factor model of Fama and French (2018) with cash-based operating profitability tends to perform better.[1]

Second, the relative performance of factor models in the presence of price impact depends on the absolute risk aversion of the investor. For instance, the high-dimensional model of DeMiguel et al. (2020) significantly outperforms the low-dimensional models *only* when spanning the investment opportunities of large (low-absolute-risk-aversion) investors. This is because high-dimensional models provide larger trading-diversification benefits, and thus, they outperform low-dimensional models at spanning the investment opportunities of large investors for whom price-impact costs are relatively more important. Overall, accounting for price impact results in a nuanced comparison of the factor models we consider—the five-factor model of Hou et al. (2021), the six-factor model of Fama and French (2018) with cash-based operating profitability, and the high-dimensional model of DeMiguel et al. (2020) are best at spanning the investment opportunities of investors with high, medium, and low absolute risk aversion, respectively.[2]

We show that, under the tenet of Barillas and Shanken (2017) that a good model should span not only the test assets but also the factors in other models, test assets are irrelevant and it suffices to compare models in terms of mean–variance utility net of price-impact costs. However, *absent* the requirement that a good model should span also the factors in other models, relative model performance in terms of test-asset spanning may differ from that in terms of mean–variance utility. Therefore, it is of interest to compare models *also* in terms of their ability to span certain test assets. Our fourth contribution is to use our statistical test to compare the different models in terms of their ability to span the 212 anomalies of Chen and Zimmermann (2022) in the presence of price impact. We find that the relative performance of factor models in terms of anomaly spanning is similar, but not identical, to that in terms of the mean–variance utility. For instance, in the absence of costs the five-factor model of Hou et al. (2021) outperforms other models in terms of both anomaly spanning and mean–variance utility. Similarly, in the presence of price-impact costs with medium absolute risk aversion, the six-factor model of Fama and French (2018) with cash-based operating profitability outperforms other models in terms of both anomaly spanning and mean–variance utility. However, the six-factor model of Fama and French (2018) with cash-based operating profitability outperforms other models in terms of anomaly spanning *also* for the case with high absolute risk aversion, for which the five-factor model of Hou et al. (2021) is better in terms of mean–variance utility.[3]

Our manuscript is closely related to Detzel et al. (2023), who compare asset-pricing models in the presence of proportional transaction costs using the maximum squared Sharpe ratio criterion of Barillas and Shanken (2017). We formally prove that the squared Sharpe ratio criterion remains valid in the presence of proportional transaction costs, and thus, we provide theoretical support for the empirical analysis of Detzel et al. (2023). We also demonstrate that the squared Sharpe ratio criterion is no longer sufficient to characterize the investment opportunity set in the presence of price-impact costs and, instead, we propose comparing factor models in terms of the mean–variance utility of returns net of price-impact costs. The different comparison methodology and our focus on price-impact costs instead of proportional transaction costs are key distinctive elements of our work.

Our work is also related to Jensen et al. (2022), who generalize the dynamic portfolio framework of Gârleanu and Pedersen (2013) to integrate machine-learning return forecasts obtained from a large set of

[1] Following Detzel, Novy-Marx, and Velikov (2023), we ignore short-selling costs in our analysis. Although the findings of Nagel (2005) and Muravyev, Pearson, and Pollet (2022) suggest that short-selling costs may affect factor model performance, we ignore short-selling costs in our analysis in order to isolate the effect of price-impact costs, which are the focus of our manuscript. This also facilitates the comparison of our results for the case with price-impact costs with those of Detzel et al. (2023) for the case with proportional transaction costs.

[2] As a robustness check, we use the bootstrap test of Fama and French (2018) and Detzel et al. (2023) to show that the out-of-sample performance of the different models is consistent with the empirical findings from our statistical tests. In addition, Section IA.7 of the Internet Appendix shows that our empirical findings are robust to considering factors constructed using the *banding* transaction-cost mitigation strategy used by Detzel et al. (2023).

[3] Barillas and Shanken (2017) also show that the relative performance of factor models in terms of test-asset spanning and squared Sharpe ratio may be different in the absence of trading costs once one drops the requirement that the better model should also span the factors in the other model—see the last paragraph on page 1317 and section 1.2 of Barillas and Shanken (2017).

firm characteristics. Like us, Jensen et al. (2022) account for the price-impact costs that are relevant to "market participants with a substantial fraction of aggregate assets under management, such as large pension funds or other professional asset managers". A key distinctive feature of our work is that our focus is not to use machine learning to exploit a large number of characteristics, but rather to propose a rigorous methodology to compare existing asset-pricing models in terms of their ability to span the investment opportunity set in the presence of price impact.

There is a large literature that proposes statistical tests to compare asset-pricing models in the absence of transaction costs (Avramov and Chao, 2006; Kan and Robotti, 2009; Kan et al., 2013; Barillas and Shanken, 2018; Goyal et al., 2018; Fama and French, 2018; Ferson et al., 2019; Chib et al., 2020; Kan et al., 2019). In contrast to these papers, we propose a statistical methodology that accounts for the effect of price-impact costs when comparing asset-pricing models.

Finally, our work is related to the literature on the profitability of factor strategies (Korajczyk and Sadka, 2004; Novy-Marx and Velikov, 2016; Frazzini et al., 2018; Chen and Velikov, 2023; Barroso and Detzel, 2021). Most of these papers study the profitability of individual-factor strategies. However, DeMiguel et al. (2020) show that the trades in the underlying stocks required to rebalance different factors often cancel out, and thus the trading cost of exploiting the factors in a model is lower when the factors are combined.[4] In this manuscript, instead of studying the profitability of the individual factor strategies, we explicitly account for the effect of trading diversification when we compare low- and high-dimensional factor models in the presence of price-impact costs.

The rest of the manuscript is organized as follows. Section 2 proposes mean–variance utility net of price-impact costs as a criterion to compare factor models. Sections 3 and 4 discuss the statistical tests used to perform pairwise and multiple model comparisons in the presence of price impact. Sections 5 and 6 compare the empirical performance of ten factor models in terms of mean–variance utility net of price-impact costs and in terms of anomaly spanning. Section 7 concludes. Appendix A contains the proofs of all theoretical results with the exception of Proposition 4, which is proven and discussed in Appendix B. The Internet Appendix contains several robustness checks and additional information.

2. A criterion to compare models with trading costs

In this section, we propose a novel criterion to compare factor models in the presence of price-impact costs. Section 2.1 gives the notation and assumptions. Section 2.2 reviews the squared Sharpe ratio criterion proposed by Barillas and Shanken (2017) to compare factor models in the absence of trading costs, and in Section 2.3 we prove that this criterion is also valid in the presence of *proportional* transaction costs. In Section 2.4, however, we show that the squared Sharpe ratio criterion is no longer sufficient to characterize the achievable investment opportunity set in the presence of price-impact costs, and thus, in Section 2.5 we propose comparing factor models in terms of their mean–variance utility net of trading costs. Section 2.6 shows that there is a close relation between the mean–variance utility net of price-impact cost criterion and the alpha *net of price impact*. Finally, Section 2.7 shows that the relative performance of factor models in the presence of price impact depends in general on the investor's absolute risk aversion.

2.1. Notation and assumptions

We first describe the notation we use in our analysis. We consider a market with N stocks whose return vector at time t is $r_t \in \mathbb{R}^N$ and a risk-free asset with return $r_{f,t} \in \mathbb{R}$. Let $X_t \in \mathbb{R}^{N\times K}$ be the matrix whose columns contain the weights of the K factor portfolios at time t. Then, the vector of returns of the K factors at time $t+1$ is

$$F_{t+1} = X_t^\top (r_{t+1} - r_{f,t+1} e) \in \mathbb{R}^K, \tag{1}$$

where e is the N-dimensional vector of ones. Every factor we consider is a return in excess of the risk-free rate. In particular, every factor (other than the market) is the return of a long-short portfolio of stocks with one dollar invested in the long leg and one dollar in the short leg, and thus, its returns equal its excess returns. The market factor is also the return of a long-short portfolio because it is the market return in excess of the risk-free rate, and thus, its investment in the long leg is equal to that in the short leg once we account for its negative investment in the risk-free asset.

Let $\mu = E[F_t]$ and $\Sigma = \text{var}(F_t)$ be the mean and covariance matrix of factor returns. Then, the mean–variance factor portfolio, $\theta^* \in \mathbb{R}^K$, is the maximizer to the following problem:

$$\max_{\theta} \qquad \theta^\top \mu - f(\theta) - \frac{\gamma}{2}\theta^\top \Sigma \theta, \tag{2}$$

where the kth component of θ is the *dollar-amount* allocated to the kth factor, $\theta^\top \mu$ is the expected portfolio return, $f(\theta)$ is the trading cost associated with the portfolio θ, $\theta^\top \Sigma\theta$ is the portfolio return variance, and γ is the absolute risk-aversion parameter. Note that because the factors are returns in excess of the risk-free rate, we do not need to impose a budget constraint on the mean–variance factor portfolio weights. Thus, like the portfolio proposed by Gârleanu and Pedersen (2013), our mean–variance factor portfolio depends on the investor's endowment only through her absolute risk aversion, which is the ratio of the investor's relative risk-aversion parameter to her endowment.

A couple of comments are in order. First, Sections 2.3 and 2.4 consider specific examples of proportional transaction costs and price-impact costs. Second, consistent with the asset-pricing literature on factor model comparison (Gibbons et al., 1989; Kan and Robotti, 2008; Barillas and Shanken, 2017, 2018; Barillas et al., 2020; Detzel et al., 2023), we consider an *unconditional* mean–variance portfolio of the factors in a model. This is not a limitation because prominent asset-pricing models include factors constructed using time-varying characteristics such as profitability and momentum that encapsulate conditioning information. Moreover, one can also use conditioning variables to generate *managed* versions of popular asset-pricing factors and include them as additional factors in the unconditional mean–variance portfolio.[5]

We now state the assumptions required in our theoretical analysis. First, we require that the factor returns are not perfectly colinear.

Assumption 2.1. The covariance matrix of the factor returns Σ is positive definite.

Second, we make the following assumption for the functional form of trading costs.

Assumption 2.2. The trading-cost function $f(\theta)$ is continuous in θ, $f(0) = 0$, and $f(\theta) > 0$ for all $\theta \neq 0$.

Assumption 2.2 is satisfied by most popular trading-cost models, such as proportional and quadratic trading-cost models. In particular, prominent asset-pricing models include factors constructed using time-varying firm characteristics, and thus, investing in these factors requires

[4] Other papers also provide empirical evidence that combining factors can reduce trading costs (Barroso and Santa-Clara, 2015; Frazzini et al., 2015; Novy-Marx and Velikov, 2016).

[5] For instance, Moreira and Muir (2017) consider volatility-managed factors and DeMiguel et al. (2024) incorporate the volatility-managed factors together with the unmanaged factors in an unconditional mean–variance portfolio.

the investor to rebalance her portfolio regularly, incurring strictly positive trading costs. Finally, the following assumption rules out the trivial case in which it is not optimal to invest in any of the factors.

Assumption 2.3. The set $S = \{\theta | \theta^\top \mu - f(\theta) > 0\}$ is non-empty.

2.2. The case without trading costs

In the absence of trading costs, the mean–variance portfolio θ^* of the factors is the solution to Problem (2) for the case with $f(\theta) = 0$. One can recover all portfolios on the efficient frontier by solving the problem for different values of γ. The following proposition reviews a well-known property of the efficient frontier; see, for instance, Campbell (2017, Section 2.2.6).

Proposition 1. *Let Assumption 2.1 hold and consider an investor with absolute risk aversion $\gamma > 0$. Then, the unique maximizer to the mean–variance problem (2) in the absence of transaction costs is*

$$\theta^* = \Sigma^{-1}\mu/\gamma, \tag{3}$$

with mean–variance utility $MVU^\gamma = \mu^\top \Sigma^{-1}\mu/(2\gamma)$, and squared Sharpe ratio $SR^2 = 2\gamma MVU^\gamma$. Thus, the efficient frontier is a straight line in the mean–standard-deviation diagram because every mean–variance portfolio delivers the same maximum Sharpe ratio, $SR = \sqrt{\mu^\top \Sigma^{-1}\mu}$.

Proposition 1 shows that, in the absence of trading costs, the Sharpe ratio of any mean–variance portfolio of the factors in the model is a sufficient statistic to characterize the investment opportunity set spanned by the model. Thus, the model that best spans the investment opportunity set is the one whose factors attain the highest squared Sharpe ratio as noted by Barillas and Shanken (2017).

2.3. The case with proportional trading costs

We first provide a general definition of proportional-trading-cost function.

Definition 1 (*Proportional-Trading-Cost Function*)**.** A proportional-trading-cost function $f(\theta)$ is one that satisfies Assumption 2.2 and is homogeneous of degree one, that is,

$$f(c\theta) = c f(\theta) \quad \text{for all } \theta \text{ and } c \geq 0. \tag{4}$$

We now give a popular example of proportional-trading-cost function used (among others) by DeMiguel et al. (2020) and Detzel et al. (2023). We start by defining the rebalancing-trade matrix of the K factors at time t as

$$\tilde{X}_t = X_t - \mathrm{diag}(e + r_t) X_{t-1}, \tag{5}$$

where e is the N-dimensional vector of ones and $\mathrm{diag}(v)$ is a diagonal matrix whose diagonal contains the elements in vector v. Note that the element in the nth row and kth column of $\tilde{X}_t$ is the rebalancing trade on stock n required at time t to hold the kth factor portfolio. To see this, note that the kth factor portfolio weight on stock n changes from $x_{n,k,t-1}(1 + r_{n,t})$ before rebalancing at time t to $x_{n,k,t}$ after rebalancing, where $x_{n,k,t}$ is the kth factor portfolio weight on the nth stock at time t, that is, the element in the nth row and kth column of X_t. Then, the rebalancing trade required at time t to hold the factor portfolio θ can be written as $\Delta w = \tilde{X}_t \theta$, and thus, the proportional-trading-cost function can be defined as

$$f(\theta) = E\left[\|K_t \tilde{X}_t \theta\|_1\right], \tag{6}$$

where $\|v\|_1 = \sum_{i=1}^N |v_i|$ is the L_1-norm of vector $v \in \mathbb{R}^N$ and $K_t \in \mathbb{R}^{N\times N}$ is a diagonal matrix whose nth element, $\kappa_{n,t} > 0$, is the transaction-cost parameter of stock n at time t.[6]

Solving Problem (2) with a proportional-trading-cost function for different values of the risk-aversion parameter γ, one can recover the efficient frontier in the presence of proportional trading costs. In the following proposition, we prove that this efficient frontier is a straight line in the mean–standard-deviation diagram.[7]

Proposition 2. *Let $f(\theta)$ be a proportional-trading-cost function. Then, the efficient frontier in the presence of proportional trading costs is a straight line in the mean–standard-deviation diagram, and all portfolios on the efficient frontier deliver the same maximum Sharpe ratio of returns net of proportional trading costs, $SR_{PTC} < SR = \sqrt{\mu^\top \Sigma^{-1}\mu}$, where SR is the maximum Sharpe ratio in the absence of trading costs.*

Proposition 2 shows that, similar to the case without trading costs, the investment opportunity set spanned by the factors in the presence of proportional trading costs is fully characterized by the Sharpe ratio of returns net of costs. Thus, Proposition 2 demonstrates that the maximum squared Sharpe ratio criterion remains valid to compare factor models in the presence of proportional costs, and thus, it provides theoretical support for the empirical analysis in Detzel et al. (2023). However, proportional costs ignore the price impact of large trades, which affects the portfolio choices of large investors and thus the overall achievable investment opportunity set. In the next section, we show that the squared Sharpe ratio criterion is no longer sufficient to characterize the investment opportunity set in the presence of price-impact costs.

2.4. The case with price-impact costs

We now consider the case with price-impact costs. First, we provide a general definition of price-impact-cost function.

Definition 2 (*Price-Impact-Cost Function*)**.** A price-impact-cost function $f(\theta)$ satisfies Assumption 2.2 and the following inequality:

$$f(c\theta) > c f(\theta) \quad \text{for all } \theta \neq 0 \text{ and } c > 1. \tag{8}$$

We now specify the price-impact-cost function that we use in our analysis. A common assumption in the literature is that the impact on prices from large trades is linear in the amount traded (Korajczyk and Sadka, 2004; Novy-Marx and Velikov, 2016). Under this assumption, the *price impact* of rebalancing the factor portfolio at time t is:

$$\mathrm{PI}_t = D_t \Delta w_t = D_t \tilde{X}_t \theta, \tag{9}$$

where $\theta \in \mathbb{R}^K$ is the factor portfolio in dollars, $\tilde{X}_t$ is the rebalancing-trade matrix defined in (5), $\Delta w_t = \tilde{X}_t \theta$ is the rebalancing trade required

[6] Detzel et al. (2023) consider the proportional-trading-cost function (6) as a robustness check in section 6.2 of their manuscript. In their main analysis, Detzel et al. (2023) use the following proportional-trading-cost function:

$$f(\theta) = E\left[\sum_{n=1}^N \kappa_{n,t} \sum_{k=1}^K |\tilde{x}_{n,k,t}\theta_k|\right], \tag{7}$$

where $\tilde{x}_{n,k,t}$ is the rebalancing trade of factor k on stock n at time t, which is the element in the nth row and kth column of the rebalancing-trade matrix $\tilde{X}_t$. An advantage of the proportional-trading-cost function (6) compared to (7) is that it aggregates the rebalancing trades across the K factors and thus accounts for the trading-diversification benefits from combining multiple factors. DeMiguel et al. (2020) find that the trades in the underlying stocks required to rebalance different factors often net out, and therefore exploiting multiple factors simultaneously reduces trading costs.

[7] The mean–standard-deviation diagram for the case with proportional trading costs depicts in the horizontal axis the standard deviation of portfolio returns, and in the vertical axis the mean of portfolio returns net of proportional trading costs.

to rebalance the factor portfolio θ at time t, and $D_t \in \mathbb{R}^{N\times N}$ is a diagonal matrix whose nth element, $d_{n,t} > 0$, is the price-impact-cost parameter (i.e., Kyle's lambda) of stock n at time t. Then, the price-impact *cost*, in dollars, required to rebalance the factor portfolio θ at time t is half of the scalar product of the price impact $\mathrm{PI}_t = D_t\tilde{X}_t\theta$ and the rebalancing trade $\Delta w_t = \tilde{X}_t\theta$:

$$f_t(\theta) = \frac{1}{2}\theta^\top \tilde{X}_t^\top D_t \tilde{X}_t\theta. \tag{10}$$

To simplify notation, let

$$\Lambda_t = \tilde{X}_t^\top D_t \tilde{X}_t \in \mathbb{R}^{K\times K} \tag{11}$$

be the price-impact matrix at time t, and $\Lambda = E[\Lambda_t]$ the expected price-impact matrix, which is assumed to be positive definite. Then, the quadratic price-impact-cost function is

$$f(\theta) = E\left[\frac{\theta^\top \Lambda_t \theta}{2}\right] = \frac{\theta^\top \Lambda\theta}{2}, \tag{12}$$

which gives the expected price-impact costs from trading the factor portfolio θ. It is straightforward to show that this function satisfies Definition 2 and accounts for trading diversification across factors.

The mean–variance problem (2) for the case with quadratic price-impact costs can then be rewritten as

$$\max_\theta \quad \theta^\top\mu - \frac{1}{2}\theta^\top\Lambda\theta - \frac{\gamma}{2}\theta^\top\Sigma\theta,$$

where θ is the factor portfolio in dollars, $\theta^\top\mu$ is the expected factor portfolio return, $\theta^\top\Sigma\theta$ is the portfolio variance, and $\theta^\top\Lambda\theta/2$ is the quadratic price-impact cost. Thus, the mean–variance portfolio is

$$\theta^* = \frac{1}{\gamma}(\Sigma + \Lambda/\gamma)^{-1}\mu, \tag{13}$$

and the investor's mean–variance utility net of price-impact costs is

$$\mathrm{MVU}^\gamma = \frac{\mu^\top(\Sigma+\Lambda/\gamma)^{-1}\mu}{2\gamma}, \tag{14}$$

which is *not* proportional to the squared Sharpe ratio in the absence of costs. More precisely, price-impact costs affect the investor's portfolio choice and utility nonlinearly, by replacing the matrix Σ in (3) with the matrix $(\Sigma + \Lambda/\gamma)$, which depends on γ.

Solving Problem (2) with a price-impact-cost function for different values of γ, one can recover the efficient frontier in the presence of price-impact costs. The following proposition shows that the efficient frontier in the presence of price-impact costs is strictly concave in the mean–standard-deviation diagram.

Proposition 3. *Let $f(\theta)$ be a price-impact-cost function. Then, the efficient frontier in the presence of price-impact costs is strictly concave. In addition, the Sharpe ratio of returns net of price-impact costs of any portfolio on the efficient frontier, SR^γ_{PIC}, is lower than the maximum Sharpe ratio in the absence of trading costs, that is, $SR^\gamma_{PIC} < SR = \sqrt{\mu^\top\Sigma^{-1}\mu}$.*

The intuition behind Proposition 3 is that, while the mean and standard deviation of the portfolio returns grow proportionally with the dollar amount invested, the price-impact costs grow faster than linearly, and thus, the efficient frontier in the presence of price-impact costs is *strictly concave*. Consequently, the squared Sharpe ratio is no longer a sufficient criterion to compare factor models in the presence of price-impact costs because the achievable investment opportunity set of a factor model is not fully characterized by a *single* slope in the mean–standard-deviation diagram as in the absence of trading costs or the presence of proportional trading costs.

Fig. 2 illustrates the efficient frontiers attained by the factors of a model for the cases without trading costs, with proportional trading costs, and with price-impact costs. The frontiers for the cases with proportional costs and with price-impact costs are below that for the case without costs. Moreover, while the efficient frontier is a straight line in the cases without costs and with proportional trading costs, in the presence of price-impact costs, the efficient frontier is strictly concave, and thus the investment opportunity set in this case cannot be summarized by a single Sharpe ratio because every mean–variance portfolio has a different Sharpe ratio.

2.5. Mean–variance utility as a comparison criterion

In the previous section we showed that, in the presence of price-impact costs, the efficient frontier is strictly concave and thus a single Sharpe ratio no longer characterizes the achievable investment opportunity set as in the cases without costs or with proportional transaction costs. Thus, we cannot compare asset-pricing models in the presence of price-impact costs using the squared Sharpe ratio criterion because this metric is no longer a sufficient statistic to describe the extent to which the factors of a model span the achievable investment opportunity set. Instead, in this section we propose comparing factor models in terms of mean–variance utility net of price-impact costs.

