The Review of Financial Studies

A Transaction-Cost Perspective on the Multitude of Firm Characteristics

Victor DeMiguel
London Business School

Alberto Martín-Utrera
New Jersey Institute of Technology

Francisco J. Nogales
Universidad Carlos III de Madrid

Raman Uppal
EDHEC Business School, EDHEC Scientific Beta Research Chair, and CEPR

Received March 9, 2018; editorial decision April 20, 2019 by Editor Andrew Karolyi. Authors have furnished an Internet Appendix, which is available on the Oxford University Press Web site next to the link to the final published paper online.

We thank the Editor (Andrew Karolyi) and two anonymous referees for valuable feedback. We gratefully acknowledge comments from Torben Andersen, Turan Bali, Pedro Barroso, Kerry Back, Malcolm Baker, Alexandre Belloni, Jonathan Berk, Harjoat Bhamra, John Birge, Michael Brandt, Svetlana Bryzgalova, Andrea Buraschi, Veronika Czellar, Serge Darolles, Alex Edmans, Wayne Ferson, Lorenzo Garlappi, René Garcia, Francisco Gomes, Amit Goyal, Nick Hirschey, Christian Julliard, Petri Jylha, Bige Kahraman, Nishad Kapadia, Ralph Koijen, Robert Kosowski, Apostolos Kourtis, Serhiy Kozak, Anton Lines, Abraham Lioui, Raphael Markellos, Lionel Martellini, Spyros Mesomeris, Maurizio Montone, Narayan Naik, Stefan Nagel, Andreas Neuhierl, Ľuboš Pástor, Barbara Ostdiek, Anna Pavlova, Julien Penasse, Ludovic Phalippou, Ilaria Piatti, Jeffrey Pontiff, Riccardo Rebonato, Scott Richardson, Thierry Roncalli, Shrihari Santosh, James Sefton, Georgios Skoulakis, Stephen Taylor, Nikolaos Tessaromatis, Grigory Vilkov, Christian Wagner, Michael Weber, Dacheng Xiu, Paolo Zaffaroni, Frank Zhang, and Lu Zhang and seminar participants at American Finance Association, Cass Business School, Citi Global Quant Research Conference, Conference on Computational and Financial Econometrics, Deutsche Bank Global Quantitative Conference, Durham University Business School, Edhec Business School, European Finance Association, Frankfurt School of Finance and Management, French Finance Association, Frontiers of Factor Investing Conference (Lancaster), HEC Paris, Imperial College Business School, INET Econometric Seminar Series (Oxford), INFORMS Annual Meeting, INQUIRE Europe, INQUIRE UK, Invesco, KU Leuven, Lancaster University Management School, London Business School, London School of Economics, New Methods for the Empirical Analysis of Financial Markets Conference, Northern Finance Association, Norwich Business School, Rice University, Saïd Business School, Stevanovich Center at University of Chicago, Universitat Pompeu Fabra, Université Catolique de Louvain, Université de Nantes, University College Dublin, Vienna University of Economics and Business, World Symposium on Investment Research (Montreal), 5th Luxembourg Asset Management Summit, 10th International Conference on Computational and Financial Econometrics (Seville), 30th Annual Seminar of the London Quant Group (Oxford), and XXIV Finance Forum (Madrid). Nogales acknowledges support from the Spanish government [for projects MTM2013-44902-P and MTM2017-88797-P] and the UC3M-BS Institute of Financial Big Data. Supplementary data can be found on *The Review of Financial Studies* web site. Send correspondence to Victor DeMiguel, London Business School, 26 Sussex Place, London NW1 4SA, United Kingdom; telephone: +44-20-7000-8831. E-mail: avmiguel@london.edu.

The Review of Financial Studies 33 (2020) 2180–2222

doi:10.1093/rfs/hhz085

We investigate how transaction costs change the number of characteristics that are *jointly* significant for an investor's optimal portfolio and, hence, how they change the dimension of the cross-section of stock returns. We find that transaction costs increase the number of significant characteristics from 6 to 15. The explanation is that, as we show theoretically and empirically, combining characteristics reduces transaction costs because the trades in the underlying stocks required to rebalance different characteristics often cancel out. Thus, transaction costs provide an economic rationale for considering a larger number of characteristics than that in prominent asset-pricing models. (*JEL* G11)

Hundreds of variables have been proposed to explain the cross-section of stock returns (see, for instance, Harvey, Liu, and Zhu 2015; McLean and Pontiff 2016; Hou, Xue, and Zhang 2017). This abundance of cross-sectional predictors leads Cochrane (2011, p. 1060) to ask, "Which characteristics really provide independent information about average returns? Which are subsumed by others?" Likewise, Goyal (2012, p. 23) states that "these days one has a multitude of variables that seem to explain the cross-sectional pattern of returns. The amount of independent information in these variables is unclear as no study to date [...] has conducted a comprehensive study to analyze the joint impact of these variables."

Cochrane and Goyal challenge researchers to characterize the *dimension* of the cross-section of stock returns by identifying a small set of characteristics that subsume the rest.[1] Several papers address this challenge in the *absence* of transaction costs; these include Feng, Giglio, and Xiu (2017), Freyberger, Neuhierl, and Weber (2018), Green, Hand, and Zhang (2017), Kelly, Pruitt, and Su (2018), Kozak, Nagel, and Santosh (forthcoming), and Messmer and Audrino (2017). However, transaction costs matter for the dimension of the cross-section because they affect the number of characteristics that are *jointly* significant for an investor's optimal portfolio. To address this gap in the literature, our objective is to study how transaction costs affect the dimension of the cross-section.

We build on the insightful work of Novy-Marx and Velikov (2016) who propose a generalization of Jensen's alpha that takes transaction costs into account, and compute the "generalized alpha" of 23 characteristics with respect to the four factors in Fama and French (1993) and Carhart (1997). Novy-Marx and Velikov (2016) find that in the presence of transaction costs the number of characteristics that have a significant generalized alpha is *smaller* than in the absence of transaction costs.

However, Novy-Marx and Velikov (2016) test the significance of a *single* characteristic at a time. We, in contrast, consider all characteristics

[1] Note that, as explained by Cochrane (2011), the cross-section of stock returns can be always explained by a single-factor model, where the mean-variance efficient portfolio is the single factor. However, the mean-variance efficient portfolio is not observable, and estimating it from individual stock-return data is difficult because of estimation error. Therefore, the literature has focused on finding a small number of observable firm characteristics that explain the cross-section of stock returns.

simultaneously and, as a result, find that the number of characteristics that are jointly significant for an investor's portfolio in the presence of transaction costs is *larger* than in the absence of transaction costs. The explanation for this is that, as we show theoretically and empirically, combining characteristics *reduces* transaction costs, and hence increases the investor's utility, because the trades in the underlying stocks required to rebalance different characteristics often cancel each other out.[2] Essentially, combining characteristics allows one to *diversify trading*, just as combining them allows one to diversify risk. As a consequence, our work shows that the impact of transaction costs is *smaller* when considering characteristics jointly than when considering them one at a time, as in Novy-Marx and Velikov (2016).

We first quantify the benefits from trading diversification in a simple manner by comparing the average trading volume (turnover) required to exploit an *equally weighted* portfolio of characteristics simultaneously with that required to exploit them in isolation. Analytically, we show that the turnover required to rebalance an equally weighted portfolio of K characteristics is about $1/\sqrt{K}$ of that required to rebalance the characteristics in isolation. Empirically, we find that while the average monthly turnover required to exploit a characteristic in isolation is 24.09%, the turnover required to exploit an equally weighted combination of characteristics is only 6.71%; that is, trading diversification delivers a 72.15% reduction in turnover. Note that a reduction in turnover will translate into a reduction in transaction costs regardless of the particular manner in which transaction costs are modeled.

We then turn to our main research question of how transaction costs affect the dimension of the cross-section. To answer this question, we study how many firm-specific characteristics matter *jointly* from a *portfolio perspective*; that is, from the perspective of an investor who cares not only about average returns, but also about portfolio risk and transaction costs.[3] To do this, we extend the "parametric portfolios" in Brandt, Santa-Clara, and Valkanov (2009) and use them as an alternative method to the traditional regression approaches, which cannot answer our main research question because they either ignore transaction costs or consider characteristics one at a time.[4]

Parametric portfolios are obtained by adding to a benchmark portfolio a linear combination of the long-short portfolios associated with each characteristic considered. To determine which characteristics are jointly significant, we use

[2] For instance, assume that rebalancing a momentum portfolio requires buying \$3,000 of the Apple stock, whereas rebalancing a value portfolio requires selling \$2,000 of Apple. Then, rebalancing a combination of these two characteristics requires buying only \$1,000 of Apple.

[3] We are agnostic about whether or not a particular characteristic is a proxy for the loading on a common risk factor; instead, we directly account for risk via the mean-variance utility of the investor.

[4] Although the distinguishing feature of our approach is that it accounts for *transaction costs*, in Appendix A we also characterize the theoretical relation of the parametric-portfolio approach to cross-sectional and time-series regressions in the *absence* of transaction costs.

a "screen-and-clean" method to test which characteristics have parametric-portfolio weights that are significantly different from zero. We then use this test to compare the number of characteristics that are jointly significant in the absence and presence of transaction costs.

