Comparing Factor Models with Price-Impact Costs*

Sicong (Allen) Li Victor DeMiguel Alberto Martín-Utrera

January 15, 2023

Abstract

The Sharpe-ratio criterion proposed by Barillas and Shanken (2017) to compare factor models is insufficient in the presence of *price-impact costs* because the efficient frontier spanned by the factors is nonlinear. Instead, we propose a statistical test to compare factor models in terms of mean-variance utility net of price-impact costs. Empirically, model performance depends not only on the turnover required to rebalance the factor portfolios, but also on the liquidity of the stocks traded and the absolute risk-aversion parameter. The q-factor model, the six-factor Fama-French model, and a high-dimensional model outperform for high, medium and low absolute risk aversion, respectively.

Keywords: trading costs, mean-variance utility, statistical test.

*Li, London Business School, e-mail: `ali@london.edu`, DeMiguel, London Business School, e-mail: `avmiguel@london.edu`; Martín-Utrera, Iowa State University, e-mail: `amutrera@iastate.edu`. We are grateful for comments from Pedro Barroso, Svetlana Bryzgalova, Andrew Chen, Andrew Detzel, Raymond Kan, Nicos Savva, Stephen Schaefer, Raman Uppal, Mihail Velikov, Guofu Zhou, Paolo Zaffaroni and seminar participants at AFA PhD poster session, FMA Annual Meeting, INFORMS Annual Meeting, International Young Finance Scholars' Conference, and Trans-Atlantic Doctoral Conference.

1 Introduction

Barillas and Shanken (2017) show that the squared Sharpe ratio is a sufficient statistic to compare factor models *in the absence of trading costs* and Detzel, Novy-Marx, and Velikov (2021) use the squared Sharpe ratio of *net* returns to compare models in the presence of *proportional transaction costs*. We show, however, that the squared Sharpe ratio criterion is no longer a sufficient statistic *in the presence of price-impact costs* that grow faster than linearly in the amount traded because, in this case, the efficient frontier spanned by the factors in the model is *nonlinear*. Instead, we propose comparing models in terms of mean-variance utility net of price-impact costs—which represent the lion's share of the trading costs incurred by institutional investors—and develop a formal statistical test to compare nested and non-nested factor models. Empirically, we find that model performance in the presence of price-impact costs depends not only on the portfolio turnover required to rebalance the factors, but also on the liquidity of the stocks that have to be traded and the absolute risk-aversion parameter.

A popular approach to compare asset-pricing models is the GRS test of Gibbons, Ross, and Shanken (1989), which evaluates the ability of the factors in the model to span the efficient frontier of certain test assets. Specifically, 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 not only to the factors in the model, but also the test assets. Moreover, Barillas and Shanken (2017) show that the model whose factors produce the largest squared Sharpe ratio is also the one that best spans the efficient frontier of the test assets, and thus, test assets are irrelevant and instead it is sufficient to compare factor models in terms of their squared Sharpe ratio.

Detzel et al. (2021) point out that one has to account for *trading costs* when comparing factor models because 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. Therefore, Detzel et al. (2021) compare factors models in terms of their squared Sharpe ratio of returns net of *proportional* transaction costs.

In this paper we propose a methodological framework to compare factor models in the presence of *price-impact costs*, which are more substantial than proportional costs for the large institutional investors that manage most of the capital in financial markets. Indeed, Gârleanu and Pedersen (2022) show that institutional investors held around 50% of the US equity market in 2017 and Edelen, Evans, and Kadlec (2007) show that price-impact costs represent around 65% of the total trading costs of mutual funds, whereas proportional transaction costs associated with bid-ask spreads represent only 17%. Therefore, a relevant criterion to compare factor models must account for price-impact costs.

Our contribution is threefold. First, we show that in the presence of price-impact costs that grow faster than linearly with the amount traded the efficient frontier is strictly concave, and therefore, the squared Sharpe ratio criterion used by Barillas and Shanken (2017) and Detzel et al. (2021) is no longer sufficient to compare factor models. This is because each portfolio in the efficient frontier has a different Sharpe ratio of returns net of price-impact costs, and therefore a single Sharpe ratio does not characterize the investment opportunity set. Instead, we propose using *mean-variance utility net of price-impact costs* as the comparison criterion. This criterion is economically motivated because it captures the lion's share of the trading costs faced by the institutional investors that manage most of the capital in financial markets. In addition, our criterion is equivalent to the squared Sharpe ratio in the cases with proportional costs and without trading costs. We generalize the result of Barillas and Shanken (2017) to show that test assets are irrelevant for model-comparison purposes *also* in the presence of price-impact costs, and therefore it is sufficient to compare factor models in terms of their maximum mean-variance utility net of price-impact costs.

Our second contribution is to develop a statistical methodology to test the significance of the difference between the mean-variance utilities net of price-impact costs of two factor models. In particular, we derive two different asymptotic distributions that allow us to compare two factor models for the cases when they are nested or non-nested. Our approach extends the tests of Kan and Robotti (2009) and Barillas, Kan, Robotti, and Shanken (2020) to address the case with price-impact costs. We also develop closed-form expressions for the variance of the asymptotic distribution and use them 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 when the mean-variance portfolio returns of the two models are positively correlated, the mean-variance portfolio return of each model is positively correlated with the rebalancing trades of the portfolio of the other model, and when the rebalancing trades of the two portfolios are positively correlated.

Our third contribution is to use our statistical test to compare the empirical performance of five different factor models. We consider four prominent low-dimensional models: the q-factor model of Hou, Xue, and Zhang (2015), HXZ4, the four-factor model of Fama and French (1993) and Carhart (1997), FFC4, the five-factor model of Fama and French (2015), FF5, and the six-factor model of Fama and French (2018), FF6. 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 fifth 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 costs, model performance depends not only on the portfolio turnover required to trade the factors in the model, as pointed out by Detzel et al. (2021) for the case with proportional costs, but also on the liquidity of the stocks traded. In particular, we find that, compared to their FF6 counterparts, the HXZ4 investment and profitability factors not only involve higher portfolio turnover, but also require trading stocks with smaller market capitalization, which are more illiquid and subject to larger price-impact costs. As a result, while in the absence of trading costs the four-factor model of Hou et al. (2015) outperforms the six-factor model of Fama and French (2018), in the presence of price-impact costs the six-factor model of Fama and French (2018) tends to perform better.

Second, the relative performance of factor models in the presence of price-impact costs depends on the absolute risk-aversion parameter. For instance, the high-dimensional model of DeMiguel et al. (2020) significantly outperforms the low-dimensional models of Hou et al. (2015) and Fama and French (2018) *only* for the case with low absolute risk aversion. This is because absolute risk aversion determines the relative weight of variance risk versus mean return in the utility function. A lower absolute risk-aversion parameter implies that investors are more willing to take on larger positions to increase their mean return at the expense of higher return variance. However, by increasing their positions, investors also have to trade more, and hence face higher price-impact costs. Thus, high-dimensional models, which provide larger trading-diversification benefits, tend to outperform low-dimensional models for low absolute risk aversion because in this case price-impact costs are relatively more important. Overall, accounting for price-impact costs results in a more nuanced comparison of the various factor models we consider—the q-factor model of Hou et al. (2015), the six factor model of Fama and French (2018), and the high-dimensional model of DeMiguel et al. (2020) are the best performing models for high, medium, and low absolute risk aversion, respectively. As a robustness check, we discuss in Section 5.5 the results from the out-of-sample bootstrap tests proposed by Fama and French (2018) and used by Detzel et al. (2021) and we observe that the results are consistent with the empirical findings from our statistical tests.

Our work is closely related to Detzel et al. (2021), who compare prominent asset-pricing models in the presence of proportional transaction costs using the maximum squared Sharpe ratio criterion of Barillas and Shanken (2017). We show that the squared Sharpe ratio criterion is no longer sufficient in the presence of price-impact costs and, instead, we compare 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.

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, Robotti, and Shanken, 2013; Barillas and Shanken, 2018; Goyal, He, and Huh, 2018; Fama and French, 2018; Ferson, Siegel, and Wang, 2019; Chib, Zeng, and Zhao, 2020; Kan, Wang, and Zheng, 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.

Our work is also related to the literature on the profitability of factor strategies (Korajczyk and Sadka, 2004; Novy-Marx and Velikov, 2016; Frazzini, Israel, and Moskowitz, 2018; Chen and Velikov, 2022; 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 smaller when the factors are combined.[1] 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 describes the data. Section 3 proposes mean-variance utility net of price-impact costs as a criterion to compare factor models. Section 4 develops a formal statistical test to compare factor models in the presence of price-impact costs. Section 5 compares empirically five factor models from the literature. Section 6 concludes. Appendix A contains the proofs of all theoretical results. The Internet Appendix contains several robustness checks and additional information.

2 Data

We download data for the 28 tradable factors listed in Table 1. Our sample spans the period from January 1980 to December 2020. We consider nine factors included in prominent low-dimensional asset-pricing models. In particular, we construct the market (MKT), size (SMB), value (HML), profitability (RMW) and investment (CMA) factors of Fama and French (2015), the momentum (UMD) factor of Carhart (1997), and the profitability (ROE), investment (IA), and size (ME) factors of Hou et al. (2015). We construct the market factor as the excess return of the value-weighted market portfolio and the rest of the factors as the returns of value-weighted long-short portfolios obtained from double or triple sorts on firm characteristics following the procedure in the papers that originally proposed the factors.

DeMiguel et al. (2020) provide an economic rationale based on trading costs to consider high-dimensional factor models. Moreover, in their Appendix IA.2, they propose a model containing 20 factors, including the market, that are statistically significant in the presence of price-impact costs. Therefore, we construct the 19 factors (other than the market) in the model of DeMiguel et al. (2020) 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, Santa-Clara, and Valkanov (2009) and DeMiguel

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

Table 1: List of characteristics considered

This table lists the 28 factors we consider. Panel A lists nine factors that replicate those in prominent asset-pricing models, including the market. Other than the market, each of these factors are constructed as value-weighted portfolios obtained from double or triple sorts on firm characteristics. Panel B lists 19 factors constructed using value-weighted 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 number, 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 date and journal of publication, respectively.

#	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	Robust-minus-weak: size-neutral portfolio that is long stocks with high operating profitability and is short stocks with low operating profitability.	RMW	Fama & French	2015, JFE
5	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
6	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
7	Return on equity: portfolio that is long stocks with high profitability and is short stocks with low profitability.	ROE	Hou, Xue & Zhang	2015, RFS
8	Investment: portfolio that is long stocks with high investment and is short stocks with low investment.	IA	Hou, Xue & Zhang	2015, RFS
9	Size: portfolio that is long stocks with low market capitalization and is short stocks with large market capitalization.	ME	Hou, Xue & Zhang	2015, RFS

Table 1 continued

#	Definition	Acronym	Author(s)	Date and journal
Panel B: Factors constructed from single sorts				
10	Asset growth: Annual percent change in total assets	agr	Cooper, Gulen & Schill	2008, JF
11	Cash productivity: Fiscal year-end market capitalization plus long term debt minus total assets divided by cash and equivalents	cashpr	Chandrashekar & Rao	2009 WP
12	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
13	Change in shares outstanding: Annual percent change in shares outstanding	chcsho	Pontiff & Woodgate	2008, JF
14	Convertible debt indicator: An indicator equal to 1 if company has convertible debt obligations	convind	Valta	2016 JFQA
15	Change in common shareholder equity: Annual percent change in equity book value	egr	Richardson, Sloan, Soliman & Tuna	2005, JAE
16	Earnings to price: Annual income before extraordinary items divided by end of fiscal year market cap	ep	Basu	1977, JF
17	Gross profitability: Revenues minus cost of goods sold divided by lagged total assets	gma	Novy-Marx	2013, JFE
18	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
19	Industry momentum: Equal weighted average industry 12-month returns	indmom	Moskowitz & Grinblatt	1999, JF
20	Financial-statements score: Sum of 9 indicator variables to form fundamental health score	ps	Piotroski	2000, JAR
21	R&D to market cap: R&D expense divided by end-of-fiscal-year market cap	rd_mve	Guo, Lev & Shi	2006, JBFA
22	Return volatility: Standard deviation of daily returns from month $t-1$	retvol	Ang, Hodrick, Xing & Zhang	2006, JF
23	Return on assets: Income before extraordinary items divided by one quarter lagged total assets	roaq	Balakrishnan, Bartov & Faurel	2010, JAE
24	Annual sales growth: Annual percent change in sales	sgr	Lakonishok, Shleifer & Vishny	1994, JF
25	Volatility of share turnover: Monthly standard deviation of daily share turnover	std_turn	Chordia, Subrahmanyan & Anshuman	2001, JFE
26	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
27	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
28	Zero trading days: Turnover weighted number of zero trading days for most recent month	zerotrade	Liu	2006, JFE

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 with a value of the characteristic above the 70th percentile and a short value-weighted portfolio of stocks with a value of characteristic below the 30th percentile.

3 Comparing factor models with trading costs

In this section, we propose a novel criterion to compare factor models in the presence of price-impact costs. Section 3.1 gives the notation and assumptions that we use for the analysis. Section 3.2 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 3.3 we demonstrate that this criterion is also valid in the presence of *proportional* transaction costs. In Section 3.4, however, we show that the squared Sharpe ratio criterion is no longer sufficient in the presence of price-impact costs and in Section 3.5 we propose comparing factor models in terms of their mean-variance utility net of trading costs.

3.1 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 $r_t \in \mathbb{R}^N$. Let $X_t \in \mathbb{R}^{N\times K}$ be the matrix whose columns contain the portfolio weights of the K factors at time t. Then, the vector of returns of the K factors at time $t+1$ is

$$R_{t+1} = X_t^\top r_{t+1} \in \mathbb{R}^K. \tag{1}$$

In addition, $\mu = E[R_t]$ and $\Sigma = \mathrm{var}(R_t)$ are the mean and the covariance matrix of factor returns, respectively.

We now define the mean-variance portfolio of the factors, $\theta^* \in \mathbb{R}^K$, as 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,[2] $f(\theta)$ is the trading costs associated with θ, $\theta^\top \Sigma \theta$ is the portfolio variance, and γ is the absolute risk-aversion parameter.[3]

We define the rebalancing-trade matrix of the K factors at time t as

$$\tilde{X}_t = X_t - \text{diag}(e + r_t) X_{t-1}, \tag{3}$$

where e is the N-dimensional vector of ones and $\text{diag}(x)$ is a diagonal matrix whose main diagonal contains the elements in vector x. Thus, the element in the nth row and kth column of $\tilde{X}_t$ is the rebalancing trade of factor k on stock n at time t.

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

Assumption 3.1 *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 3.2 *The trading-cost function $f(\theta)$ is such that $f(0) = 0$, $f(\theta) > 0$ for all $\theta \neq 0$, and $f(\theta)$ is continuous in θ.*

Assumption 3.2 is satisfied by most popular trading-cost models, such as proportional and quadratic trading-cost models. Finally, the following assumption rules out the trivial case in which it is not optimal to invest in any of the factors.

Assumption 3.3 *The set $S = \{\theta | \theta^\top \mu - f(\theta) > 0\}$ is non-empty.*

3.2 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) 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).

[2]More precisely, $\theta^\top \mu$ should be termed as the expected payoff. However, for exposition purposes we use the term *expected return*, which is consistent with the existing literature.

[3]The absolute risk-aversion parameter can be defined as $\gamma = \bar{\gamma}/b$, where $\bar{\gamma}$ is the relative risk-aversion parameter and b is the investor's endowment in dollars; see, for instance, Gârleanu and Pedersen (2013).

Proposition 1 *Let Assumption 3.1 hold and let $\gamma > 0$, then the unique maximizer to problem (2) with $f(\theta) = 0$ is:*

$$\theta^* = \frac{1}{\gamma}\Sigma^{-1}\mu. \tag{4}$$

Moreover, the efficient frontier is a straight line in the mean-standard-deviation diagram, and all portfolios on the efficient frontier deliver the 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 of the model. Thus, the model that best spans the efficient frontier is the one whose factors attain the highest squared Sharpe ratio as noted by Barillas and Shanken (2017).