Barillas and Shanken (2017) posit that when comparing two factor models, the better model should be able to span not only the investment opportunity set offered by the tests assets, but also that offered by the factors in the other model. In particular, let us consider two models with factors F_A and F_B and a set of test assets Π. In the absence of price-impact costs, Barillas and Shanken (2017) show that model A is better than model B if

$$SR^2([\Pi, F_A, F_B]) - SR^2(F_A) < SR^2([\Pi, F_A, F_B]) - SR^2(F_B), \tag{15}$$

where $SR^2(x)$ is the squared Sharpe ratio delivered by the assets in vector x. In particular, they explain that the two sides of Inequality (15) measure the *misspecification* of models A and B, and thus, model A is considered better (less misspecified) than model B because an investor with access to the factors in model A obtains a lower Sharpe ratio improvement by having access to the test assets and the factors in the other model than an investor with access to the factors in model B. This inequality is equivalent to

$$SR^2(F_A) > SR^2(F_B), \tag{16}$$

and thus Barillas and Shanken (2017) show that the test assets Π are irrelevant for model comparison, and it is sufficient to compare models in terms of squared Sharpe ratio, which measures the ability of factor models to span the investment opportunity set.

In the absence of trading costs or in the presence of proportional transaction costs, the efficient frontier is a straight line in the mean–standard-deviation diagram, as shown in Propositions 1 and 2. Thus, the portfolios in the efficient frontier that maximize the investor's mean–variance utility are equivalent to those that maximize the Sharpe ratio. In contrast, in the presence of price-impact costs the efficient frontier is strictly concave and hence the portfolios that maximize mean–variance utility are not equivalent to those that maximize Sharpe ratio. Thus, model comparison via the squared Sharpe ratio criterion as in (15) is no longer consistent with the optimal choices of investors that determine the asset-pricing equilibrium. To address this issue, we propose measuring model misspecification in terms of mean–variance utility net of price-impact costs. Thus, applying the logic of Barillas and Shanken (2017), model A is better than model B if

$$\mathrm{MVU}^\gamma([\Pi, F_A, F_B]) - \mathrm{MVU}^\gamma(F_A) < \mathrm{MVU}^\gamma([\Pi, F_A, F_B]) - \mathrm{MVU}^\gamma(F_B), \tag{17}$$

where $\mathrm{MVU}^\gamma(x)$ is the maximum mean–variance utility net of price-impact costs of an investor with absolute risk aversion γ who has access to the assets in x.[8] Therefore, we have that model A is better than model B if

$$\mathrm{MVU}^\gamma(F_A) > \mathrm{MVU}^\gamma(F_B), \tag{18}$$

[8] Note that the same argument can be made for investor utility functions other than the mean–variance utility that we consider for our empirical work.

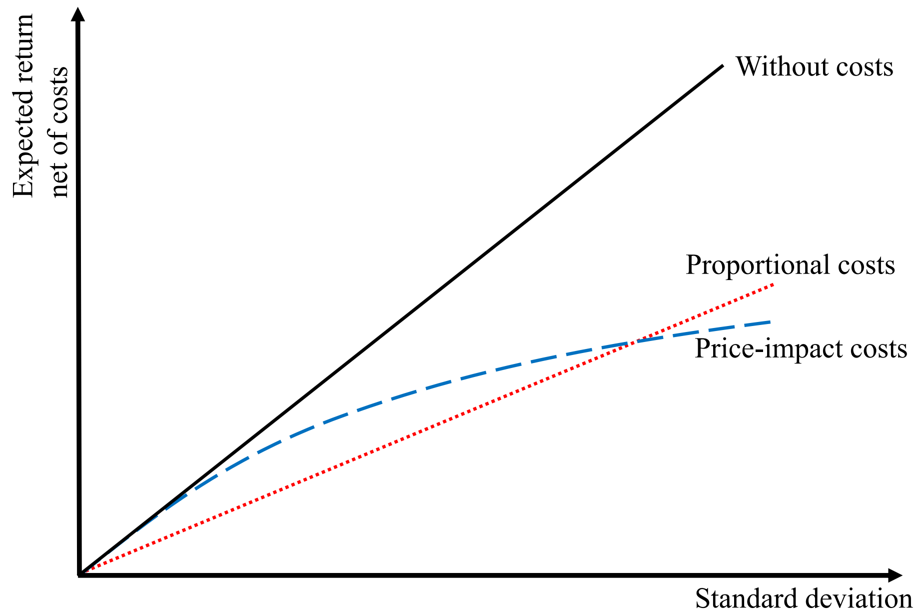

Fig. 2. Efficient frontiers for different trading-cost functions.
This figure illustrates the efficient frontiers of a factor model in the presence of different trading-cost functions. The black solid, red dotted, and blue dashed lines depict the efficient frontiers in the absence of trading costs, the presence of proportional costs, and the presence of price-impact costs, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

which shows that test assets are irrelevant *also* when comparing factor models in terms of mean–variance utility net of price-impact costs. Consequently, the best model is the one whose factors generate the highest mean–variance utility net of price-impact costs, and thus, is best at spanning the achievable investment opportunity set.

2.6. Relation between mean–variance utility and alpha

In the absence of trading costs, the squared Sharpe ratio criterion proposed by Barillas and Shanken (2017) to compare factor models is closely related to the traditional alpha criterion. In particular, Gibbons et al. (1989) show that a quadratic form of the alpha measures the increase in the squared Sharpe ratio that an investor can achieve by having access to the test assets, in addition to the factors in the model. In this section, we show that the mean–variance utility net of price-impact cost criterion that we propose is also closely related to the alpha *net of price impact*. To do this, in the following proposition, which we prove and discuss in Appendix B, we generalize the result by Gibbons et al. (1989) to the case with quadratic price-impact costs.

Proposition 4. *Consider an investor with absolute risk aversion γ who faces the quadratic price-impact costs defined in (12). Then, the increase in the mean–variance utility net of price-impact costs of the investor when she has access to a set of test assets R in addition to the factors F in a model is:*

$$MVU^{\gamma}([F,R]) - MVU^{\gamma}(F) = \left(\alpha^{net}\right)^{\top} H_{\gamma}^{-1} \alpha^{net}, \tag{19}$$

where H_{γ} is a positive-definite matrix that depends on the investor's absolute risk aversion, and α^{net} is the net alpha of the test assets with respect to the factors in the model:

$$\alpha^{net} = \underbrace{\alpha}_{\substack{\text{gross}\\ \text{alpha}}} - \underbrace{\left(\Lambda_{R,F} - \beta^{\top}\Lambda_{F,F}\right)\theta^{*}}_{\substack{\text{price-impact}\\ \text{adjustment}}}, \tag{20}$$

where α and β are the intercept and slope obtained from an OLS regression of the test asset returns on the factors in the model, θ^* is the investor's mean–variance portfolio of the factors in the model, $\Lambda_{F,F} = E[(\tilde{X}_t^F)^{\top} D_t \tilde{X}_t^F]$ is the expected price-impact matrix for the factors in the model, and $\Lambda_{R,F} = E[(\tilde{X}_t^R)^{\top} D_t \tilde{X}_t^F]$ is the expected price-impact matrix for the test assets when the investor is also holding the factors in the model.

A couple of comments are in order. First, Appendix B.2 shows that for the case with no trading costs, Proposition 4 implies the result in equation (23) of Gibbons et al. (1989), which shows that in the absence of trading costs the increase in the squared Sharpe ratio of the investor when she has access to the test assets in addition to the factors in the model is a quadratic form of the gross alpha.

Second, Appendix B.3 shows that the net alpha (α^{net}) defined in (20) is the incremental return net of price-impact costs that an investor with absolute risk-aversion γ can achieve by making a marginal investment in the test assets when she is already holding the mean–variance portfolio of the factors in the model. In other words, the net alpha is a generalization of the traditional alpha to the case with price-impact costs.

2.7. Model performance and absolute risk aversion

Note that the net alpha α^{net} depends on the investor's absolute risk aversion via her mean–variance portfolio $\theta^* = (\Sigma_{F,F} + \Lambda_{F,F}/\gamma)^{-1}\mu_F/\gamma$, where μ_F and Σ_F are the mean and covariance matrix of the factors in the model. Thus, in general the net alpha is different for each investor. Moreover, the matrix H_{γ} also depends on γ. Consequently, Eqs. (19) and (20) in Proposition 4 show that the relative performance of two factor models in the presence of price impact may depend in general on the investor's absolute risk aversion, which determines the importance of portfolio risk relative to the average portfolio return net of price-impact costs.

This is illustrated in Fig. 1 in the introduction, which depicts the investment opportunity set spanned by two different factor models A and B, where the factors in model A generate a higher Sharpe ratio in the absence of trading costs, but also generate higher price-impact costs as the amount traded increases, compared to the factors in model B. Then, model B is better at spanning the investment opportunities of large investors with low absolute risk aversion, while model A is better at spanning the investment opportunities of investors with high absolute risk aversion. This is because investors with low absolute risk aversion are willing to take on larger investment positions to increase their mean return at the expense of higher return variance. However, by increasing their positions, they also increase the amount they trade, and thus, face higher price-impact costs. Consequently, model B is better at spanning their investment opportunities because its factors generate lower price-impact costs.

3. Pairwise model comparison

We now develop a formal statistical methodology to compare two factor models in the presence of price-impact costs. In Section 3.1, we derive two asymptotic distributions for the difference in mean–variance utility net of price-impact costs of two factor models. Section 3.2 describes how these two asymptotic distributions can be used to compare two factors models for the cases where they are nested, non-nested without overlapping factors, and non-nested with overlapping factors. Finally, in Section 3.3, we develop a closed-form expression for the variance of the asymptotic distribution and use it to study how the statistical properties of factor models affect the power of our proposed pairwise model comparison test.

3.1. Two asymptotic distributions

We assume price-impact costs are quadratic as in (12). Also, for simplicity we make Assumption 3.1, but it can be relaxed by adjusting the variance of the asymptotic distribution.

Assumption 3.1. The factor returns F_t, the matrix $\Sigma_t = (F_t-\mu)(F_t-\mu)^\top$, and the price-impact matrix Λ_t in Eq. (11) are serially uncorrelated.

We now derive two asymptotic distributions in Propositions 5 and 6 for the difference between the sample mean–variance utilities net of price-impact costs of two factor models.

Proposition 5. *Let Assumptions 2.1–2.3 and 3.1 hold. Then, the asymptotic distribution of the sample estimator for the mean–variance utility net of price-impact costs in (14) is*

$$\sqrt{T}(\widehat{MVU^\gamma} - MVU^\gamma) \overset{A}{\sim} N(0, \frac{E[h_t^2]}{4\gamma^2}), \tag{21}$$

provided that $E[h_t^2] > 0$, *where*

$$h_t = 2\mu^\top(\Sigma + \Lambda/\gamma)^{-1}(F_t - \mu) - \mu^\top(\Sigma + \Lambda/\gamma)^{-1}(\Sigma_t + \Lambda_t/\gamma)(\Sigma + \Lambda/\gamma)^{-1}\mu + \mu^\top(\Sigma + \Lambda/\gamma)^{-1}\mu. \tag{22}$$

In addition, the asymptotic distribution of the difference between the sample mean–variance utilities net of price-impact costs of two factor models A and B is

$$\sqrt{T}([\widehat{MVU^\gamma_A} - \widehat{MVU^\gamma_B}] - [MVU^\gamma_A - MVU^\gamma_B]) \overset{A}{\sim} N(0, \frac{E\left[(h_{t,A} - h_{t,B})^2\right]}{4\gamma^2}), \tag{23}$$

provided that $E\left[(h_{t,A}-h_{t,B})^2\right] > 0$, *where* $h_{t,A}$ *and* $h_{t,B}$ *are given by Eq. (22) applied to models A and B.*

A couple of comments are in order. First, Proposition 5 generalizes the analysis of Barillas et al. (2020), who provide an asymptotic distribution for the difference in squared Sharpe ratios of two models in the absence of costs. Second, Proposition 5 shows that the distribution in (23) can be used to compare two factor models provided that the variance of the asymptotic distribution is strictly greater than zero. However, the variance is zero under the null hypothesis, $\mathrm{MVU}^\gamma_A = \mathrm{MVU}^\gamma_B$, in two cases: (i) when model A nests model B and the additional factors of model A are redundant and (ii) when models A and B overlap (share common factors) and the non-overlapping factors of both models are redundant.[9] For these two cases, one cannot use Proposition 5 to test whether two models generate the same mean–variance utility net of price-impact costs. However, Proposition 6 provides an alternative asymptotic distribution that can be used in these two cases. Section 3.2 discusses how Propositions 5 and 6 can be used to compare nested or non-nested factor models.

Proposition 6. *Let Assumptions 2.1–2.3 and 3.1 hold. Consider two nested models A and B containing factors* $F_A = [F_1, F_2]$ *and* $F_B = F_1$, *where* F_1 *and* F_2 *contain* K_1 *and* K_2 *mutually exclusive factors. Partition the matrix* $\Sigma_A + \Lambda_A/\gamma$ *as*

$$\Sigma_A + \Lambda_A/\gamma = \begin{bmatrix} \Sigma_{11} + \Lambda_{11}/\gamma & \Sigma_{12} + \Lambda_{12}/\gamma \\ \Sigma_{21} + \Lambda_{21}/\gamma & \Sigma_{22} + \Lambda_{22}/\gamma \end{bmatrix},$$

where $\Sigma_{22}+\Lambda_{22}/\gamma \in \mathbb{R}^{K_2\times K_2}$. *Then, under the null hypothesis that* $MVU^\gamma_A = MVU^\gamma_B$, *the asymptotic distribution of the difference between the sample mean–variance utilities net of price-impact costs of the two models A and B is given by*

$$T(\widehat{MVU^\gamma_A} - \widehat{MVU^\gamma_B}) \overset{A}{\sim} \sum_{i=1}^{K_2} \xi_i x_i, \tag{24}$$

where x_i *for* $i = 1, \dots, K_2$ *are independent chi-square random variables with one degree of freedom, and* ξ_i *for* $i = 1, \dots, K_2$ *are the eigenvalues of matrix*

$$\frac{E[l_t l_t^\top]_{22} W}{2\gamma}, \tag{25}$$

where

$$W = (\Sigma_{22} + \Lambda_{22}/\gamma) - (\Sigma_{21} + \Lambda_{21}/\gamma)(\Sigma_{11} + \Lambda_{11}/\gamma)^{-1}(\Sigma_{12} + \Lambda_{12}/\gamma) \quad \text{and} \tag{26}$$

$$l_t = (\Sigma_A + \Lambda_A/\gamma)^{-1} F_{A,t} - (\Sigma_A + \Lambda_A/\gamma)^{-1}(\Sigma_{A,t} + \Lambda_{A,t}/\gamma)(\Sigma_A + \Lambda_A/\gamma)^{-1}\mu_A. \tag{27}$$

This proposition is related to proposition 2 of Kan and Robotti (2009), which compares nested factor models in terms of their Hansen–Jagannathan distance in the absence of trading costs. We extend their result to compare nested factor models in terms of mean–variance utility net of price-impact costs.[10]

3.2. Comparing models with any nesting structure

We now show how to compare two factor models with any nesting structure using Propositions 5 and 6. We consider three cases: (i) non-nested factor models without overlapping factors, (ii) nested factor models, and (iii) non-nested factor models with overlapping factors.

When models A and B are non-nested and have no overlapping factors, the variance of the asymptotic distribution in (23) is strictly greater than zero. Therefore, one can directly apply Proposition 5 and reject the null hypothesis $\mathrm{MVU}^\gamma_A = \mathrm{MVU}^\gamma_B$ when $\sqrt{T}(\widehat{\mathrm{MVU}^\gamma_A} - \widehat{\mathrm{MVU}^\gamma_B})$ is greater (less) than, for instance, the 97.5th (2.5th) percentile of the distribution on the right-hand side of (23).

However, as explained in the previous section, one cannot use Proposition 5 to compare nested factor models because under the null hypothesis where the extra factors of the larger model are redundant, the variance of the distribution in (23) is zero. Instead, we use Proposition 6 and reject the null hypothesis $\mathrm{MVU}^\gamma_A = \mathrm{MVU}^\gamma_B$ when $T(\widehat{\mathrm{MVU}^\gamma_A} - \widehat{\mathrm{MVU}^\gamma_B})$ is greater than, for instance, the 95th percentile of

[9] To see that the asymptotic variance in Proposition 5 is zero under the null hypothesis for these two cases, note that the mean–variance portfolios for factor models A and B are identical in these two cases. Thus, when applying Eq. (22) to models A and B, we have that $E[(h_{t,A} - h_{t,B})^2] = 0$. Barillas et al. (2020) also deal with these two cases when comparing models in the absence of transaction costs.

[10] Note that to compare nested models in the *absence* of trading costs, one can either use Proposition 6 with $\Lambda = 0$, or run time-series regressions of the additional factors of the larger model on the common factors of the two models, and apply the GRS test to assess whether the non-common factors contribute to expanding the investment opportunity set of the common factors. Section IA.1 of the Internet Appendix compares these two approaches in the absence of trading costs and shows that the two methods deliver similar results.

the distribution on the right-hand side of (24), in which case the larger model A performs significantly better than the smaller model B.

Comparing two non-nested models with overlapping factors is more complicated because, as Barillas et al. (2020) point out, the null hypothesis may hold in two ways: (i) the two models have the same mean–variance utility net of price-impact costs as the common factors of the two models, and (ii) the two models have identical mean–variance utility net of price-impact costs that is *higher* than that of their common factors. In the first case, the extra factors of both models are redundant and Proposition 5 cannot be applied because the variance of the distribution in (23) is zero. Thus, we test whether the null hypothesis holds using Proposition 6 where we define a nesting model containing all factors of models A and B, and a nested model containing only the common factors of models A and B. If this test does not reject the null, the two models are statistically indistinguishable in the first way. However, if this test rejects its null, then the null hypothesis does not hold in the first way, but it may still hold in the second way, which can be tested using Proposition 5 because in this case the asymptotic variance in (23) is greater than zero.

Finally, to empirically characterize the asymptotic distribution in Proposition 5, one can replace h_t in (22) with its sample counterpart, $\hat{h}_t$, which guarantees that $\sum_{t=1}^{T}(\hat{h}_{t,A}-\hat{h}_{t,B})^2/T$ is a consistent estimator of $E[(h_{t,A}-h_{t,B})^2]$. Similarly, to empirically characterize the asymptotic distribution in Proposition 6, one can replace $E[l_t l_t^\top]_{22}$ and W in (25) with their sample counterparts to obtain consistent estimators of the eigenvalues ξ_i in (24).

3.3. The asymptotic variance

In this section, we derive closed-form expressions for the asymptotic variance in Proposition 5, and use them to study how the statistical properties of factor models affect the power of our proposed pairwise model comparison test. Our main finding is that it is easier to reject the null hypothesis that the mean–variance utilities net of price-impact costs of two models are equal not only when the mean–variance portfolio returns of the two models are positively correlated as shown by Barillas et al. (2020) for the case without trading costs, but also when the mean–variance portfolio return of each model is highly correlated with the rebalancing trades of the portfolio of the other model, and when the rebalancing trades of the two portfolios are highly correlated.

Let the *matrix of scaled rebalancing trades* at time t be

$$\tilde{Y}_t = \frac{D_t^{1/2}\tilde{X}_t}{\sqrt{\gamma}} \in \mathbb{R}^{N\times K},$$

where D_t, defined in (9), is the diagonal matrix whose nth element, $d_{n,t}$, is the price-impact parameter of stock n at time t. Note that

$$E[\tilde{Y}_t^\top \tilde{Y}_t] = E\left[\frac{\tilde{X}_t^\top D_t \tilde{X}_t}{\gamma}\right] = \frac{\Lambda}{\gamma}.$$

Let $\tilde{y}_{n,t} \in \mathbb{R}^K$ be the nth row of matrix $\tilde{Y}_t$, which contains the scaled rebalancing trades on the nth stock required by the K factors at time t.

For simplicity, we assume that the factor returns F_t and $\tilde{y}_{n,t}$ are normally distributed, but similar results can be derived for the case where they are elliptically distributed.

Assumption 3.2. The factor returns F_t follow a multivariate normal distribution with mean μ and covariance matrix Σ. In addition, each vector $\tilde{y}_{n,t}$ for $n = 1,\dots,N$ follows a multivariate normal distribution with zero mean and covariance matrix Λ_n/γ.

The following proposition gives the closed-form expressions for the asymptotic variance of the sample mean–variance utility net of price-impact costs of a factor model and that of the difference between the sample mean–variance utilities of two models. For notational simplicity, we define $u_t = \mu^\top(\Sigma+\Lambda/\gamma)^{-1}F_t \in \mathbb{R}$, which is proportional to the mean–variance factor portfolio return at time t, and $v_{n,t} = \mu^\top(\Sigma+\Lambda/\gamma)^{-1}\tilde{y}_{n,t} \in \mathbb{R}$, which is proportional to the total scaled rebalancing trade on stock n at time t of the mean–variance factor portfolio.

Proposition 7. *Let Assumptions 2.1–2.3, 3.1, and 3.2 hold. Then,*

$$E[h_t^2] = 4\mathrm{var}(u_t) + 2\left[\mathrm{var}(u_t)\right]^2 + 4\sum_{n=1}^{N}\left[\mathrm{cov}(u_t, v_{n,t})\right]^2 + 2\sum_{i=1}^{N}\sum_{j=1}^{N}\left[\mathrm{cov}(v_{i,t}, v_{j,t})\right]^2. \tag{28}$$

Moreover, given two factor models A and B, we have

$$E[(h_{t,A}-h_{t,B})^2] = E[h_{t,A}^2] + E[h_{t,B}^2] - 2E[h_{t,A}h_{t,B}], \tag{29}$$

where $E[h_{t,A}^2]$ and $E[h_{t,B}^2]$ are given by applying (28) to models A and B, and

$$E[h_{t,A}h_{t,B}] = 4\mathrm{cov}(u_t^A, u_t^B) + 2\left[\mathrm{cov}(u_t^A, u_t^B)\right]^2 + 2\sum_{i=1}^{N}\sum_{j=1}^{N}\left[\mathrm{cov}(v_{i,t}^A, v_{j,t}^B)\right]^2 + 2\sum_{n=1}^{N}\left(\left[\mathrm{cov}(u_t^A, v_{n,t}^B)\right]^2 + \left[\mathrm{cov}(u_t^B, v_{n,t}^A)\right]^2\right). \tag{30}$$

Eq. (28) shows that the asymptotic variance of the sample mean–variance utility net of price-impact costs increases not only with the variance of the mean–variance portfolio returns, $\mathrm{var}(u_t)$, as shown by Barillas et al. (2020) for the case without trading costs, but also with the squared covariance between the mean–variance portfolio returns and the rebalancing trades for each stock in the mean–variance portfolio, $[\mathrm{cov}(u_t, v_{n,t})]^2$, and with the squared covariance between the rebalancing trades for different firms in the mean–variance portfolio, $[\mathrm{cov}(v_{i,t}, v_{j,t})]^2$.