We find that in the absence of transaction costs, of the 51 characteristics that we consider, only a small number—about 6—are significant. Moreover, in contrast to what one would observe if evaluating characteristics in isolation, we find that transaction costs *increase* the number of jointly significant characteristics from 6 to 15, thus increasing the dimension of the cross-section.[5] This is because the benefits of trading diversification are large when combining characteristics to maximize the investor's expected utility: we find empirically that the marginal transaction cost of trading the stocks underlying a characteristic is reduced by 65% on average when characteristics are combined optimally in the parametric portfolio.

Our findings have implications for asset-pricing theories based on stochastic discount factors (SDFs) because the investor's first-order optimality condition determines not only her optimal portfolio but also the associated SDF, as shown in Appendix B. Thus, our work shows that transaction costs provide a rationale for considering a larger number of characteristics than that in prominent asset-pricing models.

To alleviate data-mining concerns raised in the literature,[6] we also undertake an *out-of-sample* analysis. We find that the out-of-sample performance of the parametric portfolios in the presence of transaction costs can be significantly improved by exploiting a large number of characteristics instead of the small number typically considered in popular asset-pricing models.[7]

We now discuss how our work is related to the literature. Several papers use *cross-sectional regressions* to study the dimension of the cross-section because they allow one to test which characteristics are *jointly* significant (see, for instance, Green, Hand, and Zhang 2017; Freyberger, Neuhierl, and Weber 2018; Messmer and Audrino 2017). Light, Maslov, and Rytchkov (2017) use an information-aggregation technique based on the three-pass regression filter in Kelly and Pruitt (2015) to aggregate multiple characteristics into a few composite variables that predict the cross-section of expected stock returns. Whereas all of these papers ignore transaction costs, we focus on the effect of transaction costs. Another difference is that while cross-sectional regressions

---

[5] In Section IA.7 of the Online Appendix, we show that our main insight is robust to considering a larger set of 100 characteristics. For this larger data set, we find that while 7 characteristics are significant in the absence of transaction costs, 15 are significant in the presence of transaction costs.

[6] See, for example, Fama (1991), Kogan and Tian (2013), Harvey, Liu, and Zhu (2015), Bryzgalova (2015), McLean and Pontiff (2016), Linnainmaa and Roberts (2018), and Chordia, Goyal, and Saretto (2017).

[7] The out-of-sample Sharpe ratio of returns net of transaction costs from exploiting 51 characteristics is around 100% larger than that from exploiting the 3 traditional characteristics considered in Brandt, Santa-Clara, and Valkanov (2009) and 25% higher than that from exploiting a set of 4 characteristics that include investment and profitability characteristics, highlighted in Hou, Xue, and Zhang (2014) and Fama and French (2016).

focus on mean returns, our portfolio approach accounts for *both* mean and variance of returns.

The *time-series approach* regresses the return of a characteristic-based long-short portfolio on the returns of a few commonly accepted factors. If the intercept (or alpha) is statistically significant, then the return on the characteristic is not fully explained by the commonly accepted factors. Gibbons, Ross, and Shanken (1989) show that this approach captures the trade-off between mean return and risk. Novy-Marx and Velikov (2016) extend time-series regressions to capture transaction costs. The focus of the time-series approach on the regression *intercept* implies that it evaluates the significance of a *single* characteristic at a time. This is a limitation because, as we show in Appendix A.2, the significance results depend on the *sequence in which variables are selected*. In contrast, our portfolio approach considers *all* characteristics simultaneously.[8]

Some papers combine elements from both cross-sectional and time-series regressions (see, for instance, Back, Kapadia, and Ostdiek 2015; Baker, Luo, and Taliaferro 2017; Feng, Giglio, and Xiu 2017). Just as for time-series regressions, the inference in these papers also depends on the *sequence* in which characteristics are tested.

Finally, the *stochastic discount factor* (SDF) approach is closely related to our portfolio approach because the first-order optimality condition of the investor determines not only her optimal portfolio but also the associated SDF. Kozak, Nagel, and Santosh (forthcoming) propose a robust SDF and finds that a small number of principle components predict the cross-section better than a small number of characteristics. Two main differences separate this paper from our work. First, we study the impact of *transaction costs* on the dimension of the cross-section of stock returns. Second, while Kozak, Nagel, and Santosh (forthcoming) focus on *prediction*, our work focuses on *inference*, because we wish to study how transaction costs affect the number of characteristics that are jointly *significant*. Our main finding is that transaction costs *increase* the number of characteristics that are jointly significant. Thus, our work provides another rationale for considering a larger number of characteristics than that in prominent asset-pricing models.

Several papers study the transaction costs associated with trading *individual* characteristics. Korajczyk and Sadka (2004) find that momentum can be exploited on only a modest scale. Novy-Marx and Velikov (2016) find that simple transaction-cost mitigation strategies such as introducing a buy/hold spread can substantially reduce transaction costs. Chen and Velikov (2017) show that if, in addition to transaction costs, one accounts for post-publication

[8] We show analytically in Appendix A.2 that our approach of testing the significance of the characteristics for mean-variance parametric portfolios is *equivalent* to testing the significance of each *slope* in a particular time-series regression; that is, our significance test is equivalent to a t-test of explanatory variables in a multiple regression.

decay, the profitability of anomaly-based trading strategies is substantially diminished. These papers use publicly available data sets to estimate the trading costs of an average investor. In contrast, Frazzini, Israel, and Moskowitz (2015) use proprietary data and find that the trading costs associated with exploiting size, momentum, and book to market can be quite small for large institutional investors.

Other papers have also found that combining characteristics helps to reduce transaction costs. Frazzini, Israel, and Moskowitz (2015, p. 23) explain that "value and momentum trades tend to offset each other, resulting in lower turnover which has real transaction costs benefits." Barroso and Santa-Clara (2015, p. 1043) consider currency portfolios based on 6 characteristics and explain that "transaction costs depend crucially on the time-varying interaction between characteristics." Novy-Marx and Velikov (2016) study "filtering," a cost mitigation technique that allows investors trading one strategy to opportunistically take small positions in another at effectively *negative* trading costs. The distinguishing feature of our work is that we consider a large number of characteristics *jointly* to show how transaction costs lead to an increase in the dimension of the cross-section.

1. Data

We combine U.S. stock-market information from CRSP, Compustat, and I/B/E/S, covering the period from January 1980 to December 2014. We start by compiling data on the 100 firm-specific characteristics considered in Green, Hand, and Zhang (2017),[9] but drop characteristics with a large proportion of missing observations to ensure that our results are reliable. Specifically, we drop characteristics with more than 5% of missing observations for more than 5% of firms with CRSP returns available for the entire sample from 1980 to 2014. In addition, we drop characteristics without any observations for more than 1% of these firms. Table 1 lists the resultant 51 characteristics together with their definitions, the name of the author(s) who identified them, and the date and journal of publication.

Our initial database contains every firm traded on the NYSE, AMEX, and NASDAQ exchanges. We then remove firms with negative book-to-market ratios. As in Brandt, Santa-Clara, and Valkanov (2009), we also remove firms below the 20th percentile of market capitalization because these are very small firms that are difficult to trade. Our final data set contains 51 firm-specific characteristics for a total of 17,930 firms of which an average of 3,071 firms have return data in a particular month.

---

[9] Like in Green, Hand, and Zhang (2017), when constructing monthly characteristics at time t, we assume that annual (quarterly) accounting data are available at the end of month $t-1$ if the firm's fiscal year ended at least 6 (4) months earlier.