3.3 The case with proportional trading costs

We first define the proportional-trading-cost function.

Definition 1 (Proportional-trading-cost function) *A proportional trading-cost function $f(\theta)$ is one that satisfies Assumption 3.2 and is homogeneous of degree one, that is,*

$$f(c\theta) = cf(\theta) \quad \text{for all } \theta \text{ and } c \geq 0. \tag{5}$$

Two examples of proportional trading-cost functions that satisfy this definition are those used by Detzel et al. (2021) and DeMiguel et al. (2020). Detzel et al. (2021) use the following proportional trading-cost function:

$$f(\theta) = E\Big[\sum_{n=1}^{N} \kappa_{n,t} \sum_{k=1}^{K} |\tilde{x}_{n,k}^t \theta_k|\Big], \tag{6}$$

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 $\tilde{X}_t$ defined in (3), and $\kappa_{n,t} > 0$ is the proportional trading-cost parameter of the nth stock at time t. DeMiguel et al. (2020) use the following proportional trading-cost function:

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

An advantage of the proportional trading-cost function (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.[4]

Solving problem (2) with proportional trading costs for different values of the risk-aversion parameter γ, one can recover the efficient frontier in the presence of proportional trading costs, which, as the following proposition shows, is a straight line in the mean-standard-deviation diagram.[5]

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, $SR_p < SR = \sqrt{\mu^\top \Sigma^{-1} \mu}$, 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 efficient frontier spanned by the factors in the presence of proportional trading costs is fully characterized by the Sharpe ratio of returns net of costs of any mean-variance portfolio on the efficient frontier. Thus, the model comparison criterion based on the maximum squared Sharpe ratio remains valid. Detzel et al. (2021) use this criterion to compare the empirical performance of several prominent factor models in the presence of proportional costs. However, proportional costs ignore the price impact of large trades. In the next section, we show that the squared Sharpe ratio criterion is not sufficient in the presence of price-impact costs.

3.4 The case with price-impact costs

We now consider the case with price-impact costs. First, we define the price-impact cost function.

[4] As a robustness check, Detzel et al. (2021) also consider the proportional trading-cost function (7) in Section 6.2.

[5] 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.

Definition 2 (Price-impact-cost function) *A price-impact-cost function $f(\theta)$ satisfies Assumption 3.2 and the following inequality:*

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

Solving problem (2) with a price-impact cost function for different values of γ, one can recover the efficient frontier in the presence of of price-impact costs.

Proposition 3 *Let $f(\theta)$ be a price-impact cost function. Then, the efficient frontier in the presence of price-impact costs is a strictly concave function in the mean-standard-deviation diagram, and all portfolios on the efficient frontier deliver different Sharpe ratios of returns net of price-impact costs. In addition, the Sharpe ratios of returns net of price-impact costs of the portfolios on the efficient frontier are all smaller than the maximum Sharpe ratio in the absence of trading costs, that is, $SR_{PIC}(\gamma) < SR = \sqrt{\mu^\top \Sigma^{-1} \mu}$ for all γ.*

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 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.

We now specify the particular price-impact-cost function we use in this manuscript. A common assumption in the literature is that the impact on prices from large trades is linear in the amount traded and thus price-impact costs are quadratic (Korajczyk and Sadka, 2004; Novy-Marx and Velikov, 2016). Under this assumption, the price-impact cost, in dollars, required to rebalance the factor portfolio θ at time t is

$$\frac{1}{2}\theta^\top \tilde{X}_t^\top D_t \tilde{X}_t \theta, \tag{9}$$

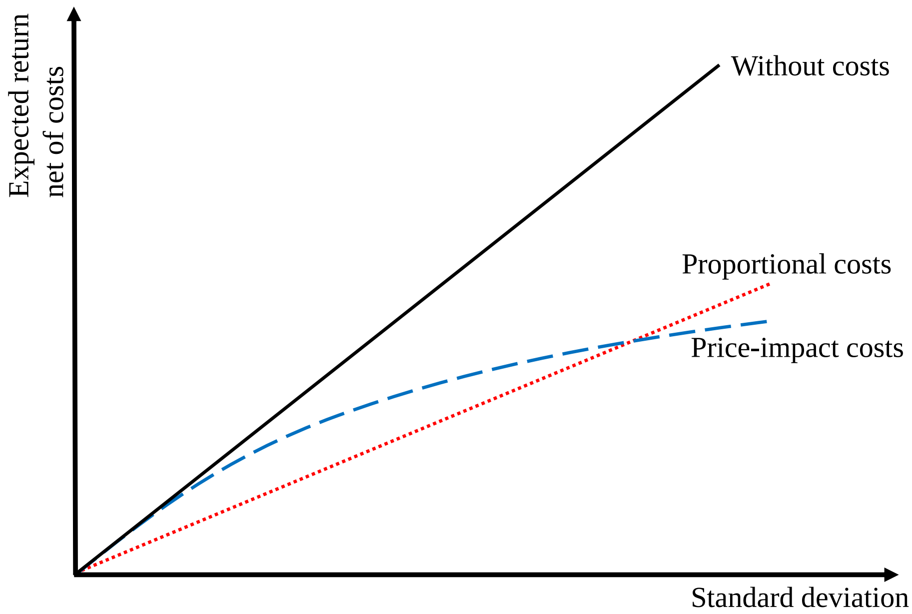
where $D_t \in \mathbb{R}^{N\times N}$ is a diagonal matrix whose nth element, $d_{n,t} > 0$, is the price-impact-cost parameter of stock n at time t.

For notational simplicity, let us define

$$\Lambda_t^* = \tilde{X}_t^\top D_t \tilde{X}_t \in \mathbb{R}^{K\times K} \tag{10}$$

Figure 1: Efficient frontiers for different trading-cost functions

This figure illustrates the efficient frontiers of the factor model in the presence of different trading-cost functions. The black solid line, the red dotted line, and the blue dashed line depict the efficient frontiers in the absence of trading costs, presence of proportional costs, and presence of price-impact costs, respectively.



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

$$f(\theta) = E\left[\frac{\theta^\top \Lambda_t^* \theta}{2}\right] = \frac{\theta^\top \Lambda^* \theta}{2}, \tag{11}$$

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

In summary, Figure 1 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.

3.5 Mean-variance utility as a comparison criterion

In the presence of price-impact costs, the efficient frontier is strictly concave and thus a single squared Sharpe ratio no longer characterizes the efficient frontier as in the cases without transaction costs or with proportional transaction costs. Therefore, 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 sufficient to describe the extent to which the factors of a model span the efficient frontier. Instead, we propose comparing factor models in terms of mean-variance utility net of trading costs.

Barillas and Shanken (2017) posit that the preferred factor model should be able to span not only the investment opportunity set offered by the tests assets, but also 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

$$SR^2([\Pi, f_A, f_B]) - SR^2(f_A) < SR^2([\Pi, f_A, f_B]) - SR^2(f_B), \tag{12}$$

where $SR^2(x)$ is the squared Sharpe ratio delivered by the assets in vector x. This indicates that 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

$$SR^2(f_A) > SR^2(f_B), \tag{13}$$

and thus Barillas and Shanken (2017) show that it is sufficient to compare models in terms of squared Sharpe ratio.

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. Therefore, the portfolios in the efficient frontier that maximize an investor's mean-variance utility are equivalent to those that maximize the Sharpe ratio. However, in the presence of price-impact costs, the Sharpe ratio is no longer a sufficient measure of utility because investors may find it optimal to form a portfolio of factors that does not maximize Sharpe ratio. Accordingly, we propose comparing factor models in terms of investor's utility:

$$U^\gamma([\Pi, f_A, f_B]) - U^\gamma(f_A) < U^\gamma([\Pi, f_A, f_B]) - U^\gamma(f_B), \tag{14}$$

where γ is the absolute risk aversion of the investor, and $U^\gamma(x)$ is the maximum mean-variance utility in the presence of trading costs of an investor with access to the assets in x, that is, the optimal value of the objective function in (2). Applying the logic of Barillas and Shanken (2017) to the case with price-impact costs, we have that model A is better than model B if

$$U^\gamma(f_A) > U^\gamma(f_B), \tag{15}$$

which shows that test assets are irrelevant *also* when comparing factor models in terms of mean-variance utility net of trading costs. Consequently, the preferred model is the one whose factors generate the highest mean-variance utility net of trading costs.

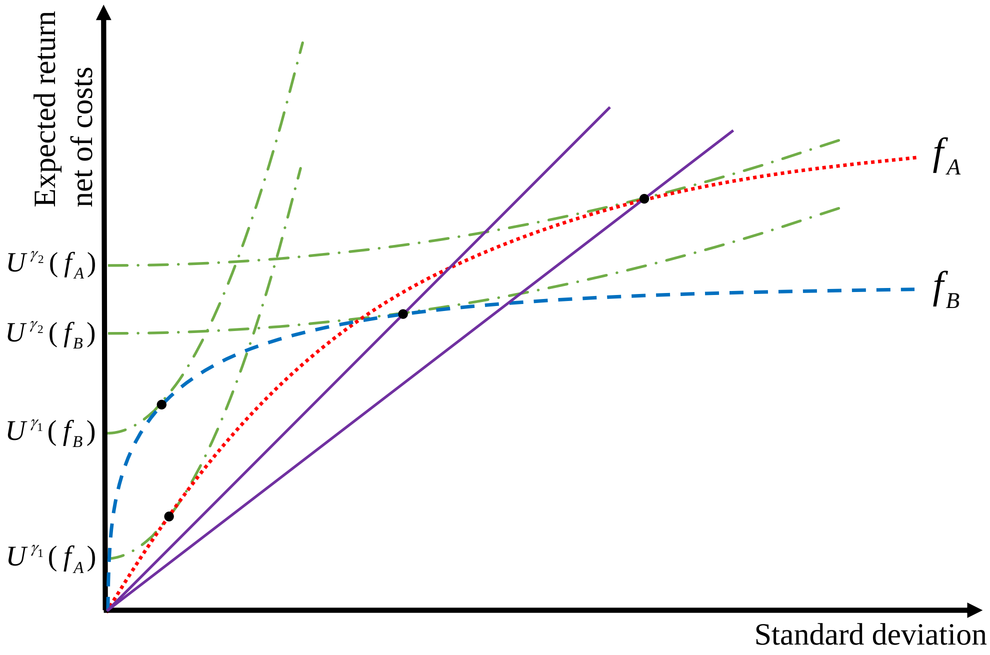
Note that the relative performance of two factor models in the presence of price-impact costs depends on the absolute risk-aversion parameter. This is because the investor's absolute risk-aversion parameter determines the importance of portfolio risk relative to the average portfolio return net of price-impact costs. For instance, consider two factor models, A and B, and assume that the factors in model B generate a higher Sharpe ratio in the absence of trading costs, but they also generate higher price-impact costs as the amount traded increases. Then, it is possible that investors with high absolute risk aversion prefer model B, while those with low absolute risk aversion prefer model A.[6] In particular, investors with low absolute risk aversion are willing to take on larger 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, low absolute risk-aversion investors may prefer the factors in model A.

This example is illustrated in Figure 2, which depicts the efficient frontiers in the presence of price-impact costs of model A (red dotted line) and model B (blue dashed line). The figure also depicts the indifference curves of two investors with different absolute risk-aversion parameters (green dash-dotted lines). The first investor has higher absolute risk aversion, and thus her indifference curves are steeper. The investor's optimal portfolio corresponds with the tangent between the investor's indifference curve and the efficient frontier. The figure shows that $U^{\gamma_1}(f_B) > U^{\gamma_1}(f_A)$, but $U^{\gamma_2}(f_A) > U^{\gamma_2}(f_B)$; that is, the first investor

[6]As mentioned in Footnote 3, the absolute risk-aversion parameter is often defined as the ratio of the relative risk-aversion parameter to the investor's endowment; see, for instance, Gârleanu and Pedersen (2013). Thus, retail investors are likely to have high absolute risk aversion, and thus may prefer model B, while institutional investors are likely to have lower absolute risk aversion, and thus may prefer model A.

Figure 2: Investment opportunity sets of models with factors f_A and f_B

This figure illustrates the efficient frontiers net of price-impact costs of model A (red dotted line) and model B (blue dashed line). The figure also depicts the indifference curves of the two investors (green dash-dotted lines). The first investor has higher absolute risk aversion ($\gamma_1 > \gamma_2$), and thus her indifference curves are steeper. The optimal portfolio of an investor is the tangent between her indifference curves and the efficient frontier.



prefers model B and the second prefers model A. This is because the low-risk-aversion investor —that is, the second investor— is willing to take on higher risk aiming to maximize net average returns. Accordingly, the price-impact costs from exploiting the factors in model A are much lower than those from exploiting the factors in model B, and thus the investor can obtain a larger average portfolio return net of transaction costs using the factors of model A instead of those of model B.

Note that the second investor prefers model A over model B even though the tangent portfolio for model A delivers a lower Sharpe ratio of returns net of price-impact costs than model B– see the slope of the straight purple lines. This example illustrates one of our main insights that, in the presence of price-impact costs, the Sharpe ratio criterion is insufficient and is not equivalent to the mean-variance utility criterion.

4 Statistical tests

In this section, we develop a formal statistical methodology to test whether two factor models generate the same mean-variance utility net of price-impact costs.

4.1 The case without trading costs

To set the stage for our main result, we first consider the case without trading costs. For this case, Proposition 1 shows that the squared Sharpe ratio of any mean-variance portfolio is $SR^2 = \mu^\top \Sigma^{-1} \mu$. Moreover, it is easy to show that the mean-variance utility generated by a mean-variance portfolio is proportional to the squared Sharpe ratio,

$$U^\gamma = \frac{\mu^\top \Sigma^{-1} \mu}{2\gamma} = \frac{SR^2}{2\gamma}. \tag{16}$$

Therefore, as explained in Section 3.5, the mean-variance utility and the squared Sharpe ratio criteria to compare factor models are equivalent in the absence of trading costs.

Barillas et al. (2020) derive the asymptotic distribution of the sample squared Sharpe ratio generated by the factors of a model as well as that of the difference of the sample squared Sharpe ratios generated by the different factors of two models. For completeness, we state their result after introducing the following assumption.

Assumption 4.1 *Factor returns are serially independent.*

Assumption 4.1 is made for simplicity, and in Appendix A.3 we show that it can be relaxed by adjusting the variance of the asymptotic distribution of the sample squared Sharpe ratio that we introduce in the following proposition.

Proposition 4 (Barillas et al., 2020) *Let Assumptions 3.1, 3.3 and 4.1 hold. Then, the asymptotic distribution of $\widehat{SR}^2$, the sample estimator of SR^2, is*

$$\sqrt{T}(\widehat{SR}^2 - SR^2) \overset{A}{\sim} N(0, E[h_t^2]), \tag{17}$$

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

$$h_t = 2\mu^\top \Sigma^{-1}(R_t - \mu) - \mu^\top \Sigma^{-1} \Sigma_t \Sigma^{-1} \mu + SR^2, \tag{18}$$

and $\Sigma_t = (R_t - \mu)(R_t - \mu)^\top$. *In addition, the asymptotic distribution of the difference between the sample squared Sharpe ratios of two factor models A and B is*

$$\sqrt{T}\left([\widehat{SR}_A^2 - \widehat{SR}_B^2] - [SR_A^2 - SR_B^2]\right) \overset{A}{\sim} N\left(0, E\big[(h_{t,A} - h_{t,B})^2\big]\right), \tag{19}$$

provided that $E\big[(h_{t,A} - h_{t,B})^2\big] > 0$, *where* $h_{t,A}$ *and* $h_{t,B}$ *are given by applying equation* (18) *to models* A *and* B, *respectively.*

4.2 The case with price-impact costs

In this section, we derive two asymptotic distributions for the difference in mean-variance utility net of price-impact costs of two factor models. We then show 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.