Eqs. (29) and (30) show that, similar to the case without costs, the asymptotic variance of the difference between the sample mean–variance utilities net of price-impact costs of two models increases with the variance of the mean–variance portfolio return for each of the two models, and decreases with the covariance of the mean–variance portfolio returns for the two models, provided that $\mathrm{cov}(u_t^A, u_t^B) > -1$. In addition, the asymptotic variance of the difference decreases with the squared covariance between the mean–variance portfolio return of one model and the rebalancing trades for each stock in the mean–variance portfolio of the other model, $[\mathrm{cov}(u_t^A, v_{n,t}^B)]^2$ and $[\mathrm{cov}(u_t^B, v_{n,t}^A)]^2$, and with the squared covariance between the rebalancing trades of the stocks in the mean–variance portfolios of the two models, $[\mathrm{cov}(v_{i,t}^A, v_{j,t}^B)]^2$.

Consequently, it is easier to reject the null hypothesis that the mean–variance utilities net of price-impact costs of two models are equal when the mean–variance portfolio returns of the two models are highly positively correlated, the mean–variance portfolio return of each model is highly correlated with the rebalancing trades of the portfolio of the other model, and the rebalancing trades of the two portfolios are highly correlated.[11]

4. Multiple model comparison

In the previous section, we proposed a statistical test to compare the performance of *two* factor models in the presence of price-impact costs, generalizing the pairwise tests of Kan and Robotti (2009) and Barillas et al. (2020). Empirically, one may also want to test whether a benchmark model delivers a mean–variance utility net of price-impact costs at least as high as that of any of the multiple models in an alternative set. In this section, we describe how to use the multiple model comparison test of Barillas et al. (2020) to compare multiple models in the presence of price-impact costs. The exposition closely follows that of Barillas et al. (2020) and we refer the interested reader to Kan et al. (2013) for the derivation of the test. Sections 4.1 and 4.2 deal with the cases of nested and non-nested models, respectively.

[11] To estimate the asymptotic variances, one can plug the sample estimators $\hat{\mu}$, $\hat{\Sigma}$, and $\hat{\Lambda}_n$ into the closed-form expressions in Proposition 7.

4.1. Nested models

Suppose we have a benchmark model and a set of alternative models that nest the benchmark model. To conduct multiple model comparison, we first form a "large" model that includes all factors of the alternative models. In this case, it is straightforward to show that the benchmark model delivers a mean–variance utility net of price-impact costs at least as high as that of any of the alternative models if and only if it delivers the same mean–variance utility net of price-impact costs as the large model. Because the large model nests the benchmark model, we can use Proposition 6 in Section 3.1 to test the null hypothesis that the benchmark model and the large model deliver the same mean–variance utility net of price-impact costs. If the test rejects the null hypothesis, then we conclude that the benchmark model is dominated by one or more of the alternative models. Otherwise, we fail to reject the hypothesis that the benchmark model performs as well as any of the alternative models.

4.2. Non-nested models

To compare multiple non-nested models, we use the multiple non-nested model comparison test of Kan et al. (2013), which is based on the multivariate inequality test of Wolak (1987, 1989). Suppose there is a benchmark model 0 and p alternative models indexed from $i = 1$ to p. Let MVU_i^γ be the mean–variance utility net of price-impact costs of model i and let $\delta_i = \text{MVU}_0^\gamma - \text{MVU}_i^\gamma$ from $i = 1$ to p. We would like to test the null hypothesis that the benchmark model 0 delivers a mean–variance utility net of price-impact costs at least as high as that of the p alternative models:

$$H_0 : \delta \geq \mathbf{0}_p, \tag{31}$$

where $\delta = (\delta_1, \ldots, \delta_p)$. Thus, the alternative hypothesis is that there is at least another model with a higher mean–variance utility net of price-impact costs than the benchmark model 0.

Let $\hat{\delta}$ be the sample counterpart of δ and assume $\hat{\delta}$ is asymptotically normally distributed with mean δ and covariance matrix $\Sigma_{\hat{\delta}}$.[12] Moreover, let $\tilde{\delta}$ be the minimizer to the following quadratic program:

$$\min_{\delta} \quad \left(\hat{\delta} - \delta\right)^\top \hat{\Sigma}_{\hat{\delta}}^{-1} \left(\hat{\delta} - \delta\right), \tag{32}$$

$$\text{s.t.} \quad \delta \geq \mathbf{0}_p, \tag{33}$$

where $\hat{\Sigma}_{\hat{\delta}}$ is a consistent estimator of $\Sigma_{\hat{\delta}}$.[13] Then, the likelihood-ratio test statistic of the null hypothesis in (31) is

$$LR = T \left(\hat{\delta} - \tilde{\delta}\right)^\top \hat{\Sigma}_{\hat{\delta}}^{-1} \left(\hat{\delta} - \tilde{\delta}\right). \tag{34}$$

A large value of LR suggests that the null hypothesis does not hold. Wolak (1989) characterizes the asymptotic distribution of LR under the null hypothesis and the Internet Appendix of Kan et al. (2013) proposes numerical methods to calculate the p-value of LR.

Following Kan et al. (2013) and Barillas et al. (2020), when comparing a benchmark model with a set of alternative models, we first remove any alternative models i that are nested by the benchmark model because $\delta_i \geq 0$ holds by construction for these models. We also remove any alternative models that nest the benchmark model because the asymptotic normality assumption for $\hat{\delta}$ does not hold under the null hypothesis $\delta_i = 0$, and instead for these models we conduct a separate multiple nested model comparison test as described in Section 4.1. Finally, if any of the remaining alternative models is nested by another remaining alternative model, we remove the "nested" model because its mean–variance utility net of price-impact costs cannot be higher than that of the "nesting" model. We then perform the multiple non-nested model comparison test on the remaining models using the likelihood-ratio statistic in (34). We reject the null that the benchmark model generates a mean–variance utility net of price-impact costs at least as high as any of the alternative models if the p-value from either the nested or non-nested model comparison test is significant.

5. Empirical results: mean–variance utility

In this section, we use the statistical tests of Sections 3 and 4 to compare the empirical performance of ten factor models in terms of mean–variance utility net of price-impact costs. Section 5.1 lists the ten factor models we consider and describes the data we use to construct their factors. Section 5.2 describes how we estimate the price-impact costs incurred by different stocks. Section 5.3 reports factor summary statistics. Section 5.4 reports the pairwise model comparison results based on the statistical test developed in Section 3. Section 5.5 reports the multiple model comparison results based on the statistical test of Barillas et al. (2020) described in Section 4. Finally, as a robustness check, Section 5.6 compares the out-of-sample performance of the different models using the bootstrap approach of Fama and French (2018).

5.1. Factor models and data

Table 1 lists the ten factor models we consider, ordered by increasing number of factors. We consider nine *low-dimensional* factor models: the CAPM model of Sharpe (1964) and Lintner (1965), the four-factor model of Hou et al. (2015), HXZ4, the four-factor model of Fama and French (1993) and Carhart (1997), FFC4, the five-factor model of Hou et al. (2021), HMXZ5, the five-factor model of Fama and French (2015), FF5, the six-factor model of Fama and French (2018), FF6, and the six-factor model of Barillas and Shanken (2018), BS6. In addition, Fama and French (2018) and Ball et al. (2016) show that using cash-based operating profitability instead of accrual-based operating profitability can improve model performance, and thus, following Detzel et al. (2023) we consider versions of the five- and six-factor Fama–French models constructed using cash operating profitability, FF5c and FF6c. Finally, to evaluate the trading-diversification benefits from combining a large number of factors, we consider a *high-dimensional* factor model containing the 20 factors that DeMiguel et al. (2020, section IA.2) find statistically significant in the presence of price-impact costs, DMNU20.

To construct the factors associated with the aforementioned ten factor models, we download data for the 31 tradable factors listed in Table 2. Our sample spans the period from January 1980 to December 2020. We replicate the construction of twelve factors included in prominent low-dimensional asset-pricing models using the same procedure as in the papers that originally proposed them. In particular, we construct the market (MKT), size (SMB), value (HML), profitability (RMW), and investment (CMA) factors of Fama and French (2015) as well as a monthly value (HMLm) factor rebalanced monthly instead of annually and a profitability (RMWc) factor based on cash-based operating profitability instead of accrual-based operating profitability, the momentum (UMD) factor of Carhart (1997), the profitability (ROE), investment (IA), and size (ME) factors of Hou et al. (2015), and the expected growth (EG) factor of Hou et al. (2021).[14] Finally, we construct the

[12] Using arguments similar to those in the online appendix of Kan et al. (2013), it is straightforward to show that a sufficient condition for the asymptotic normality of $\hat{\delta}$ is that the mean–variance portfolios of the $p + 1$ models are all different.

[13] It is straightforward to show, by generalizing Proposition 5 to the case with multiple models, that a consistent estimator of $\Sigma_{\hat{\delta}}$ is the sample counterpart of $E[\delta_{h,t}\delta_{h,t}^\top]/(4\gamma^2)$, where $\delta_{h,t}^\top = \left(h_{t,0} - h_{t,1}, h_{t,0} - h_{t,2}, \ldots, h_{t,0} - h_{t,p}\right)$ and $h_{t,i}$ is given by Eq. (22) applied to model i.

[14] Section IA.7 of the Internet Appendix shows that our empirical findings are robust to considering factors constructed using the *banding* transaction-cost mitigation strategy used by Detzel et al. (2023). We are grateful to Detzel et al. (2023) for allowing us to use their replication code.

Table 1
List of factor models considered.
This table lists the factor models we consider, ordered by increasing number of factors. The first column gives the acronym of the model, the second column the number of factors in the model (K), the third and fourth columns the authors who proposed the model, and the date and journal of publication, respectively. The last column lists the acronyms of the factors in the model.

Acronym	K	Authors	Date, journal	Factor acronyms
CAPM	1	Sharpe and Lintner	1964, JF and 1965, JF	MKT
HXZ4	4	Hou, Xue & Zhang	2015, RFS	MKT, ROE, IA, ME
FFC4	4	Fama & French and Carhart	1993, JFE and 1997, JF	MKT, SMB, HML, UMD
HMXZ5	5	Hou, Mo, Xue & Zhang	2021, RF	MKT, ROE, IA, ME, EG
FF5	5	Fama & French	2015, JFE	MKT, SMB, HML, RMW, CMA
FF5c	5	Fama & French	2015, JFE	MKT, SMB, HML, RMWc, CMA
FF6	6	Fama & French	2018, JFE	MKT, SMB, HML, RMW, CMA, UMD
FF6c	6	Fama & French	2018, JFE	MKT, SMB, HML, RMWc, CMA, UMD
BS6	6	Barillas & Shanken	2018, JF	MKT, SMB, HMLm, ROE, IA, UMD
DMNU20	20	DeMiguel, Martin-Utrera, Nogales & Uppal	2020, RFS	MKT, agr, cashpr, chatoia, chcsho, convind, egr, ep, gma, idiovol, indmom, ps, rd_mve, retvol, roaq, sgr, std_turn, sue, turn, zerotrade

19 factors (other than the market) in the high-dimensional DMNU20 model as the returns on value-weighted long-short portfolios obtained from single sorts on 19 firm characteristics. In particular, we start with a database that contains every firm traded on the NYSE, AMEX, and NASDAQ exchanges. We then drop firms with negative book-to-market or with market capitalization below the 20th cross-sectional percentile as in Brandt et al. (2009) and DeMiguel et al. (2020). We then rank stocks at the beginning of every month based on a particular firm characteristic and build a long value-weighted portfolio of stocks whose characteristic is above the 70th percentile and a short value-weighted portfolio of stocks below the 30th percentile.[15]

5.2. Estimating price-impact cost parameters

We explain in this section how we estimate the price-impact parameter (Kyle's lambda) of the nth stock in month t, the quantity $d_{n,t}$ defined below Eq. (9), which is required for the computation of the price-impact costs incurred by the factors. Following Novy-Marx and Velikov (2016), we use the Trade and Quote (TAQ) data from December 2003 to December 2020 to estimate $d_{n,t}$ by regressing daily stock returns on daily order flows:

$$r_{n,\tau} = \alpha_n + d_{n,t}\text{OrderFlow}_{n,\tau} + \varepsilon_{n,\tau}, \tag{35}$$

where $r_{n,\tau}$ is the return of stock n on day τ and $\text{OrderFlow}_{n,\tau}$ is the order flow of stock n on day τ.[16] For the earlier part of our sample from January 1980 to December 2003, we estimate $d_{n,t}$ following DeMiguel et al. (2020, appendix IA.2) who rely on the empirical results of Novy-Marx and Velikov (2016) based on Trade and Quote (TAQ) data.[17]

As in Korajczyk and Sadka (2004) and Novy-Marx and Velikov (2016), we express all quantities, including the optimal factor portfolio θ, in terms of market capitalization at the end of our sample (December 2020). To make price-impact costs comparable over the entire estimation window from 1980 to 2020, we scale the price-impact parameter, $d_{n,t}$, by multiplying it with the ratio of the *aggregate* market capitalization in month t to that in December 2020.[18]

5.3. Factor summary statistics

Table 3 reports summary statistics for the 31 factors listed in Table 2. The first column gives the acronym of the factor. The second and third columns give the average monthly *gross* factor return and its t-statistic. The fourth, fifth, and sixth columns give the average monthly *net-of-price-impact-costs* factor return, its t-statistic, and the factor's monthly price-impact cost (PIC), when one invests one billion dollars in each leg of the factor. The seventh and eighth columns give the factor's monthly turnover (TO) and the factor's capacity, defined as the total investment that can be allocated to each leg of the factor before price-impact costs erode its gross return entirely. The ninth column reports the average of the monthly trade-weighted market capitalization, and the last column reports the average of the trade-weighted market capitalization at the end of June. Average returns and turnovers are reported in percentage. Price-impact costs are reported in basis points. Investment positions, capacity, and trade-weighted market capitalization are reported in terms of market capitalization at the end of our sample, which spans January 1980 to December 2020.

Consistent with the findings of Detzel et al. (2023), we find that, among the factors constructed from double and triple sorts, factors that are rebalanced monthly (HMLm, UMD, ROE, IA, ME, EG) have monthly turnovers ranging between 20.20% to 52.38% that are much higher than those of factors that are rebalanced annually (SMB, HML, RMW, RMWc, CMA), which range between 7.54% and 14.44%. As a

[15] Section IA.4 of the Internet Appendix shows that our findings are robust to considering an alternative DMNU20 model whose factors (other than the market) are constructed using double sorts (instead of single sorts) and without dropping firms with market capitalization below the 20th cross-sectional percentile.

[16] Order flow is defined as the dollar value of the difference between the buyer- and seller-initiated trades. The daily order flow data is obtained from the Millisecond Trade and Quote (TAQ) dataset, and the trades are signed using the algorithm of Lee and Ready (1991). The price-impact parameters are estimated monthly using daily observations from the previous year.

[17] Specifically, Novy-Marx and Velikov (2016) show that the R-squared of a cross-sectional regression of log price-impact parameters on log market capitalization is 70% and the slope is statistically indistinguishable from minus one. This suggests that a good approximation to the cross-sectional variation of price-impact parameters is to assume they are inversely proportional to market capitalization. Therefore, for months between December 1993 and December 2003, we use figure 4 in Novy-Marx and Velikov (2016) to recover estimates of the cross-sectional average price elasticity of stock supply, defined as the product of the estimated price impact per dollar traded and market capitalization, and estimate the price-impact parameter of stock n in month t as the ratio of the average price elasticity of supply in month t to the market capitalization of stock n in month t. In addition, we estimate the price-impact parameter of stock n in month t from January 1980 to December 1993 as the ratio of 6.5 to the market capitalization of stock n in month t, where 6.5 is the time-series average of the average cross-sectional price elasticity.

[18] Section IA.8 of the Internet Appendix shows the relative performance of the ten factor models is robust to estimating the price-impact parameters using the results of Frazzini et al. (2018).

Table 2
List of factors considered.
This table lists the 31 factors we consider. Panel A lists twelve factors that replicate those in the prominent low-dimensional asset-pricing models listed in Table 1. Except for the market factor, each of these factors is constructed as value-weighted portfolios obtained from double or triple sorts on firm characteristics. Panel B lists 19 factors constructed using value-weighted long-short portfolios from single sorts on characteristics that, together with the market factor, compose the 20-factor model of DeMiguel et al. (2020). The first column gives the factor index, the second column gives the factor's definition, the third column gives the acronym and the fourth and fifth columns give the authors who analyzed them and the publication date and journal.

#	Definition	Acronym	Author(s)	Date and journal
Panel A: Market factor and factors constructed from double and triple sorts				
1	Market: value-weighted portfolio of all tradable stocks in US equity markets.	MKT	Sharpe	1964, JF
2	Small-minus-big: value-neutral portfolio that is long stocks with small market capitalization and is short stocks with large market capitalization.	SMB	Fama & French	1993, JFE
3	High-minus-low: size-neutral portfolio that is long stocks with high book-to-market ratios and is short stocks with low book-to-market ratios.	HML	Fama & French	1993, JFE
4	High-minus-low (monthly): size-neutral portfolio that is long stocks with high book-to-market ratios and is short stocks with low book-to-market ratios, rebalanced monthly instead of annually.	HMLm	Fama & French	1993, JFE
5	Robust-minus-weak: size-neutral portfolio that is long stocks with high accruals-based operating profitability and is short stocks with low accruals-based operating profitability.	RMW	Fama & French	2015, JFE
6	Robust-minus-weak (cash based): size-neutral portfolio that is long stocks with high cash-based operating profitability and is short stocks with low cash-based operating profitability.	RMWc	Fama & French	2015, JFE
7	Conservative-minus-aggressive: size-neutral portfolio that is long stocks with high investment and is short stocks with low investment.	CMA	Fama & French	2015, JFE
8	Momentum: portfolio that is long stocks with the largest return over the past 12 months, skipping the last month, and is short stocks with the lowest return over the past 12 months, skipping the last month.	UMD	Carhart	1997, JF
9	Return on equity: portfolio that is long stocks with high profitability and is short stocks with low profitability.	ROE	Hou, Xue & Zhang	2015, RFS
10	Investment: portfolio that is long stocks with high investment and is short stocks with low investment.	IA	Hou, Xue & Zhang	2015, RFS
11	Size: portfolio that is long stocks with low market capitalization and is short stocks with large market capitalization.	ME	Hou, Xue & Zhang	2015, RFS
12	Expected growth: portfolio that is long stocks with high expected investment growth and is short stocks with low expected investment growth.	EG	Hou, Mo, Xue & Zhang	2021, RF
Panel B: Factors constructed from single sorts				
13	Asset growth: Annual percent change in total assets	agr	Cooper, Gulen & Schill	2008, JF
14	Cash productivity: Fiscal year-end market capitalization plus long term debt minus total assets divided by cash and equivalents	cashpr	Chandrashekar & Rao	2009, WP
15	Industry adjusted change in asset turnover: 2-digit SIC fiscal-year mean adjusted change in sales divided by average total assets	chatoia	Soliman	2008, TAR
16	Change in shares outstanding: Annual percent change in shares outstanding	chcsho	Pontiff & Woodgate	2008, JF
17	Convertible debt indicator: An indicator equal to 1 if company has convertible debt obligations	convind	Valta	2016, JFQA
18	Change in common shareholder equity: Annual percent change in equity book value	egr	Richardson, Sloan, Soliman & Tuna	2005, JAE
19	Earnings to price: Annual income before extraordinary items divided by end of fiscal year market cap	ep	Basu	1977, JF
20	Gross profitability: Revenues minus cost of goods sold divided by lagged total assets	gma	Novy-Marx	2013, JFE
21	Idiosyncratic return volatility: Standard deviation of residuals of weekly returns on weekly equal weighted market returns for 3 years prior to month-end	idiovol	Ali, Hwang & Trombley	2003, JFE
22	Industry momentum: Equal weighted average industry 12-month returns	indmom	Moskowitz & Grinblatt	1999, JF
23	Financial-statements score: Sum of 9 indicator variables to form fundamental health score	ps	Piotroski	2000, JAR
24	R&D to market cap: R&D expense divided by end-of-fiscal-year market cap	rd_mve	Guo, Lev & Shi	2006, JBFA
25	Return volatility: Standard deviation of daily returns from month $t-1$	retvol	Ang, Hodrick, Xing & Zhang	2006, JF
26	Return on assets: Income before extraordinary items divided by one quarter lagged total assets	roaq	Balakrishnan, Bartov & Faurel	2010, JAE
27	Annual sales growth: Annual percent change in sales	sgr	Lakonishok, Shleifer & Vishny	1994, JF
28	Volatility of share turnover: Monthly standard deviation of daily share turnover	std_turn	Chordia, Subrahmanyan & Anshuman	2001, JFE
29	Unexpected quarterly earnings: Unexpected quarterly earnings divided by fiscal-quarter-end market cap. Unexpected earnings is I/B/E/S actual earnings minus median forecasted earnings if available, else it is the seasonally differenced quarterly earnings before extraordinary items from Compustat quarterly file	sue	Rendelman, Jones & Latane	1982, JFE
30	Share turnover: Average monthly trading volume for most recent 3 months scaled by number of shares outstanding in current month	turn	Datar, Naik & Radcliffe	1998, JFM
31	Zero trading days: Turnover weighted number of zero trading days for most recent month	zerotrade	Liu	2006, JFE

result, the annually rebalanced factors have, on average, lower price-impact costs and higher capacity than the monthly rebalanced factors. In particular, the average monthly price-impact cost and capacity of the five annually rebalanced factors are 1.66 basis points and 15.66 billion dollars, respectively, while those of the six monthly rebalanced factors are 5.65 basis points and 6.96 billion dollars, respectively. However, we also find that the relative performance of factors in terms of turnover is different from that in terms of price-impact cost. For instance, while UMD is the factor with the highest turnover, ROE is the factor with the highest price-impact cost. Specifically, for the case where one invests one billion dollars in each leg of the factors, the monthly price-impact cost of UMD is around eight basis points, but that of ROE is almost eleven basis points.