Table 1
List of characteristics considered

#	Characteristic and definition	Acronym	Author(s)	Date and journal
1	Abnormal volume in earnings announcement: Average daily trading volume for 3 days around earnings announcement minus average daily volume for 1-month ending 2 weeks before earnings announcement divided by 1-month average daily volume. Earnings announcement day from Compustat quarterly	aeavol	Lerman, Livnat, and Mendenhall	2007, WP
2	Asset growth: Annual percentage change in total assets	agr	Cooper, Gulen, and Schill	2008, JF
3	Bid-ask spread: Monthly average of daily bid-ask spread divided by average of daily spread	baspread	Amihud, and Mendelson	1989, JF
4	Beta: Estimated market beta from weekly returns and equal weighted market returns for 3 years ending month $t-1$ with at least 52 weeks of returns	beta	Fama and MacBeth	1973, JPE
5	Book to market: Book value of equity divided by end of fiscal-year market capitalization	bm	Rosenberg, Reid, and Lanstein	1985, JPM
6	Industry adjusted book to market: Industry adjusted book-to-market ratio	bm_ia	Asness, Porter, and Stevens	2000, WP
7	Cash productivity: Fiscal year-end market capitalization plus long term debt minus total assets divided by cash and equivalents	cashpr	Chandrashekar and Rao	2009 WP
8	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
9	Change in shares outstanding: Annual percentage change in shares outstanding	chcsho	Pontiff and Woodgate	2008, JF
10	Industry adjusted change in employees: Industry-adjusted change in number of employees	chempia	Asness, Porter, and Stevens	1994, WP
11	Change in 6-month momentum: Cumulative returns from months $t-6$ to $t-1$ minus months $t-12$ to $t-7$	chmom	Gettleman and Marks	2006 WP
12	Industry adjusted change in profit margin: 2-digit SIC fiscal-year mean adjusted change in income before extraordinary items divided by sales	chpmia	Soliman	2008, TAR
13	Change in tax expense: Percent change in total taxes from quarter $t-4$ to t	chtx	Thomas and Zhang	2011 JAR
14	Convertible debt indicator: An indicator equal to 1 if company has convertible debt obligations	convind	Valta	2016 JFQA
15	Dollar trading volume in month $t-2$: Natural log of trading volume times price per share from month $t-2$	dolvol	Chordia, Subrahmanyan, and Anshuman	2001, JFE
16	Dividends-to-price: Total dividends divided by market capitalization at fiscal year-end	dy	Litzenberger and Ramaswamy	1982, JF
17	3-day return around earnings announcement: Sum of daily returns in three days around earnings announcement. Earnings announcement from Compustat quarterly file	ear	Kishore, Brandt, Santa-Clara, and Venkatachalam	2008, WP
18	Change in common shareholder equity: Annual percentage change in book value of equity	egr	Richardson, Sloan, Soliman, and Tuna	2005, JAE
19	Earnings to price: Annual income before extraordinary items divided by end of fiscal year market cap	ep	Basu	1977, JF
20	Gross profitability: Revenues minus cost of goods sold divided by lagged total assets	gma	Novy-Marx	2013 JFE
21	Industry sales concentration: Sum of squared percentage of sales in industry for each company	herf	Hou and Robinson	2006, JF
22	Employee growth rate: Percent change in number of employees	hire	Bazdresch, Belo, and Lin	2014 JPE
23	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, and Trombley	2003, JFE
24	Industry momentum: Equal weighted average industry 12-month returns	indmom	Moskowitz and Grinblatt	1999, JF

(*Continued*)

Table 1
(Continued)

#	Characteristic and definition	Acronym	Author(s)	Date and journal
25	Leverage: Total liabilities divided by fiscal year-end market capitalization	lev	Bhandari	1988, JF
26	Change in long-term debt: Annual percentage change in total liabilities	lgr	Richardson, Sloan, Soliman, and Tuna	2005, JAE
27	12-month momentum: 11-month cumulative returns ending one month before month-end	mom12m	Jegadeesh	1990, JF
28	1-month momentum: 1-month cumulative return	mom1m	Jegadeesh	1990, JF
29	36-month momentum: Cumulative returns from months $t-36$ to $t-13$	mom36m	De Bondt and Thaler	1985, JF
30	6-month momentum: 5-month cumulative returns ending one month before month-end	mom6m	Jegadeesh and Titman	1990, JF
31	Market capitalization: Natural log of market capitalization at end of month $t-1$	mve	Banz	1981, JFE
32	Industry-adjusted firm size: 2-digit SIC industry-adjusted fiscal year-end market capitalization	mve_ia	Asness, Porter, and Stevens	2000, WP
33	Δ% CAPEX - industry Δ% AR: 2-digit SIC fiscal-year mean adjusted percentage change in capital expenditures	pchcapx_ia	Abarbanell and Bushee	1998, TAR
34	Δ% gross margin - Δ% sales: Percent change in gross margin minus percentage change in sales	pchgm_pchsale	Abarbanell and Bushee	1998, TAR
35	Δ% sales - Δ% AR: Annual percentage change in sales minus annual percentage change in receivables	pchsale_pchrect	Abarbanell and Bushee	1998, TAR
36	Price delay: The proportion of variation in weekly returns for 36 months ending in month t explained by 4 lags of weekly market returns incremental to contemporaneous market return	pricedelay	Hou and Moskowitz	2005, RFS
37	Financial-statements score: Sum of 9 indicator variables to form fundamental health score	ps	Piotroski	2000, JAR
38	R&D to market cap: R&D expense divided by end-of-fiscal-year market capitalization	rd_mve	Guo, Lev, and Shi	2006, JBFA
39	Return volatility: Standard deviation of daily returns from month $t-1$	retvol	Ang, Hodrick, Xing, and Zhanf	2006, JF
40	Return on assets: Income before extraordinary items divided by one quarter lagged total assets	roaq	Balakrishnan, Bartov, and Faurel	2010, JAE
41	Revenue surprise: Sales from quarter t minus sales from quarter $t-4$ divided by fiscal-quarter-end market capitalization	rsup	Kama	2009, JBFA
42	Sales to cash: Annual sales divided by cash and cash equivalents	salecash	Ou and Penman	1989, JAE
43	Sales to inventory: Annual sales divided by total inventory	saleinv	Ou and Penman	1989, JAE
44	Sales to receivables: Annual sales divided by accounts receivable	salerec	Ou and Penman	1989, JAE
45	Annual sales growth: Annual percentage change in sales	sgr	Lakonishok, Shleifer, and Vishny	1994, JF
46	Volatility of dollar trading volume: Monthly standard deviation of daily dollar trading volume	std_dolvol	Chordia, Subrahmanyan, and Anshuman	2001, JFE
47	Volatility of share turnover: Monthly standard deviation of daily share turnover	std_turn	Chordia, Subrahmanyan, and Anshuman	2001, JFE
48	Cashflow volatility: Standard deviation for 16 quarters of cash flows divided by sales	stdcf	Huang	2009, JEF
49	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, and Latane	1982, JFE
50	Share turnover: Average monthly trading volume for most recent 3 months scaled by number of shares outstanding in current month	turn	Datar, Naik, and Radcliffe	1998, JFM
51	Zero trading days: Turnover weighted number of zero trading days for most recent month	zerotrade	Liu	2006, JFE

This table lists the characteristics we consider, ordered alphabetically by acronym. The first column gives the number of the characteristic, the second column gives the characteristic'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. Our definitions and acronyms match those in Green, Hand, and Zhang (2017).

We cross-sectionally winsorize each characteristic; that is, we replace extreme observations that are beyond a certain threshold with the value of the threshold. Specifically, we set equal to the third (first) quartile plus (minus) three times the interquartile range any observations that are above (below) this threshold.[10]

Finally, like in Brandt, Santa-Clara, and Valkanov (2009), we standardize each characteristic so that it has a cross-sectional mean of zero and standard deviation of one. The resultant standardized characteristic is a long-short portfolio that goes long stocks whose characteristic is above the cross-sectional average, and short stocks whose characteristic is below the cross-sectional average.

2. Methodology

This section explains how we extend the parametric-portfolio methodology in Brandt, Santa-Clara, and Valkanov (2009) in order to study how transaction costs change the number of characteristics that are jointly significant for an investor's portfolio. We also will describe below the screen-and-clean test used to evaluate whether the parametric-portfolio weight corresponding to each characteristic is significant. In addition, Appendix A compares analytically and empirically our methodological approach based on parametric portfolios with cross-sectional and time-series regressions.

2.1 Mean-variance parametric portfolios

Parametric portfolios use a set of firm-specific characteristics to *tilt* the benchmark portfolio toward stocks that help to increase the investor's utility. The portfolios are obtained by adding to the benchmark portfolio a linear combination of long-short portfolios obtained by standardizing K firm-specific characteristics cross-sectionally. The resultant parametric portfolio at time t, $w_t(\theta)\in\mathbb{R}^{N_t}$, can be written as

$$w_t(\theta)=w_{b,t}+(x_{1,t}\theta_1+x_{2,t}\theta_2+\ldots+x_{K,t}\theta_K)/N_t, \tag{1}$$

where $w_{b,t}\in\mathbb{R}^{N_t}$ is the *benchmark portfolio* at time t, $x_{k,t}\in\mathbb{R}^{N_t}$ is the long-short portfolio obtained by standardizing the kth firm-specific characteristic at time t, θ_k is the weight of the kth characteristic in the parametric portfolio, and N_t is the number of firms at time t.[11] Like in Brandt, Santa-Clara, and Valkanov (2009), we consider a portfolio that is

[10] This winsorization is the one used in the 2014 version of Green, Hand, and Zhang (2017). Section IA.9 of the Online Appendix shows that our findings are robust to winsorizing the data at the 1st and 99th cross-sectional percentiles, as in the published version of Green, Hand, and Zhang (2017).

[11] The weights of the characteristics in the parametric portfolio are scaled by the number of stocks N_t so that they are meaningful for the case with a varying number of stocks. Without this scaling parameter, increasing the number of stocks while keeping the weights fixed would result in more aggressive portfolio allocations.

fully invested in risky assets.[12] The parametric portfolio can also be written in compact matrix notation by defining $X_t \in \mathbb{R}^{N_t \times K}$ to be the matrix whose kth column is $x_{k,t}$:

$$w_t(\theta) = w_{b,t} + X_t \theta / N_t, \tag{2}$$

where $\theta \in \mathbb{R}^K$ is the *parameter vector*, whose kth component is the weight of the kth characteristic θ_k, and $X_t\theta/N_t$ is the characteristic portfolio at time t.