4.2.1 Notation and an assumption

To develop our statistical test, we assume price-impact costs are quadratic as in Equation (11). Then, the mean-variance problem (2) can be rewritten as

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

where $\Lambda = \Lambda^*/\gamma$. Thus, the mean-variance portfolio is

$$\theta^* = \frac{1}{\gamma}(\Sigma + \Lambda)^{-1}\mu \tag{20}$$

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

$$U_\Lambda^\gamma = \frac{\mu^\top(\Sigma + \Lambda)^{-1}\mu}{2\gamma}, \tag{21}$$

which is *not* proportional to the squared Sharpe ratio of the factors in the absence of trading costs. More precisely, price-impact costs affect the investor's utility in a nonlinear way by replacing the matrix Σ in (16) with the matrix $(\Sigma + \Lambda) = (\Sigma + \Lambda^*/\gamma)$, which depends on γ.

For simplicity, we make the following assumption, which, like Assumption 4.1, can also be relaxed by adjusting the asymptotic variance of the mean-variance utility net of price-impact costs.

Assumption 4.2 *Each column of the rebalancing-trade matrix* $\tilde{X}_t$ *is serially independent.*

4.2.2 Two asymptotic distributions

In this section, we derive two different 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 3.1–3.3 and 4.1–4.2 hold. Then, the asymptotic distribution of the sample estimator of the mean-variance utility net of price-impact costs in (21) is*

$$\sqrt{T}(\hat{U}^{\gamma}_{\Lambda} - U^{\gamma}_{\Lambda}) \overset{A}{\sim} N(0, \frac{E[h^2_{t,\Lambda}]}{4\gamma^2}), \tag{22}$$

provided that $E[h^2_{t,\Lambda}] > 0$, where

$$h_{t,\Lambda} = 2\mu^{\top}(\Sigma + \Lambda)^{-1}(R_t - \mu) - \mu^{\top}(\Sigma + \Lambda)^{-1}(\Sigma_t + \Lambda_t)(\Sigma + \Lambda)^{-1}\mu + \mu^{\top}(\Sigma + \Lambda)^{-1}\mu, \tag{23}$$

and $\Lambda_t = \Lambda^*_t/\gamma$. 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

$$\sqrt{T}([\hat{U}^{\gamma}_{\Lambda,A} - \hat{U}^{\gamma}_{\Lambda,B}] - [U^{\gamma}_{\Lambda,A} - U^{\gamma}_{\Lambda,B}]) \overset{A}{\sim} N(0, \frac{E\big[(h_{t,\Lambda,A} - h_{t,\Lambda,B})^2\big]}{4\gamma^2}), \tag{24}$$

provided that $E\big[(h_{t,\Lambda,A} - h_{t,\Lambda,B})^2\big] > 0$, where $h_{t,\Lambda,A}$ and $h_{t,\Lambda,B}$ are given by applying equation (23) to models A and B, respectively.

Proposition 5 shows that the distribution in Equation (24) can be used to compare factor models provided that the variance of the asymptotic distribution is strictly greater than zero. However, the variance is zero under the null hypothesis, $U^{\gamma}_{\Lambda,A} = U^{\gamma}_{\Lambda,B}$, in two cases.[7] First, when model A nests model B and the extra factors of model A are redundant, and second, when models A and B overlap (i.e., share common factors) and the extra factors of both models are redundant. Therefore, we provide in Proposition 6 another asymptotic distribution whose variance is nonzero for the case with nested models. Section 4.2.3 discusses how Propositions 5 and 6 can be used to compare nested or nonnested factor models.

Proposition 6 *Let Assumptions 3.1–3.3 and 4.1–4.2 hold. Consider the nested models A and B containing factors $f_A = [f_1, f_2]$ and $f_B = f_1$, respectively, where f_1 and f_2 contain K_1*

[7] Barillas et al. (2020) discuss a similar issue for the case without transaction costs.

and K_2 mutually exclusive factors, respectively. The partition matrix of $\Sigma_A+\Lambda_A$ is defined as

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

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

$$T(\hat{U}^{\gamma}_{\Lambda,A} - \hat{U}^{\gamma}_{\Lambda,B}) \overset{A}{\sim} \sum_{i=1}^{K_2} \xi_i x_i, \tag{25}$$

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

$$\frac{E[l_t l_t^{\top}]_{22} W}{2\gamma}, \tag{26}$$

where

$$W = (\Sigma_{22} + \Lambda_{22}) - (\Sigma_{21} + \Lambda_{21})(\Sigma_{11} + \Lambda_{11})^{-1}(\Sigma_{12} + \Lambda_{12}) \quad \text{and} \tag{27}$$

$$l_t = (\Sigma_A + \Lambda_A)^{-1} R_{A,t} - (\Sigma_A + \Lambda_A)^{-1}(\Sigma_{A,t} + \Lambda_{A,t})(\Sigma_A + \Lambda_A)^{-1}\mu_A. \tag{28}$$

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.[8]

4.2.3 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.

[8]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 expand 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.

When models A and B are non-nested and have no overlapping factors, the variance of the asymptotic distribution in (24) is strictly greater than zero. Therefore, one can directly use Proposition 5 and reject the null hypothesis $U^{\gamma}_{\Lambda,A} = U^{\gamma}_{\Lambda,B}$ when $\sqrt{T}(\hat{U}^{\gamma}_{\Lambda,A} - \hat{U}^{\gamma}_{\Lambda,B})$ is greater (less) than, for instance, the 97.5th (2.5th) percentile of the probability density function of the right-hand side of (24).

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 (24) is zero. Therefore, we use Proposition 6 instead. The null hypothesis $U^{\gamma}_{\Lambda,A} = U^{\gamma}_{\Lambda,B}$ is rejected when $T(\hat{U}^{\gamma}_{\Lambda,A} - \hat{U}^{\gamma}_{\Lambda,B})$ is greater than, for instance, the 95th percentile of the probability density function of the distribution on the right-hand side of (25), 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 the same utility net of price-impact costs and it is *higher* than that of the 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 (24) is zero. Therefore, we test whether the null hypothesis holds using Proposition 6 where we define as the first model the one that contains all factors of models A and B, and as the second model the one that contains 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 may still hold in the second way, which can be tested using Proposition 5 because in this case the asymptotic variance in (24) is greater than zero.

Finally, to empirically characterize the asymptotic distribution in Proposition 5, one can replace $h_{t,\Lambda}$ in (23) with its sample counterpart, $\hat{h}_{t,\Lambda}$, which guarantees that $\sum_{t=1}^{T}(\hat{h}_{t,\Lambda,A} - \hat{h}_{t,\Lambda,B})^2/T$ is a consistent estimator of $E[(h_{t,\Lambda,A} - h_{t,\Lambda,B})^2]$. Similarly, to empirically characterize the asymptotic distribution in Proposition 6, one can replace $E[l_t l_t^\top]_{22}$ and W in (26) with their sample counterparts to obtain consistent estimators of the eigenvalues ξ_i.

4.3 The asymptotic variance

In this section, we obtain closed-form expressions for the asymptotic variances given in Proposition 5, and use them to study how the statistical properties of factor models affect the power of our proposed 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 when the mean-variance portfolio returns of the two models are positively correlated, the mean-variance portfolio return of each model is positively correlated with the rebalancing trades of the portfolio of the other model, and when the rebalancing trades of the two portfolios are positively correlated.

4.3.1 The case without trading costs

To set the stage for the case with price-impact costs, we first review the closed-form expressions provided by Barillas et al. (2020) for the asymptotic variances in Proposition 4 for the case without trading costs. For simplicity, we assume factor returns are normally distributed, but similar results can be derived for the more general case in which factor returns are elliptically distributed.

Assumption 4.3 *Factor returns follow a multivariate normal distribution with mean μ and covariance matrix Σ.*

Proposition 7 (Barillas et al., 2020) *Let Assumptions 3.1, 3.3, 4.1, 4.2, and 4.3 hold, then the asymptotic variance of the sample squared Sharpe ratio is*

$$E[h_t^2] = 4\mu^\top \Sigma^{-1}\mu + 2(\mu^\top \Sigma^{-1}\mu)^2. \tag{29}$$

Given two factor models A and B, we have that the asymptotic variance of the difference between their sample squared Sharpe ratios is

$$E[(h_{A,t} - h_{B,t})^2] = E[h_{A,t}^2] + E[h_{B,t}^2] - 2E[h_{A,t}h_{B,t}], \tag{30}$$

where $E[h_{A,t}^2]$ and $E[h_{B,t}^2]$ are given by applying (29) to models A and B, respectively, and

$$\begin{aligned} E[h_{A,t}h_{B,t}] &= 4\mu_A^\top \Sigma_A^{-1} E\big[(R_{A,t} - \mu_A)(R_{B,t} - \mu_B)^\top\big]\Sigma_B^{-1}\mu_B \\ &\quad + 2\big(\mu_A^\top \Sigma_A^{-1} E\big[(R_{A,t} - \mu_A)(R_{B,t} - \mu_B)^\top\big]\Sigma_B^{-1}\mu_B\big)^2. \end{aligned} \tag{31}$$

Equation (29) shows that the asymptotic variance of the sample squared Sharpe ratio of a model, $E[h_t^2]$, increases quadratically in the variance of the mean-variance portfolio return, $(\mu^\top \Sigma^{-1}\mu/\gamma^2)$. Then, Equations (30) and (31) show that the asymptotic variance of the *difference* between the estimated squared Sharpe ratios 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 returns of the mean-variance portfolios for the two models. That is, it is easier to reject the null hypothesis that the squared Sharpe ratios of two models are equal when the returns of the mean-variance portfolios of the two models are positively correlated.

4.3.2 The case with price-impact costs

In this section, we consider the closed-form expressions of the asymptotic variances in Proposition 5. 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

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

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 $\tilde{y}_{n,t}$ is normally distributed, although similar results can be derived for the more general case in which $\tilde{y}_{n,t}$ is elliptically distributed.

Assumption 4.4 *Each vector $\tilde{y}_{n,t}$ for $n = 1,\dots,N$ follows a multivariate normal distribution with zero mean and covariance matrix Λ_n.*

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 $u_t = \mu^\top(\Sigma+\Lambda)^{-1}R_t \in \mathbb{R}$, which is proportional to the mean-variance factor portfolio return at time t, and $v_{n,t} = \mu^\top(\Sigma+\Lambda)^{-1}\tilde{y}_{n,t} \in \mathbb{R}$, which is proportional to the total scaled rebalancing trade on stock n at time t of the mean-variance factor portfolio.

Proposition 8 *Let Assumptions 3.1–3.3 and 4.1–4.4 hold. Then,*

$$E[h_{t,\Lambda}^2] = 4\text{var}(u_t) + 2\left[\text{var}(u_t)\right]^2 + 4\sum_{n=1}^{N}\left[\text{cov}(u_t, v_{n,t})\right]^2 + 2\sum_{i=1}^{N}\sum_{j=1}^{N}\left[\text{cov}(v_{i,t}, v_{j,t})\right]^2. \quad (32)$$

Given two factor models A and B, we have

$$E[(h_{t,\Lambda,A} - h_{t,\Lambda,B})^2] = E[h_{t,\Lambda,A}^2] + E[h_{t,\Lambda,B}^2] - 2E[h_{t,\Lambda,A}h_{t,\Lambda_B}], \quad (33)$$

where $E[h_{t,\Lambda,A}^2]$ *and* $E[h_{t,\Lambda,B}^2]$ *are given by applying (32) to models A and B, respectively, and*

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

$$+ 2\sum_{n=1}^{N}\left(\left[\text{cov}(u_t^A, v_{n,t}^B)\right]^2 + \left[\text{cov}(u_t^B, v_{n,t}^A)\right]^2\right). \quad (34)$$

Equation (32) shows that the asymptotic variance of the sample mean-variance utility net of price-impact costs increases with the variance of the mean-variance portfolio returns, $\text{var}(u_t)$, similar to the case without transaction costs, but it also increases with the covariance between the returns and the rebalancing trades on each firm of the mean-variance portfolio, $\text{cov}(u_t, v_{n,t})$, and with the covariance between the rebalancing trades on different firms of the mean-variance portfolio, $\text{cov}(v_{i,t}, v_{j,t})$.

Equations (33) and (34) show that the asymptotic variance of the difference between the estimated 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, $\text{cov}(u_t^A, u_t^B)$, similar to the case without costs. In addition, the asymptotic variance of the difference decreases with the covariance between the mean-variance portfolio return of one model and the rebalacing trades of the mean-variance portfolio of the other model, $\text{cov}(u_t^A, v_{n,t}^B)$ and $\text{cov}(u_t^B, v_{n,t}^A)$, and with the covariance between the rebalacing trades of the mean-variance portfolios of the two models, $\text{cov}(v_{i,t}^A, v_{j,t}^B)$. That is, 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 positively correlated, the mean-variance portfolio return of each model is positively correlated with the rebalancing trades of the portfolio of the other model, and when the rebalancing trades of the two portfolios are positively correlated.[9]

5 Empirical results

In this section, we use the asymptotic distributions derived in the previous section to compare the empirical performance of five factor models in the presence of price-impact costs. Section 5.1 describes how we estimate the price-impact cost incurred by different portfolios. Section 5.2 reports summary statistics for the 28 factors listed in Table 1. Section 5.3 describes the factor models we compare. Section 5.4 compares the different factor models using the statistical tests introduced in Section 4. Finally, as a robustness check Section 5.5 compares the out-of-sample performance of the different factor models using the bootstrap approach of Fama and French (2018).

5.1 Calibration of price-impact costs

We explain in this section how we calibrate the price-impact cost model in Equation (11), which is required for the computation of price-impact costs incurred by the factors. In particular, we estimate the value of the price-impact parameter of the nth stock at time t, $d_{n,t}$ 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. Novy-Marx and Velikov (2016) show that the R-squared of a cross-sectional regression of log transaction-cost parameters on log market capitalization is 70% and the slope is statistically indistinguishable from minus one. This suggests that a reasonable approximation to the cross-sectional variation of price-impact cost parameters is to assume they are inversely proportional to market capitalization. Moreover, Novy-Marx and Velikov (2016) show that the average price elasticity of supply, defined as the product between the transaction costs parameter and market capitalization is about 6.5. Based on this evidence we model the price-impact cost parameter of the nth firm at time t as $d_{n,t} = 6.5/me_{n,t}$, where $me_{n,t}$ is the market capitalization of the nth firm at time t.

[9]Note that 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 Propositions 7 and 8.

Table 2: Factor summary statistics

This table reports the several summary statistics of the factors. The first column gives the acronym of the factor. The second and third columns give the average monthly *gross* return of the factor and its t-statistic, respectively. The fourth and fifth columns give the average monthly *net-of-price-impact-costs* return of the factor and its t-statistic, respectively, when one invests one billion dollars on each leg of the factor. The sixth column gives the factor's monthly turnover (TO), and the seventh column gives the factor's monthly price-impact cost (PIC). The eighth 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, both in billions of dollars. Average returns, turnovers, and price-impact costs are reported in percentage. Our sample spans the period from January 1980 to December 2020.