To understand the difference in the relative performance of factors in terms of turnover and price-impact cost, the last two columns of Table 3 report the average trade-weighted market capitalization (in billions of dollars) of the different factors listed in Table 2. In particular, for each factor we compute the monthly trade-weighted

Table 3
Factor summary statistics.
This table reports several summary statistics of the factors listed in Table 2. The first column gives the acronym of the factor. The second and third columns give the average monthly *gross* factor return and its *t*-statistic. The fourth, fifth, and sixth columns give the average monthly *net-of-price-impact-costs* factor return, its *t*-statistic, and the factor's monthly price-impact cost (PIC), when one invests one billion dollars in each leg of the factor. The seventh and eighth columns give the factor's monthly turnover (TO) and the factor's capacity, defined as the total investment that can be allocated to each leg of the factor before price-impact costs erode its gross return entirely. The ninth column reports the average of the monthly trade-weighted market capitalization, and the last column reports the average of the trade-weighted market capitalization at the end of June. Average returns and turnovers are reported in percentage. Price-impact costs are reported in basis points. Investment positions, capacity, and trade-weighted market capitalization are reported in terms of market capitalization at the end of our sample, which spans January 1980 to December 2020.

Factor	Gross returns (%)		Net returns (%)		Costs (bp), turnover (%), and capacity ($B)			Trade-weighted market cap ($B)	
	Average	*t*-stat	Average	*t*-stat	PIC	TO	Capacity	Monthly	June
Panel A: Market and factors constructed from double and triple sorts									
MKT	0.70	3.46	0.70	3.46	0.01	2.15	–	160.97	146.74
SMB	0.10	0.73	0.09	0.69	0.53	7.54	18.74	69.70	43.86
HML	0.14	1.04	0.13	0.93	1.51	9.94	9.36	71.83	46.49
HMLm	0.18	1.05	0.13	0.79	4.32	21.20	4.07	56.83	46.84
RMW	0.37	3.60	0.36	3.45	1.50	10.11	24.84	64.63	54.43
RMWc	0.38	4.39	0.36	4.14	2.08	12.10	18.10	75.02	65.92
CMA	0.19	2.37	0.17	2.03	2.66	14.44	7.24	79.87	82.70
UMD	0.58	2.89	0.49	2.49	8.04	52.38	7.15	86.85	83.59
ROE	0.53	4.48	0.42	3.55	10.88	37.92	4.90	59.60	52.57
IA	0.27	3.19	0.22	2.56	5.31	26.10	5.18	70.66	64.91
ME	0.15	1.12	0.13	0.98	1.84	20.20	8.05	69.92	56.38
EG	0.44	4.55	0.40	4.19	3.53	21.49	12.38	57.16	52.18
Panel B: Factors constructed from single sorts									
agr	0.12	0.97	0.11	0.88	1.11	15.03	10.52	142.94	164.99
cashpr	0.03	0.24	0.03	0.21	0.35	8.14	9.72	164.96	168.72
chatoia	0.15	1.79	0.14	1.64	1.30	16.33	11.83	166.35	178.34
chcsho	0.25	2.37	0.24	2.27	1.05	14.06	23.68	164.33	186.20
convind	0.09	0.95	0.08	0.91	0.38	6.27	23.11	148.39	137.27
egr	0.16	1.40	0.14	1.30	1.07	14.94	14.50	145.78	167.80
ep	0.21	1.24	0.20	1.16	1.40	14.83	14.98	123.87	145.20
gma	0.20	1.44	0.19	1.43	0.20	6.67	99.17	169.59	135.04
idiovol	0.01	0.03	−0.01	−0.05	2.21	11.40	0.40	77.05	78.75
indmom	0.17	1.07	0.14	0.91	2.52	41.45	6.71	179.78	173.57
ps	0.09	0.96	0.07	0.77	1.80	16.85	5.15	152.00	179.35
rd_mve	0.55	2.83	0.52	2.71	2.32	11.66	23.50	162.90	187.75
retvol	0.26	0.99	0.09	0.32	17.35	83.83	1.49	105.13	89.57
roaq	0.18	1.17	0.15	0.98	2.85	24.74	6.26	111.30	100.76
sgr	0.13	0.99	0.12	0.91	1.12	15.28	11.85	157.74	179.76
std_turn	0.16	0.83	0.09	0.46	7.17	77.89	2.28	120.65	102.09
sue	0.20	1.86	0.12	1.16	7.46	45.62	2.65	112.75	88.56
turn	0.07	0.36	0.05	0.24	2.41	29.27	2.97	171.17	160.43
zerotrade	0.22	1.14	0.13	0.68	9.03	62.61	2.48	176.73	167.21

market capitalization of the stocks traded by the factor and report the time-series average. Table 3 shows that, as expected, the factor that trades in the largest, and thus, most liquid stocks is the market (MKT). Specifically, the average firm traded by the MKT factor has a market capitalization of 160.97 billion dollars. In contrast, the average market capitalization of the stocks traded by the return on equity (ROE) and the investment (IA) factors of Hou et al. (2015) is only 59.60 and 70.66 billion dollars, respectively, and that of the expected growth (EG) factor of Hou et al. (2021) is the smallest at 57.16 billion dollars. The low market capitalization of the average stock traded by the ROE factor explains why the price-impact cost of ROE is much higher than that of UMD, even though UMD has a substantially higher turnover. Finally, Panel B shows that the trade-weighted market capitalization of the factors constructed from single sorts is substantially larger than that of the factors obtained from double and triple sorts. This is because the factors obtained from single sorts assign a much lower weight to small stocks compared to factors obtained from double or triple sorts, which use market capitalization as one of the sorting variables. As a result, although the monthly turnover of the factors from single sorts is comparable to that of the factors from double and triple sorts, their price-impact costs are generally lower.

In summary, the results in this section show that the price-impact costs incurred by the different factors depend not only on the turnover required to rebalance them, which was highlighted by Detzel et al. (2023) as an important driver in the context of *proportional* transaction costs, but also on the size and liquidity of the stocks traded.

5.4. Pairwise model comparison

We now compare the ten models listed in Table 1 in terms of mean–variance utility net of price-impact costs using the pairwise model comparison test developed in Section 3. Like Gârleanu and Pedersen (2013), we consider a base case with an absolute risk-aversion parameter of $\gamma = 10^{-9}$, which corresponds to an investor with a relative risk-aversion parameter of five and an endowment of five billion dollars. In addition, we consider cases where the investor has the same relative risk-aversion parameter, but her endowment is ten times larger or smaller than in the base case; that is, when $\gamma = 10^{-10}$ or $\gamma = 10^{-8}$. For a constant relative risk-aversion level, a lower absolute risk-aversion parameter implies a larger investor, and thus price-impact costs play a more important role in the investor's mean–variance utility.

Note that the CAPM model is nested in all other models, HXZ4 is nested in HMXZ5, FFC4 and FF5 are nested in FF6, and FFC4 and FF5c are nested in FF6c. Thus, we use Proposition 6 to compare these nested models. Also, all models have one common market factor. Therefore, following our discussion in Section 3.2, we compare non-nested models with overlapping factors in two stages. First, we use Proposition 6 to test whether a model with all factors in the two models yields the same utility net of price-impact costs as a model with only the common factors. If the test does not reject the null, then the two models are statistically indistinguishable.[19] If the test rejects the null, we then

[19] For every non-nested pairwise model comparison, we report the p-value for the second-stage test provided that the first-stage test is significant at

Table 4
Pairwise model comparison without price-impact costs.
This table reports the p-values for all pairwise model comparisons in the absence of trading costs, that is, when the expected price-impact matrix $\Lambda = 0$. Panel A reports the scaled sample mean–variance utility of each of the ten factor models in the absence of trading costs. Panel B reports the p-value for the difference in mean–variance utility between each row and column model. The p-value is computed using Proposition 5 when the row and column models overlap and Proposition 6 when the row model is nested in the column model.

Panel A: Mean–variance utilities without trading costs

2γMVU$^{\gamma}$	CAPM	HXZ4	FFC4	HMXZ5	FF5	FF5c	FF6	FF6c	BS6	DMNU20
	0.0217	0.1316	0.0586	0.2129	0.1129	0.1508	0.1277	0.1603	0.1535	0.1247

Panel B: p-values

	CAPM	HXZ4	FFC4	HMXZ5	FF5	FF5c	FF6	FF6c	BS6	DMNU20
CAPM		0.000	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000
HXZ4			0.003	0.000	0.245	0.278	0.433	0.167	0.082	0.434
FFC4				0.000	0.039	0.005	0.000	0.000	0.000	0.044
HMXZ5					0.004	0.055	0.006	0.071	0.061	0.028
FF5						0.022	0.027	0.023	0.105	0.381
FF5c							0.160	0.067	0.470	0.270
FF6								0.035	0.163	0.470
FF6c									0.415	0.203
BS6										0.260

implement the second-stage test that uses Proposition 5 to compare the two models.[20]

To understand how price-impact costs affect the relative performance of the ten factor models, we first compare their performance in the *absence* of price-impact costs. Panel A in Table 4 reports the scaled sample mean–variance utility of each model in the absence of price-impact costs and Panel B reports the p-values for all pairwise comparisons. To facilitate the comparison of utilities across different values of the absolute risk-aversion parameter, we report all mean–variance utilities scaled by multiplying them by 2γ. Also, to simplify notation, herein we use the symbol MVU$^{\gamma}$ instead of $\widehat{\text{MVU}}^{\gamma}$ to refer to the sample mean–variance utility. Our main observation is that in the absence of price-impact costs, HMXZ5 is the best model. To see this, note first that the mean–variance utility delivered by the HMXZ5 model is the highest among all ten models we consider. Moreover, the differences between the utility generated by the factors in the HMXZ5 model and those derived from other more parsimonious models with fewer factors (CAPM, HXZ4, and FFC4) are statistically significant at the 1% level. Finally, the mean–variance utility of the HMXZ5 model is also significantly higher than those of the FF5, FF6, and DMNU20 models at the 5% confidence level, and those of the FF5c, FF6c, and BS6 models at the 10% confidence level. Overall, we conclude that the HMXZ5 model best spans the investment opportunity set in the absence of costs.

Table 5 reports the performance of the ten models in the *presence* of price impact for our base case with absolute risk aversion $\gamma = 10^{-9}$. Our main finding is that price-impact costs change the relative performance of the different models: While HMXZ5 was the best model in the absence of trading costs, its mean–variance utility is lower than those of the FF5, FF5c, FF6, FF6c, BS6, and DMNU20 models in the presence of price-impact costs. Moreover, the difference between the mean–variance utilities net of price-impact costs of the DMNU20 and HMXZ5 models is significant at the 10% confidence level, and the p-value for the difference between the utilities of the FF6c and HMXZ5 models is just above 10%, at 12.3%.

The explanation for the poor performance of the HMXZ5 model in the presence of price impact is not only that its investment and profitability factors require higher turnover than those of the Fama–French models, but also that they require trading stocks with smaller market capitalization, and thus, less liquid. For instance, the seventh column of Table 3 shows that the monthly turnovers of the ROE and IA factors included in HMXZ5 are 37.92 and 26.10%, while those of the RMWc and CMA factors included in FF5c and FF6c are only 12.10 and 14.44%. Also, the ninth column of Table 3 shows that the trade-weighted market capitalizations of the stocks of the ROE and IA factors are 59.60 and 70.66 billion dollars, while those of the RMWc and CMA factors are 75.02 and 79.87 billion dollars. Thus, the investment and profitability factors of the HMXZ5 model require trading smaller (less liquid) stocks compared to the RMWc and CMA factors in FF5c and FF6c. Similarly, the expected growth factor (EG) in HMXZ5 has a monthly turnover of 21.49% and a trade-weighted market capitalization of only 57.16 billion dollars, which is the second lowest of all 31 factors listed in Table 2. Thus, investing in the factors in the HMXZ5 model incurs high price-impact costs.

Table 5 also shows that the cash-profitability six-factor Fama–French model (FF6c) is the best low-dimensional model in the presence of price-impact costs because it significantly outperforms the CAPM, HXZ4, FFC4, and FF5c models at the 1% confidence level, the BS6 model at the 5% level, and the FF5 model at the 10% level.[21] Also,

the 5% confidence level. Out of all non-nested pairwise comparisons in the manuscript, we find that there is a single comparison (HXZ4 versus BS6 for $\gamma = 10^{-10}$ in Table IA.19) for which the first-stage test fails to reject the null hypothesis at the 5% level. For this comparison, we report the p-value for the first-stage test 0.061, which is less significant than that for the second-stage test 0.028. Out of all non-nested pairwise comparisons in the Internet Appendix, we find that there is a single comparison (HXZ4 versus BS6 for $\gamma = 10^{-10}$ in Table IA.19) for which the first-stage test fails to reject the null hypothesis at the 5% level. For this comparison, we report the p-value for the first-stage test 0.056, which is less significant than that for the second-stage test 0.036.

[20] In detail, the p-values are computed as follows. Assume without loss of generality that the sample mean–variance utilities net of price-impact costs for models A and B satisfy $\text{MVU}^{\gamma}_{A} > \text{MVU}^{\gamma}_{B}$ (to simplify the exposition we drop the hat symbol for sample mean–variance utilities in this footnote). Then, we compute the p-value as the integral over the values greater than $\text{MVU}^{\gamma}_{A} - \text{MVU}^{\gamma}_{B}$ of the probability density function in (23) if the two models are non-nested and of the probability density function in (24) if they are nested. Like Barillas et al. (2020), we use the bias-adjusted values of MVU^{γ}_{A} and MVU^{γ}_{B} when comparing non-nested factor models using Proposition 5. This is because the asymptotic distribution in (23) fails to capture the finite-sample bias in estimates of mean–variance utility. Section IA.2 of the Internet Appendix details the procedure we use to adjust the bias. However, when using Proposition 6 to compare nested factor models, we use the raw values of MVU^{γ}_{A} and MVU^{γ}_{B} because the asymptotic distribution in (24) adequately captures the finite-sample bias of the sample mean–variance utility. This is also demonstrated by the bootstrap experiments in Section IA.3 of the Internet Appendix.

[21] This result is counterintuitive because the FF6c model is obtained by adding the momentum factor to FF5c and trading the momentum factor incurs high price-impact costs as illustrated in the sixth column of Table 3. The explanation for this is twofold. First, as shown in the second column of Table 3,

Table 5
Pairwise model comparison with price-impact costs.
This table reports the p-values for all pairwise model comparisons in the presence of price-impact costs for the base case with absolute risk-aversion parameter $\gamma = 10^{-9}$. Panel A reports the scaled sample mean–variance utility net of price-impact costs of each of the ten factor models. Panel B reports the p-value for the difference in mean–variance utility net of price-impact costs between each row and column model. The p-value is computed using Proposition 5 when the row and column models overlap and Proposition 6 when the row model is nested in the column model.

Panel A: Mean–variance utilities net of price-impact costs

$2\gamma \text{MVU}^{\gamma}$	CAPM	HXZ4	FFC4	HMXZ5	FF5	FF5c	FF6	FF6c	BS6	DMNU20
	0.0216	0.0402	0.0407	0.0571	0.0596	0.0627	0.0730	0.0756	0.0594	0.0868

Panel B: p-values

	CAPM	HXZ4	FFC4	HMXZ5	FF5	FF5c	FF6	FF6c	BS6	DMNU20
CAPM		0.001	0.003	0.000	0.000	0.000	0.000	0.000	0.000	0.000
HXZ4			0.479	0.000	0.057	0.025	0.010	0.004	0.008	0.015
FFC4				0.117	0.089	0.045	0.000	0.000	0.000	0.021
HMXZ5					0.440	0.351	0.180	0.123	0.429	0.081
FF5						0.332	0.002	0.078	0.496	0.106
FF5c							0.175	0.002	0.392	0.139
FF6								0.357	0.095	0.257
FF6c									0.050	0.302
BS6										0.104

Table 6
Pairwise model comparison with price-impact costs for $\gamma = 10^{-10}$.
This table reports the p-values for all pairwise model comparisons in the presence of price-impact costs for the case with low absolute risk aversion $\gamma = 10^{-10}$. Panel A reports the scaled sample mean–variance utility net of price-impact costs of each of the ten factor models. Panel B reports the p-value for the difference in mean–variance utility net of price-impact costs between each row and column model. The p-value is computed using Proposition 5 when the row and column models overlap and Proposition 6 when the row model is nested in the column model.

Panel A: Mean–variance utilities net of price-impact costs

$2\gamma \text{MVU}^{\gamma}$	CAPM	HXZ4	FFC4	HMXZ5	FF5	FF5c	FF6	FF6c	BS6	DMNU20
	0.0215	0.0237	0.0246	0.0261	0.0297	0.0281	0.0323	0.0307	0.0263	0.0488

Panel B: p-values

	CAPM	HXZ4	FFC4	HMXZ5	FF5	FF5c	FF6	FF6c	BS6	DMNU20
CAPM		0.021	0.048	0.001	0.004	0.007	0.001	0.002	0.014	0.000
HXZ4			0.327	0.000	0.069	0.073	0.019	0.015	0.061	0.007
FFC4				0.296	0.066	0.070	0.000	0.000	0.073	0.008
HMXZ5					0.221	0.290	0.095	0.109	0.475	0.012
FF5						0.185	0.000	0.331	0.179	0.020
FF5c							0.036	0.000	0.252	0.016
FF6								0.181	0.042	0.032
FF6c									0.036	0.025
BS6										0.011

although the high-dimensional DMNU20 model achieves higher mean–variance utility than the FF6c model, the difference between the utilities of the FF6c and DMNU20 models is not significant (p-value of 30.2%), and thus FF6c is preferable because of its parsimony. Overall, the pairwise comparisons show that while the HMXZ5 model was the best at spanning the investment opportunity set in the absence of costs, the FF6c model is best at spanning the achievable investment opportunity set in the presence of price-impact costs.

The finding that DMNU20 does *not* significantly outperform FF6c for the base case with $\gamma = 10^{-9}$ is surprising because DeMiguel et al. (2020) find that in the presence of trading costs, high-dimensional models are likely to perform well because the benefits from trading diversification increase with the number of factors. To shed light on this result, we consider a case with a lower absolute risk aversion $\gamma = 10^{-10}$, which corresponds to an investor with the same relative risk aversion as in our base case, but with an endowment ten times higher than that in the base case. For this level of absolute risk aversion, price-impact costs should play a more important role and we expect that the high-dimensional DMNU20 model should dominate other factor models because of the benefits from trading diversification. Table 6 confirms this intuition: the high-dimensional model DMNU20 significantly outperforms every low-dimensional model at the 5% confidence level.[22] Among the low-dimensional models, FF6 achieves the highest mean–variance utility net of price-impact costs, with the p-value for the difference being significant at the 10% level for every low-dimensional model except FF6c.

Finally, Table 7 reports the results for the case with a higher absolute risk-aversion parameter, $\gamma = 10^{-8}$, which corresponds to an investor with the same relative risk aversion as in the base case, but

the momentum factor achieves the second-highest average gross return among the 31 factors we consider. Second, even though momentum is expensive when traded in isolation, it is much cheaper to trade in *combination* with the other five factors in the FF6c model because of trading diversification (DeMiguel et al., 2020). Indeed, Section IA.6 of the Internet Appendix reports summary statistics of the optimal portfolio weights for the different factor models, and shows that trading diversification greatly reduces the price-impact cost incurred by FF6c.

[22] An alternative explanation for the favorable performance of the DMNU20 model for the case with low absolute risk aversion is that DMNU20 includes factors constructed using single sorts after dropping stocks with market capitalization below the 20th cross-sectional percentile, which, as shown in Table 3, trade stocks with higher market capitalization than the Fama–French factors obtained using double sorts on the entire cross section of stocks. However, Section IA.4 of the Internet Appendix shows that the results in Table 6 are robust to considering a DMNU20 model whose factors are constructed using *double* sorts and including all stocks.

Table 7
Pairwise model comparison with price-impact costs for $\gamma = 10^{-8}$.
This table reports the p-values for all pairwise model comparisons in the presence of price-impact costs for the case with high absolute risk aversion $\gamma = 10^{-8}$. Panel A reports the scaled sample mean–variance utility net of price-impact costs of each of the ten factor models. Panel B reports the p-value for the difference in mean–variance utility net of price-impact costs between each row and column model. The p-value is computed using Proposition 5 when the row and column models overlap and Proposition 6 when the row model is nested in the column model.

Panel A: Mean–variance utilities net of price-impact costs

	CAPM	HXZ4	FFC4	HMXZ5	FF5	FF5c	FF6	FF6c	BS6	DMNU20
$2\gamma\text{MVU}^{\gamma}$	0.0217	0.0958	0.0556	0.1517	0.0980	0.1255	0.1145	0.1382	0.1249	0.1086

Panel B: p-values

	CAPM	HXZ4	FFC4	HMXZ5	FF5	FF5c	FF6	FF6c	BS6	DMNU20
CAPM		0.000	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000
HXZ4			0.015	0.000	0.456	0.108	0.154	0.035	0.027	0.336
FFC4				0.001	0.053	0.008	0.000	0.000	0.000	0.047
HMXZ5					0.027	0.180	0.082	0.313	0.187	0.087
FF5						0.034	0.015	0.027	0.145	0.364
FF5c							0.295	0.031	0.492	0.308
FF6								0.052	0.297	0.425
FF6c									0.290	0.192
BS6										0.319

with an endowment ten times *lower* than that in the base case. For this case, price-impact costs are less important, and thus, we expect the relative performance of the different models to be similar to that in the *absence* of costs. Table 7 confirms this intuition: the HMXZ5 model delivers the highest mean–variance utility net of price-impact costs among all ten models. Moreover, the differences between the utility generated by the factors in the HMXZ5 model and those derived from other more parsimonious models with fewer factors (CAPM, HXZ4, and FFC4) are statistically significant at the 1% level. Finally, the mean–variance utility of the HMXZ5 model is also significantly higher than that of the FF5 model at the 5% level and the FF6 and DMNU20 models at the 10% level. Overall, HMXZ5 is the best model just as in the case without trading costs.

In summary, the pairwise model comparisons show that accounting for price-impact costs results in a more nuanced comparison of the various factor models we consider—HMXZ5, FF6c, and DMNU20 are the best models at spanning the achievable investment opportunities of investors with high, medium, and low absolute risk aversion, respectively.