The return of the parametric portfolio at time $t+1$, which we denote as $r_{p,t+1}(\theta)$, can thus be rewritten as

$$r_{p,t+1}(\theta) = w_{b,t}^\top r_{t+1} + \theta^\top X_t^\top r_{t+1}/N_t = r_{b,t+1} + \theta^\top r_{c,t+1}, \tag{3}$$

where $r_{t+1} \in \mathbb{R}^{N_t}$ is the return vector at time $t+1$, $r_{b,t+1} = w_{b,t}^\top r_{t+1}$ is the benchmark portfolio return at time $t+1$, and $r_{c,t+1} = X_t^\top r_{t+1}/N_t$ is the *characteristic return vector* at time $t+1$, which contains the returns of the long-short portfolios corresponding to the K characteristics scaled by the number of firms N_t.[13] Equation (3) shows that the parametric-portfolio return is the benchmark-portfolio return plus the return of the characteristic portfolio.

We assume that the investor optimizes mean-variance utility. The advantages of mean-variance utility, as we will show below, are that it allows us to identify the marginal contribution of each characteristic to the investor's utility and to compare analytically the parametric-portfolio weights to the results from time-series and cross-sectional regressions.[14] In particular, we assume the investor solves the following problem:

$$\min_{\theta} \quad \frac{\gamma}{2} \mathrm{var}_t[r_{p,t+1}(\theta)] - E_t[r_{p,t+1}(\theta)], \tag{4}$$

where γ is the risk-aversion parameter and $\mathrm{var}_t[r_{p,t+1}(\theta)]$ and $E_t[r_{p,t+1}(\theta)]$ are the variance and mean of the parametric-portfolio return, respectively.

Given T historical observations of returns and characteristics, the following proposition shows that the parametric-portfolio problem can be formulated as a tractable quadratic optimization problem.

Proposition 1. The mean-variance parametric-portfolio problem in (4) is equivalent to

$$\min_{\theta} \quad \underbrace{(\gamma/2)\theta^\top \widehat{\Sigma}_c \theta}_{var(char)} + \underbrace{\gamma \theta^\top \widehat{\sigma}_{bc}}_{cov(bench)} - \underbrace{\theta^\top \widehat{\mu}_c}_{mean}, \tag{5}$$

where $\widehat{\Sigma}_c$ and $\widehat{\mu}_c$ are the sample covariance matrix and mean of the characteristic-return vector r_c, and $\widehat{\sigma}_{bc}$ is the sample vector of covariances

[12] Consequently, the parametric-portfolio weights on the stocks sum to one. Because the weights on the stocks in the long-short portfolios sum to zero, this implies that the parametric weight on the benchmark portfolio must equal one.

[13] Note that we use only lagged values of characteristics to build portfolios; thus, the returns of the characteristic portfolio formed at time t, $X_t\theta/N_t$ are evaluated using the return at time $t+1$; that is, $\theta^\top X_t^\top r_{t+1}/N_t$.

[14] We have also run our empirical analysis for power utility, and the main insights are unchanged.

between the benchmark portfolio return r_b and the characteristic-return vector r_c.

Proposition 1 shows that the mean-variance parametric-portfolio problem finds the vector θ with the optimal trade-off among the variance of the characteristic portfolio return, $(\gamma/2)\theta^\top\widehat{\Sigma}_c\theta$; the covariance of the characteristic portfolio return with the benchmark portfolio return, $\gamma\theta^\top\widehat{\sigma}_{bc}$; and the mean characteristic portfolio return, $\theta^\top\widehat{\mu}_c$.

2.2 Transaction costs

We consider an investor who faces proportional transaction costs that decrease with firm size and over time, as parameterized in Brandt, Santa-Clara, and Valkanov (2009) and Hand and Green (2011). Sections IA.1 and IA.2 of the Online Appendix, respectively, show that our findings are robust to estimating proportional transaction costs from daily price data and to considering quadratic transaction costs, which are often used to model the price-impact costs of large investors.

Let the proportional transaction-cost parameter for the ith stock at time t be

$$\kappa_{i,t} = y_t z_{i,t}, \tag{6}$$

where y_t and $z_{i,t}$ capture the variation of the transaction-cost parameter with time and firm size, respectively. Following Brandt, Santa-Clara, and Valkanov (2009) and Hand and Green (2011), we assume y_t decreases linearly from 3.3 in January 1980 to 1.0 in January 2002, and after that it remains at 1.0.[15] We set $z_{i,t} = 0.006 - 0.0025 \times me_{i,t}$, where $me_{i,t}$ is the market capitalization of firm i at time t after being normalized cross-sectionally so that it takes values between zero and one. This functional form results in proportional transaction costs in the 1980s of about 180 basis points for the smallest firms and 100 basis points for the largest firms, and after 2002 of about 60 basis points for the smallest firms and 35 basis points for the largest firms.

Given T historical observations of returns and characteristics, the transaction cost associated with implementing the parametric portfolios can be estimated as

$$\mathrm{TC}(\theta) = \frac{1}{T-1}\sum_{t=1}^{T-1} \| \Lambda_t (w_{t+1}(\theta) - w_t^+(\theta)) \|_1, \tag{7}$$

where the transaction-cost matrix at time t, Λ_t, is the diagonal matrix whose ith diagonal element contains $\kappa_{i,t}$, $\| a \|_1 = \sum_{i=1}^N \| a_i \|$ is the 1-norm of the

15 Brandt, Santa-Clara, and Valkanov (2009) define y_t so that transaction costs in 1974 are four times larger than in 2002. Therefore, if we decrease y_t uniformly until 1980, we would have a starting value for y_t approximately equal to 3.3.

N-dimensional vector a, and w_t^+ is the portfolio before rebalancing at time $t+1$, that is,

$$w_t^+ = \left(w_{b,t} + X_t \times \theta / N_t\right) \circ (e_t + r_{t+1}), \tag{8}$$

where e_t is the N_t-dimensional vector of ones and $x \circ y$ is the Hadamard or componentwise product of vectors x and y. Combining (5) and (7), the mean-variance parametric-portfolio problem *with* transaction costs is

$$\min_{\theta} \quad \underbrace{(\gamma/2)\theta^\top \widehat{\Sigma}_c \theta}_{var(char)} + \underbrace{\theta^\top \gamma \widehat{\sigma}_{bc}}_{cov(bench)} - \underbrace{\theta^\top \widehat{\mu}_c}_{mean} + \underbrace{\mathrm{TC}(\theta).}_{transaction\ costs} \tag{9}$$

2.3 Understanding why a characteristic matters

To understand why particular characteristics are significant from a portfolio perspective, it is useful to consider the first-order optimality conditions for the mean-variance parametric-portfolio problem with transaction costs in (9).

By decomposing the variance of the characteristic portfolio return, $\theta^\top \widehat{\Sigma}_c \theta$, into a term associated with the characteristic *own-variances*, $\theta^\top \mathrm{diag}(\widehat{\Sigma}_c)\theta$, and a term associated with the characteristic covariances, $\theta^\top(\widehat{\Sigma}_c - \mathrm{diag}(\widehat{\Sigma}_c))\theta$, where $\mathrm{diag}(\widehat{\Sigma}_c)$ is the diagonal matrix whose kth diagonal element contains the variance of the kth characteristic return, the mean-variance parametric-portfolio problem with transaction costs can be rewritten as

$$\min_{\theta} \underbrace{(\gamma/2)\theta^\top \mathrm{diag}(\widehat{\Sigma}_c)\theta}_{own-var(char)} + \underbrace{(\gamma/2)\theta^\top(\widehat{\Sigma}_c - \mathrm{diag}(\widehat{\Sigma}_c))\theta}_{cov(char)} + \underbrace{\theta^\top \gamma \widehat{\sigma}_{bc}}_{cov(bench)} - \underbrace{\theta^\top \widehat{\mu}_c}_{mean} + \underbrace{\mathrm{TC}(\theta).}_{tran.\ costs} \tag{10}$$

Note that the transaction-cost term $\mathrm{TC}(\theta)$ is a convex function of the parameter θ, but it is not differentiable at values of θ for which $w_{i,t+1}(\theta) = w_{i,t}^+(\theta)$ for some i and t. Therefore, the optimality conditions must be formally defined in terms of the subdifferential $\partial \mathrm{TC}(\theta)$.[16]

Proposition 2. The first-order optimality conditions for problem (10) are

$$0 \in \underbrace{\gamma \mathrm{diag}(\widehat{\Sigma}_c)\theta}_{own-var(char)} + \underbrace{\gamma(\widehat{\Sigma}_c - \mathrm{diag}(\widehat{\Sigma}_c))\theta}_{cov(char.)} + \underbrace{\gamma \widehat{\sigma}_{bc}}_{cov(bench.)} - \underbrace{\widehat{\mu}_c}_{mean} + \underbrace{\partial \mathrm{TC}(\theta),}_{costs} \tag{11}$$

where the kth component of the subdifferential of the transaction-cost term is

$$\partial_{\theta_k} \mathrm{TC}(\theta) = \frac{1}{T-1} \sum_{t=1}^{T-1} \mathrm{sign}(w_{t+1}(\theta) - w_t^+(\theta))^\top (\Lambda_t[(X_{t+1})_{\bullet,k} - (X_t)_{\bullet,k} \circ (e_t + r_{t+1})]), \tag{12}$$

where $A_{\bullet,k}$ is the kth column of matrix A, and

$$\mathrm{sign}(w_{i,t+1}(\theta) - w_{i,t}^+(\theta)) = \begin{cases} +1 & \text{if} \quad w_{i,t+1}(\theta) > w_{i,t}^+(\theta), \\ -1 & \text{if} \quad w_{i,t+1}(\theta) < w_{i,t}^+(\theta), \\ [-1,1] & \text{if} \quad w_{i,t+1}(\theta) = w_{i,t}^+(\theta). \end{cases} \tag{13}$$

[16] See Rockafellar (2015) for an extensive treatment of subdifferentials.