	Gross returns		Net returns		Turnover/cost		Trade-weighted market cap	
	Average (%)	t-statistic	Average (%)	t-statistic	TO (%)	PIC (%)	Monthly	June
Panel A: Market and factors constructed from double and triple sorts								
MKT	0.705	3.466	0.705	3.464	2.27	0.000	63.020	62.035
SMB	0.086	0.638	0.051	0.377	8.08	0.035	23.967	16.047
HML	0.163	1.198	0.065	0.475	10.84	0.098	25.321	16.711
RMW	0.348	3.283	0.242	2.262	10.76	0.106	24.428	19.196
CMA	0.240	2.663	0.051	0.531	15.35	0.188	32.450	28.020
UMD	0.557	2.746	0.050	0.245	52.08	0.508	29.522	23.662
ROE	0.521	4.398	−0.321	−2.540	35.42	0.842	20.564	17.467
IA	0.286	3.313	−0.128	−1.223	24.62	0.413	23.031	20.790
ME	0.147	1.109	0.003	0.023	19.20	0.144	22.266	17.772
Panel B: Factors constructed from single sorts								
agr	0.163	1.367	0.075	0.620	15.17	0.089	55.326	51.258
cashpr	0.013	0.092	−0.008	−0.060	7.98	0.021	54.286	54.029
chatoia	0.165	2.103	0.061	0.766	16.40	0.104	58.503	52.970
chcsho	0.297	2.855	0.221	2.108	13.91	0.076	56.506	58.969
convind	0.098	1.035	0.080	0.844	6.25	0.018	49.039	36.753
egr	0.164	1.498	0.085	0.772	15.04	0.079	54.814	51.196
ep	0.213	1.188	0.122	0.681	14.36	0.091	44.656	43.493
gma	0.220	1.676	0.205	1.560	6.71	0.015	56.476	47.696
idiovol	0.203	0.731	0.013	0.050	11.35	0.189	32.156	29.671
indmom	0.210	1.349	0.043	0.273	40.72	0.167	59.755	61.809
ps	0.160	1.697	0.034	0.354	18.10	0.126	53.010	52.715
rd_mve	0.409	2.392	0.336	1.960	10.77	0.073	53.523	53.488
retvol	0.388	1.491	−1.037	−3.829	83.40	1.425	37.243	30.184
roaq	0.272	1.843	0.103	0.696	25.71	0.169	36.826	33.040
sgr	0.100	0.744	0.016	0.122	15.20	0.083	57.893	54.336
std_turn	0.088	0.459	−0.497	−2.562	78.79	0.585	42.055	35.805
sue	0.238	2.310	−0.199	−1.858	45.51	0.437	36.803	31.938
turn	0.020	0.098	−0.165	−0.821	28.70	0.185	54.467	49.985
zerotrade	0.221	1.121	−0.609	−3.014	61.95	0.830	59.806	48.398

5.2 Factor summary statistics

Table 2 reports summary statistics for the 28 factors listed in Table 1. The first column gives the acronym of the factor. The second and the third columns give the average monthly *gross* return of the factor and its t-statistic, respectively. The fourth and the fifth columns give the factor's average monthly return *net of price-impact costs* and its t-statistic, respectively, when one invests one billion dollars on each leg of the factor. The sixth column gives the factor's monthly turnover (TO), and the seventh column gives the factor's monthly price-impact cost (PIC). The eighth 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, turnovers, and price-impact costs are reported in percentage.

Consistent with the findings of Detzel et al. (2021), we find that, among the factors constructed from double and triple sorts, factors that are rebalanced monthly (UMD, ROE, IA, ME) have turnovers that are much higher than those of factors that are rebalanced annually (SMB, HML, RMW, CMA). However, we also find that the relative performance of factors in terms of turnover is different from that in terms of price-impact costs. For instance, while UMD is the factor with the highest turnover, ROE is the factor with the highest price-impact costs.

To understand the difference in the relative performance of factors in terms of turnover and price-impact cost, the last two columns of Table 2 report the average trade-weighted market capitalization (in billions of dollars) of the different factors listed in Table 1. In particular, for each factor we compute the monthly trade-weighted market capitalization of the stocks traded by the factor and report the time-series average. Table 2 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 about 63 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 20.5 and 23 billion dollars, respectively. The low market capitalization of the average stock traded by the ROE factor explains why the price-impact cost of ROE is much larger than the price-impact cost of UMD, even though UMD has a substantially larger turnover.

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. (2021) as an important driver in the context of *proportional* transaction costs, but also on the size and liquidity of the stocks traded.

Table 3: List of factor models considered

This table lists the factor models we consider, ordered in 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 give 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
HXZ4	4	Hou, Xue & Zhang	2015, RFS	MKT, ROE, IA, ME
FFC4	4	Fama & French and Carhart	1993, JFE and 1997, JOF	MKT, SMB, HML, UMD
FF5	5	Fama & French	2015, JFE	MKT, SMB, HML, RMW, CME
FF6	6	Fama & French	2018, JFE	MKT, SMB, HML, RMW, CME, UMD
DMNU20	20	DeMiguel, Martin-Utrera, Nogales & Uppal	2020, RFS	MKT, agr, cashpr, chatoia, chcsho, convind, egr, ep, gma, idiovol, ind-mom, ps, rd_mve, retvol, roaq, sgr, std_turn, sue, turn, zerotrade

5.3 Factor models

Table 3 lists the five factor models we consider. In particular, we consider four popular *low-dimensional* factor models: 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 Fama and French (2015), FF5, and the six-factor model of Fama and French (2018), FF6. In addition, we consider a *high-dimensional* factor model, using the 20 factors that DeMiguel et al. (2020) find statistically significant in the presence of price-impact costs, DMNU20. We consider this high-dimensional model to evaluate the trading-diversification benefits from combining a large number of factors.

5.4 Model comparison using our proposed statistical tests

In this section, we compare the performance of the factor models listed in Table 3 in the presence of price-impact costs using the statistical tests developed in Section 4. Like Gârleanu and Pedersen (2013), we consider a base case with an absolute risk-aversion parameter of 5×10^{-9}, which corresponds to an institutional investor with a relative risk-aversion parameter of five and an endowment of one billion dollars. For comparison, we also consider cases where the investor has the same relative risk-aversion parameter, but her endowment is twenty times larger or smaller than in the base case; that is, when $\gamma = 2.5 \times 10^{-10}$ or $\gamma = 1 \times 10^{-7}$. For a constant relative risk-aversion level, a lower absolute risk-aversion parameter implies a larger endowment, and therefore price-impact costs play a more important role in the investor's mean-variance utility.

Note that both FFC4 and FF5 are nested by the FF6 model, and thus we use the distribution in Proposition 6 to compare FFC4 and FF5 with FF6. Also, all models we consider have one common market factor. Therefore, following our discussion in Section 4.2.3, we compare non-nested models in two stages. First, we use the distribution in Proposition 6 to test whether a model with all factors in the two models has 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.[10] If the test rejects the null, we then implement a second-stage test that uses the distribution in Proposition 5 to compare the two models.[11]

To understand how price-impact costs affect the relative performance of the five factor models, we first compare their performance in the *absence* of price-impact costs. Panel A in Table 4 reports the 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. Our main observation is that in the absence of price-impact costs, HXZ4 is the best model. To see this, note first that the mean-variance utility delivered by the factors in the HXZ4 model is higher than those delivered by the factors in the other three low-dimensional models (FFC4, FF5, and FF6). Moreover, the difference between the utilities provided by the factors in the HXZ4 model and the FFC4 model is statistically significant. In contrast, the differences between the utility derived from the factors in the HXZ4 model and those derived from the factors in the FF5 and FF6 models are not statistically significant. However, HXZ4 is the preferred model because it contains fewer factors than FF5 and FF6, and thus, it is more parsimonious. Finally, although the high-dimensional model DMNU20 achieves a sample mean-variance

[10]For every non-nested model comparison, we find in unreported results that the first-stage test rejects the null hypothesis at the 1% level, and thus we have to perform the second-stage test.

[11]In detail, the p-values are computed as follows. Assume without loss of generality that the mean-variance utilities net of price-impact costs for models A and B satisfy $\hat{U}^{\gamma}_{\Lambda,A} > \hat{U}^{\gamma}_{\Lambda,B}$. Then, we compute the p-value as the integral over the values greater than $\hat{U}^{\gamma}_{\Lambda,A} - \hat{U}^{\gamma}_{\Lambda,B}$ of the probability density function in (24) if the two models are non-nested and of the probability density function in (25) if they are nested. Like Barillas et al. (2020), we use the bias-adjusted values of $\hat{U}^{\gamma}_{\Lambda,A}$ and $\hat{U}^{\gamma}_{\Lambda,B}$ when comparing non-nested factor models using Proposition 5. This is because the asymptotic distribution in (24) 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 $\hat{U}^{\gamma}_{\Lambda,A}$ and $\hat{U}^{\gamma}_{\Lambda,B}$ because the asymptotic distribution in (25) adequately captures the finite-sample bias as demonstrated by the bootstrap experiments in Section IA.3 of the Internet Appendix.

Table 4: Significance of difference in mean-variance utility without price-impact costs

This table reports the significance of the difference between the mean-variance utilities of the row and column models in the absence of trading costs. Panel A reports the scaled sample mean-variance utility of each of the five factor models in the absence of trading costs for the baseline case with absolute risk-aversion parameter $\gamma = 5 \times 10^{-9}$. Panel B reports the p-value for the difference in mean-variance utility for every pairwise model comparison. The p-value is computed using Proposition 5 when the row and column models are non-nested and Proposition 6 when the row model is nested in the column model.

Panel A: Mean-variance utilities without trading costs

	HXZ4	FFC4	FF5	FF6	DMNU20
$2\gamma\hat{U}^{\gamma}$	0.1328	0.0542	0.1007	0.1131	0.1570

Panel B: p-values

	HXZ4	FFC4	FF5	FF6	DMNU20
HXZ4		0.002	0.100	0.175	0.279
FFC4			0.049	0.000	0.005
FF5				0.036	0.074
FF6					0.132

utility that is higher than that delivered by the factors in the HXZ4 model, the difference in utilities is not statistically significant, and thus HXZ4 is again the preferred model because of its parsimony.

Table 5 reports the performance of the five models in the *presence* of price-impact costs for our base-case absolute risk-aversion parameter $\gamma = 5 \times 10^{-9}$. Our main finding is that price-impact costs change the relative performance of the different models. While HXZ4 was the best model in the absence of trading costs, HXZ4 delivers the lowest mean-variance utility *net* of price-impact costs. Moreover, HXZ4 is significantly outperformed by both FF5 and FF6. The explanation for the poor performance of the HXZ4 model in the presence of price-impact costs is not only that its investment and profitability factors require higher turnover than those corresponding to the FF5 and FF6 models as shown in the sixth column of Table 2, but also that they require trading stocks with smaller market capitalization, and thus, less liquid as shown in eighth column of Table 2. FF6 emerges as the best low-dimensional model in the presence of price-impact costs because it significantly outperforms HXZ4, FFC4, and FF5.[12] Finally, although the high-dimensional model DMNU20 achieves

[12]This result is counterintuitive because the FF6 model is obtained by adding the momentum factor to FF5 and trading the momentum factor incurs high price-impact costs. However, even though momentum is

Table 5: Significance of difference in mean-variance utility with price-impact costs

This table reports the significance of the difference between the mean-variance utilities net of price-impact costs of the row and column models for the baseline case with absolute risk-aversion parameter $\gamma = 5 \times 10^{-9}$. Panel A reports the scaled sample mean-variance utility net of price-impact costs of each of the five factor models. Panel B reports the p-value for the difference in mean-variance utility net of price impact costs for every pairwise model comparison. The p-value is computed using Proposition 5 when the row and column models are non-nested and Proposition 6 when the row model is nested in the column model.

Panel A: Mean-variance utilities for $\gamma = 5 \times 10^{-9}$

	HXZ4	FFC4	FF5	FF6	DMNU20
$2\gamma\hat{U}^{\gamma}_{\Lambda}$	0.0334	0.0366	0.0532	0.0638	0.0921

Panel B: p-values

	HXZ4	FFC4	FF5	FF6	DMNU20
HXZ4		0.353	0.039	0.010	0.003
FFC4			0.082	0.000	0.006
FF5				0.002	0.030
FF6					0.075

higher sample mean-variance utility than the FF6 model, the difference of utilities between the FF6 model and the DMNU20 model is not statistically significant at the 5% level, and thus FF6 is the preferred model because of its parsimony.

The finding that DMNU20 does *not* significantly outperform FF6 for the base case with $\gamma = 5 \times 10^{-9}$ is surprising because DeMiguel et al. (2020) find that in the presence of trading costs, high-dimensional models are likely to perform well because of the benefits of trading diversification across factors. To shed light over this result, we consider a case with a lower absolute risk aversion $\gamma = 2.5 \times 10^{-10}$, which corresponds to an investor with the same relative risk aversion as in our base case, but with an endowment 20 times larger than that in the base case. For this case, one expects price-impact costs to play a more important role and DMNU20 to dominate other factor models. The results in Table 6 confirm this intuition: the high-dimensional model DMNU20 significantly outperforms every low-dimensional model at the 5% confidence level. Among low-dimensional models, FF6 is again the best model as it significantly outperforms HXZ4, FFC4, and FF5.

expensive when traded in isolation, it is a lot cheaper to trade in *combination* with the other five factors in the FF6 model because of trading diversification (DeMiguel et al., 2020).

Table 6: Significance of difference in mean-variance utility with costs for $\gamma = 2.5 \times 10^{-10}$

This table reports the significance of the difference between the mean-variance utilities net of price-impact costs of the row and column models for the case with absolute risk-aversion parameter $\gamma = 2.5 \times 10^{-10}$. Panel A reports the scaled sample mean-variance utility net of price-impact costs of each of the five factor models. Panel B reports the p-value for the difference in mean-variance utility net of price impact costs for every pairwise model comparison. The p-value is computed using Proposition 5 when the row and column models are non-nested and Proposition 6 when the row model is nested in the column model.

Panel A: Mean-variance utilities for $\gamma = 2.5 \times 10^{-10}$

	HXZ4	FFC4	FF5	FF6	DMNU20
$2\gamma\hat{U}^{\gamma}_{\Lambda}$	0.0218	0.0223	0.0246	0.0256	0.0373

Panel B: p-values

	HXZ4	FFC4	FF5	FF6	DMNU20
HXZ4		0.313	0.089	0.038	0.006
FFC4			0.068	0.000	0.005
FF5				0.001	0.009
FF6					0.012

Finally, Table 7 reports the results for the case with a larger absolute risk-aversion parameter, $\gamma = 1 \times 10^{-7}$, which corresponds to an investor with a relative risk aversion of five as in the base case, but with an endowment 20 times *smaller* than that in the base case. For such small endowments, one would expect the relative performance of the different models to be quite similar to that in the *absence* of costs. Table 7 confirms that this is indeed the case: HXZ4 outperforms FFC4 and FF5, with the difference being statistically significant for FFC4. Also, although FF6 and DMNU20 deliver a higher mean-variance utility net of price-impact costs than HXZ4, the difference between the utilities of these two models and HXZ4 is not statistically significant. Thus, HXZ4 emerges as the preferred model just as in the case without trading costs.

In summary, accounting for price-impact costs results in a more nuanced comparison of the various factor models we consider—the HXZ4, FF6, and DMNU20 models are the best performing for high, medium, and low absolute risk aversion, respectively.

Table 7: Significance of difference in mean-variance utility with costs for $\gamma = 1 \times 10^{-7}$

This table reports the significance of the difference between the mean-variance utilities net of price-impact costs of the row and column models for the case with absolute risk-aversion parameter $\gamma = 1 \times 10^{-7}$. Panel A reports the scaled sample mean-variance utility net of price-impact costs of each of the five factor models. Panel B reports the p-value for the difference in mean-variance utility net of price impact costs for every pairwise model comparison. The p-value is computed using Proposition 5 when the row and column models are non-nested and Proposition 6 when the row model is nested in the column model.

Panel A: Mean-variance utilities for $\gamma = 1 \times 10^{-7}$

	HXZ4	FFC4	FF5	FF6	DMNU20
$2\gamma\hat{U}^{\gamma}_{\Lambda}$	0.1013	0.0525	0.0932	0.1067	0.1308

Panel B: p-values

	HXZ4	FFC4	FF5	FF6	DMNU20
HXZ4		0.006	0.332	0.380	0.163
FFC4			0.056	0.000	0.008
FF5				0.025	0.108
FF6					0.215

5.5 Model comparison using out-of-sample bootstrap tests

In the previous section, we compared factor models using our proposed statistical tests, which address the main asset-pricing question: is the mean-variance utility in the presence of price-impact costs of a model significantly higher than that of another? 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 proposed by Fama and French (2018) and used by Detzel et al. (2021).

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 bootstrap optimal portfolio weights

Table 8: Bootstrap out-of-sample utility net of price-impact costs

Panel A reports the average scaled out-of-sample (OOS) mean-variance utility net of price-impact costs across 100,000 bootstrap samples of each factor model under the baseline case with absolute risk-aversion parameter $\gamma = 5 \times 10^{-9}$. Panel B reports the frequency with which the row model outperforms the column model out-of-sample across the 100,000 bootstrap samples.