5.5. Multiple model comparison

In the previous section, we discussed the results from all pairwise comparisons of the ten models. In this section, we discuss the results from the multiple model comparisons obtained using the test of Barillas et al. (2020) described in Section 4. As mentioned above, the CAPM, HXZ4, FFC4, FF5, and FF5c models are nested by at least another model. Therefore, as discussed in Section 4, for these models we perform both nested and non-nested multiple model comparison tests, and reject the null that the benchmark model generates a mean–variance utility net of price-impact costs at least as high as that of any other model if the p-value from either the nested or non-nested test is significant.

Table 8 reports the p-values for the multiple model comparisons for the base case with price-impact costs and absolute risk aversion $\gamma = 10^{-9}$. The first and second columns report the acronym of the benchmark model and its scaled sample mean–variance utility net of price-impact costs ($2\gamma\text{MVU}^{\gamma}$). The third, fourth, and fifth columns report the number of alternative models considered in the multiple *non-nested* model comparison (n), the value of the likelihood-ratio statistic (LR), and the p-value for the multiple non-nested model comparison. The sixth and seventh columns report the number of alternative models considered in the multiple *nested* model comparison (m) and the p-value for the multiple nested model comparison.

Table 8 confirms the finding from the pairwise model comparisons that the cash-profitability six-factor Fama–French model (FF6c) performs relatively well for our base case with absolute risk-aversion $\gamma = 10^{-9}$. For instance, FF6c has the lowest likelihood-ratio statistic (0.27) among the low-dimensional models, and the multiple model comparison test cannot reject the null that FF6c achieves a mean–variance utility net of price-impact costs at least as high as that of any other model. To see this, note that the FF6c model is not nested by any other model and the p-value for the multiple non-nested model comparison is 0.668.

Table 8
Multiple model comparison with price-impact costs.
This table reports the p-values for the multiple model comparisons in the presence of price-impact costs for the base case with absolute risk-aversion parameter $\gamma = 10^{-9}$. The first and second columns report the acronym of the benchmark model and its scaled sample mean–variance utility net of price-impact costs ($2\gamma\text{MVU}^{\gamma}$). The third, fourth, and fifth columns report the number of alternative models considered in the multiple *non-nested* model comparison (n), the value of the likelihood-ratio statistic (LR), and the p-value for the multiple non-nested model comparison. The sixth and seventh columns report the number of alternative models considered in the multiple *nested* model comparison (m) and the p-value for the multiple nested model comparison.

Benchmark model	$2\gamma\text{MVU}^{\gamma}$	n	LR	p-value (non-nested)	m	p-value (nested)
CAPM	0.0216				9	0.000
HXZ4	0.0402	4	9.83	0.005	1	0.000
FFC4	0.0407	5	15.95	0.000	2	0.000
HMXZ5	0.0571	4	2.46	0.156		
FF5	0.0596	4	2.80	0.161	1	0.002
FF5c	0.0627	4	1.54	0.317	1	0.002
FF6	0.0730	4	0.52	0.553		
FF6c	0.0756	4	0.27	0.668		
BS6	0.0594	4	3.49	0.134		
DMNU20	0.0868	4	0.00	0.668		

Table 8 also shows that the CAPM, HXZ4, FFC4, FF5, and FF5c models are rejected by the multiple nested model comparison. However, the multiple model comparison cannot reject the HMXZ5, FF6, BS6, and DMNU20 models, all of which have p-values for the multiple non-nested model comparison that are not significant at the 10% level. This result is consistent with the pairwise model comparison results in Table 5, which show that FF6c does not significantly outperform HMXZ5, FF6, and DMNU20 at the 10% confidence level. Nonetheless, overall FF6c is the best model for our base case with $\gamma = 10^{-9}$ because of its high sample mean–variance utility net of price-impact costs, its high p-value for the multiple model comparison test, and its parsimony compared to the high-dimensional DMNU20 model.

Tables IA.2 and IA.3 in the Internet Appendix report the multiple model comparison results for the cases with lower and higher absolute risk aversion. The findings are consistent with those from the pairwise comparisons. For instance, Table IA.2 shows that the high-dimensional DMNU20 model performs relatively well for the case with low absolute risk aversion $\gamma = 10^{-10}$. In particular, the multiple non-nested model comparison test cannot reject DMNU20, with a large p-value of 0.588. Also, although the multiple model comparison test cannot reject the FF6 and FF6c models, the p-values for these two models are only slightly above 10%. The good performance of DMNU20 for the case with low absolute risk aversion is again consistent with the finding by DeMiguel et al. (2020) that the benefits from trading diversification increase with the number of factors in a model.[23] Finally, Table IA.3 shows that for the case with high absolute risk aversion $\gamma = 10^{-8}$, HMXZ5 performs relatively well. In particular, the multiple non-nested model comparison test cannot reject the HMXZ5 model, with a large p-value of 0.772. Moreover, although the multiple model comparison test cannot reject the FF6, FF6c, BS6, and DMNU20 models, HMXZ5 is the model with the smallest number of factors out of all models that are not rejected, and thus, is preferable because of its parsimony.

5.6. Out-of-sample model comparison

In the previous sections, we compared factor models using our proposed statistical tests, which address the following question: Is the mean–variance utility in the presence of price-impact costs of a model significantly higher than that of other models? As a robustness check, we now address a different question that is relevant for investment management: Are the utility gains of a superior factor model achievable out of sample? To do this, we use the out-of-sample bootstrap test used by Fama and French (2018) and Detzel et al. (2023).

This bootstrap test guarantees that disjoint sets of observations are used for the in-sample and out-of-sample calculations. For each bootstrap sample, we carry out a four-step procedure. First, for every pair of consecutive months, we randomly assign one month to the set of in-sample (IS) observations and the other to the set of out-of-sample (OOS) observations. Second, within the IS set, we bootstrap with replacement a set with the same number of observations as the original sample, and allocate the corresponding partner months to the OOS set. Third, we use the factor returns and the factor-rebalancing trades of the months in the bootstrap IS set to calculate the optimal portfolio weights of each model using Eq. (13).[24] Fourth, we apply the optimal portfolio weights from the third step to the bootstrap OOS set to obtain the OOS mean–variance utility net of price-impact costs for each factor model. We repeat these four steps 100,000 times, and thus, obtain 100,000 observations of the OOS mean–variance utility net of price-impact costs for each model. Finally, we compare models in terms of their average mean–variance utility, the frequency with which one model outperforms another model, and the frequency with which each model outperforms every other model across the bootstrap samples. This procedure not only guarantees that the IS and OOS sets for each bootstrap sample are disjoint, but also prevents the IS and OOS sets from having substantially different time-series properties because they are obtained from pairs of consecutive months.

Table 9 reports the out-of-sample performance of the ten models in the presence of price-impact costs for the base case with absolute risk aversion $\gamma = 10^{-9}$. Panel A reports the out-of-sample average scaled mean–variance utility net of price-impact costs of each factor model, Panel B reports the out-of-sample frequency with which the row model outperforms the column model, and Panel C reports the out-of-sample frequency with which each model outperforms every other model across the bootstrap samples. As expected, the average *out-of-sample* mean–variance utilities in Panel A of Table 9 are much lower than the *in-sample* utilities in Panel A of Table 5 because of estimation error. However, the out-of-sample relative performance of the various models is generally consistent with their in-sample relative performance.[25]

Note that the out-of-sample frequencies in Panel B of Table 9 are larger than the p-values based on our pairwise model comparison test in Panel B of Table 5. This is not surprising because even if a model has a significantly higher mean–variance utility than another, it may deliver a lower out-of-sample mean–variance utility in a particular bootstrap sample because of estimation error. Nonetheless, the out-of-sample bootstrap pairwise comparison results in Panel B of Table 9 are generally consistent with those from the statistical test in Panel B of Table 5. In particular, we observe that, out of sample, HMXZ5 outperforms FF6 and FF6c only on 40.2% and 35.6% of the bootstrap samples, respectively. This is consistent with the finding in Panel B of Table 5 that the FF6 and FF6c models deliver higher mean–variance utility net of price-impact costs than the HMXZ5 model. In addition, FF6c outperforms the CAPM, HXZ4, FFC4, FF5c, and BS6 models on around 85%, 83%, 94%, 78%, and 78% of the bootstrap samples, respectively, which is consistent with the finding in Panel B of Table 5 that the FF6c model significantly outperforms these models. Finally, FF6c outperforms DMNU20 on 72.2% of the bootstrap samples. This result is compatible with our finding in Panel B of Table 5 that FF6c and DMNU20 are statistically indistinguishable. The explanation for the poor *out-of-sample* performance of DMNU20 compared to FF6c is that estimation error affects the performance of the high-dimensional DMNU20 model more heavily than that of the low-dimensional FF6c model.

The out-of-sample bootstrap *multiple* model comparison results in Panel C of Table 9 are also consistent with those from the statistical test in Table 8. In particular, FF6c is the best model, outperforming every other model on 27.1% of the bootstrap samples, followed by the DMNU20, HMXZ5, and FF6 models, which outperform every other model on 18.5%, 18.2%, and 17.5% of the bootstrap samples, respectively.

In summary, the out-of-sample bootstrap test confirms the main finding from our statistical tests in Sections 5.4 and 5.5 that, in the base case with absolute risk-aversion parameter $\gamma = 10^{-9}$, the FF6c model emerges as the best model. Moreover, the out-of-sample test shows that the gains from using the FF6c factor model can actually be realized out of sample. Section IA.5 of the Internet Appendix shows that the findings from the out-of-sample bootstrap test are also consistent with the findings from our statistical test for the cases with lower and higher absolute risk-aversion parameters. For instance, Panel C of Table IA.7 shows that for the case with low absolute risk aversion, the high-dimensional DMNU20 model outperforms every other model on 56.1% of the bootstrap samples. Also, Panel C of Table IA.8 shows that for the case with high absolute risk aversion, the HMXZ5 model outperforms every other model on 58.7% of the bootstrap samples.

[23] Table IA.6 in Section IA.4 of the Internet Appendix shows that the results in Table IA.2 are robust to considering a DMNU20 model whose factors are obtained from *double* sorts and without dropping stocks with market capitalization below the 20th cross-sectional percentile.

[24] We estimate the vector of factor mean returns, μ, and the price-impact cost matrix, Λ, using their sample counterparts. For the covariance matrix of factor returns, Σ, we use the shrinkage estimator of Ledoit and Wolf (2004) to alleviate the impact of estimation error on the out-of-sample performance of the different models.

[25] The only exception is the DMNU20 model, whose out-of-sample relative performance is worse than its in-sample relative performance. In particular, while the average out-of-sample mean–variance utility net of price-impact costs of DMNU20 is the second lowest among all models, it delivers the highest in-sample mean–variance utility net of price-impact costs. The explanation for this is that estimation error impacts the out-of-sample performance of the high-dimensional DMNU20 model more heavily than that of the low-dimensional models.

Table 9
Bootstrap out-of-sample utility net of price-impact costs.
This table reports the out-of-sample performance of the ten models in the presence of price-impact costs for the base case with absolute risk-aversion parameter $\gamma = 10^{-9}$, using the bootstrap test of Fama and French (2018) with 100,000 bootstrap samples. Panel A reports the out-of-sample average scaled mean–variance utility net of price-impact costs of each factor model, Panel B reports the out-of-sample frequency with which the row model outperforms the column model, and Panel C reports the out-of-sample frequency with which each model outperforms every other model across the bootstrap samples.

Panel A: Average mean–variance utility net of price-impact costs

	CAPM	HXZ4	FFC4	HMXZ5	FF5	FF5c	FF6	FF6c	BS6	DMNU20
$2\gamma\text{MVU}^{\gamma}$	0.0122	0.0238	0.0173	0.0395	0.0338	0.0371	0.0435	0.0461	0.0357	0.0165

Panel B: Frequency row model outperforms column model

	CAPM	HXZ4	FFC4	HMXZ5	FF5	FF5c	FF6	FF6c	BS6	DMNU20
CAPM		0.173	0.336	0.079	0.200	0.170	0.166	0.147	0.159	0.421
HXZ4			0.593	0.101	0.271	0.225	0.197	0.171	0.202	0.513
FFC4				0.166	0.231	0.181	0.084	0.059	0.039	0.469
HMXZ5					0.554	0.495	0.402	0.356	0.558	0.652
FF5						0.375	0.209	0.225	0.466	0.614
FF5c							0.306	0.217	0.527	0.640
FF6								0.397	0.726	0.704
FF6c									0.776	0.722
BS6										0.627

Panel C: Frequency column model performs best

	CAPM	HXZ4	FFC4	HMXZ5	FF5	FF5c	FF6	FF6c	BS6	DMNU20
	0.022	0.002	0.002	0.182	0.042	0.071	0.175	0.271	0.047	0.185

6. Empirical results: anomaly spanning

In Section 5, we compared models in terms of the mean–variance utility net of price-impact costs generated by the factors in each model. As discussed in Section 2.5, this criterion is sufficient to compare two models under the tenet of Barillas and Shanken (2017) that the better model should not only span the investment opportunity set of the test assets, but also that of the factors in the other model. Nevertheless, it is also of interest to compare factor models *solely* in terms of their ability to span certain test assets. *Absent* the requirement that a factor model has to span the factors in the other model, relative model performance in terms of test-asset spanning may differ from that in terms of mean–variance utility. To see this, note that model A is better than model B at spanning the test-asset returns, Π, if

$$\text{MVU}^{\gamma}([\Pi, F_A]) - \text{MVU}^{\gamma}(F_A) < \text{MVU}^{\gamma}([\Pi, F_B]) - \text{MVU}^{\gamma}(F_B),$$

where F_A and F_B are the returns of the factors in models A and B. Because in general

$$\text{MVU}^{\gamma}([\Pi, F_A]) \neq \text{MVU}^{\gamma}([\Pi, F_B]),$$

we have that the ranking of models in terms of their ability to span the test assets may differ from that in terms of mean–variance utility.[26]

In this section, we use the statistical test developed in Proposition 6 of Section 3 to compare factors models solely in terms of their ability to span the 212 anomalies in the dataset by Chen and Zimmermann (2022) using an experiment similar to that used by Detzel et al. (2023) for the case with proportional transaction costs.[27] Figs. 3 and 4 report the results for the case without costs and for the base case with price-impact costs and absolute risk aversion $\gamma = 10^{-9}$. For each of the ten factor models in Table 1 with factor returns F_i and each anomaly with returns Π_j, we calculate the mean–variance utility (net of price-impact costs) of the factor model $\text{MVU}^{\gamma}(F_i)$ and that of the factor model augmented with the anomaly $\text{MVU}^{\gamma}([F_i, \Pi_j])$. We then calculate the relative utility improvement, $\text{MVU}^{\gamma}([F_i, \Pi_j])/\text{MVU}^{\gamma}(F_i) - 1$, and the p-value of the utility improvement using Proposition 6. Panel A in each figure illustrates the percentiles of the distribution for the relative utility improvement across the anomalies for each model, and Panel B the percentiles of the distribution for the p-value of utility improvement across the anomalies for each model. To facilitate interpretation, both panels depict the results for only five of the best-performing models: HXZ4, HMXZ5, FF6, FF6c, and DMNU20.

Fig. 3 shows that, in the absence of costs, the HMXZ5 model performs relatively well at spanning the anomalies. To see this, note that Panel A of Fig. 3 shows that for any level of relative utility improvement Δ on the vertical axis, the proportion of anomalies that generate a utility improvement smaller than Δ is larger for the HMXZ5 model than for any other model. Panel B of Fig. 3 shows that for any significance level α on the vertical axis, the proportion of anomalies that generate an α-significant improvement to mean–variance utility is smaller for the HMXZ5 model than for the FF6 and DMNU20 models. Also, comparing the HMXZ5 model to the HXZ4 and FF6c models, the three models are similar in terms of the proportion of anomalies that generate an α-significant improvement to their mean–variance utility. Overall, Fig. 3 demonstrates that the HMXZ5 model performs relatively well at anomaly spanning in the absence of trading costs, consistent with the results from comparing models in terms of mean–variance utility discussed in Table 4 of Section 3.

Fig. 4 shows that for our base case with price-impact costs and absolute risk aversion $\gamma = 10^{-9}$, the FF6c model performs relatively well at spanning the anomalies. To see this, note that Panel B of Fig. 4 shows that for any significance level α on the vertical axis, the proportion of anomalies that generate an α-significant improvement to mean–variance utility is smaller for FF6c than for any other model. Also, Panel A of Fig. 4 shows that for any level of relative utility improvement Δ on the vertical axis, the proportion of anomalies that

[26] Barillas and Shanken (2017) also show that the relative performance of factor models in terms of test-asset spanning and squared Sharpe ratio may be different in the absence of trading costs once one drops the requirement that the better model should also span the factors in the other model—see the last paragraph on page 1317 and section 1.2 of their paper. For instance, Barillas and Shanken (2017) show that although a two-factor model with the market and SMB factors has a higher squared Sharpe ratio than the CAPM model, it delivers a higher pricing error for the loser decile portfolio based on past-year returns than the CAPM model. These two results are reconciled by the fact that the CAPM model does not explain the returns of the SMB factor.

[27] We use the replication code of Detzel et al. (2023) to download data for the 212 anomalies of Chen and Zimmermann (2022) for the period from January 1980 to December 2020. For some anomalies data is not available for the entire sample from January 1980 to December 2020, and thus, we perform the statistical test on the sample of months for which we have data for the anomaly.

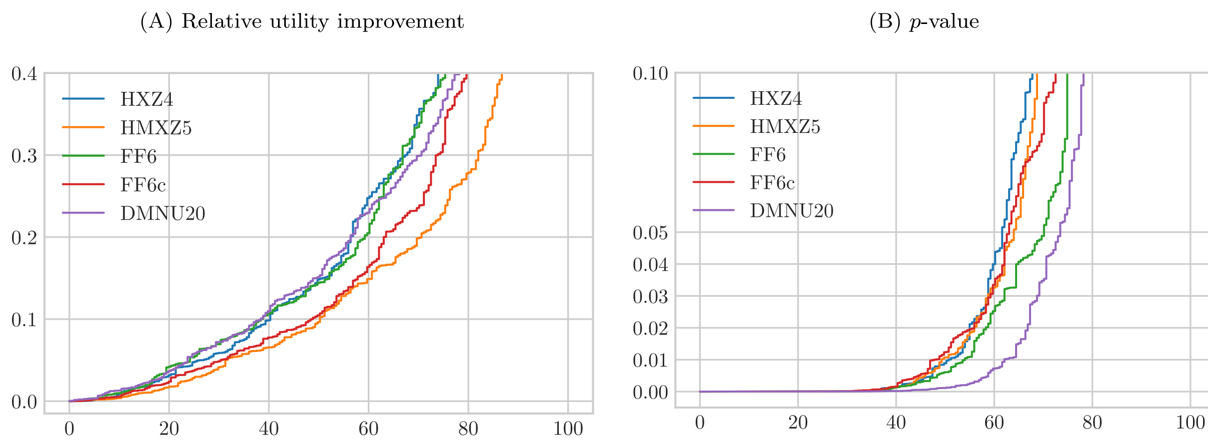

Fig. 3. Significance of relative utility improvement without costs.
This figure compares factor models in terms of their ability to span the 212 anomalies in the dataset by Chen and Zimmermann (2022) for the case without costs. For each of the ten factor models in Table 1 with factor returns F_i and each anomaly with returns Π_j, we calculate the sample mean–variance utility in the absence of costs of the factor model $\mathrm{MVU}^{\gamma}(F_i)$ and that of the factor model augmented with the anomaly $\mathrm{MVU}^{\gamma}([F_i, \Pi_j])$. We then calculate the relative utility improvement, $\mathrm{MVU}^{\gamma}([F_i, \Pi_j])/\mathrm{MVU}^{\gamma}(F_i) - 1$, and the p-value of the utility improvement using Proposition 6. Panel A illustrates the percentiles of the distribution for the relative utility improvement across the anomalies for each model, and Panel B the percentiles of the distribution for the p-value of utility improvement across the anomalies for each model. To facilitate interpretation, both panels depict the results for only five of the best-performing models (HXZ4, HMXZ5, FF6, FF6c, and DMNU20) and the y-axes for Panels A and B are truncated at 0.4 and 0.1, respectively.

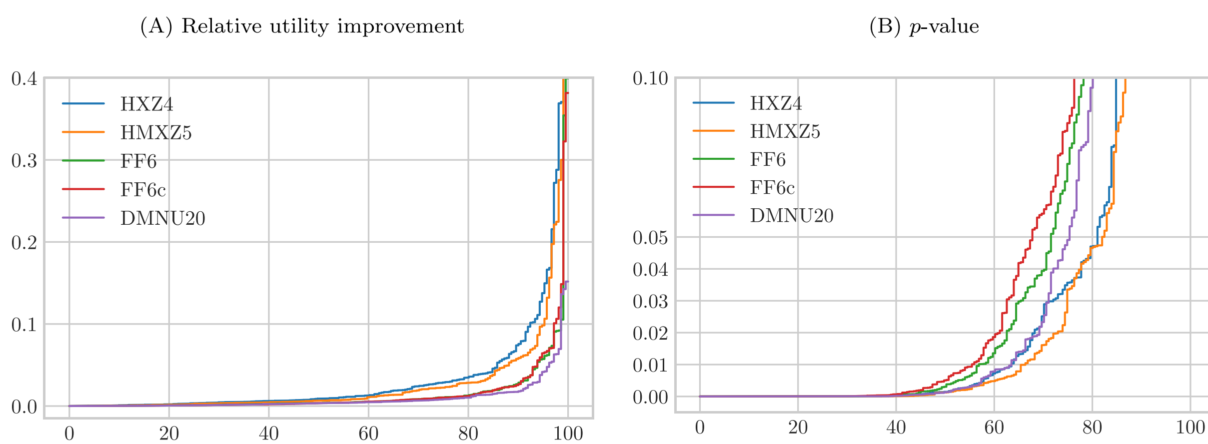

Fig. 4. Significance of relative utility improvement with price-impact costs ($\gamma = 10^{-9}$).
This figure compares factor models in terms of their ability to span the 212 anomalies in the dataset by Chen and Zimmermann (2022) for the base case with price-impact costs and absolute risk aversion $\gamma = 10^{-9}$. For each of the ten factor models in Table 1 with factor returns F_i and each anomaly with returns Π_j, we calculate the sample mean–variance utility net of price-impact costs of the factor model $\mathrm{MVU}^{\gamma}(F_i)$ and that of the factor model augmented with the anomaly $\mathrm{MVU}^{\gamma}([F_i, \Pi_j])$. We then calculate the relative utility improvement, $\mathrm{MVU}^{\gamma}([F_i, \Pi_j])/\mathrm{MVU}^{\gamma}(F_i) - 1$, and the p-value of the utility improvement using Proposition 6. Panel A illustrates the percentiles of the distribution for the relative utility improvement across the anomalies for each model, and Panel B the percentiles of the distribution for the p-value of utility improvement across the anomalies for each model. To facilitate interpretation, both panels depict the results for only five of the best-performing models (HXZ4, HMXZ5, FF6, FF6c, and DMNU20) and the y-axes for Panels A and B are truncated at 0.4 and 0.1, respectively.

generate a utility improvement smaller than Δ is larger for the FF6c model than for the HXZ4 and HMXZ5 models. Comparing DMNU20 to FF6c and FF6, the three models are similar in terms of the proportion of anomalies that generate a utility improvement smaller than Δ. Overall, Fig. 4 demonstrates that the FF6c model performs relatively well at anomaly spanning for our base case with price-impact costs and $\gamma = 10^{-9}$, consistent with the results from comparing models in terms of mean–variance utility discussed in Table 5 of Section 3.