The first-order optimality conditions in (11) allow us to identify the *marginal* contribution of each characteristic to the investor's mean-variance utility. Specifically, the kth component of the right-hand side in (11) is the marginal contribution of the kth characteristic to the parametric-portfolio mean-variance utility; that is, the marginal change to mean-variance utility associated with a unit increase in the weight that the parametric portfolio assigns to the kth characteristic. Moreover, the five terms on the right-hand side of (11) are: the marginal contributions of the kth characteristic to the characteristic own-variance, $\gamma \mathrm{diag}(\widehat{\Sigma}_c)\theta$; the characteristic covariance with the other characteristics, $\gamma(\widehat{\Sigma}_c - \mathrm{diag}(\widehat{\Sigma}_c))\theta$; the covariance between the characteristic and the benchmark portfolio, $\gamma\widehat{\sigma}_{bc}$; the characteristic portfolio mean, $-\widehat{\mu}_c$; and the subdifferential of the transaction-cost function, $\partial \mathrm{TC}(\theta)$.

Finally, to gauge the size of the trading-diversification benefit associated with combining characteristics, it will be useful to compute the marginal contribution to transaction costs of trading the kth characteristic in isolation (i.e., without the benchmark or any other characteristics), which is

$$\partial^{iso}_{\theta_k}\mathrm{TC}(\theta) = \frac{1}{T-1}\sum_{t=1}^{T-1} \| \Lambda_t[(X_{t+1})_{\bullet,k} - (X_t)_{\bullet,k} \circ (e_t + r_{t+1})] \|_1 . \tag{14}$$

Straightforward algebra shows that the marginal contribution to transaction costs of trading the kth characteristic in isolation given in (14) is larger in general than that of trading it in combination given in (12).

2.4 The regularized parametric portfolios

To deal with the large number of characteristics in our data set, we develop a new class of parametric portfolios, which we term *regularized parametric portfolios*. These portfolios are obtained by imposing a lasso[17] constraint on the parametric portfolio. This constraint reduces the impact of estimation error and acts as a variable-selection method that helps to reduce problem dimensionality, a feature that makes the regularized parametric portfolios suitable for the first stage of the screen-and-clean significance test described in Section 2.5.

The regularized parametric portfolios are obtained by solving problem (9) subject to the lasso constraint,

$$\min_{\theta} \quad \frac{\gamma}{2}\theta^\top \widehat{\Sigma}_c \theta + \theta^\top \gamma \widehat{\sigma}_{bc} - \theta^\top \widehat{\mu}_c + \mathrm{TC}(\theta), \tag{15}$$

$$\text{s.t.} \quad \| \theta \|_1 \leq \delta, \tag{16}$$

where $\| \theta \|_1 = \sum_{k=1}^{K} |\theta_k|$ is the 1-norm of θ and δ is the threshold parameter. To gain intuition about δ, note that for $\delta = \infty$, we recover the standard parametric

17 The term "lasso" originated as the acronym for least absolute shrinkage and selection operator. Tibshirani (1996) originally proposed the lasso in the context of statistical learning, and it has become a prominent tool in the age of machine learning. See Hastie, Tibshirani, and Wainwright (2015) for an in-depth treatment of the lasso and DeMiguel et al. (2009a) for a Bayesian interpretation of the lasso constraint in the context of portfolio choice.

portfolios, and for $\delta=0$, we recover the benchmark portfolio. Thus, as one increases δ, the regularized parametric portfolios change from the benchmark portfolio toward the unregularized parametric portfolio.

2.5 Screen-and-clean significance test

We now explain how to test whether the parametric-portfolio weights corresponding to the different characteristics are significantly different from zero. Because we consider a large number of characteristics, it is desirable to use a variable-selection method such as lasso to reduce the number of characteristics before testing for significance. However, Chatterjee and Lahiri (2011) show that it is challenging to test for significance in the presence of a lasso constraint. To address this challenge, we use a *two-stage* screen-and-clean method, similar to the methods proposed in Wasserman and Roeder (2009), Meinshausen and Yu (2009), and Meinshausen, Meier, and Buhlmann (2009).

In the first stage, we *screen* the characteristics by using the regularized parametric portfolios. Specifically, we employ fivefold cross-validation, as explained in Hastie, Tibshirani, and Wainwright (2015, section 2.3), to select the lasso threshold δ that optimizes the mean-variance criterion.[18] Using the resultant optimal lasso threshold, we compute the regularized parametric portfolios and "screen" or remove any characteristics with a zero parameter, thus reducing problem dimensionality and paving the way for the second (clean) stage.

In the second stage, we *clean* the characteristics that were not removed in the first stage. That is, we compute the parametric portfolios using the characteristics that survived the first stage, but now *without* a lasso constraint, thus circumventing the concerns highlighted in Chatterjee and Lahiri (2011). We then apply a bootstrap method to establish which of these characteristics have parametric-portfolio weights that are significantly different from zero. Specifically, we apply the percentile-interval method (Hastie, Tibshirani, and Wainwright 2015, section 6.2) to establish significance of the surviving characteristics. First, we generate 1,000 bootstrap samples from the original data set using sampling with replacement. Second, we estimate the optimal parametric portfolio for each bootstrap sample. Third, we declare as significant those characteristics whose estimated parameter is strictly positive (strictly negative) for at least 95% of the bootstrap samples, and compute the p-value as the proportion of bootstrap samples for which the parameter is nonpositive (nonnegative).[19]

---

18 In particular, we divide the sample into five equal intervals. For j from 1 to 5, we remove the jth interval from the sample and use the remaining sample returns to compute the regularized parametric portfolio for several values of δ. We then evaluate the return of the resultant portfolios on the jth-interval. After completing this process for each of the five intervals, we have out-of-sample portfolio returns for the entire sample for each value of δ. Finally, we compute the mean-variance utility of these out-of-sample returns and select the value of δ that optimizes mean-variance utility.

19 We have repeated the tests using the stationary bootstrap in Politis and Romano (1994), which takes serial dependence into account, and we have found that the results are robust.

We now explain how our significance test relates to several regularization approaches used in the literature to identify the characteristics that are jointly relevant. For instance, Freyberger, Neuhierl, and Weber (2018) and Messmer and Audrino (2017) use "adaptive lasso," and Kozak, Nagel, and Santosh (forthcoming) use "elastic net." These regularization methods are similar to the first (screen) stage of our approach because they employ cross-validation to maximize *out-of-sample fit*. However, because of our focus on *significance*, unlike these papers, our analysis includes a *second* (clean) stage that performs a bootstrap significance test on the parametric portfolios of those characteristics that survived the first (screen) stage.[20]

Another alternative is to use a *sequential* bootstrap method to test the significance of adding one more characteristic to an existing parametric portfolio. This approach would be similar to the methodology proposed in Harvey and Liu (2018) in the context of sequential factor selection. However, a sequential significance test would not capture the risk- and trading-diversification benefits from adding *several* characteristics simultaneously. This is crucial because both risk and transaction costs depend critically on how characteristics are combined.[21]

We conclude this section with some comments on the robustness of our significance test. First, the screen-and-clean significance test is unlikely to suffer from the type of overfitting bias documented in Novy-Marx (2016) and Rytchkov and Zhong (2017) because it tests the *marginal* significance of each characteristic when considered jointly with the others, and thus follows exactly the recommendation in Novy-Marx (2016, p. 3) that "*the marginal contribution of each individual signal should be evaluated individually*" (see Section IA.6 of the Online Appendix for a more detailed discussion). Second, our main finding that transaction costs increase the number of significant characteristics is robust to using alternative significance tests and data samples, as we show in Sections IA.3, IA.4, and IA.8 of the Online Appendix. This is because our main finding is obtained by *comparing* the number of significant characteristics for the cases with and without transaction costs, and therefore, any differences due to the method or sample are likely to wash out. Third, because the characteristics that we consider were discovered in the literature for their ability to explain the cross-section of *expected* stock returns rather than something related to transaction costs and trading diversification, it is unlikely that our findings

20 The adaptive lasso and elastic net could be used instead of lasso for the first (screen) stage of our significance method. Indeed, in Section IA.3 of the Online Appendix, we repeat the screen-and-clean significance test, but employ elastic net for the first (screen) stage, and our results are robust.

21 Note that lasso can be interpreted also as a sequential procedure, because as one increases the lasso threshold δ, the regularized parametric portfolios assign a nonzero weight to a larger number of characteristics. However, the lasso does not suffer from the limitation of a purely sequential procedure because it allows for characteristics to drop out of the active set as the lasso threshold increases (see Efron and Hastie 2016, section 16.4.) More importantly, we employ the lasso only in the first (screen) stage of the significance test. The second (clean) stage is carried out on the *unregularized* parametric portfolios and tests the *joint* significance of all the characteristics that survived the screen stage.

about the impact of transaction costs on the dimension of the cross-section are driven by data mining.