Panel A: Average mean-variance utilities

	HXZ4	FFC4	FF5	FF6	DMNU20
$2\gamma\hat{U}^{\gamma}_{\Lambda}$	0.0181	0.0145	0.0282	0.0357	0.0299

Panel B: Frequency row model outperforms column model

	HXZ4	FFC4	FF5	FF6	DMNU20
HXZ4		0.523	0.259	0.207	0.356
FFC4			0.233	0.105	0.334
FF5				0.217	0.445
FF6					0.518

of each factor model using Equation (20).[13] 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 obtain 100,000 observations of the OOS mean-variance utility net of price-impact costs for each factor model. Finally, we compare models in terms of average mean-variance utility and the frequency with which one model outperforms another 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 8 reports the out-of-sample bootstrap results for the base case with absolute risk aversion $\gamma = 5 \times 10^{-9}$. Panel A reports the average mean-variance utility net of price-impact costs of each model and Panel B reports the frequency with which the row model outperforms the column model across the bootstrap samples.[14] As expected, the average *out-of-sample* mean-variance utilities of the different models in Panel A of Table 8 are much lower than the

[13]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.

[14]Section IA.4 of the Internet Appendix reports the results for the cases with $\gamma = 2.5 \times 10^{-10}$ and $\gamma = 1 \times 10^{-7}$.

in-sample utilities in Panel A of Table 5 because of the impact of estimation error. However, the out-of-sample relative performance of the various factor models is generally consistent with that in sample.[15]

Note that the frequencies in Panel B of Table 8 are larger than the p-values based on our statistical tests in Panel B of Table 5. This is not surprising because even if a factor model has a significantly larger mean-variance utility than another, it may achieve a smaller out-of-sample mean-variance utility in a particular bootstrap sample because of estimation error. Nonetheless, the results in Panel B of Table 8 are consistent with those in Panel B of Table 5. In particular, we observe that, out of sample, HXZ4 outperforms FF5, FF6, and DMNU20 only on 25.9%, 20.7%, and 35.6% of the bootstrap samples, respectively. This is consistent with the finding in Panel B of Table 5 that the FF5, FF6, and DMNU20 models are significantly better than the HXZ4 model. In addition, FF6 outperforms the FFC4 and FF5 models on around 80% of the bootstrap samples, which is consistent with the finding in Panel B of Table 5 that the FF6 model outperforms all other low-dimensional models. Finally, the FF6 model outperforms the DMNU20 model on 51.8% of the bootstrap samples, which again is coherent with our finding in Panel B of Table 8 that FF6 and DMNU20 are statistically indistinguishable.

In summary, the out-of-sample bootstrap tests confirm the main finding from our statistical tests in Table 5 that, in the base case with absolute risk-aversion parameter $\gamma = 5 \times 10^{-9}$, the FF6 model emerges as the best low-dimensional model. Moreover, the out-of-sample test shows that the gains from using the FF6 factor model can actually be realized out of sample. Section IA.4 of the Internet Appendix shows that the findings from the out-of-sample bootstrap tests are also consistent with the findings from our statistical tests for the cases with larger and smaller absolute risk-aversion parameters.

[15]There are two pairwise comparisons of factor models for which the out-of-sample performance results differ from those in sample. First, the average out-of-sample mean-variance utility net of costs of HXZ4 is higher than that of FFC4, whereas the in-sample utility of FFC4 is better. This is not surprising because the two models have similar in-sample mean-variance utility net of costs and our statistical test in Panel B of Table 5 shows that HXZ4 and FFC4 are statistically indistinguishable. Second, the out-of-sample performance of FF6 is better than that of DMNU20, whereas the in-sample performance of DMNU20 was better. Again, this is not surprising as our statistical test shows that the two models are statistically indistinguishable and the performance of the high-dimensional DMNU20 model is likely to be more impacted by estimation error out of sample than that of the FF6 model.

6 Conclusion

We show that comparing factor models in terms of their squared Sharpe ratio is no longer sufficient in the presence of price-impact costs because the investment opportunity set spanned by a factor model is no longer linear. Instead, we propose comparing factor models in terms of mean-variance utility net of price-impact costs and develop a formal statistical methodology to compare nested and non-nested factor models. 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 costs is a more nuanced exercise than in the absence of trading costs.

Empirically, we find that while in the absence of trading costs the four-factor model of Hou et al. (2015) outperforms other low-dimensional models, in the presence of price-impact costs the six-factor model of Fama and French (2018) is preferred. 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 costs are important enough for the trading diversification benefits of combining a large number of factors to dominate other effects such as the impact of estimation error. More broadly, we find that model performance depends not only on the portfolio turnover required to rebalance the factors, but also on the liquidity of the stocks traded and the absolute risk-aversion parameter.

References

Avramov, Doron, and John C. Chao, 2006, An exact Bayes test of asset pricing models with application to international markets, *Journal of Business* 79, 293–324.

Barillas, Francisco, Raymond Kan, Cesare Robotti, and Jay Shanken, 2020, Model comparison with Sharpe ratios, *Journal of Financial and Quantitative Analysis* 55, 1840–1874.

Barillas, Francisco, and Jay Shanken, 2017, Which alpha?, *Review of Financial Studies* 30, 1316–1338.

Barillas, Francisco, and Jay Shanken, 2018, Comparing asset pricing models, *Journal of Finance* 73, 715–754.

Barroso, Pedro, and Andrew Detzel, 2021, Do limits to arbitrage explain the benefits of volatility-managed portfolios?, *Journal of Financial Economics* 140, 744–767.

Barroso, Pedro, and Pedro Santa-Clara, 2015, Beyond the carry trade: Optimal currency portfolios, *Journal of Financial and Quantitative Analysis* 50, 1037–1056.

Brandt, Michael W., Pedro Santa-Clara, and Rossen Valkanov, 2009, Parametric portfolio policies: Exploiting characteristics in the cross-section of equity returns, *Review of Financial Studies* 22, 3411–3447.

Campbell, John Y., 2017, *Financial decisions and markets: A course in asset pricing* (Princeton University Press, Princeton, NJ).

Carhart, Mark M., 1997, On persistence in mutual fund performance, *Journal of Finance* 52, 57–82.

Chen, Andrew, and Mihail Velikov, 2022, Zeroing in on the expected returns of anomalies, forthcoming in *Journal of Financial and Quantitative Analysis*.

Chib, Siddhartha, Xiaming Zeng, and Lingxiao Zhao, 2020, On comparing asset pricing models, *Journal of Finance* 75, 551–577.

DeMiguel, Victor, Alberto Martin-Utrera, Francisco J. Nogales, and Raman Uppal, 2020, A transaction-cost perspective on the multitude of firm characteristics, *Review of Financial Studies* 33, 2180–2222.

Detzel, Andrew L., Robert Novy-Marx, and Mihail Velikov, 2021, Model selection with transaction costs, *Available at SSRN* 3805379.

Edelen, Roger M., Richard B. Evans, and Gregory B. Kadlec, 2007, Scale effects in mutual fund performance: The role of trading costs, *Available at SSRN* 951367.

Fama, Eugene F., and Kenneth R. French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.

Fama, Eugene F., and Kenneth R. French, 2015, A five-factor asset pricing model, *Journal of Financial Economics* 116, 1–22.

Fama, Eugene F., and Kenneth R. French, 2018, Choosing factors, *Journal of Financial Economics* 128, 234–252.

Ferson, Wayne E., Andrew F. Siegel, and Junbo L. Wang, 2019, Asymptotic variances for tests of portfolio efficiency and factor model comparisons with conditioning information, *Available at SSRN* 3330663.

Frazzini, Andrea, Ronen Israel, and Tobias J. Moskowitz, 2015, Trading costs of asset pricing anomalies, Fama-Miller working paper.

Frazzini, Andrea, Ronen Israel, and Tobias J. Moskowitz, 2018, Trading costs, *Available at SSRN* 3229719.

Gârleanu, Nicolae, and Lasse H. Pedersen, 2013, Dynamic trading with predictable returns and transaction costs, *Journal of Finance* 68, 2309–2340.

Gibbons, Michael R., Stephen A. Ross, and Jay Shanken, 1989, A test of the efficiency of a given portfolio, *Econometrica* 57, 1121–1152.

Goyal, Amit, Zhongzhi Lawrence He, and Sahn-Wook Huh, 2018, Distance-based metrics: A Bayesian solution to the power and extreme-error problems in asset-pricing tests, Swiss Finance Institute Research Paper.

Gârleanu, Nicolae, and Lasse Heje Pedersen, 2022, Active and passive investing: Understanding Samuelson's dictum, forthcoming in *Review of Asset Pricing Studies*.

Hayashi, Fumio, 2000, *Econometrics* (Princeton University Press).

Hou, Kewei, Chen Xue, and Lu Zhang, 2015, Digesting anomalies: An investment approach, *Review of Financial Studies* 28, 650–705.

Jobson, J. D., and B. Korkie, 1980, Estimation for markowitz efficient portfolios, *Journal of the American Statistical Association* 75, 544–554.

Kan, Raymond, and Cesare Robotti, 2009, Model comparison using the Hansen-Jagannathan distance, *Review of Financial Studies* 22, 3449–3490.

Kan, Raymond, Cesare Robotti, and Jay Shanken, 2013, Pricing model performance and the two-pass cross-sectional regression methodology, *Journal of Finance* 68, 2617–2649.

Kan, Raymond, Xiaolu Wang, and Xinghua Zheng, 2019, In-sample and out-of-sample Sharpe ratios of multi-factor asset pricing models, *Available at SSRN* 3454628.

Korajczyk, Robert A., and Ronnie Sadka, 2004, Are momentum profits robust to trading costs?, *Journal of Finance* 59, 1039–1082.

Ledoit, Olivier, and Michael Wolf, 2004, A well-conditioned estimator for large-dimensional covariance matrices, *Journal of Multivariate Analysis* 88, 365–411.

Maruyama, Yosihito, and Takashi Seo, 2003, Estimation of moment parameter in elliptical distributions, *Journal of the Japan Statistical Society* 33, 215–229.

Novy-Marx, Robert, and Mihail Velikov, 2016, A taxonomy of anomalies and their trading costs, *Review of Financial Studies* 29, 104–147.

Ross, Stephen, 1976, The arbitrage theory of capital asset pricing, *Journal of Economic Theory* 13, 341–360.

A Proofs of all results

This appendix contains the proofs of all novel propositions in the manuscript. For expositional purposes, the manuscript also contains three propositions from the existing literature, whose proofs can be found in Campbell (2017, Section 2.2.6) for Proposition 1 and in Barillas et al. (2020) for Propositions 4 and 7.

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 3.3, we have that the set $S = \{\theta | \theta^\top \mu - f(\theta) \geq 0\}$ is nonempty. Moreover, by Assumption 3.2, $f(\theta)$ 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 3.3, 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 absolute risk aversion is γ/c, then there exists θ_2 such that

$$\theta_2^\top \mu - f(\theta_2) - \frac{\gamma}{2c}\theta_2^\top \Sigma\theta_2 > c\theta_1^\top \mu - f(c\theta_1) - \frac{\gamma}{2c}c\theta_1^\top \Sigma c\theta_1, \tag{35}$$

which is equivalent to

$$\frac{\theta_2^\top}{c}\mu - f\left(\frac{\theta_2}{c}\right) - \frac{\gamma}{2}\frac{\theta_2^\top}{c}\Sigma\frac{\theta_2}{c} > \theta_1^\top \mu - f(\theta_1) - \frac{\gamma}{2}\theta_1^\top \Sigma\theta_1, \tag{36}$$

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

$$\theta_1^\top \Sigma\theta_1 = \theta_2^\top \Sigma\theta_2, \tag{37}$$

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

$$c\theta_1^\top \mu - cf(\theta_1) = \theta_2^\top \mu - f(\theta_2). \tag{38}$$

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, then we must have that

$$(c\theta_1^\top)\Sigma(c\theta_1) > \theta_2^\top \Sigma\theta_2. \tag{39}$$

Thus,

$$c\theta_1^\top \mu - cf(\theta_1) - \frac{\gamma}{2c}(c\theta_1^\top)\Sigma(c\theta_1) < \theta_2^\top \mu - f(\theta_2) - \frac{\gamma}{2c}\theta_2^\top \Sigma\theta_2, \tag{40}$$

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_p. The Sharpe ratio of $c\theta^*$, a maximizer for the case with absolute risk aversion γ/c is

$$\frac{c\theta^{*\top}\mu - f(c\theta^*)}{c\sqrt{\theta^{*\top}\Sigma\theta^*}} = \frac{\theta^{*\top}\mu - f(\theta^*)}{\sqrt{\theta^{*\top}\Sigma\theta^*}}, \tag{41}$$

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 3.2 we have that $f(\theta) > 0$ for any $\theta \neq 0$, and thus,

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

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

$$f(c'\theta) < c' f(\theta) \quad \text{for } \theta \neq 0 \text{ and } 0 < c' < 1. \tag{42}$$

We first prove that (8) implies (42). Let $\theta' = c\theta$ with $c > 1$. Then (8) becomes

$$\frac{1}{c} f(\theta') > f\big(\frac{1}{c}\theta'\big). \tag{43}$$

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

$$c' f(\theta') > f(c'\theta'), \tag{44}$$

which is (42). Using a similar argument, it is straightforward to show that (42) 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

$$\theta^{*\top}\mu - f(\theta^*) - \frac{c\gamma}{2}\theta^{*\top}\Sigma\theta^* \leq (\theta_c^*)^\top\mu - f(\theta_c^*) - \frac{c\gamma}{2}(\theta_c^*)^\top\Sigma\theta_c^*, \tag{45}$$

$$(\theta_c^*)^\top\mu - f(\theta_c^*) - \frac{\gamma}{2}(\theta_c^*)^\top\Sigma\theta_c^* \leq \theta^{*\top}\mu - f(\theta^*) - \frac{\gamma}{2}\theta^{*\top}\Sigma\theta^*. \tag{46}$$

Combining these two inequalities yields

$$\frac{\gamma}{2}(\theta^{*\top}\Sigma\theta^* - (\theta_c^*)^\top\Sigma\theta_c^*) \leq \theta^{*\top}\mu - f(\theta^*) - (\theta_c^*)^\top\mu + f(\theta_c^*) \leq \frac{c\gamma}{2}(\theta^{*\top}\Sigma\theta^* - (\theta_c^*)^\top\Sigma\theta_c^*). \tag{47}$$

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 (47), 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

$$\frac{(\theta_c^*)^\top\mu - f(\theta_c^*)}{\sqrt{(\theta_c^*)^\top\Sigma\theta_c^*}} \leq \frac{\theta^{*\top}\mu - f(\theta^*)}{\sqrt{\theta^{*\top}\Sigma\theta^*}}, \tag{48}$$

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

When $(\theta_c^*)^\top\Sigma\theta_c^* = \theta^{*\top}\Sigma\theta^*$, (47) implies that $\theta^{*\top}\mu - f(\theta^*) = (\theta_c^*)^\top\mu - f(\theta_c^*)$, and thus (48) 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 (48) with strict inequality, we prove by contradiction that

$$(\theta_c^*)^\top\mu - f(\theta_c^*) < c(\theta^{*\top}\mu - f(\theta^*)). \tag{49}$$

Suppose (49) does not hold and thus $\theta^{*\top}\mu - f(\theta^*) \leq ((\theta_c^*)^\top\mu - f(\theta_c^*))/c$, then

$$\begin{aligned}\theta^{*\top}\mu - f(\theta^*) - \frac{\gamma}{2}\theta^{*\top}\Sigma\theta^* &\leq \frac{1}{c}(\theta_c^*)^\top\mu - \frac{1}{c}f(\theta_c^*) - \frac{\gamma}{2}\frac{(\theta_c^*)^\top}{c}\Sigma\frac{\theta_c^*}{c} \\ &< \frac{1}{c}(\theta_c^*)^\top\mu - f(\frac{1}{c}\theta_c^*) - \frac{\gamma}{2}\frac{(\theta_c^*)^\top}{c}\Sigma\frac{\theta_c^*}{c},\end{aligned} \tag{50}$$

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 $(\theta_c^*)^\top\Sigma\theta_c^* > \theta^{*\top}\Sigma\theta^*$, (49) holds.