Figs. 3 and 4 show that for the cases without costs and with price-impact costs and base case absolute risk aversion $\gamma = 10^{-9}$, the relative performance of the ten factors models in terms of anomaly spanning is similar to that in terms of mean–variance utility. However, as pointed out at the beginning of this section, this does not necessarily have to be the case because when comparing factor models in terms of anomaly spanning, we drop the requirement that the better factor model should span not only the anomalies, but also the factors in the other model. Indeed, Figures IA.4 and IA.5 in the Internet Appendix show that the relative performance of the ten models in terms of anomaly spanning is, in general, different from that in terms of mean–variance utility for the cases with low and high absolute risk aversion. For example, Figure IA.4 shows that for the case with low absolute risk aversion $\gamma = 10^{-10}$, while DMNU20 outperforms FF6 and FF6c in terms of the proportion of anomalies that generate a utility improvement smaller than Δ, FF6 and FF6c outperform DMNU20 in terms of the proportion of anomalies that generate an α-significant improvement. This result is not entirely consistent with that in Table 6 that DMNU20 is the best model in terms of mean–variance utility for the case with low absolute risk aversion $\gamma = 10^{-10}$. Also, Figure IA.5 shows that the FF6c model outperforms the HMXZ5 model for the case with high absolute risk aversion $\gamma = 10^{-8}$ in terms of anomaly spanning, a result that contrasts with that from Table 7 that HMXZ5 is the best model in terms of mean–variance utility for the case with $\gamma = 10^{-8}$. Overall, the results in this section show that the relative model performance in terms of anomaly spanning is similar, but not identical to that in terms of mean–variance utility.[28]

7. Conclusion

We show that the squared Sharpe ratio criterion is no longer sufficient to compare asset-pricing factor models in the presence of price impact because the efficient frontier spanned by a factor model is strictly concave. Instead, we propose comparing factor models in terms of the mean–variance utility net of price-impact costs generated by their factors, and develop a formal statistical test to compare two factor models for the cases when they are nested or non-nested. Importantly, we observe that the relative performance of factor models depends on the absolute risk-aversion parameter, and thus comparing factor models in the presence of price impact is a more nuanced exercise than in the absence of trading costs.

Empirically, we find that while in the absence of trading costs the five-factor model of Hou et al. (2021) outperforms other models, in the presence of price-impact costs the six-factor model of Fama and French (2018) with cash-based operating profitability performs better. We also find that the high-dimensional model of DeMiguel et al. (2020) significantly outperforms the low-dimensional models *only* for the case with low absolute risk aversion, where price impact is important enough for the trading diversification benefits of combining a large number of factors to dominate other effects such as estimation error. Thus, an implication of our work is that different benchmark factor models should be used to evaluate the performance of investment strategies designed for different investors, depending on their absolute risk aversion.

CRediT authorship contribution statement

Sicong Li: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Victor DeMiguel:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Alberto Martín-Utrera:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

Authors have nothing to declare.

Data availability

Comparing Factor Models with Price-Impact Costs (Original data) (Mendeley Data)

Acknowledgment

We thank the AQR Asset Management Institute at LBS for funding the license to use the Trade and Quote (TAQ) dataset.

[28] Section IA.9 of the Internet Appendix shows that the relative model performance in terms of two alternative criteria (out-of-sample cumulative returns net of price-impact costs and Sharpe ratio of returns net of price-impact costs) is also similar, but not identical, to that in terms of mean–variance utility net of price-impact costs.

Appendix A. Proofs of all results

This appendix contains the proofs of all novel propositions in the manuscript, except for Proposition 4, which is proven and discussed in Appendix B. For expositional purposes, Proposition 1 states a well-known result that is proven, for instance, in Campbell (2017, Section 2.2.6).

A.1. Proof of Proposition 2

Note that the proportional-trading-cost function given in Definition 1 is not convex in general and this complicates the proof, which consists of two parts. Part (i) shows that there exists a nonzero maximizer to the mean–variance problem. Part (ii) shows that the efficient frontier is a straight line.

Part (i): existence of a nonzero maximizer to mean–variance problem

We first show that for any absolute risk-aversion parameter γ, the objective function of Problem (2) has a nonzero maximizer and its maximum is strictly positive.

Denote the mean–variance utility in Problem (2) as

$$g_\gamma(\theta) = \theta^\top \mu - f(\theta) - \frac{\gamma}{2}\theta^\top \Sigma\theta.$$

By Assumption 2.3, we have that the set $S = \{\theta | \theta^\top \mu - f(\theta) \geq 0\}$ is nonempty. Moreover, by Assumption 2.2, $f(\theta)$ is continuous in S, and hence, S is compact. Furthermore, $g_\gamma(\theta)$ is also continuous in S, and thus, by the extreme-value theorem we have that there exists $\theta^* \in S$ such that $g_\gamma(\theta^*) \geq g_\gamma(\theta)$ for all $\theta \in S$. Also, by Assumption 2.3, we know that there are values of θ in S such that $g_\gamma(\theta) > 0$. Therefore, the maximum value, $g_\gamma(\theta^*)$, must be strictly positive. Consequently, $\theta^* \neq 0$ because $g_\gamma(0) = 0$.

Part (ii): the efficient frontier is a straight line

We first show by contradiction that if θ_1 is a maximizer for the case with absolute risk aversion γ, then for any $c > 0$ we have that $c\theta_1$ is a maximizer for the case with absolute risk aversion γ/c. Suppose $c\theta_1$ is not a maximizer for the case with absolute risk aversion γ/c, then there exists θ_2 such that

$$\theta_2^\top \mu - f(\theta_2) - \frac{\gamma}{2c}\theta_2^\top \Sigma\theta_2 > c\theta_1^\top \mu - f(c\theta_1) - \frac{\gamma}{2c}c\theta_1^\top \Sigma c\theta_1, \tag{A.1}$$

which is equivalent to

$$\frac{\theta_2^\top}{c}\mu - f\left(\frac{\theta_2}{c}\right) - \frac{\gamma}{2}\frac{\theta_2^\top}{c}\Sigma\frac{\theta_2}{c} > \theta_1^\top \mu - f(\theta_1) - \frac{\gamma}{2}\theta_1^\top \Sigma\theta_1, \tag{A.2}$$

which contradicts θ_1 being a maximizer for the case with absolute risk aversion γ. Note that this argument also shows that if θ_1 is a maximizer for the case with absolute risk aversion γ, then $c\theta_1$ with $c > 0$ is *not* a maximizer for the case with absolute risk aversion γ.

Next, we show by contradiction that given two maximizers θ_1 and θ_2 for the case with absolute risk aversion γ, we must have

$$\theta_1^\top \Sigma\theta_1 = \theta_2^\top \Sigma\theta_2, \tag{A.3}$$

and thus $\theta_1^\top \mu - f(\theta_1) = \theta_2^\top \mu - f(\theta_2)$. To see this, suppose without loss of generality that $\theta_2^\top \Sigma\theta_2 > \theta_1^\top \Sigma\theta_1$. Because both θ_1 and θ_2 are maximizers, by Part (i), we have $\theta_2^\top \mu - f(\theta_2) > \theta_1^\top \mu - f(\theta_1) > 0$. Thus, there exists $c > 1$, such that

$$c\theta_1^\top \mu - cf(\theta_1) = \theta_2^\top \mu - f(\theta_2). \tag{A.4}$$

Moreover, because we have shown that for $c > 0$, we have that $c\theta_1$ is not a maximizer for the case with absolute risk aversion γ, we must have that

$$(c\theta_1^\top)\Sigma(c\theta_1) > \theta_2^\top \Sigma\theta_2. \tag{A.5}$$

Thus,

$$c\theta_1^\top \mu - cf(\theta_1) - \frac{\gamma}{2c}(c\theta_1^\top)\Sigma(c\theta_1) < \theta_2^\top \mu - f(\theta_2) - \frac{\gamma}{2c}\theta_2^\top \Sigma\theta_2, \tag{A.6}$$

which contradicts $c\theta_1$ being optimal for the case with absolute risk aversion is γ/c. Therefore, $\theta_1^\top \Sigma \theta_1 = \theta_2^\top \Sigma \theta_2$ and $\theta_2^\top \mu - f(\theta_2) = \theta_1^\top \mu - f(\theta_1)$, and thus, any two maximizers θ_1 and θ_2 for the case with absolute risk aversion γ must have the same Sharpe ratio.

We now show that the efficient frontier is a straight line by showing every efficient portfolio has the same Sharpe ratio, SR_{PTC}. The Sharpe ratio of $c\theta^*$, a maximizer for the case with absolute risk aversion γ/c, is

$$\frac{c\theta^{*\top}\mu - f(c\theta^*)}{c\sqrt{\theta^{*\top}\Sigma\theta^*}} = \frac{\theta^{*\top}\mu - f(\theta^*)}{\sqrt{\theta^{*\top}\Sigma\theta^*}}, \tag{A.7}$$

which is also the Sharpe ratio of θ^*. Therefore, every efficient portfolio has the same Sharpe ratio of returns net of proportional trading costs, and thus the efficient frontier is a straight line starting at the origin of the standard deviation-mean diagram. Moreover, by Assumption 2.2 we have that $f(\theta) > 0$ for any $\theta \neq 0$, and thus,

$$SR_{PTC} = \frac{\theta^{*\top}\mu - f(\theta^*)}{\sqrt{\theta^{*\top}\Sigma\theta^*}} < \frac{\theta^{*\top}\mu}{\sqrt{\theta^{*\top}\Sigma\theta^*}} \leq SR.$$

A.2. Proof of Proposition 3

The proof consists of two parts. Part (i) provides an alternative condition to define a price-impact-cost function. Part (ii) shows that the efficient frontier is strictly concave.

Part (i): an alternative condition to define a price-impact-cost function

Definition 2 states that a price-impact-cost function must satisfy condition (8). We now show that this condition is equivalent to

$$f(c'\theta) < c' f(\theta) \quad \text{for } \theta \neq 0 \text{ and } 0 < c' < 1. \tag{A.8}$$

We first prove that (8) implies (A.8). Let $\theta' = c\theta$ with $c > 1$. Then (8) becomes

$$\frac{1}{c} f(\theta') > f\left(\frac{1}{c}\theta'\right). \tag{A.9}$$

If we define $c' = 1/c \in (0,1)$, then the previous inequality becomes

$$c' f(\theta') > f(c'\theta'), \tag{A.10}$$

which is (A.8). Using a similar argument, it is straightforward to show that (A.8) implies (8).

Part (ii): the efficient frontier is concave

Part (i) of the proof of Proposition 2 shows that for any γ, there exists a nonzero maximizer to Problem (2). Let θ^* and θ_c^* be the maximizers to Problem (2) for the cases with absolute risk aversion γ and $c\gamma$, respectively, where $0 < c < 1$. We first show that the variance of portfolio θ_c^* is greater than or equal to that of portfolio θ^*. We then show that the Sharpe ratio of θ_c^* is strictly lower than that of θ^* when the variance of θ_c^* is strictly greater than that of θ^*, and thus the efficient frontier is strictly concave.

Step 1: the variance of θ_c^* is greater than or equal to that of θ^*.

We show by contradiction that $(\theta_c^*)^\top \Sigma \theta_c^* \geq \theta^{*\top}\Sigma\theta^*$. Suppose $(\theta_c^*)^\top \Sigma \theta_c^* < \theta^{*\top}\Sigma\theta^*$. The optimality of θ^* and θ_c^* for the cases with absolute risk aversion γ and $c\gamma$, respectively, implies that

$$\theta^{*\top}\mu - f(\theta^*) - \frac{c\gamma}{2}\theta^{*\top}\Sigma\theta^* \leq (\theta_c^*)^\top\mu - f(\theta_c^*) - \frac{c\gamma}{2}(\theta_c^*)^\top\Sigma\theta_c^*, \tag{A.11}$$

$$(\theta_c^*)^\top\mu - f(\theta_c^*) - \frac{\gamma}{2}(\theta_c^*)^\top\Sigma\theta_c^* \leq \theta^{*\top}\mu - f(\theta^*) - \frac{\gamma}{2}\theta^{*\top}\Sigma\theta^*. \tag{A.12}$$

Combining these two inequalities yields

$$\begin{aligned}\frac{\gamma}{2}(\theta^{*\top}\Sigma\theta^* - (\theta_c^*)^\top\Sigma\theta_c^*) &\leq \theta^{*\top}\mu - f(\theta^*) - (\theta_c^*)^\top\mu + f(\theta_c^*) \\ &\leq \frac{c\gamma}{2}(\theta^{*\top}\Sigma\theta^* - (\theta_c^*)^\top\Sigma\theta_c^*).\end{aligned} \tag{A.13}$$

Because we have assumed that $(\theta_c^*)^\top\Sigma\theta_c^* < \theta^{*\top}\Sigma\theta^*$ and $0 < c < 1$, the leftmost term is strictly greater than the rightmost term in (A.13), and thus we have a contradiction. Therefore, we must have that $(\theta_c^*)^\top\Sigma\theta_c^* \geq \theta^{*\top}\Sigma\theta^*$.

Step 2: the Sharpe ratio of the portfolio θ_c^* is not greater than that of θ^*.

We show that

$$\frac{(\theta_c^*)^\top\mu - f(\theta_c^*)}{\sqrt{(\theta_c^*)^\top\Sigma\theta_c^*}} \leq \frac{\theta^{*\top}\mu - f(\theta^*)}{\sqrt{\theta^{*\top}\Sigma\theta^*}}, \tag{A.14}$$

and the equality holds only when $(\theta_c^*)^\top\Sigma\theta_c^* = \theta^{*\top}\Sigma\theta^*$.

When $(\theta_c^*)^\top\Sigma\theta_c^* = \theta^{*\top}\Sigma\theta^*$, (A.13) implies that $\theta^{*\top}\mu - f(\theta^*) = (\theta_c^*)^\top\mu - f(\theta_c^*)$, and thus (A.14) holds with equality.

When $(\theta_c^*)^\top\Sigma\theta_c^* > \theta^{*\top}\Sigma\theta^*$, let $(\theta_c^*)^\top\Sigma\theta_c^* = c^2\theta^{*\top}\Sigma\theta^*$ where $c > 1$. To prove (A.14) with strict inequality, we prove by contradiction that

$$(\theta_c^*)^\top\mu - f(\theta_c^*) < c(\theta^{*\top}\mu - f(\theta^*)). \tag{A.15}$$

Suppose (A.15) does not hold and thus $\theta^{*\top}\mu - f(\theta^*) \leq ((\theta_c^*)^\top\mu - f(\theta_c^*))/c$, then

$$\begin{aligned}\theta^{*\top}\mu - f(\theta^*) - \frac{\gamma}{2}\theta^{*\top}\Sigma\theta^* &\leq \frac{1}{c}(\theta_c^*)^\top\mu - \frac{1}{c}f(\theta_c^*) - \frac{\gamma}{2}\frac{(\theta_c^*)^\top}{c}\Sigma\frac{\theta_c^*}{c} \\ &< \frac{1}{c}(\theta_c^*)^\top\mu - f\left(\frac{1}{c}\theta_c^*\right) - \frac{\gamma}{2}\frac{(\theta_c^*)^\top}{c}\Sigma\frac{\theta_c^*}{c},\end{aligned} \tag{A.16}$$

where the second inequality comes from Part (i). This contradicts θ^* being a maximizer for the case with absolute risk aversion is γ. Thus, when $(\theta_c^*)^\top\Sigma\theta_c^* > \theta^{*\top}\Sigma\theta^*$, (A.15) holds. Dividing both sides of (A.15) by $\sqrt{(\theta_c^*)^\top\Sigma\theta_c^*} = c\sqrt{\theta^{*\top}\Sigma\theta^*}$, (A.14) holds with strict inequality. Thus, the efficient frontier is strictly concave. Moreover, since $f(\theta) > 0$ for any $\theta \neq 0$, both sides of (A.14) are less than the Sharpe ratio in the absence of trading costs, SR.

A.3. Proof of Proposition 5

The proof consists of two parts. Part (i) derives the asymptotic distribution of the sample mean–variance utility net of price-impact costs of a factor model. Part (ii) derives the asymptotic distribution of the difference between the sample mean–variance utilities net of price-impact costs of two factor models. For ease of notation, we drop the superscript γ from MVU^γ throughout this proof.

Part (i): asymptotic distribution of sample mean–variance utility of one model

The proof of Part (i) contains two steps. We first show that the sample mean–variance utility of a model is asymptotically normally distributed and then derive the variance of the asymptotic normal distribution.

Step 1: $\sqrt{T}(\widehat{MVU} - MVU)$ is asymptotically normally distributed. We extend the notation in the proof of Proposition 2 of Barillas et al. (2020) to the case with price-impact costs. In particular, let

$$\varphi = [\mu, \text{vec}(\Sigma), \text{vec}(\Lambda/\gamma)] \in \mathbb{R}^{K+2K^2}, \tag{A.17}$$

$$\hat{\varphi} = [\hat{\mu}, \text{vec}(\hat{\Sigma}), \text{vec}(\hat{\Lambda}/\gamma)] \in \mathbb{R}^{K+2K^2}, \tag{A.18}$$

$$r_t(\varphi) = [F_t - \mu, \text{vec}(\Sigma_t - \Sigma), \text{vec}((\Lambda_t - \Lambda)/\gamma)] \in \mathbb{R}^{K+2K^2}. \tag{A.19}$$

Under standard regularity conditions,[29] the central limit theorem implies that,

$$\sqrt{T}(\hat{\varphi} - \varphi) \overset{A}{\sim} N(0, S_0), \quad \text{where } S_0 = \sum_{j=-\infty}^{\infty} E[r_t(\varphi) r_{t+j}^\top(\varphi)].$$

Using the delta method, we have that

$$\sqrt{T}(\widehat{\text{MVU}} - \text{MVU}) \overset{A}{\sim} N\left(0, \frac{\partial \text{MVU}}{\partial \varphi^\top} S_0 \frac{\partial \text{MVU}}{\partial \varphi}\right). \tag{A.20}$$

Step 2: variance of asymptotic normal distribution, $h_t(\varphi)$.

[29] For example, we could assume that the returns and the rebalancing trades are stationary and ergodic, and the corresponding Gordin's condition is satisfied, as in Proposition 6.10 of Hayashi (2000).

Let

$$h_t(\varphi) = 2\gamma \frac{\partial \text{MVU}}{\partial \varphi^\top} r_t(\varphi), \tag{A.21}$$

then (A.20) can be rewritten as

$$\sqrt{T}(\widehat{\text{MVU}} - \text{MVU}) \overset{A}{\sim} N(0, W), \quad \text{where } W = \sum_{j=-\infty}^{\infty} E\left[\frac{h_t(\varphi)h_{t+j}(\varphi)}{4\gamma^2}\right]. \tag{A.22}$$

Assumption 3.1 implies that $h_t(\varphi)$ is serially uncorrelated, and thus, we have that

$$W = E\left[\frac{h_t^2(\varphi)}{4\gamma^2}\right]. \tag{A.23}$$

Also, note that

$$\frac{\partial \text{MVU}}{\partial \mu} = \frac{1}{\gamma}(\Sigma + \Lambda/\gamma)^{-1}\mu = \theta^*,$$

$$\frac{\partial \text{MVU}}{\partial \Sigma} = \gamma \frac{\partial \text{MVU}}{\partial \Lambda} = -\frac{1}{2\gamma}(\Sigma + \Lambda/\gamma)^{-1}\mu\mu^\top(\Sigma + \Lambda/\gamma)^{-1} = -\frac{\gamma}{2}\theta^*\theta^{*\top},$$

and thus,

$$\frac{\partial \text{MVU}}{\partial \text{vec}(\Sigma)} = \gamma \frac{\partial \text{MVU}}{\partial \text{vec}(\Lambda)} = -\frac{\gamma}{2}\theta^* \otimes \theta^*,$$

where $\otimes$ denotes the Kronecker product. Plugging these partial derivatives in the definition of $h_t(\varphi)$ in (A.21), we have that

$$\begin{aligned} h_t(\varphi) &= 2\gamma\left[\frac{\partial \text{MVU}}{\partial \mu^\top}(F_t - \mu) + \frac{\partial \text{MVU}}{\partial \text{vec}(\Sigma)^\top}\text{vec}(\Sigma_t - \Sigma) + \frac{\partial \text{MVU}}{\partial \text{vec}(\Lambda)^\top}\text{vec}(\Lambda_t - \Lambda)\right] \\ &= 2\gamma\theta^{*\top}(F_t - \mu) - \gamma^2\theta^{*\top}\Sigma_t\theta^* - \gamma\theta^{*\top}\Lambda_t\theta^* + \gamma^2\theta^{*\top}\Sigma\theta^* + \gamma\theta^{*\top}\Lambda\theta^* \\ &= \mu^\top(\Sigma + \Lambda/\gamma)^{-1}(2F_t - \mu) - \mu^\top(\Sigma + \Lambda/\gamma)^{-1}(\Sigma_t + \Lambda_t/\gamma)(\Sigma + \Lambda/\gamma)^{-1}\mu, \end{aligned} \tag{A.24}$$

which completes the first part of the proof.