3. Trading Diversification

We now characterize analytically and empirically the magnitude of the trading-diversification benefits obtained by combining characteristics. We do this by comparing the average trading volume (turnover) required to exploit characteristics in combination with that required to exploit them in isolation. Note that the reduction in turnover that we characterize in this section will result in a reduction in transaction costs *regardless* of the particular manner in which transaction costs are modeled. Indeed, the analysis in Section 5 and Sections IA.1 and IA.2 of the Online Appendix shows that, in the presence of either proportional or quadratic transaction costs, the benefits of trading diversification lead to a substantial reduction in the transaction costs associated with the optimal investor's portfolio.

3.1 Analytical results

To simplify the exposition, in this section we focus on the case where the investor holds an *equally weighted* portfolio of the characteristics, but all results can be extended to the case of a generic portfolio of characteristics.

Proposition 3 characterizes the reduction in turnover obtained by combining characteristics. The intuition underlying this proposition is that, just as we get diversification of risk when we combine stocks, we get trading diversification when we combine characteristics. To see this, note that rebalancing the long-short portfolio associated with *each* characteristic requires trading in the *same* set of underlying stocks. Thus, exploiting multiple characteristics allows one to cancel out some of the trades in the underlying stocks required to rebalance the characteristic long-short portfolios. For instance, if to rebalance a characteristic long-short portfolio we need to buy a particular stock, whereas to rebalance another characteristic we need to sell the same stock, then the net amount of trading required to exploit these two characteristics in combination will be smaller than that required to exploit them in isolation.

Proposition 3. Assume that the trades in the ith stock required to rebalance $K > 1$ different characteristics, that is, the quantities

$$\text{trade}_{i,k} = (X_{t+1})_{i,k} - (X_t)_{i,k}(1 + r_{i,t+1}), \qquad k = 1, 2, \ldots, K \tag{17}$$

are jointly distributed as a multivariate Normal distribution with zero mean and positive-definite covariance matrix Ω. Then:

1. The ratio of the average trading volume (turnover) in the ith stock required to rebalance an equally weighted portfolio of the K

characteristics to that required to rebalance the K characteristics in isolation is

$$\frac{\text{turnover}(\text{trade}_i^{ew})}{\text{turnover}(\text{trade}_i^{iso})} = \frac{\sqrt{e^\top \Omega e}}{\sum_{k=1}^{K} \sqrt{\Omega_{kk}}} < 1,$$

where $e \in \mathbb{R}^K$ is the vector of ones, Ω_{kk} is the variance of $\text{trade}_{i,k}$,

$$\text{turnover}(\text{trade}_i^{ew}) = E\left[\frac{1}{K}\left|\sum_{k=1}^{K} \text{trade}_{i,k}\right|\right], \quad \text{and}$$

$$\text{turnover}(\text{trade}_i^{iso}) = E\left[\frac{1}{K}\sum_{k=1}^{K} \left|\text{trade}_{i,k}\right|\right].$$

2. If, in addition, the covariance matrix Ω is symmetric with respect to all K characteristics, that is, if the variances and correlations between the trades in the ith stock required to rebalance the K different characteristics are all equal to σ^2 and ρ, respectively, then[22]

$$\frac{\text{turnover}(\text{trade}_i^{ew})}{\text{turnover}(\text{trade}_i^{iso})} = \sqrt{\frac{1+\rho(K-1)}{K}} < 1. \tag{18}$$

3. If, in addition, the correlations between the trades in the ith stock required to rebalance the K different characteristics are all zero ($\rho = 0$), then

$$\frac{\text{turnover}(\text{trade}_i^{ew})}{\text{turnover}(\text{trade}_i^{iso})} = \frac{1}{\sqrt{K}} < 1.$$

Part 1 of Proposition 3 shows that, provided the covariance matrix of the rebalancing trades is positive definite (and thus, the rebalancing trades between some of the characteristics are not perfectly correlated), combining characteristics will result in trading diversification and a reduction in turnover. Also, Part 2 of Proposition 3 shows that trading diversification increases with the number of characteristics and decreases with the correlation between the rebalancing trades of different characteristics.

3.2 Empirical results

We now evaluate empirically the benefits from trading diversification. Figure 1 compares the monthly turnover required to exploit the 51 characteristics in isolation with that required to exploit them in an equally weighted combination.[23] The figure shows that the trading-diversification benefits of

22 Note that in (18) the term $1+\rho(K-1)$ is strictly positive because of the assumption that Ω is positive definite.

23 For this section only, we have adjusted the sign of every characteristic so that its associated long-short portfolio produces positive average returns. The marginal contributions to turnover are computed using Equation (12) for the case in which the transaction-cost matrix Λ_t is replaced by the identity matrix and for an equally weighted portfolio of the 51 characteristics without the benchmark; that is, $w_t = X_t e/(51 N_t)$, where e is the vector of ones.

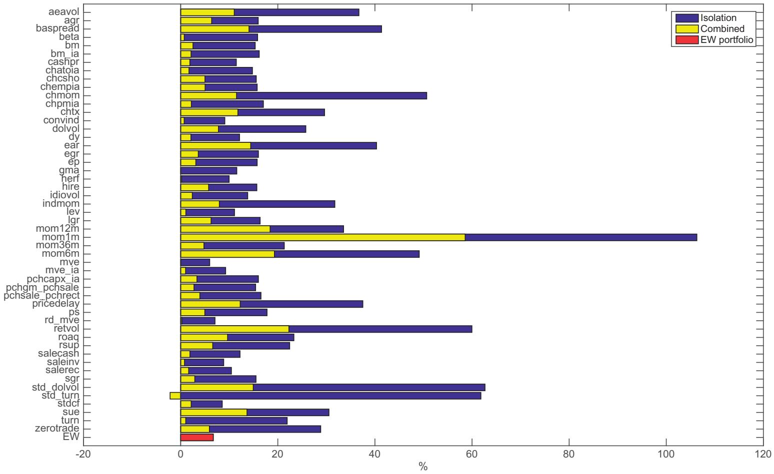

Figure 1
Marginal contribution to turnover of characteristics traded in isolation and in equally weighted portfolio
This figure compares the average trading volume (turnover) required to exploit the 51 characteristics in isolation with that required to exploit them in an equally weighted portfolio. The horizontal axis gives the turnover in percentage, and the vertical axis gives the acronyms of the characteristics and the equally weighted portfolio (EW). The blue bars represent the turnover required to exploit each of the characteristics in isolation (Isolation), the yellow bars represent the marginal contribution to turnover of each characteristic in an equally weighted portfolio (Combined), and the red bar represents the turnover of the equally weighted portfolio of the 51 characteristics (EW portfolio).

combining characteristics are large empirically. While the average monthly turnover required to exploit the 51 characteristics in isolation is 24.09%, the turnover required to exploit an equally weighted combination of them is only 6.71%; that is, trading diversification delivers a 72.15% reduction in turnover.[24]

Note that this 72.15% reduction in turnover is similar in magnitude to that predicted by Part 3 of Proposition 3 for the symmetric case with *zero correlation* between rebalancing trades across characteristics: $1-1/\sqrt{K}=1-1/\sqrt{51}\approx 86\%$. Indeed, Figure 2 gives a heatmap of the correlations between the rebalancing trades for the 51 characteristics for a particular stock and shows that many of the correlations are close to zero.[25] Moreover, we find that the average correlation between rebalancing trades across the 51 characteristics and the entire universe of stocks is 5.47%, not very different from zero. This explains why the empirical benefits from trading diversification are so large and in line with those predicted by Part 3 of Proposition 3.

24 In fact, Figure 1 shows that the turnover required to exploit each characteristic in isolation (blue bars) is much larger than the marginal contribution to turnover of each characteristic in an equally weighted combination (yellow bars). Most strikingly, for the volatility of share turnover (*std_turn*) characteristic, the marginal contribution to turnover in an equally weighted combination is *negative*, implying that including this characteristic results in an absolute reduction in turnover of the portfolio.

25 We have produced heatmaps for several stocks as well as the heatmap for the average correlations across stocks, and the insights are similar.

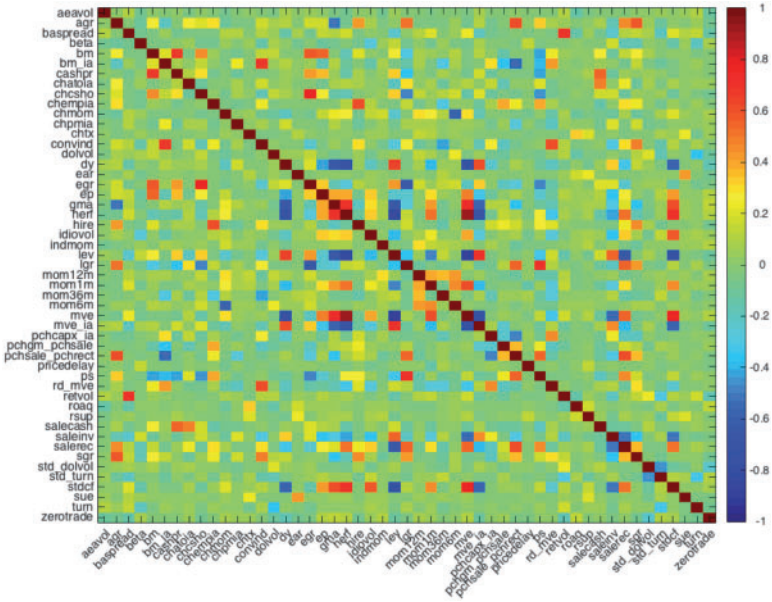

Figure 2
Correlations between rebalancing trades of different characteristics
This figure depicts a heatmap of the correlations between the rebalancing trades for the 51 characteristics for a particular stock.