Dividing both sides of (49) by $\sqrt{(\theta_c^*)^\top \Sigma \theta_c^*} = c\sqrt{\theta^{*\top}\Sigma\theta^*}$, (48) holds with strict inequality. Therefore, the efficient frontier is strictly concave. Moreover, since $f(\theta) > 0$ for any $\theta \neq 0$ both sides of (48) 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 U_Λ^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 second derive the variance of the asymptotic normal distribution.

Step 1: $\sqrt{T}(\hat{U}_\Lambda - U_\Lambda)$ 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

$$\varphi = [\mu, \mathrm{vec}(\Sigma), \mathrm{vec}(\Lambda)] \in \mathbb{R}^{K+2K^2}, \tag{51}$$

$$\hat{\varphi} = [\hat{\mu}, \mathrm{vec}(\hat{\Sigma}), \mathrm{vec}(\hat{\Lambda})] \in \mathbb{R}^{K+2K^2}, \tag{52}$$

$$r_t(\varphi) = [R_t - \mu, \mathrm{vec}(\Sigma_t - \Sigma), \mathrm{vec}(\Lambda_t - \Lambda)] \in \mathbb{R}^{K+2K^2}. \tag{53}$$

Under standard regularity conditions[16], 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

$$\sqrt{T}(\hat{U}_\Lambda - U_\Lambda) \overset{A}{\sim} N(0, \frac{\partial U_\Lambda}{\partial \varphi^\top} S_0 \frac{\partial U_\Lambda}{\partial \varphi}). \tag{54}$$

[16] 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)

Step 2: variance of asymptotic normal distribution, $h_{t,\Lambda}(\varphi)$.

Let

$$h_{t,\Lambda}(\varphi) = 2\gamma \frac{\partial U_\Lambda}{\partial \varphi^\top} r_t(\varphi), \tag{55}$$

then (54) can be rewritten as

$$\sqrt{T}(\hat{U}_\Lambda - U_\Lambda) \overset{A}{\sim} N(0, W), \quad \text{where } W = \sum_{j=-\infty}^{\infty} E\left[\frac{h_{t,\Lambda}(\varphi) h_{t+j,\Lambda}(\varphi)}{4\gamma^2}\right]. \tag{56}$$

Assumptions 4.1 and 4.2 imply that the factor returns and the rebalancing trades are serially independent, and thus, we have that

$$W = E\left[\frac{h^2_{t,\Lambda}(\varphi)}{4\gamma^2}\right]. \tag{57}$$

Also, note that

$$\begin{aligned}
\frac{\partial U_\Lambda}{\partial \mu} &= \frac{1}{\gamma}(\Sigma + \Lambda)^{-1}\mu = \theta^*, \\
\frac{\partial U_\Lambda}{\partial \Sigma} &= \frac{\partial U_\Lambda}{\partial \Lambda} = -\frac{1}{2\gamma}(\Sigma + \Lambda)^{-1}\mu\mu^\top(\Sigma + \Lambda)^{-1} = -\frac{\gamma}{2}\theta^*\theta^{*\top},
\end{aligned}$$

and thus,

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

where $\otimes$ denotes the Kronecker product. Plugging these partial derivatives in the definition of $h_{t,\Lambda}(\varphi)$ in (55), we have that

$$\begin{aligned}
h_{t,\Lambda}(\varphi) &= 2\gamma\left[\frac{\partial U_\Lambda}{\partial \mu^\top}(R_t - \mu) + \frac{\partial U_\Lambda}{\partial \text{vec}(\Sigma)^\top}\text{vec}(\Sigma_t - \Sigma) + \frac{\partial U_\Lambda}{\partial \text{vec}(\Lambda)^\top}\text{vec}(\Lambda_t - \Lambda)\right] \\
&= 2\gamma\theta^{*\top}(R_t - \mu) - \gamma^2\theta^{*\top}\Sigma_t\theta^* - \gamma^2\theta^{*\top}\Lambda_t\theta^* + \gamma^2\theta^{*\top}\Sigma\theta^* + \gamma^2\theta^{*\top}\Lambda\theta^* \\
&= \mu^\top(\Sigma + \Lambda)^{-1}(2R_t - \mu) - \mu^\top(\Sigma + \Lambda)^{-1}(\Sigma_t + \Lambda_t)(\Sigma + \Lambda)^{-1}\mu,
\end{aligned} \tag{58}$$

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

$$\sqrt{T}\big([\hat{U}_{\Lambda,A} - \hat{U}_{\Lambda,B}] - [U_{\Lambda,A} - U_{\Lambda,B}]\big) \overset{A}{\sim} N\left(0, \frac{\partial(U_{\Lambda,A} - U_{\Lambda,B})}{\partial \varphi^\top} S_0 \frac{\partial(U_{\Lambda,A} - U_{\Lambda,B})}{\partial \varphi}\right). \tag{59}$$

By Assumptions 4.1 and 4.2, we have that

$$\sqrt{T}\big([\hat{U}_{\Lambda,A} - \hat{U}_{\Lambda,B}] - [U_{\Lambda,A} - U_{\Lambda,B}]\big) \overset{A}{\sim} N\left(0, E\left[\frac{(h_{t,\Lambda,A} - h_{t,\Lambda,B})^2}{4\gamma^2}\right]\right), \tag{60}$$

where $h_{t,\Lambda,A}$ and $h_{t,\Lambda,B}$ are obtained by applying Equation (55) 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 $U_{\Lambda,A} = U_{\Lambda,B}$ holds and equation (58) suggests that $h_{t,\Lambda,A} = h_{t,\Lambda,B}$ for all t, and thus the variance in (60), $E[(h_{t,\Lambda,A} - h_{t,\Lambda,B})^2/(4\gamma^2)] = 0$. Consequently, the distribution in (60) is not applicable to perform a statistical test in these cases. Instead, in these cases we use the asympotic 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 U_Λ^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

$$\sqrt{T}(\hat{\theta}_A^* - \theta_A^*) \overset{A}{\sim} N(0, \frac{E[l_t l_t^\top]}{\gamma^2}), \tag{61}$$

where

$$l_t = (\Sigma_A + \Lambda_A)^{-1} R_{A,t} - (\Sigma_A + \Lambda_A)^{-1}(\Sigma_{A,t} + \Lambda_{A,t})(\Sigma_A + \Lambda_A)^{-1}\mu_A \in \mathbb{R}^{K_1+K_2}. \tag{62}$$

Part (ii): expression for $U_{\Lambda,A} - U_{\Lambda,B}$ as a function of θ_2^*.

$$
\begin{aligned}
U_{\Lambda,A} - U_{\Lambda,B} &= \frac{1}{2\gamma}\left[\mu_1^\top, \mu_2^\top\right]\begin{bmatrix}\Sigma_{11}+\Lambda_{11} & \Sigma_{12}+\Lambda_{12}\\ \Sigma_{21}+\Lambda_{21} & \Sigma_{22}+\Lambda_{22}\end{bmatrix}^{-1}\begin{bmatrix}\mu_1\\ \mu_2\end{bmatrix} - \frac{1}{2\gamma}\left[\mu_1^\top, \mu_2^\top\right]\begin{bmatrix}(\Sigma_{11}+\Lambda_{11})^{-1} & 0\\ 0 & 0\end{bmatrix}\begin{bmatrix}\mu_1\\ \mu_2\end{bmatrix}\\
&= \frac{\gamma}{2}\theta_A^{*\top}\begin{bmatrix}\Sigma_{11}+\Lambda_{11} & \Sigma_{12}+\Lambda_{12}\\ \Sigma_{21}+\Lambda_{21} & \Sigma_{22}+\Lambda_{22}\end{bmatrix}\theta_A^*\\
&- \frac{\gamma}{2}\theta_A^{*\top}\begin{bmatrix}\Sigma_{11}+\Lambda_{11} & \Sigma_{12}+\Lambda_{12}\\ \Sigma_{21}+\Lambda_{21} & \Sigma_{22}+\Lambda_{22}\end{bmatrix}\begin{bmatrix}(\Sigma_{11}+\Lambda_{11})^{-1} & 0\\ 0 & 0\end{bmatrix}\begin{bmatrix}\Sigma_{11}+\Lambda_{11} & \Sigma_{12}+\Lambda_{12}\\ \Sigma_{21}+\Lambda_{21} & \Sigma_{22}+\Lambda_{22}\end{bmatrix}\theta_A^*\\
&= \frac{\gamma}{2}\theta_A^{*\top}\begin{bmatrix}\Sigma_{11}+\Lambda_{11} & \Sigma_{12}+\Lambda_{12}\\ \Sigma_{21}+\Lambda_{21} & \Sigma_{22}+\Lambda_{22}\end{bmatrix}\theta_A^*\\
&- \frac{\gamma}{2}\theta_A^{*\top}\begin{bmatrix}\Sigma_{11}+\Lambda_{11} & \Sigma_{12}+\Lambda_{12}\\ \Sigma_{21}+\Lambda_{21} & (\Sigma_{21}+\Lambda_{21})(\Sigma_{11}+\Lambda_{11})^{-1}(\Sigma_{12}+\Lambda_{12})\end{bmatrix}\theta_A^*\\
&= \frac{\gamma}{2}\theta_2^{*\top}\left[(\Sigma_{22}+\Lambda_{22}) - (\Sigma_{21}+\Lambda_{21})(\Sigma_{11}+\Lambda_{11})^{-1}(\Sigma_{12}+\Lambda_{12})\right]\theta_2^*\\
&= \frac{\gamma}{2}\theta_2^{*\top}W\theta_2^*, \qquad (63)
\end{aligned}
$$

where $W = (\Sigma_{22}+\Lambda_{22}) - (\Sigma_{21}+\Lambda_{21})(\Sigma_{11}+\Lambda_{11})^{-1}(\Sigma_{12}+\Lambda_{12})$. Replacing the population parameters in Equation (63) with their sample counterparts we have that

$$
\hat{U}_{\Lambda,A} - \hat{U}_{\Lambda,B} = \frac{\gamma}{2}\hat{\theta}_2^{*\top}\hat{W}\hat{\theta}_2^*, \quad \text{where} \quad \hat{W} \xrightarrow{a.s.} W. \qquad (64)
$$

Part (iii): asymptotic distribution for $T(\hat{U}_{\Lambda,A} - \hat{U}_{\Lambda,B})$.

We now use (61) and (64) to derive the asymptotic distribution for $T(\hat{U}_{\Lambda,A} - \hat{U}_{\Lambda,B})$. 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 (61) we have that that $z \sim N(0, I_{K_2})$, where I_{K_2} is a K_2-dimensional identity matrix. Thus, from Equation (64) we have that

$$
\begin{aligned}
T(\hat{U}_{\Lambda,A} - \hat{U}_{\Lambda,B}) &= \frac{\gamma}{2}T\hat{\theta}_2^{*\top}\hat{W}\hat{\theta}_2^*\\
&\overset{A}{\sim} \frac{1}{2\gamma}z^\top(E[l_t l_t^\top]_{22})^{\frac{1}{2}}W(E[l_t l_t^\top]_{22})^{\frac{1}{2}}z. \qquad (65)
\end{aligned}
$$

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 $i = 1, \ldots, K_2$. 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 (65) can be rewritten as

$$T(\hat{U}_{\Lambda,A} - \hat{U}_{\Lambda,B}) \overset{A}{\sim} \bar{z}^\top \Xi \bar{z} = \sum_{i=1}^{K_2} \xi_i x_i,$$

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

A.5 Proof of Proposition 8

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, $E[h_{t,\Lambda}^2]/(4\gamma^2)$, and then simplify this expression.

Step 1: express $E[h_{t,\Lambda}^2]$ 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 (23), we have that

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

Therefore,

$$\begin{aligned} E[h_{t,\Lambda}^2] =& E\Bigg[4(u_t - \bar{u})^2 - 4(u_t - \bar{u})^3 - 4(u_t - \bar{u}) \sum_{n=1}^{N} v_{n,t}^2 + 4(u_t - \bar{u})\bar{u} \\ & + (u_t - \bar{u})^4 + 2(u_t - \bar{u})^2 \sum_{n=1}^{N} v_{n,t}^2 - 2(u_t - \bar{u})^2 \bar{u} \\ & + \left(\sum_{n=1}^{N} v_{n,t}^2\right)^2 - 2\bar{u} \sum_{n=1}^{N} v_{n,t}^2 + \bar{u}^2\Bigg]. \end{aligned} \tag{66}$$

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

$$E[X_i X_j X_k] = 0, \tag{67}$$

$$E[X_i X_j X_k X_l] = (\sigma_{ij}\sigma_{kl} + \sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk}), \tag{68}$$

where σ_{ab} is the covariance between X_a and X_b. Because $(u_t - \bar{u})$ and $v_{n,t}$ for $n = 1, \ldots, N$ are jointly normally distributed, using Equation (67), we can drop the third-order moments from Equation (66) to obtain

$$\begin{aligned} E[h_{t,\Lambda}^2] =& E\Bigg[4(u_t - \bar{u})^2 + (u_t - \bar{u})^4 + 2(u_t - \bar{u})^2 \sum_{n=1}^{N} v_{n,t}^2 - 2(u_t - \bar{u})^2 \bar{u} \\ &+ \left(\sum_{n=1}^{N} v_{n,t}^2\right)^2 - 2\bar{u} \sum_{n=1}^{N} v_{n,t}^2 + \bar{u}^2\Bigg]. \end{aligned} \tag{69}$$

Step 2: simplify (69). Using Equation (68), we can rewrite the terms on the right-hand side of Equation (69) as

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

Plugging these equations into (69), we have that

$$E[h_{t,\Lambda}^2] = 4\text{var}(u_t) + 3\left[\text{var}(u_t)\right]^2 + 2\left(\text{var}(u_t) \sum_{n=1}^{N} \text{var}(v_{n,t}) + 2 \sum_{n=1}^{N} \left[\text{cov}(u_t, v_{n,t})\right]^2\right)$$

$$
- 2\text{var}(u_t)\left(\text{var}(u_t) + \sum_{n=1}^{N}\text{var}(v_{n,t})\right) + \sum_{i=1}^{N}\sum_{j=1}^{N}\left(\text{var}(v_{i,t})\text{var}(v_{j,t}) + 2\left[\text{cov}(v_{i,t}, v_{j,t})\right]^2\right)
$$

$$
- 2\sum_{n=1}^{N}\text{var}(v_{n,t})\left(\text{var}(u_t) + \sum_{n=1}^{N}\text{var}(v_{n,t})\right) + \left(\text{var}(u_t) + \sum_{n=1}^{N}\text{var}(v_{n,t})\right)^2
$$

$$
= 4\text{var}(u_t) + 2\left[\text{var}(u_t)\right]^2 - \left(\sum_{n=1}^{N}\text{var}(v_{n,t})\right)^2 + 4\sum_{n=1}^{N}\left[\text{cov}(u_t, v_{n,t})\right]^2
$$

$$
+ \sum_{i=1}^{N}\sum_{j=1}^{N}\left(\text{var}(v_{i,t})\text{var}(v_{j,t}) + 2\left[\text{cov}(v_{i,t}, v_{j,t})\right]^2\right)
$$

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

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

$$
\frac{E[(h_{t,\Lambda,A} - h_{t,\Lambda,B})^2]}{4\gamma^2} = \frac{1}{4\gamma^2}\left(E[h_{t,\Lambda,A}^2] + E[h_{t,\Lambda,B}^2] - 2E[h_{t,\Lambda,A}h_{t,\Lambda,B}]\right). \tag{70}
$$