Part (ii): asymptotic distribution of difference between utilities of two models

Following the same steps as in Part (i), we have that

$$\sqrt{T}\left([\widehat{\text{MVU}}_A - \widehat{\text{MVU}}_B] - [\text{MVU}_A - \text{MVU}_B]\right) \overset{A}{\sim} N\left(0, \frac{\partial(\text{MVU}_A - \text{MVU}_B)}{\partial \varphi^\top} S_0 \frac{\partial(\text{MVU}_A - \text{MVU}_B)}{\partial \varphi}\right).$$

By Assumption 3.1, we have that

$$\sqrt{T}\left([\widehat{\text{MVU}}_A - \widehat{\text{MVU}}_B] - [\text{MVU}_A - \text{MVU}_B]\right) \overset{A}{\sim} N\left(0, E\left[\frac{(h_{t,A} - h_{t,B})^2}{4\gamma^2}\right]\right), \tag{A.25}$$

where $h_{t,A}$ and $h_{t,B}$ are obtained by applying Eq. (A.21) to models A and B, respectively. This completes the proof.

Remark. When model A nests model B and the extra factors of model A are redundant, or when models A and B share common factors and the extra factors of both models are redundant, the two models have the same optimal factor portfolio. In either case, the null hypothesis $\text{MVU}_A = \text{MVU}_B$ holds and Eq. (A.24) suggests that $h_{t,A} = h_{t,B}$ for all t, and thus the variance in (A.25), $E[(h_{t,A} - h_{t,B})^2/(4\gamma^2)] = 0$. Consequently, the distribution in (A.25) is not applicable to perform a statistical test in these cases. Instead, in these cases we use the asymptotic distribution in Proposition 6.

A.4. Proof of Proposition 6

Let the mean–variance portfolio in the presence of price-impact costs for model A be $\theta_A^* = [\theta_1^*, \theta_2^*]$. Note that the null hypothesis that models A and B have the same mean–variance utility holds if and only if $\theta_2^* = 0$. Using this condition, we prove this proposition in three parts. Part (i) derives the asymptotic distribution of the sample factor portfolio $\hat{\theta}_A^*$. Part (ii) provides an expression for the difference between the mean–variance utilities net of price-impact costs of models A and B as a function of θ_2^*. Part (iii) uses the asymptotic distribution of $\hat{\theta}_2^*$ to derive the asymptotic distribution of the difference between the sample mean–variance utilities net of price-impact costs of models A and B. Similar to the proof of Proposition 5, we drop the superscript γ from MVU^γ throughout this proof.

Part (i): asymptotic distribution for $\hat{\theta}_A^*$.

Following similar steps as those in Part (i) of the proof of Proposition 5, the asymptotic distribution of $\hat{\theta}_A^*$ is

$$\sqrt{T}(\hat{\theta}_A^* - \theta_A^*) \overset{A}{\sim} N(0, \frac{E[l_t l_t^\top]}{\gamma^2}), \tag{A.26}$$

where

$$l_t = (\Sigma_A + \Lambda_A/\gamma)^{-1}F_{A,t} - (\Sigma_A + \Lambda_A/\gamma)^{-1}(\Sigma_{A,t} + \Lambda_{A,t}/\gamma)(\Sigma_A + \Lambda_A/\gamma)^{-1}\mu_A \in \mathbb{R}^{K_1+K_2}. \tag{A.27}$$

Part (ii): expression for $\text{MVU}_A - \text{MVU}_B$ as a function of θ_2^*.

The difference $\text{MVU}_A - \text{MVU}_B$ can be written as

$$\begin{aligned} &= \frac{1}{2\gamma}\left[\mu_1^\top, \mu_2^\top\right]\begin{bmatrix}\Sigma_{11} + \Lambda_{11}/\gamma & \Sigma_{12} + \Lambda_{12}/\gamma \\ \Sigma_{21} + \Lambda_{21}/\gamma & \Sigma_{22} + \Lambda_{22}/\gamma\end{bmatrix}^{-1}\begin{bmatrix}\mu_1 \\ \mu_2\end{bmatrix} \qquad (A.28) \\ &- \frac{1}{2\gamma}\left[\mu_1^\top, \mu_2^\top\right]\begin{bmatrix}(\Sigma_{11} + \Lambda_{11}/\gamma)^{-1} & 0 \\ 0 & 0\end{bmatrix}\begin{bmatrix}\mu_1 \\ \mu_2\end{bmatrix} \\ &= \frac{\gamma}{2}\theta_A^{*\top}\begin{bmatrix}\Sigma_{11} + \Lambda_{11}/\gamma & \Sigma_{12} + \Lambda_{12}/\gamma \\ \Sigma_{21} + \Lambda_{21}/\gamma & \Sigma_{22} + \Lambda_{22}/\gamma\end{bmatrix}\theta_A^* \\ &- \frac{\gamma}{2}\theta_A^{*\top}\begin{bmatrix}\Sigma_{11} + \Lambda_{11}/\gamma & \Sigma_{12} + \Lambda_{12}/\gamma \\ \Sigma_{21} + \Lambda_{21}/\gamma & \Sigma_{22} + \Lambda_{22}/\gamma\end{bmatrix} \\ &\times \begin{bmatrix}(\Sigma_{11} + \Lambda_{11}/\gamma)^{-1} & 0 \\ 0 & 0\end{bmatrix}\begin{bmatrix}\Sigma_{11} + \Lambda_{11}/\gamma & \Sigma_{12} + \Lambda_{12}/\gamma \\ \Sigma_{21} + \Lambda_{21}/\gamma & \Sigma_{22} + \Lambda_{22}/\gamma\end{bmatrix}\theta_A^* \\ &= \frac{\gamma}{2}\theta_A^{*\top}\begin{bmatrix}\Sigma_{11} + \Lambda_{11}/\gamma & \Sigma_{12} + \Lambda_{12}/\gamma \\ \Sigma_{21} + \Lambda_{21}/\gamma & \Sigma_{22} + \Lambda_{22}/\gamma\end{bmatrix}\theta_A^* \\ &- \frac{\gamma}{2}\theta_A^{*\top}\begin{bmatrix}\Sigma_{11} + \Lambda_{11}/\gamma & \Sigma_{12} + \Lambda_{12}/\gamma \\ \Sigma_{21} + \Lambda_{21}/\gamma & (\Sigma_{21} + \Lambda_{21}/\gamma)(\Sigma_{11} + \Lambda_{11}/\gamma)^{-1}(\Sigma_{12} + \Lambda_{12}/\gamma)\end{bmatrix}\theta_A^* \\ &= \frac{\gamma}{2}\theta_2^{*\top}\left[(\Sigma_{22} + \Lambda_{22}/\gamma) - (\Sigma_{21} + \Lambda_{21}/\gamma)(\Sigma_{11} + \Lambda_{11}/\gamma)^{-1}(\Sigma_{12} + \Lambda_{12}/\gamma)\right]\theta_2^* \\ &= \frac{\gamma}{2}\theta_2^{*\top} W\theta_2^*, \end{aligned} \tag{A.29}$$

where $W = (\Sigma_{22} + \Lambda_{22}/\gamma) - (\Sigma_{21} + \Lambda_{21}/\gamma)(\Sigma_{11} + \Lambda_{11}/\gamma)^{-1}(\Sigma_{12} + \Lambda_{12}/\gamma)$. Replacing the population parameters in Eq. (A.29) with their sample counterparts we have that

$$\widehat{\text{MVU}}_A - \widehat{\text{MVU}}_B = \frac{\gamma}{2}\hat{\theta}_2^{*\top}\hat{W}\hat{\theta}_2^*, \quad \text{where} \quad \hat{W} \overset{a.s.}{\to} W. \tag{A.30}$$

Part (iii): asymptotic distribution for $T(\widehat{\text{MVU}}_A - \widehat{\text{MVU}}_B)$.

We now use (A.26) and (A.30) to derive the asymptotic distribution for $T(\widehat{\text{MVU}}_A - \widehat{\text{MVU}}_B)$. Let

$$z = \lim_{T\to\infty}\sqrt{T}\left(\frac{E[l_t l_t^\top]_{22}}{\gamma^2}\right)^{-\frac{1}{2}}\hat{\theta}_2^*.$$

Under the null hypothesis that $\theta_2^* = 0$, from the asymptotic distribution in (A.26) we have that $z \sim N(0, I_{K_2})$, where I_{K_2} is a K_2-dimensional identity matrix. Thus, from Eq. (A.30) we have that

$$\begin{aligned} T(\widehat{\text{MVU}}_A - \widehat{\text{MVU}}_B) &= \frac{\gamma}{2}T\hat{\theta}_2^{*\top}\hat{W}\hat{\theta}_2^* \\ &\overset{A}{\sim} \frac{1}{2\gamma}z^\top(E[l_t l_t^\top]_{22})^{\frac{1}{2}}W(E[l_t l_t^\top]_{22})^{\frac{1}{2}}z. \end{aligned} \tag{A.31}$$

Let $Q\Xi Q^\top$ be the eigenvalue decomposition of $(E[l_t l_t^\top]_{22})^{\frac{1}{2}} W(E[l_t l_t^\top]_{22})^{\frac{1}{2}}/2\gamma$, where Q is the orthogonal matrix whose columns contain the eigenvectors and Ξ is a diagonal matrix whose diagonal elements contain the eigenvalues ξ_i for $i = 1, \dots, K_2$. Note the eigenvalues in the diagonal of Ξ are also the eigenvalues of $E[l_t l_t^\top]_{22}W/2\gamma$. Let

$\bar{z} = Q^\top z \sim N(0, I_{K_2})$, then (A.31) can be rewritten as

$$T(\widehat{\mathrm{MVU}}_A - \widehat{\mathrm{MVU}}_B) \overset{A}{\sim} \bar{z}^\top \Xi \bar{z} = \sum_{i=1}^{K_2} \xi_i x_i,$$

where x_i for $i = 1, \dots, K_2$ are independent chi-square random variables with one degree of freedom.

A.5. Proof of Proposition 7

The proof consists of two parts. Part (i) derives a closed-form expression for the asymptotic variance of the sample mean–variance utility of a factor model. Part (ii) derives a closed-form expression for the asymptotic variance of the difference between the sample mean–variance utilities of two factor models.

Part (i): closed-form asymptotic variance of the mean–variance utility of a model

We first provide a closed-form expression for the asymptotic variance of the sample mean–variance utility of a model, $E[h_t^2]/(4\gamma^2)$, and then simplify this expression.

Step 1: express $E[h_t^2]$ as a function of u_t, $v_{n,t}$, and $\bar{u} = E[u_t]$.

Plugging $\bar{u}, u_t$, and $v_{n,t}$ into (22), we have that

$$h_t = 2(u_t - \bar{u}) - \left[(u_t - \bar{u})^2 + \sum_{n=1}^N v_{n,t}^2\right] + \bar{u}.$$

Therefore,

$$\begin{aligned} E[h_t^2] =& E\Bigg[4(u_t - \bar{u})^2 - 4(u_t - \bar{u})^3 - 4(u_t - \bar{u}) \sum_{n=1}^N v_{n,t}^2 + 4(u_t - \bar{u})\bar{u} \\ &+ (u_t - \bar{u})^4 + 2(u_t - \bar{u})^2 \sum_{n=1}^N v_{n,t}^2 - 2(u_t - \bar{u})^2\bar{u} \\ &+ \left(\sum_{n=1}^N v_{n,t}^2\right)^2 - 2\bar{u} \sum_{n=1}^N v_{n,t}^2 + \bar{u}^2\Bigg]. \end{aligned} \tag{A.32}$$

Lemma 2 of Maruyama and Seo (2003) shows that if (X_i, X_j, X_k, X_l) are jointly normally distributed with zero mean, then

$$E[X_i X_j X_k] = 0, \tag{A.33}$$

$$E[X_i X_j X_k X_l] = (\sigma_{ij}\sigma_{kl} + \sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk}), \tag{A.34}$$

where σ_{ab} is the covariance between X_a and X_b. Because $(u_t - \bar{u})$ and $v_{n,t}$ for $n = 1, \dots, N$ are jointly normally distributed, using Eq. (A.33), we can drop the third-order moments from Eq. (A.32) to obtain

$$\begin{aligned} E[h_t^2] =& E\Bigg[4(u_t - \bar{u})^2 + (u_t - \bar{u})^4 + 2(u_t - \bar{u})^2 \sum_{n=1}^N v_{n,t}^2 - 2(u_t - \bar{u})^2\bar{u} \\ &+ \left(\sum_{n=1}^N v_{n,t}^2\right)^2 - 2\bar{u} \sum_{n=1}^N v_{n,t}^2 + \bar{u}^2\Bigg]. \end{aligned} \tag{A.35}$$

Step 2: simplify (A.35). Using Eq. (A.34), we can rewrite the terms on the right-hand side of Eq. (A.35) as

$$\begin{aligned} E\left[(u_t - \bar{u})^2\right] &= \mathrm{var}(u_t) = \mu^\top(\Sigma + \Lambda/\gamma)^{-1}\Sigma(\Sigma + \Lambda/\gamma)^{-1}\mu, \\ E\left[(u_t - \bar{u})^4\right] &= 3\left[\mathrm{var}(u_t)\right]^2, \\ E\left[\sum_{n=1}^N v_{n,t}^2\right] &= \sum_{n=1}^N \mathrm{var}(v_{n,t}) = \mu^\top(\Sigma + \Lambda/\gamma)^{-1}(\Lambda/\gamma)(\Sigma + \Lambda/\gamma)^{-1}\mu, \\ E\left[(u_t - \bar{u})^2 \sum_{n=1}^N v_{n,t}^2\right] &= E\left[(u_t - \bar{u})^2\right] \sum_{n=1}^N E\left[v_{n,t}^2\right] + 2\sum_{n=1}^N \left(E\left[(u_t - \bar{u})v_{n,t}\right]\right)^2 \\ &= \mathrm{var}(u_t) \sum_{n=1}^N \mathrm{var}(v_{n,t}) + 2\sum_{n=1}^N \left[\mathrm{cov}(u_t, v_{n,t})\right]^2, \\ E\left[\left(\sum_{n=1}^N v_{n,t}^2\right)^2\right] &= \sum_{i=1}^N \sum_{j=1}^N \left(\mathrm{var}(v_{i,t})\mathrm{var}(v_{j,t}) + 2\left[\mathrm{cov}(v_{i,t}, v_{j,t})\right]^2\right), \\ \bar{u} &= \mu^\top(\Sigma + \Lambda/\gamma)^{-1}\mu = \mathrm{var}(u_t) + \sum_{n=1}^N \mathrm{var}(v_{n,t}). \end{aligned}$$

Plugging these equations into (A.35), we have that

$$\begin{aligned} E[h_t^2] &= 4\mathrm{var}(u_t) + 3\left[\mathrm{var}(u_t)\right]^2 + 2\left(\mathrm{var}(u_t)\sum_{n=1}^N \mathrm{var}(v_{n,t}) + 2\sum_{n=1}^N \left[\mathrm{cov}(u_t, v_{n,t})\right]^2\right) \\ &\quad - 2\mathrm{var}(u_t)\left(\mathrm{var}(u_t) + \sum_{n=1}^N \mathrm{var}(v_{n,t})\right) \\ &\quad + \sum_{i=1}^N \sum_{j=1}^N \left(\mathrm{var}(v_{i,t})\mathrm{var}(v_{j,t}) + 2\left[\mathrm{cov}(v_{i,t}, v_{j,t})\right]^2\right) \\ &\quad - 2\sum_{n=1}^N \mathrm{var}(v_{n,t})\left(\mathrm{var}(u_t) + \sum_{n=1}^N \mathrm{var}(v_{n,t})\right) + \left(\mathrm{var}(u_t) + \sum_{n=1}^N \mathrm{var}(v_{n,t})\right)^2 \\ &= 4\mathrm{var}(u_t) + 2\left[\mathrm{var}(u_t)\right]^2 - \left(\sum_{n=1}^N \mathrm{var}(v_{n,t})\right)^2 + 4\sum_{n=1}^N \left[\mathrm{cov}(u_t, v_{n,t})\right]^2 \\ &\quad + \sum_{i=1}^N \sum_{j=1}^N \left(\mathrm{var}(v_{i,t})\mathrm{var}(v_{j,t}) + 2\left[\mathrm{cov}(v_{i,t}, v_{j,t})\right]^2\right) \\ &= 4\mathrm{var}(u_t) + 2\left[\mathrm{var}(u_t)\right]^2 + 4\sum_{n=1}^N \left[\mathrm{cov}(u_t, v_{n,t})\right]^2 + 2\sum_{i=1}^N \sum_{j=1}^N \left[\mathrm{cov}(v_{i,t}, v_{j,t})\right]^2. \end{aligned}$$

Part (ii): asymptotic variance for difference between utilities of two models

The asymptotic variance of the difference between the sample mean–variance utilities of two models is

$$\frac{E[(h_{t,A} - h_{t,B})^2]}{4\gamma^2} = \frac{1}{4\gamma^2}\left(E[h_{t,A}^2] + E[h_{t,B}^2] - 2E[h_{t,A}h_{t,B}]\right). \tag{A.36}$$

The closed-form expressions of $E[h_{t,A}^2]$ and $E[h_{t,B}^2]$ are given in Part (i), and thus we focus on finding the closed-form expression of $E[h_{t,A}h_{t,B}]$. Similar to Part (i), we first express $E[h_{t,A}h_{t,B}]$ as a function of $\bar{u}$, u_t, and $v_{n,t}$, and then simplify this expression.

Step 1: express $E[h_{t,A}h_{t,B}]$ as a function of $\bar{u}$, u_t, and $v_{n,t}$.

Because $(u_t^A - \bar{u}^A)$, $(u_t^B - \bar{u}^B)$, $v_{n,t}^A$, and $v_{n,t}^B$ for $n = 1, \dots, N$ are jointly normally distributed. Using Eq. (A.33), we have that

$$\begin{aligned} E[h_{t,A}h_{t,B}] = E\Bigg[&4\left(u_t^A - \bar{u}^A\right)\left(u_t^B - \bar{u}^B\right) + \left(u_t^A - \bar{u}^A\right)^2\left(u_t^B - \bar{u}^B\right)^2 \\ &+ \left(u_t^A - \bar{u}^A\right)^2 \sum_{n=1}^N (v_{n,t}^B)^2 + \left(u_t^B - \bar{u}^B\right)^2 \sum_{n=1}^N \left(v_{n,t}^A\right)^2 \\ &- \left(u_t^A - \bar{u}^A\right)^2 \bar{u}^B - \left(u_t^B - \bar{u}^B\right)^2 \bar{u}^A + \left(\sum_{n=1}^N (v_{n,t}^A)^2\right)\left(\sum_{n=1}^N \left(v_{n,t}^B\right)^2\right) \\ &- \bar{u}^A \sum_{n=1}^N \left(v_{n,t}^B\right)^2 - \bar{u}^B \sum_{n=1}^N \left(v_{n,t}^A\right)^2 + \bar{u}^A\bar{u}^B\Bigg]. \end{aligned} \tag{A.37}$$

Step 2: simplify (A.37). Using Eq. (A.34), we can rewrite the terms on the right-hand side of Eq. (A.37) as

$$\begin{aligned} E\left[\left(u_t^A - \bar{u}^A\right)\left(u_t^B - \bar{u}^B\right)\right] &= \mathrm{cov}\left(u_t^A, u_t^B\right), \\ E\left[\left(u_t^A - \bar{u}^A\right)^2\left(u_t^B - \bar{u}^B\right)^2\right] &= \mathrm{var}\left(u_t^A\right)\mathrm{var}\left(u_t^B\right) + 2\left[\mathrm{cov}\left(u_t^A, u_t^B\right)\right]^2, \\ E\left[\sum_{n=1}^N \left(v_{n,t}^A\right)^2\right] &= \sum_{n=1}^N \mathrm{var}\left(v_{n,t}^A\right), \\ E\left[\sum_{n=1}^N \left(v_{n,t}^B\right)^2\right] &= \sum_{n=1}^N \mathrm{var}\left(v_{n,t}^B\right), \\ E\left[(u_t^A - \bar{u}^A)^2 \sum_{n=1}^N (v_{n,t}^B)^2\right] &= \mathrm{var}(u_t^A)\sum_{n=1}^N \mathrm{var}(v_{n,t}^B) + 2\sum_{n=1}^N \left[\mathrm{cov}(u_t^A, v_{n,t}^B)\right]^2, \\ E\left[(u_t^B - \bar{u}^B)^2 \sum_{n=1}^N (v_{n,t}^A)^2\right] &= \mathrm{var}(u_t^B)\sum_{n=1}^N \mathrm{var}(v_{n,t}^A) + 2\sum_{n=1}^N \left[\mathrm{cov}(u_t^B, v_{n,t}^A)\right]^2, \\ E\left[\left(\sum_{n=1}^N (v_{n,t}^A)^2\right)\left(\sum_{n=1}^N (v_{n,t}^B)^2\right)\right] &= \sum_{i=1}^N \sum_{j=1}^N \left(\mathrm{var}(v_{i,t}^A)\mathrm{var}(v_{j,t}^B) + 2\left[\mathrm{cov}(v_{i,t}^A, v_{j,t}^B)\right]^2\right), \\ \bar{u}^A &= \mathrm{var}(u_t^A) + \sum_{n=1}^N \mathrm{var}(v_{n,t}^A), \end{aligned}$$

$$\bar{u}^B = \text{var}(u_t^B) + \sum_{n=1}^{N} \text{var}(v_{n,t}^B).$$

Plugging these equations into Eq. (A.37), we have that

$$\begin{aligned}
E[h_{t,A}h_{t,B}] &= 4\text{cov}(u_t^A,u_t^B) + \text{var}(u_t^A)\text{var}(u_t^B) + 2\left[\text{cov}(u_t^A,u_t^B)\right]^2 \\
&\quad + \text{var}(u_t^A)\sum_{n=1}^{N}\text{var}(v_{n,t}^B) + 2\sum_{n=1}^{N}\left[\text{cov}(u_t^A,v_{n,t}^B)\right]^2 \\
&\quad + \text{var}(u_t^B)\sum_{n=1}^{N}\text{var}(v_{n,t}^A) + 2\sum_{n=1}^{N}\left[\text{cov}(u_t^B,v_{n,t}^A)\right]^2 \\
&\quad - \text{var}(u_t^A)\left(\text{var}(u_t^B) + \sum_{n=1}^{N}\text{var}(v_{n,t}^B)\right) \\
&\quad - \text{var}(u_t^B)\left(\text{var}(u_t^A) + \sum_{n=1}^{N}\text{var}(v_{n,t}^A)\right) \\
&\quad + \sum_{i=1}^{N}\sum_{j=1}^{N}\left(\text{var}(v_{i,t}^A)\text{var}(v_{j,t}^B) + 2\left[\text{cov}(v_{i,t}^A,v_{j,t}^B)\right]^2\right) \\
&\quad - \left(\sum_{n=1}^{N}\text{var}(v_{n,t}^B)\right)\left(\text{var}(u_t^A) + \sum_{n=1}^{N}\text{var}(v_{n,t}^A)\right) \\
&\quad - \left(\sum_{n=1}^{N}\text{var}(v_{n,t}^A)\right)\left(\text{var}(u_t^B) + \sum_{n=1}^{N}\text{var}(v_{n,t}^B)\right) \\
&\quad + \left(\text{var}(u_t^A) + \sum_{n=1}^{N}\text{var}(v_{n,t}^A)\right)\left(\text{var}(u_t^B) + \sum_{n=1}^{N}\text{var}(v_{n,t}^B)\right) \\
&= 4\text{cov}(u_t^A,u_t^B) + 2\left[\text{cov}(u_t^A,u_t^B)\right]^2 \\
&\quad - \left(\sum_{n=1}^{N}\text{var}(v_{n,t}^A)\right)\left(\sum_{n=1}^{N}\text{var}(v_{n,t}^B)\right) \\
&\quad + 2\sum_{n=1}^{N}\left(\left[\text{cov}(u_t^A,v_{n,t}^B)\right]^2 + \left[\text{cov}(u_t^B,v_{n,t}^A)\right]^2\right) \\
&\quad + \sum_{i=1}^{N}\sum_{j=1}^{N}\left(\text{var}(v_{i,t}^A)\text{var}(v_{j,t}^B) + 2\left[\text{cov}(v_{i,t}^A,v_{j,t}^B)\right]^2\right) \\
&= 4\text{cov}(u_t^A,u_t^B) + 2\left[\text{cov}(u_t^A,u_t^B)\right]^2 + 2\sum_{i=1}^{N}\sum_{j=1}^{N}\left[\text{cov}(v_{i,t}^A,v_{j,t}^B)\right]^2 \\
&\quad + 2\sum_{n=1}^{N}\left(\left[\text{cov}(u_t^A,v_{n,t}^B)\right]^2 + \left[\text{cov}(u_t^B,v_{n,t}^A)\right]^2\right),
\end{aligned}$$

which completes the proof.