In this section, we have shown analytically and empirically that combining characteristics in an equally weighted portfolio results in a substantial reduction in *turnover* compared to trading them in isolation. In the next two sections, we show that combining characteristics in the parametric portfolio that maximizes an investor's expected utility also results in a substantial reduction in *transaction costs*.

4. How Many Characteristics Matter without Costs?

This section studies how many characteristics are jointly significant *in the absence of transaction costs* and Section 5 studies the effect of transaction costs.

We use the screen-and-clean method, described in Section 2.5, to test the significance of the characteristics on the sample containing the 319 monthly observations from May 1988 to December 2014.[26] When computing the parametric portfolios, we use the value-weighted portfolio as the benchmark

[26] Although our data set covers the period from January 1980 to December 2014, we drop the first 100 months so that the significance test is run on the exact same sample as the out-of-sample analysis in Section 6. However, Section IA.8.3 in the Online Appendix shows that our findings are robust to considering the full sample from 1980.

Table 2
Significance and marginal contributions without transaction costs

		Marginal contributions to			
Characteristic	Param.	Variance	Cov (char.)	Cov (bench.)	Mean
sue	20.12***	*0.00341*	−0.00068	−0.00019	−0.00254
retvol	−10.85***	*−0.03529*	0.02914	0.00292	0.00323
agr	−10.37**	*−0.00397*	0.00050	0.00057	0.00290
mom1m	−3.10**	*−0.00509*	0.00454	*−0.00109*	0.00164
gma	5.97**	*0.00252*	−0.00255	*0.00069*	−0.00066
beta	2.36*	*0.00971*	−0.01381	*0.00419*	−0.00008
bm_ia	6.49	*0.00337*	−0.00328	*0.00072*	−0.00081
chcsho	−5.89	*−0.00210*	*−0.00111*	0.00092	0.00228
rd_mve	6.01	*0.00215*	−0.00096	*0.00045*	−0.00164
std_turn	8.53	*0.01442*	−0.01576	*0.00214*	−0.00080
bm	3.10	*0.00264*	*0.00023*	−0.00082	−0.00205
mve	−4.02	*−0.00136*	0.00148	*−0.00034*	0.00022
mom12m	−4.42	*−0.00784*	0.01125	*−0.00066*	*−0.00275*

This table reports the significance and marginal contributions for the parametric portfolios without transaction costs, for risk-aversion parameter $\gamma = 5$. We run a screen-and-clean significance test. For the first (screen) stage, we calibrate the regularized parametric portfolios with fivefold cross-validation and find that the lasso threshold that maximizes investor's utility is $\delta = 25$. For the second (clean) stage, we run the bootstrap experiment for the parametric portfolios using those characteristics with nonzero θ's from the first stage. Characteristic p-values are computed using the percentile method discussed in Section 2.5. To compute the optimal parametric portfolio and marginal contributions, we include all characteristics with nonzero θ's for the screen stage plus the three characteristics considered in Brandt, Santa-Clara, and Valkanov (2009): size, book to market, and momentum. For each characteristic, the first column gives the acronym, the second the optimal value of the parameter and the significance asterisks, and the next four columns give the marginal contribution of the characteristic to: (i) the characteristic own-variance, (ii) the covariance of the characteristic with the other characteristics in the portfolio, (iii) the covariance of the characteristic with the benchmark portfolio, and (iv) the characteristic mean. Contributions that drive the characteristic to be nonzero are in blue sans serif font, and contributions that drive the characteristic toward zero are in *red italic* font (cf. Footnote 29). $^{*}p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$.

and assume a risk-aversion parameter $\gamma = 5$.[27] The first (screen) stage finds that 10 characteristics survive the screening. We then run the second (clean) stage on the *unregularized* parametric portfolios for these 10 characteristics to determine how many are significant.

Table 2 reports the significance of each characteristic that survived the first (screen) stage. We observe from the second column of Table 2 that, in the absence of transaction costs, 6 characteristics are significant. Five are significant at the 5% confidence level: unexpected quarterly earnings (*sue*), return volatility (*retvol*), asset growth (*agr*), 1-month momentum (*mom1m*), and gross profitability (*gma*); and one characteristic, beta, is significant at the 10% level. Our results show that, in the absence of transaction costs, a small number of characteristics is sufficient to explain the cross-section of stock returns. This is consistent with several papers in the literature. For instance, Hou, Xue, and Zhang (2014) and Fama and French (2016) show that four and five variables, respectively, are enough to explain the cross-section. Likewise, Green, Hand, and Zhang (2017) consider 94 characteristics and finds that 12 are jointly significant, Freyberger, Neuhierl, and Weber (2018) consider 62 and

[27] Section IA.10 of the Online Appendix considers other values of risk aversion: $\gamma = 2$ and 10.

find that 13 provide independent information, and Kelly, Pruitt, and Su (2018) consider 36 characteristics and find that 8 are significant.[28]

For each characteristic, the last four columns of Table 2 give the marginal contribution of the characteristic to (1) the characteristic own-variance, (2) the covariance of the characteristic with the other characteristics in the portfolio, (3) the covariance of the characteristic with the benchmark portfolio, and (4) the characteristic mean. Marginal contributions that drive the characteristic to be nonzero are in blue sans serif font, and marginal contributions that drive the characteristic toward zero are in *red italic* font.[29]

The marginal contributions reported in Table 2 show that the five characteristics significant at the 5% level matter because they help to reduce the risk of the portfolio of characteristics *and* increase its mean return.[30] In contrast, the beta characteristic is significant at the 10% level *only* because of its ability to reduce the risk of the portfolio of characteristics. To see this, note that Table 2 shows that, consistent with the findings in the existing literature (see Black 1993), the marginal contribution of *beta* to the portfolio's mean return is very small. However, the beta return has a large negative covariance with the returns of the other characteristics (marginal contribution $= -0.01381$), and this is what makes it relevant from a portfolio perspective. This is illustrated in Figure 3, which depicts the marginal contributions of the 6 significant characteristics, and shows that *beta* has a large negative marginal contribution to the covariance with the other characteristics that helps to reduce the overall portfolio risk.

Table 2 also explains why size, book to market, and momentum are *not* significant. For instance, 12-month momentum (*mom12m*) and book to market (*bm*) are not significant, even though their expected returns are large, because their returns have a very large positive covariance with the returns of the other characteristics in the portfolio. In contrast, market capitalization (*mve*) has only a small mean return, consistent with findings in the literature (see Asness et al. 2018), and hence, although *mve* helps to diversify the characteristic portfolio, the risk reduction is not sufficient to make it significant.

As discussed above, the contribution of characteristics to portfolio risk plays an important role, and thus, the correlations between the characteristic returns

---

28 In contrast to these papers, Kozak, Nagel, and Santosh (forthcoming) find that, in the absence of transaction costs, a small number of principal components predict the cross-section better than a small number of characteristics. The explanation for these contrasting results is that while Kozak, Nagel, and Santosh (forthcoming) focus on *out-of-sample fit*, most of the aforementioned papers focus on *significance*, with the exception of Freyberger, Neuhierl, and Weber (2018). This suggest that, while a large number of characteristics may help to *predict* the cross-section, not all may be statistically significant.

29 Note that for characteristics with a positive parametric-portfolio weight, negative (positive) marginal contributions help to decrease (increase) the objective function in the minimization problem (9) and thus increase (decrease) the investor's mean-variance utility. Therefore, for characteristics with positive parametric-portfolio weights, negative (positive) marginal contributions are in blue sans serif font (red *italic* font). The opposite color and font convention applies to characteristics with negative parametric-portfolio weights.

30 For instance, return volatility has large positive mean return (marginal contribution $= 0.00323$) and negative return covariance with the other characteristics (marginal contribution $= 0.02914$).