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

Step 1: express $E[h_{t,\Lambda,A}h_{t,\Lambda,B}]$ as a function of $\bar{u}$, u_t, and $v_{n,t}$.
Because $\left(u_t^A - \bar{u}^A\right)$, $\left(u_t^B - \bar{u}^B\right)$, $v_{n,t}^A$, and $v_{n,t}^B$ for $n = 1, \ldots, N$ are jointly normally distributed. Using Equation (67), we have that

$$
\begin{aligned}
E[h_{t,\Lambda,A}h_{t,\Lambda,B}] = E\Bigg[&4\left(u_t^A - \bar{u}^A\right)\left(u_t^B - \bar{u}^B\right) + \left(u_t^A - \bar{u}^A\right)^2\left(u_t^B - \bar{u}^B\right)^2 \\
&+ \left(u_t^A - \bar{u}^A\right)^2\sum_{n=1}^{N}(v_{n,t}^B)^2 + \left(u_t^B - \bar{u}^B\right)^2\sum_{n=1}^{N}\left(v_{n,t}^A\right)^2 \\
&- \left(u_t^A - \bar{u}^A\right)^2\bar{u}^B - \left(u_t^B - \bar{u}^B\right)^2\bar{u}^A + \left(\sum_{n=1}^{N}(v_{n,t}^A)^2\right)\left(\sum_{n=1}^{N}\left(v_{n,t}^B\right)^2\right) \\
&- \bar{u}^A\sum_{n=1}^{N}\left(v_{n,t}^B\right)^2 - \bar{u}^B\sum_{n=1}^{N}\left(v_{n,t}^A\right)^2 + \bar{u}^A\bar{u}^B\Bigg],
\end{aligned} \tag{71}
$$

Step 2: simplify (71). Using Equation (68), we can rewrite the terms on the right-hand side of Equation (71) as

$$E\left[\left(u_t^A - \bar{u}^A\right)\left(u_t^B - \bar{u}^B\right)\right] = \text{cov}\left(u_t^A, u_t^B\right),$$

$$E\left[\left(u_t^A - \bar{u}^A\right)^2\left(u_t^B - \bar{u}^B\right)^2\right] = \text{var}\left(u_t^A\right)\text{var}\left(u_t^B\right) + 2\left[\text{cov}\left(u_t^A, u_t^B\right)\right]^2,$$

$$E\left[\sum_{n=1}^{N}\left(v_{n,t}^A\right)^2\right] = \sum_{n=1}^{N}\text{var}\left(v_{n,t}^A\right),$$

$$E\left[\sum_{n=1}^{N}\left(v_{n,t}^B\right)^2\right] = \sum_{n=1}^{N}\text{var}\left(v_{n,t}^B\right),$$

$$E\left[(u_t^A - \bar{u}^A)^2\sum_{n=1}^{N}(v_{n,t}^B)^2\right] = \text{var}(u_t^A)\sum_{n=1}^{N}\text{var}(v_{n,t}^B) + 2\sum_{n=1}^{N}\left[\text{cov}(u_t^A, v_{n,t}^B)\right]^2,$$

$$E\left[(u_t^B - \bar{u}^B)^2\sum_{n=1}^{N}(v_{n,t}^A)^2\right] = \text{var}(u_t^B)\sum_{n=1}^{N}\text{var}(v_{n,t}^A) + 2\sum_{n=1}^{N}\left[\text{cov}(u_t^B, v_{n,t}^A)\right]^2,$$

$$E\left[\left(\sum_{n=1}^{N}(v_{n,t}^A)^2\right)\left(\sum_{n=1}^{N}(v_{n,t}^B)^2\right)\right] = \sum_{i=1}^{N}\sum_{j=1}^{N}\left(\text{var}(v_{i,t}^A)\text{var}(v_{j,t}^B) + 2\left[\text{cov}(v_{i,t}^A, v_{j,t}^B)\right]^2\right),$$

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

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

Plugging these equations into Equation (71), we have that

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

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

which completes the proof.

Internet Appendix to

Comparing Factor Models with Price-Impact Costs

This Internet Appendix contains several robustness checks and additional information. Section IA.1 compares our proposed p-values with those from the GRS test for the case with nested models and without price-impact costs. Section IA.2 discusses how we correct the upward bias in sample utilities. Section IA.3 uses bootstrap to check the finite-sample accuracy of our proposed asymptotic distributions. Section IA.4 gives the results for the out-of-sample bootstrap tests for different values of absolute risk aversion.

IA.1 Comparing Proposition 6 and the GRS test

Although our Proposition 6 is designed to compare factor models in the presence of price-impact costs, one can also use it to compare factor models in the absence of trading costs by setting $\Lambda_t = \Lambda = 0$. As a robustness check, we now compare the p-values for model comparisons in the absence of trading costs obtained using Proposition 6 and the GRS test, which is the test recommended by Barillas et al. (2020) to compare nested models in the absence of costs. Specifically, suppose model A nests model B. We first use Proposition 6 to compare the two models in terms of the maximum mean-variance utility and obtain the p-value of this test. We then let the extra factors of model A be the left-hand side test assets and let the factors of model B be the right-hand side factors, and run a time-series regression of the test assets on the factors. Then, we calculate the GRS test statistic based on the time-series alpha and obtain the p-value of this test.

Table IA.1 reports the p-values from the two tests for the two sets of nested models in our dataset. The first column lists the acronym of the nested model comparison. The second column reports the p-value of the test based on Proposition 6. The third column reports the p-value of the finite-sample GRS test in which the test statistic has an F distribution, and the fourth column reports the p-value of the asymptotic GRS test in which the test statistic has a χ^2 distribution. From this table, we find that the p-value of the GRS test (both the finite-sample version and the asymptotic version) is very similar to that of the test based on Proposition 6, although when comparing FF5 and FF6, the p-value of the test based on Proposition 6 is slightly larger, and thus, less significant than its GRS counterpart. Therefore, we conclude that the test based on Proposition 6 is very similar to the GRS test in the absence of trading costs, and it can be viewed as a generalization of the GRS test because it is also applicable to compare factor models in the presence of price-impact costs.

IA.2 Correcting the upward bias in sample utilities

As mentioned in Footnote 11 of the main body of the manuscript, the sample mean-variance utility net of price-impact costs of a factor model suffers from a small-sample upward bias as documented by Jobson and Korkie (1980) and Barillas et al. (2020). In this appendix, we show how we correct this upward bias.

We use bootstrap to estimate the upward bias of the sample mean-variance utility net of price-impact costs of each model. First, we bootstrap with replacement a sample with T^b months, and read the factor returns and scaled factor rebalancing trades of the bootstrapped months.[17] Second, we calculate the mean-variance utility net of price-impact costs of each factor model on the bootstrap sample. We then repeat the two steps for 100,000 times. For each model, we calculate its average mean-variance utility net of price-impact costs on the 100,000 bootstrap samples. The difference between the average mean-variance utility net of price-impact costs on the bootstrap samples and the utility in the original sample is our bootstrap estimator of the upward bias of each model, and we denote it as Δ_Λ. The bias-corrected mean-variance utility net of price-impact costs of a model is obtained by subtracting Δ_Λ from its mean-variance utility net of price-impact costs estimated using the original sample. In the main body of the manuscript, all reported mean-variance utility net of price-impact costs are bias corrected, and we implement bias correction when comparing factor models using Proposition 5.

IA.3 Finite-sample accuracy of asymptotic distributions

Propositions 5 and 6 provide two asymptotic distributions for the difference between the sample mean-variance utilities net of price-impact costs of two models. In this appendix, we use bootstrap simulations to check how accurately these asymptotic distributions fit their finite-sample counterparts. We set the absolute risk-aversion parameter $\gamma = 5 \times 10^{-9}$ as in our base case. To simplify notation, we drop the superscript γ from the mean-variance utility net of price-impact costs U_Λ^γ.

[17] In Sections 5.4 and 5.5, T^b is chosen to be 491, which equals to the size of our original sample.

IA.3.1 Asymptotic distribution from Proposition 5

In this section, we use bootstrap simulations to check how accurately the asymptotic distribution in Proposition 5 fits its finite-sample counterpart.

We assume that the true data generating process (DGP) is characterized by the sample estimators $\hat{\mu}, \hat{\Sigma}$, and $\hat{\Lambda}$. We use the superscript g to denote the true DGP, and use the superscript b to denote values obtained from bootstrap samples. We bootstrap with replacement 10,000 samples of T^b observations from our original sample. In other words, each bootstrap sample is generated from the true DGP. On each bootstrap sample, we estimate the mean-variance utility net of price-impact costs for every factor model, and adjust its finite-sample bias following Barillas et al. (2020) using the procedures in Appendix IA.2 to obtain U^b_Λ.[18] We then compute the following quantity on each bootstrap sample and for each pair of models A and B:

$$\sqrt{T^b}\big([U^b_{\Lambda,A} - U^b_{\Lambda,B}] - [U^g_{\Lambda,A} - U^g_{\Lambda,B}]\big), \tag{IA1}$$

where $U^g_{\Lambda,A}$ and $U^g_{\Lambda,B}$ denote the mean-variance utilities net of price-impact costs of models A and B under the true DGP g, which are known by construction. The 10,000 values of (IA1) characterize the finite-sample distribution of (IA1), and Proposition 5 characterizes the asymptotic distribution of (IA1) when $T^b \to \infty$.

Figure IA.1 compares the finite-sample distribution when $T^b = 491$ (pink histogram) and the asymptotic distribution based on Proposition 5 (blue curve) of (IA1) for all pairs of factor models that are non-nested. We observe that for most pairwise model comparisons the finite-sample distribution is very close to the asymptotic distribution. In some cases, such as the comparison of HXZ4 and FFC4, the asymptotic distribution does not fit the finite-sample distribution closely. The reason of this is that the number of observations in each bootstrap sample, $T^b = 491$, is not large enough to guarantee the convergence of the finite-sample distribution to the asymptotic distribution. To validate this argument, Figure IA.2 depicts the finite-sample distribution when $T^b = 2{,}000$ and the asymptotic distribution of (IA1). This figure shows that the asymptotic distribution provides a good fit when the number of observations in each bootstrap sample is large enough.

[18] In particular, for each model we subtract the quantity Δ_Λ defined in Appendix IA.2 from the sample mean-variance utility net of price-impact costs on each bootstrap sample, and thus the average bias-adjusted mean-variance utility net of price-impact costs over the 10,000 bootstrap samples equals to U^g_Λ.

IA.3.2 Asymptotic distribution from Proposition 6

In this section, we use bootstrap simulations to check how accurately the asymptotic distribution in Proposition 6 fits its finite-sample counterpart.

One difficulty in this experiment is that the asymptotic distribution in Proposition 6 holds only under the null hypothesis that $U_{\Lambda,A} = U_{\Lambda,B}$, but this null hypothesis does not hold in our sample for any pair of nested models. To address this, we assume that the true DGP is characterized by our original sample estimators $\hat{\Sigma}$ and $\hat{\Lambda}$, and we adjust $\hat{\mu}$ to make the null hypothesis hold under the true DGP.

We now describe how to adjust $\hat{\mu}$ using the notation in the proof of Proposition 6. Let the mean-variance portfolio estimated on the original sample of the larger model A be

$$\hat{\theta}^*_A = \frac{1}{\gamma}\left(\hat{\Sigma}_A + \hat{\Lambda}_A\right)^{-1}\begin{bmatrix}\hat{\mu}_1\\ \hat{\mu}_2\end{bmatrix} = \begin{bmatrix}\hat{\theta}^*_1\\ \hat{\theta}^*_2\end{bmatrix},$$

where $\hat{\mu}_1$ and $\hat{\mu}_2$ are the sample average returns of the factors f_1 and f_2, respectively, and $\hat{\theta}^*_1$ and $\hat{\theta}^*_2$ are the sample estimates of the mean-variance portfolio weights of model A on factors f_1 and f_2, respectively. We find a vector $c \in \mathbb{R}^{K_2}$, such that

$$\hat{\theta}^{*\prime}_A = \frac{1}{\gamma}\left(\hat{\Sigma}_A + \hat{\Lambda}_A\right)^{-1}\begin{bmatrix}\hat{\mu}_1\\ \hat{\mu}_2 - c\end{bmatrix} = \begin{bmatrix}\hat{\theta}^{*\prime}_1\\ 0\end{bmatrix}.$$

In other words, we adjust the mean returns of the extra factors f_2 by c so that the mean-variance portfolio of model A assigns zero weight to f_2. The vector c must satisfy

$$\frac{1}{\gamma}\left[\left(\hat{\Sigma}_A + \hat{\Lambda}_A\right)^{-1}\right]_{22} c = \hat{\theta}^*_2,$$

and thus it is uniquely identified because matrix $\left[(\hat{\Sigma}_A + \hat{\Lambda}_A)^{-1}\right]_{22}$ is invertible. We assume that the true DGP has the adjusted mean return vector $[\hat{\mu}_1, \hat{\mu}_2 - c]$. Note that with the adjusted mean return, the mean-variance portfolio of model A assigns zero weight to f_2. Therefore, under the true DGP, models A and B have the same mean-variance utility net of price-impact costs.

We use the superscript g to denote the true DGP, and use the superscript b to denote values obtained from the bootstrap samples. To make our original sample follow the true DGP, we adjust the sample returns of f_2 by setting $R'_{2,t} = R_{2,t} - c$ for all t. We then bootstrap with replacement from this adjusted sample to generate 10,000 bootstrap samples with T^b

observations. Thus, each bootstrap sample comes from the true DGP, which satisfies the null hypothesis of Proposition 6. On each bootstrap sample, we calculate the mean-variance utility net of price-impact costs U^b_Λ for every factor model. We do not adjust the finite-sample bias of U^b_Λ for the reasons discussed in Footnote 11. We then compute the following quantity on each bootstrap sample for every pair of nested models A and B:

$$T^b(U^b_{\Lambda,A} - U^b_{\Lambda,B}), \tag{IA2}$$

The 10,000 values of (IA2) characterize the finite-sample distribution of (IA2), and Proposition 6 characterizes the asymptotic distribution of (IA2) when $T^b \to \infty$.

Figure IA.3 compares the finite-sample distribution when $T^b = 491$ (pink histogram) and the asymptotic distribution based on Proposition 6 (blue histogram) of (IA2) for all pairs of nested models. The figure shows that the asymptotic distribution fits its finite-sample counterpart very accurately. Moreover, Figure IA.3 is based on sample mean-variance utilities net of price-impact costs that are *not* adjusted for finite-sample bias, and thus the figure demonstrates that the asymptotic distribution in Proposition 6 adequately captures the finite-sample bias.

IA.4 OOS bootstrap tests for different risk aversion

In the main body of the manuscript, we discuss the out-of-sample bootstrap test for the base case with absolute risk aversion $\gamma = 5 \times 10^{-9}$. In this section, we report the results for the cases with a higher and a lower absolute risk-aversion parameters. The exact procedure of the bootstrap test is the same as that in Section 5.5.

Tables IA.2 and IA.3 report the out-of-sample bootstrap results for the cases with absolute risk-aversion parameters $\gamma = 2.5 \times 10^{-10}$ and $\gamma = 1 \times 10^{-7}$, respectively. In each table, Panel A reports the average mean-variance utility net of price-impact costs of each model, and Panel B reports the frequency with which the row model outperforms the column model across the bootstrap samples. In both cases, we find that the average out-of-sample mean-variance utility net of price-impact costs of each model is lower than its in-sample counterpart because of the estimation error. However, the out-of-sample relative performance of the models is generally consistent with that in sample.

For the case with absolute risk aversion $\gamma = 2.5 \times 10^{-10}$, DMNU20 outperforms all other models in over 70% of the bootstrap samples, which is consistent with the in-sample results that DMNU20 significantly outperforms all other models. Although HXZ4 has a higher average out-of-sample utility than FFC4, FFC4 still outperforms HXZ4 in 41.0% of the bootstrap samples, which is consistent with the in-sample results that the two models are statistically indistinguishable. For the case with absolute risk aversion $\gamma = 1 \times 10^{-7}$, consistent with the in-sample results, HXZ4 has the highest average out-of-sample mean-variance utility net of price-impact costs, and it outperforms FFC4 in 92.4% of the bootstrap samples. Furthermore, it outperforms FF5, FF6, and DMNU20 in 70.7%, 59.8%, and 80.3% of the bootstrap samples. This confirms the in-sample results that when $\gamma = 1 \times 10^{-7}$, HXZ4 is the best-performing model.[19] In summary, the results of the out-of-sample bootstrap tests for the two cases with higher and lower absolute risk-aversion parameters than the base case confirm the main finding based on our statistical test.