Appendix B. Proof and discussion of Proposition 4

This appendix provides a proof and interpretation for Proposition 4. Appendix B.1 gives the proof, Appendix B.2 discusses the relation between Proposition 4 and the GRS test of Gibbons et al. (1989), and Appendix B.3 provides interpretation for the net alpha introduced in Proposition 4.

B.1. Proof of Proposition 4

Let the vector $S_t = (F_t^\top, R_t^\top)^\top$ stack the returns of the factors and test assets. Thus, the average of S_t is $\mu_S = (\mu_F^\top, \mu_R^\top)^\top$ and its covariance matrix is

$$\Sigma_{S,S} = \begin{bmatrix} \Sigma_{F,F} & \Sigma_{F,R} \\ \Sigma_{R,F} & \Sigma_{R,R} \end{bmatrix}.$$

Similarly, the expected price-impact matrix for S_t is

$$\Lambda_{S,S} = \begin{bmatrix} \Lambda_{F,F} & \Lambda_{F,R} \\ \Lambda_{R,F} & \Lambda_{R,R} \end{bmatrix},$$

where $\Lambda_{F,F} = E[(\tilde{X}_t^F)^\top D_t \tilde{X}_t^F]$ is the expected price-impact matrix for the factors, $\Lambda_{R,R} = E[(\tilde{X}_t^R)^\top D_t \tilde{X}_t^R]$ is the expected price-impact matrix for the test assets, and $\Lambda_{R,F} = \Lambda_{F,R}^\top = E[(\tilde{X}_t^R)^\top D_t \tilde{X}_t^F]$ is the expected price-impact matrix for the test assets when the investor is also holding the factors.

Consider an investor with absolute risk aversion γ who faces the quadratic price-impact costs defined in (12). Then, Eq. (12) implies that an investor holding a portfolio $\theta_S = [\theta_F, \theta_R]$ of the factors and test assets incurs the following expected price-impact cost:

$$f(\theta_S) = \frac{1}{2}\theta_F^\top \Lambda_{F,F}\theta_F + \frac{1}{2}\theta_R^\top \Lambda_{R,R}\theta_R + \theta_R^\top \Lambda_{R,F}\theta_F^\top. \tag{B.1}$$

The first term in the right-hand side of (B.1) is the price-impact cost associated with rebalancing the portfolio of the factors in isolation, θ_F, the second term is the price-impact cost associated with rebalancing the portfolio of the test assets in isolation, θ_R, and the third term is the price-impact cost associated with the interaction between the trades required to rebalance the portfolios of the test assets and the factors.

Eq. (14) implies that the mean–variance utility net of price-impact costs of the investor when she has access to both the test assets and factors is

$$\text{MVU}^\gamma([F,R]) = \frac{\mu_S^\top \left(\Sigma_{S,S} + \Lambda_{S,S}/\gamma\right)^{-1}\mu_S}{2\gamma}, \tag{B.2}$$

and that when she only has access to the factors is

$$\text{MVU}^\gamma(F) = \frac{\mu_F^\top \left(\Sigma_{F,F} + \Lambda_{F,F}/\gamma\right)^{-1}\mu_F}{2\gamma}. \tag{B.3}$$

To prove the proposition, we first note that for an invertible matrix

$$U = \begin{bmatrix} A & B \\ B^\top & D \end{bmatrix},$$

where A is an invertible square matrix, we have

$$U^{-1} = \begin{bmatrix} A^{-1} + A^{-1}B\left(D - B^\top A^{-1}B\right)^{-1}B^\top A^{-1} & -A^{-1}B\left(D - B^\top A^{-1}B\right)^{-1} \\ -\left(D - B^\top A^{-1}B\right)^{-1}B^\top A^{-1} & \left(D - B^\top A^{-1}B\right)^{-1} \end{bmatrix}.$$

Let U be $\Sigma_{S,S} + \Lambda_{S,S}/\gamma$, and thus A, B, and D correspond to $\Sigma_{F,F} + \Lambda_{F,F}/\gamma$, $\Sigma_{F,R} + \Lambda_{F,R}/\gamma$, and $\Sigma_{R,R} + \Lambda_{R,R}/\gamma$, respectively. In this case, we have

$$\begin{aligned}
\mu_S^\top U^{-1}\mu_S &= \mu_F^\top A^{-1}\mu_F + \mu_F^\top A^{-1}B\left(D - B^\top A^{-1}B\right)^{-1}B^\top A^{-1}\mu_F \\
&\quad - \mu_F^\top A^{-1}B\left(D - B^\top A^{-1}B\right)^{-1}\mu_R - \mu_R^\top\left(D - B^\top A^{-1}B\right)^{-1}B^\top A^{-1}\mu_F \\
&\quad + \mu_R^\top\left(D - B^\top A^{-1}B\right)^{-1}\mu_R \\
&= \left(\mu_R^\top - \mu_F^\top A^{-1}B\right)\left(D - B^\top A^{-1}B\right)^{-1}\left(\mu_R - B^\top A^{-1}\mu_F\right) \\
&\quad + \mu_F^\top A^{-1}\mu_F.
\end{aligned}$$

Thus,

$$\mu_S^\top U^{-1}\mu_S - \mu_F^\top A^{-1}\mu_F = \left(\mu_R^\top - \mu_F^\top A^{-1}B\right)\left(D - B^\top A^{-1}B\right)^{-1}\left(\mu_R - B^\top A^{-1}\mu_F\right). \tag{B.4}$$

Note that

$$B = \Sigma_{F,R} + \Lambda_{F,R}/\gamma = \left(\Sigma_{F,F} + \Lambda_{F,F}/\gamma\right)\beta^\top + \left(\Lambda_{F,R}/\gamma - \Lambda_{F,F}\beta^\top/\gamma\right),$$

where β is the slope obtained from an OLS regression of the test asset returns on the factor returns. Thus, we have

$$\begin{aligned}
\mu_R - B^\top A^{-1}\mu_F &= \mu_R - \left[\beta\left(\Sigma_{F,F} + \Lambda_{F,F}/\gamma\right) + \left(\Lambda_{R,F}/\gamma - \beta\Lambda_{F,F}/\gamma\right)\right] \\
&\quad \times \left(\Sigma_{F,F} + \Lambda_{F,F}/\gamma\right)^{-1}\mu_F \\
&= \mu_R - \beta\mu_F \\
&\quad - \left(\gamma\Lambda_{R,F}/\gamma - \gamma\beta\Lambda_{F,F}/\gamma\right)\frac{1}{\gamma}\left(\Sigma_{F,F} + \Lambda_{F,F}/\gamma\right)^{-1}\mu_F \\
&= \alpha - \left(\Lambda_{R,F} - \beta\Lambda_{F,F}\right)\theta_F^* \equiv \alpha^{\text{net}},
\end{aligned} \tag{B.5}$$

where α is the intercept obtained from regressing the test asset returns on the factor returns and the last equality follows from Eq. (13). Thus, Eqs. (B.4) and (B.5) imply that

$$\mathrm{MVU}^{\gamma}([F,R]) - \mathrm{MVU}^{\gamma}(F) = \left(\alpha^{\mathrm{net}}\right)^{\top} H_{\gamma}^{-1}\alpha^{\mathrm{net}}, \tag{B.6}$$

where

$$\begin{aligned} H_{\gamma} &= 2\gamma\left(\Sigma_{R,R} + \Lambda_{R,R}/\gamma\right) \\ &\quad - 2\gamma\left(\Sigma_{R,F} + \Lambda_{R,F}/\gamma\right)\left(\Sigma_{F,F} + \Lambda_{F,F}/\gamma\right)^{-1}\left(\Sigma_{F,R} + \Lambda_{F,R}/\gamma\right), \end{aligned} \tag{B.7}$$

which is positive definite because H_{γ} is the Schur complement of $2\gamma(\Sigma_{S,S} + \Lambda_{S,S}/\gamma)$, which is positive definite by assumption.

B.2. Relation to the GRS test

We now show that for the case without trading costs, Proposition 4 implies the result in equation (23) of Gibbons et al. (1989) that the increase in the squared Sharpe ratio of the investor when she has access to the test assets in addition to the factors in the model is a quadratic form of the gross alpha. To see this, note that for the case without trading costs ($\Lambda_{S,S} = 0$), we have that $\alpha^{\mathrm{net}} = \alpha$, $\mathrm{MVU}^{\gamma}([F,R]) = SR^2([F,R])/(2\gamma)$, $\mathrm{MVU}^{\gamma}(F) = SR^2(F)/(2\gamma)$, and $H_{\gamma} = 2\gamma\Sigma_{R,R} - 2\gamma\Sigma_{R,F}\Sigma_{F,F}^{-1}\Sigma_{F,R}$. Thus, Eq. (B.6) becomes

$$SR^2([F,R]) - SR^2(F) = \alpha^{\top}\left(\Sigma_{R,R} - \Sigma_{R,F}\Sigma_{F,F}^{-1}\Sigma_{F,R}\right)^{-1}\alpha. \tag{B.8}$$

B.3. Interpretation of the adjusted alpha

Consider an investor with absolute risk aversion γ. Then, the net alpha (α^{net}) defined in Eq. (20) of Proposition 4 is the incremental return net of price-impact costs that the investor can achieve by making a marginal investment in the test assets when she is already holding the mean–variance portfolio of the factors in the model.

To see this, consider first the case without trading costs. Assume the investor holds the mean–variance portfolio of the factors in the model $\theta^* = \Sigma_{F,F}^{-1}\mu_F/\gamma$ and M dollars of the ith test asset with return $R_{i,t}$. Then, the average return of the investor's portfolio is $(\theta^*)\top\mu_F + M\mu_{R_i}$, where μ_{R_i} is the average return of the ith asset. Moreover, the beta of the investor's portfolio with respect to the factors in the model is $\theta^* + M\beta_i$, where β_i is the beta of the ith asset with respect to the factors. Thus, the average return of the investor's portfolio explained by the factors in the model is $(\theta^* + M\beta_i)^{\top}\mu_F$, and the average return of the investor's portfolio that is not explained by the factors in the model, per dollar invested in the ith test asset is:

$$\frac{1}{M}\left[(\theta^*)^{\top}\mu_F + M\mu_{R_i} - \left(\theta^* + M\beta_i\right)^{\top}\mu_F\right] = \alpha_i,$$

which is the alpha of the ith asset with respect to the factors in the model. Importantly, in the absence of trading costs the alpha of asset i does not depend on the mean–variance portfolio of the factors θ^*.

In the presence of price-impact costs, however, the *net* alpha of the ith test asset depends on the investor's mean–variance factor portfolio θ^*, and thus, on the investor's absolute risk aversion γ. To see this, note that the price-impact cost associated with holding the portfolio of the investor is

$$\frac{1}{2}\left[(\theta^*)^{\top} \quad M\right]\Lambda_{S,S}\begin{bmatrix}\theta^* \\ M\end{bmatrix} = \frac{1}{2}(\theta^*)^{\top}\Lambda_{F,F}\theta^* + M(\theta^*)^{\top}\Lambda_{F,R_i} + \frac{M^2}{2}\Lambda_{R_i,R_i}.$$

Moreover, the beta of the investor's portfolio with respect to the factors is $\theta^* + M\beta_i$, and thus, the price-impact cost of the projection of the investor's portfolio on the factors is

$$\begin{aligned} \frac{1}{2}\left(\theta^* + M\beta_i\right)^{\top}\Lambda_{F,F}\left(\theta^* + M\beta_i\right) &= \frac{1}{2}(\theta^*)^{\top}\Lambda_{F,F}\theta^* + M(\theta^*)^{\top}\Lambda_{F,F}\beta_i \\ &\quad + \frac{M^2}{2}\beta_i^{\top}\Lambda_{F,F}\beta_i. \end{aligned}$$

Then, the average return net of price-impact costs of the investor's portfolio that is not explained by the factors in the model per dollar invested in the ith test asset is:

$$\begin{aligned} &\underbrace{\frac{1}{M}\left((\theta^*)^{\top}\mu_F + M\mu_{R_i} - \left(\frac{1}{2}(\theta^*)^{\top}\Lambda_{F,F}\theta^* + M(\theta^*)^{\top}\Lambda_{F,R_i} + \frac{M^2}{2}\Lambda_{R_i,R_i}\right)\right)}_{\text{Average net return of the investor's portfolio}} \\ &- \underbrace{\frac{1}{M}\left(\left(\theta^* + M\beta_i\right)^{\top}\mu_F - \left(\frac{1}{2}(\theta^*)^{\top}\Lambda_{F,F}\theta^* + M(\theta^*)^{\top}\Lambda_{F,F}\beta_i + \frac{M^2}{2}\beta_i^{\top}\Lambda_{F,F}\beta_i\right)\right)}_{\text{Average net return of the projection of the investor's portfolio on the factors}} \\ &= \alpha_i - \left((\theta^*)^{\top}\Lambda_{F,R_i} - (\theta^*)^{\top}\Lambda_{F,F}\beta_i\right) - \frac{M}{2}\left(\Lambda_{R_i,R_i} - \beta_i^{\top}\Lambda_{F,F}\beta_i\right). \end{aligned} \tag{B.9}$$

Furthermore, for the case where M is arbitrarily small we get

$$\lim_{M\to 0}\alpha_i - \left((\theta^*)^{\top}\Lambda_{F,R_i} - (\theta^*)^{\top}\Lambda_{F,F}\beta_i\right) - \frac{M}{2}\left(\Lambda_{R_i,R_i} - \beta_i^{\top}\Lambda_{F,F}\beta_i\right) = \alpha^{net}.$$

That is, α^{net} measures the incremental return net of price-impact costs that the investor can achieve by making a marginal investment in the test assets when she is already holding the mean–variance portfolio of the factors in the model.

Appendix C. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jfineco.2024.103949.

References

Avramov, D., Chao, J.C., 2006. An exact Bayes test of asset pricing models with application to international markets. J. Bus. 79 (1), 293–324.

Ball, R., Gerakos, J., Linnainmaa, J.T., Nikolaev, V., 2016. Accruals, cash flows, and operating profitability in the cross section of stock returns. J. Financ. Econ. 121 (1), 28–45.

Barillas, F., Kan, R., Shanken, J., 2020. Model comparison with sharpe ratios. J. Financ. Quant. Anal. 55 (6), 1840–1874.

Barillas, F., Shanken, J., 2017. Which alpha? Rev. Financ. Stud. 30 (4), 1316–1338.

Barillas, F., Shanken, J., 2018. Comparing asset pricing models. J. Finance 73 (2), 715–754.

Barroso, P., Detzel, A., 2021. Do limits to arbitrage explain the benefits of volatility-managed portfolios? J. Financ. Econ. 140 (3), 744–767.

Barroso, P., Santa-Clara, P., 2015. Beyond the carry trade: Optimal currency portfolios. J. Financ. Quant. Anal. 50 (05), 1037–1056.

Brandt, M.W., Santa-Clara, P., Valkanov, R., 2009. Parametric portfolio policies: Exploiting characteristics in the cross-section of equity returns. Rev. Financ. Stud. 22 (9), 3411–3447.

Campbell, J.Y., 2017. Financial Decisions and Markets: A Course in Asset Pricing. Princeton University Press, Princeton, NJ.

Carhart, M.M., 1997. On persistence in mutual fund performance. J. Finance 52 (1), 57–82.

Chen, A.Y., Velikov, M., 2023. Zeroing in on the expected returns of anomalies. J. Financ. Quant. Anal. 58, 968–1004.

Chen, A.Y., Zimmermann, T., 2022. Open source cross-sectional asset pricing. Crit. Finance Rev. 11 (2), 207–264.

Chib, S., Zeng, X., Zhao, L., 2020. On comparing asset pricing models. J. Finance 75 (1), 551–577.

Cochrane, J.H., 2009. Asset Pricing: Revised Edition. Princeton University Press.

DeMiguel, V., Martin-Utrera, A., Nogales, F.J., Uppal, R., 2020. A transaction-cost perspective on the multitude of firm characteristics. Rev. Financ. Stud. 33 (5), 2180–2222.

DeMiguel, V., Martin-Utrera, A., Uppal, R., 2024. A multifactor perspective on volatility-managed portfolios. J. Finance in press.

Detzel, A.L., Novy-Marx, R., Velikov, M., 2023. Model comparison with transaction costs. J. Finance 78 (3), 1743–1775.

Edelen, R.M., Evans, R.B., Kadlec, G.B., 2007. Scale effects in mutual fund performance: The role of trading costs. Available at SSRN 951367.

Fama, E.F., French, K.R., 1993. Common risk factors in the returns on stocks and bonds. J. Financ. Econ. 33, 3–56.

Fama, E.F., French, K.R., 2015. A five-factor asset pricing model. J. Financ. Econ. 116 (1), 1–22.

Fama, E.F., French, K.R., 2018. Choosing factors. J. Financ. Econ. 128 (2), 234–252.

Ferson, W.E., Siegel, A.F., Wang, J.L., 2019. Asymptotic variances for tests of portfolio efficiency and factor model comparisons with conditioning information. Available at SSRN 3330663.

Frazzini, A., Israel, R., Moskowitz, T.J., 2015. Trading costs of asset pricing anomalies. Fama-Miller working paper.

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

Gârleanu, N., Pedersen, L.H., 2013. Dynamic trading with predictable returns and transaction costs. J. Finance 68 (6), 2309–2340.

Gârleanu, N., Pedersen, L.H., 2022. Active and passive investing: Understanding Samuelson's dictum. Rev. Asset Pricing Stud. 12 (2), 389–446.

Gibbons, M.R., Ross, S.A., Shanken, J., 1989. A test of the efficiency of a given portfolio. Econometrica 57, 1121–1152.

Goyal, A., He, Z.L., Huh, S.-W., 2018. Distance-based metrics: A Bayesian solution to the power and extreme-error problems in asset-pricing tests. Swiss Finance Institute Research Paper.

Hayashi, F., 2000. Econometrics. Princeton University Press.

Hou, K., Mo, H., Zhang, L., 2021. An augmented q-factor model with expected growth. Rev. Finance 25 (1), 1–41.

Hou, K., Xue, C., Zhang, L., 2015. Digesting anomalies: An investment approach. Rev. Financ. Stud. 28 (3), 650–705.

Jensen, T.I., Kelly, B.T., Pedersen, L.H., 2022. Machine learning and the implementable efficient frontier. Available at SSRN 4187217.

Kan, R., Robotti, C., 2008. Specification tests of asset pricing models using excess returns. J. Empir. Financ. 15 (5), 816–838.

Kan, R., Robotti, C., 2009. Model comparison using the Hansen-Jagannathan distance. Rev. Financ. Stud. 22 (9), 3449–3490.

Kan, R., Robotti, C., Shanken, J., 2013. Pricing model performance and the two-pass cross-sectional regression methodology. J. Finance 68 (6), 2617–2649.

Kan, R., Wang, X., Zheng, X., 2019. In-sample and out-of-sample sharpe ratios of multi-factor asset pricing models. Available at SSRN 3454628.

Korajczyk, R.A., Sadka, R., 2004. Are momentum profits robust to trading costs? J. Finance 59 (3), 1039–1082.

Ledoit, O., Wolf, M., 2004. A well-conditioned estimator for large-dimensional covariance matrices. J. Multivariate Anal. 88, 365–411.

Lee, C.M., Ready, M.J., 1991. Inferring trade direction from intraday data. J. Finance 46 (2), 733–746.

Lintner, J., 1965. Security, risk, and maximal gains from diversification. J. Finance 20, 587–615.

Maruyama, Y., Seo, T., 2003. Estimation of moment parameter in elliptical distributions. J. Japan Statist. Soc. 33 (2), 215–229.

Moreira, A., Muir, T., 2017. Volatility-managed portfolios. J. Finance 72 (4), 1611–1644.

Muravyev, D., Pearson, N.D., Pollet, J.M., 2022. Anomalies and their short-sale costs. Available at SSRN 4266059.

Nagel, S., 2005. Short sales, institutional investors and the cross-section of stock returns. J. Financ. Econ. 78 (2), 277–309.

Novy-Marx, R., Velikov, M., 2016. A taxonomy of anomalies and their trading costs. Rev. Financ. Stud. 29 (1), 104–147.

Ross, S., 1976. The arbitrage theory of capital asset pricing. J. Econom. Theory 13 (3), 341–360.

Sharpe, W.F., 1964. Capital asset prices: A theory of market equilibrium condition of risk. J. Finance 19, 425–442.

Wolak, F.A., 1987. An exact test for multiple inequality and equality constraints in the linear regression model. J. Amer. Statist. Assoc. 82 (399), 782–793.

Wolak, F.A., 1989. Testing inequality constraints in linear econometric models. J. Econometrics 41 (2), 205–235.