[19]The result of the comparison between FFC4 and DMNU20 in this table is different from its in-sample counterpart. In Table 7, DMNU20 significantly outperforms FFC4, while FFC4 has higher out-of-sample mean-variance utility net of price-impact costs in 54.4% of the bootstrap samples. This is not surprising because the higher absolute risk-aversion parameter makes the price-impact costs matrix Λ have lower impact on the optimal portfolio, and the covariance matrix of the returns Σ has relatively higher impact on the optimal portfolio. Accurately estimating the covariance matrix of a twenty-factor model is hard and thus the performance of DMNU20 is likely to more impacted by the estimation error.

Figure IA.1: Distribution of the difference in mean-variance utilities: finite-sample distribution when $T^b = 491$ and asymptotic distribution based on Proposition 5

This figure compares the finite-sample distribution (pink histogram) and the asymptotic distribution (blue curve) of the difference in mean-variance utilities net of price-impact costs (IA1) for all pairs of models that are non-nested. The title of each sub-figure illustrates the two models for comparison. The finite-sample distribution is obtained by evaluating (IA1) on 10,000 bootstrap samples with $T^b = 491$ observations, and the asymptotic distribution is based on Proposition 5.

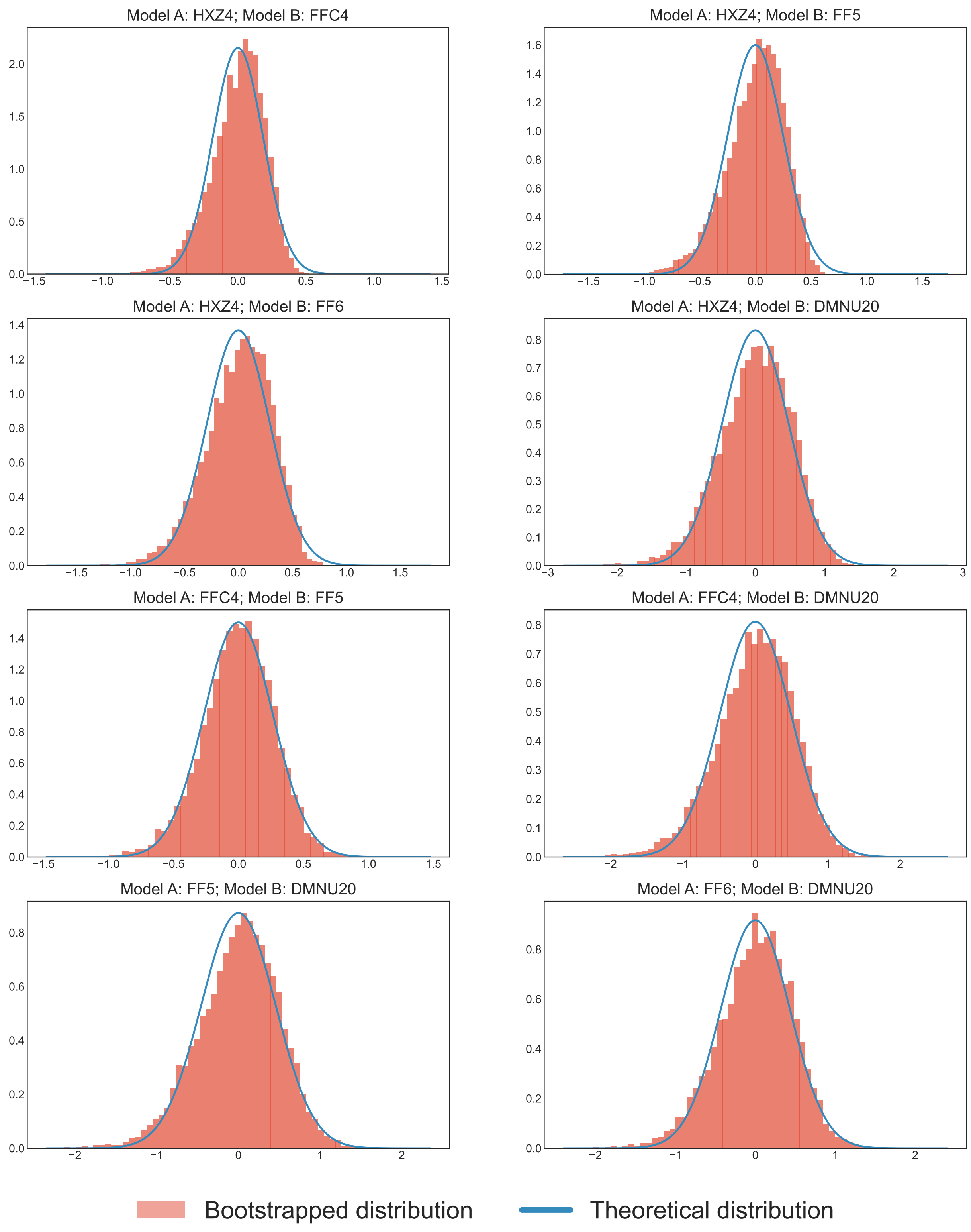


Figure IA.2: Distribution of the difference in mean-variance utilities: finite-sample distribution when $T^b = 2000$ and asymptotic distribution based on Proposition 5

This figure compares the finite-sample distribution (pink histogram) and the asymptotic distribution (blue curve) of the difference in mean-variance utilities net of price-impact costs (IA1) for all pairs of models that are non-nested. The title of each sub-figure illustrates the two models for comparison. The finite-sample distribution is obtained by evaluating (IA1) on 10,000 bootstrap samples with $T^b = 2000$ observations, and the asymptotic distribution is based on Proposition 5.

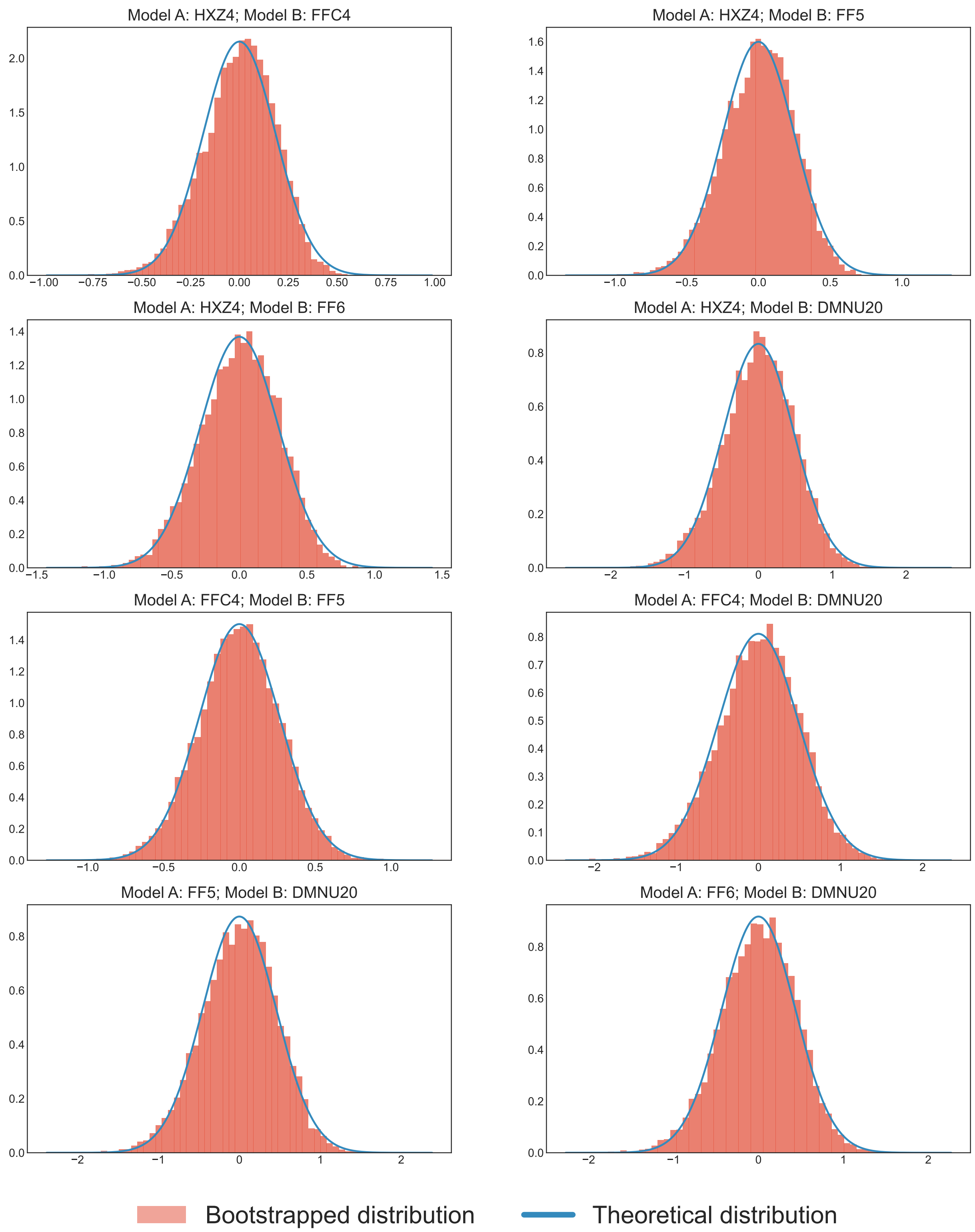


Figure IA.3: Distribution of the difference in mean-variance utilities: finite-sample distribution when $T^b = 491$ and asymptotic distributions based on Proposition 6

This figure compares compares the finite-sample distribution when $T^b = 491$ (pink histogram) and the asymptotic distribution based on Proposition 6 (blue histogram) of the difference in mean-variance utilities net of price-impact costs (IA2) for all pairs of models that are nested. The title of each sub-figure illustrates the two models for comparison. The finite-sample distribution is obtained by evaluating (IA2) on 10,000 bootstrap samples with $T^b = 491$ observations, and the asymptotic distribution is based on Proposition 6

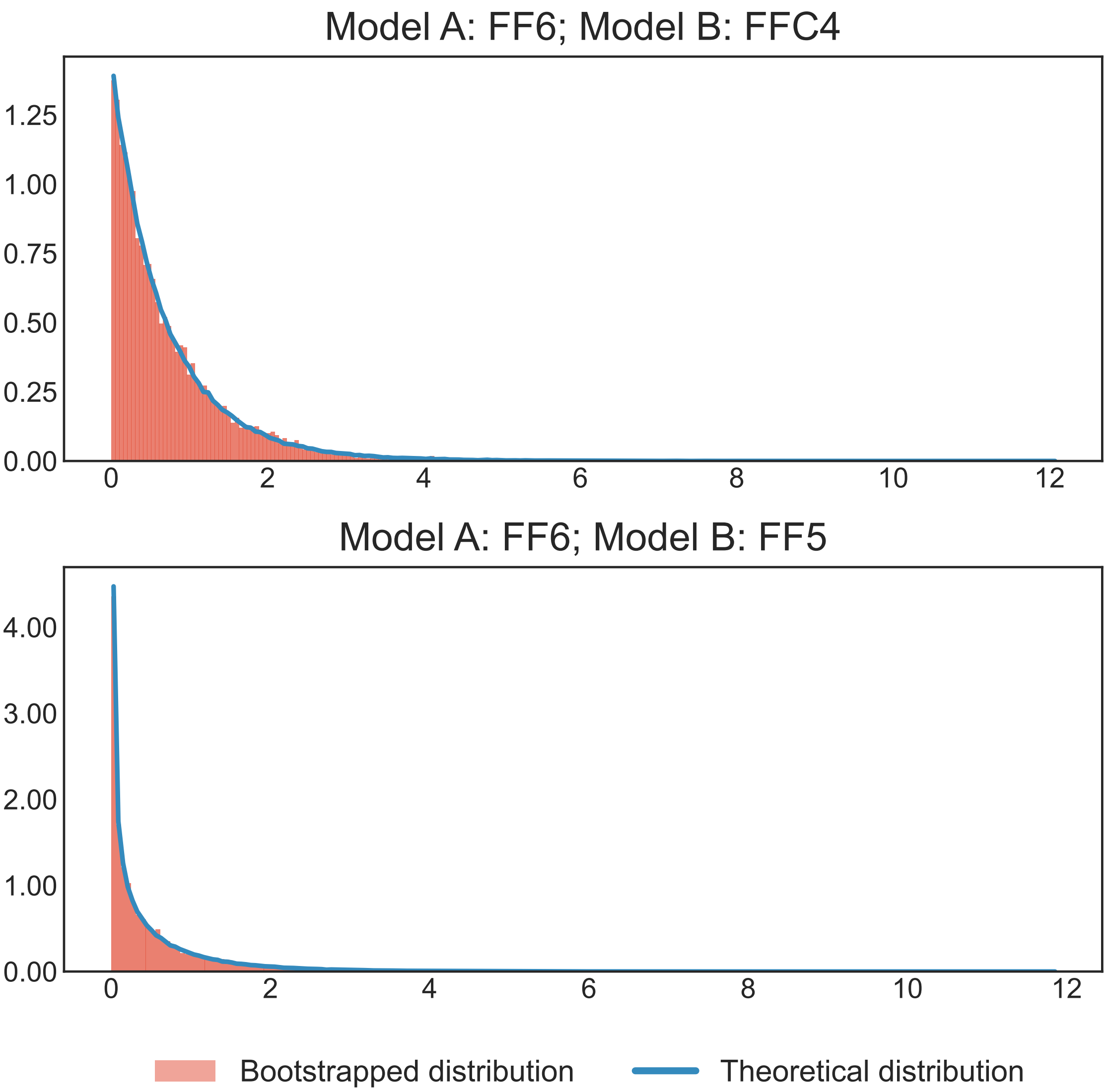


Table IA.1: Comparing p-values using Proposition 6 and the GRS test

This table reports the p-values of the test using Proposition 6 and of the GRS test for nested models in the absence of trading costs. The first column lists the acronyms of the nested models. The second column reports the p-value of the test based on Proposition 6. The third and the fourth columns report the p-value of the finite-sample GRS test and of the asymptotic GRS test, respectively.

	p-values		
	Proposition 6	Finite-sample GRS	Asymptotic GRS
FFC4 v. FF5	0.000	0.000	0.000
FF5 v. FF6	0.034	0.007	0.006

Table IA.2: Bootstrap out-of-sample (OOS) utility net of price-impact costs ($\gamma = 2.5 \times 10^{-10}$)

Panel A reports the average scaled OOS mean-variance utility net of price-impact costs across 100,000 bootstrap samples of each factor model under the baseline case with absolute risk-aversion parameter $\gamma = 2.5 \times 10^{-10}$. Panel B reports the frequency with which the row model outperforms the column model out-of-sample across the bootstrap samples.

Panel A: Average mean-variance utilities

	HXZ4	FFC4	FF5	FF6	DMNU20
$2\gamma\hat{U}_{\Lambda}^{\gamma}$	0.0114	0.0102	0.0120	0.0126	0.0176

Panel B: Frequency row model outperforms column model

	HXZ4	FFC4	FF5	FF6	DMNU20
HXZ4		0.590	0.346	0.311	0.267
FFC4			0.208	0.147	0.236
FF5				0.234	0.268
FF6					0.280

Table IA.3: Bootstrap out-of-sample (OOS) utility net of price-impact costs ($\gamma = 1 \times 10^{-7}$)

Panel A reports the average scaled OOS mean-variance utility net of price-impact costs across 100,000 bootstrap samples of each factor model under the baseline case with absolute risk-aversion parameter $\gamma = 1 \times 10^{-7}$. Panel B reports the frequency with which the row model outperforms the column model out-of-sample across the bootstrap samples.

Panel A: Average mean-variance utilities

	HXZ4	FFC4	FF5	FF6	DMNU20
$2\gamma\hat{U}_{\Lambda}^{\gamma}$	0.0753	0.0212	0.0573	0.0618	0.0036

Panel B: Frequency row model outperforms column model

	HXZ4	FFC4	FF5	FF6	DMNU20
HXZ4		0.924	0.707	0.598	0.803
FFC4			0.198	0.101	0.544
FF5				0.312	0.725
FF6					0.754