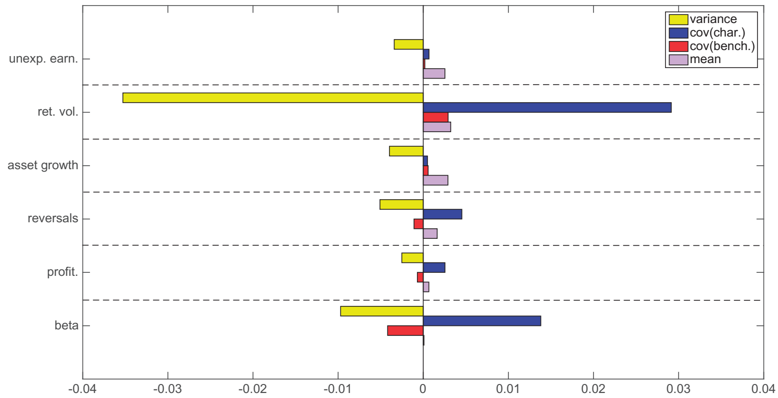

Figure 3
Marginal contributions of significant characteristics without transaction costs
This figure shows the marginal contributions to the investor's utility of the 6 significant characteristics in the absence of transaction costs. The vertical axis gives the labels of the significant characteristics: unexpected quarterly earnings (*unexp. earn.*), return volatility (*ret. vol.*), asset growth, 1-month momentum (*reversals*), gross profitability (*profit.*), and beta. The horizontal axis gives the marginal contributions of each characteristic to (1) the characteristic own-variance (yellow bars, *variance*), (2) the covariance of the characteristic with the other characteristics in the portfolio (blue bars, *cov(char.)*), (3) the covariance of the characteristic with the benchmark portfolio (red bars, *cov(bench.)*), and (4) the characteristic mean (light-purple bars, *mean*). Contributions that drive the characteristic to be nonzero are represented by positive bars, and contributions that drive the characteristic toward zero are represented by negative bars (cf. Footnote 29).

Table 3
Correlations of significant characteristics

Characteristics	sue	retvol	agr	mom1m	gma	beta	bm	mve	mom12m
Unexpected quarterly earnings (sue)	1.00	−0.43	−0.08	0.18	−0.18	−0.36	−0.05	0.41	0.45
Return volatility (retvol)	−0.43	1.00	0.22	−0.18	0.45	0.93	−0.46	−0.63	−0.17
Asset growth (agr)	−0.08	0.22	1.00	−0.33	0.56	0.33	−0.64	0.03	−0.17
1-month momentum (mom1m)	0.18	−0.18	−0.33	1.00	−0.23	−0.26	0.14	0.19	0.28
Gross profitability (gma)	−0.18	0.45	0.56	−0.23	1.00	0.54	−0.62	−0.24	−0.06
Beta (beta)	−0.36	0.93	0.33	−0.26	0.54	1.00	−0.54	−0.52	−0.21
Book to market (bm)	−0.05	−0.46	−0.64	0.14	−0.62	−0.54	1.00	−0.05	−0.08
Size (mve)	0.41	−0.63	0.03	0.19	−0.24	−0.52	−0.05	1.00	0.20
12-month momentum (mom12m)	0.45	−0.17	−0.17	0.28	−0.06	−0.21	−0.08	0.20	1.00

This table reports the correlation matrix for the returns of the 6 characteristics that are most significant in the absence of transaction costs and the returns of the three characteristics considered in Brandt, Santa-Clara, and Valkanov (2009): book to market (*bm*), size (*mve*), and momentum (*mom*12*m*).

matter. Table 3 reports the correlation matrix for the returns of the six significant characteristics and the three characteristics considered in Brandt, Santa-Clara, and Valkanov (2009): size, book to market, and momentum.

We first observe from Table 3 that the returns of the size, book to market, and momentum characteristics are not highly correlated, with their correlation coefficients being smaller than 20%. On the other hand, the returns of the 6 significant characteristics we identify are more highly correlated. To understand why these characteristics with highly correlated returns are jointly significant for portfolio choice, consider the case of return volatility and beta. The returns of these two characteristics are highly positively correlated (93%), but the mean return of beta is very small. As a consequence, the investor optimally goes long the beta characteristic to hedge the risk of her short position in the return-volatility characteristic, while preserving most of its mean return. The benefit of this strategy is illustrated in panel (a) of Figure 4, which shows the cumulative returns of a blended strategy that assigns a −50% weight to return volatility and a +50% weight to beta. This blended strategy has large cumulative returns and very low volatility.[31]

Asness, Moskowitz, and Pedersen (2013) find that the returns of value and momentum are negatively correlated and a blended strategy of these two characteristics performs well. We compare the return volatility and beta blended strategy with the value and momentum blended strategy. Panel (b) in Figure 4 shows the cumulative returns of these two blended strategies, where we have scaled them so that they have the same volatility. We find that the return-volatility and beta blend attains a cumulative return of 110%, whereas the value and momentum blend attains a cumulative return of around 80%.

Summarizing, we find that, in the absence of transaction costs, only 6 characteristics are significant and that risk diversification plays an important role in determining which characteristics are significant. We now study the role of trading diversification.

5. What is the Effect of Transaction Costs?

In this section, we examine how transaction costs influence the optimal portfolio of a utility-maximizing investor, and hence, the dimension of the cross-section. As explained in Section 2.2, we consider an investor who faces proportional transaction costs that decrease with firm size and over time, as in Brandt, Santa-Clara, and Valkanov (2009) and Hand and Green (2011). Sections IA.1 and IA.2 of the Online Appendix, respectively, show that our findings are robust to estimating proportional transaction costs from daily price data and to considering quadratic transaction costs.

Intuitively, one may expect that in the presence of transaction costs *fewer* characteristics would be significant because transaction costs can only erode the benefits from exploiting characteristics. Indeed, we find that this is the case if one were to consider each characteristic *individually*: Section IA.13 in the

[31] Our finding that, despite the high correlation between the return volatility and beta characteristics, the return-volatility characteristic commands a much higher average return than beta is consistent with results in the existing literature. As explained in Bali, Engle, and Murray (2016), return volatility and idiosyncratic volatility are very similar in the cross-section. Therefore, the high average return of the return-volatility characteristic can be traced back to the high average return of the idiosyncratic-volatility characteristic, which is documented in Ang et al. (2006). Moreover, Bali, Engle, and Murray (2016, table 15.7) show that the idiosyncratic risk characteristic commands a high average return mostly when computed from daily data over short horizons, which is how return volatility is computed in our analysis. Beta, on the other hand, is computed from weekly returns over the past 3 years and thus delivers much lower average returns (see also Liu, Stambaugh, and Yuan 2018).

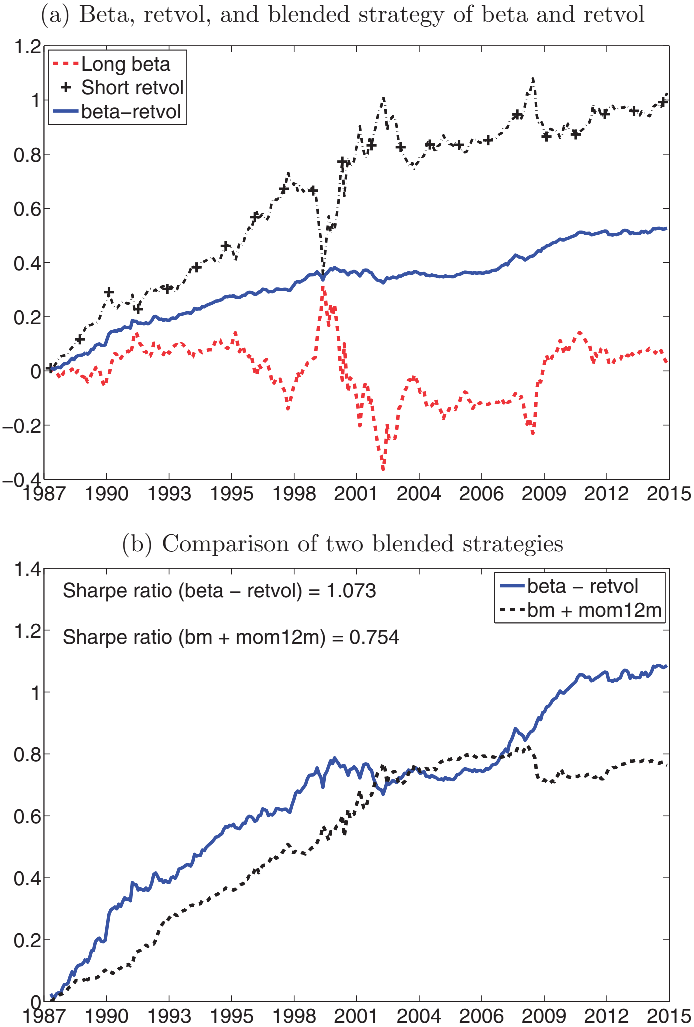

Figure 4
Cumulative returns for beta and return-volatility blended strategy
This figure shows the cumulative returns of a blended strategy that goes long beta and short return volatility. Panel (a) depicts the cumulative returns of going long beta (Long beta), of going short return volatility (Short retvol), and of a blended strategy formed by assigning 50% weight to beta and −50% to return volatility. Panel (b) compares the cumulative returns of the blended strategy that is long beta and short return volatility with those of a blended strategy that assigns 50% to book to market (*bm*) and 50% to 12-month momentum (*mom*12*m*). For comparison purposes, in panel (b) we normalize both strategies so that they have the same volatility.

Online Appendix shows that 21 characteristics are individually significant in the absence of transaction costs, but only 14 in the presence transaction costs. However, when considered *jointly*, we find that the number of characteristics

Table 4
Significance and marginal contributions with transaction costs

		Marginal contributions to					Isolation
Characteristic	Param.	Variance	Cov (char.)	Cov (bench.)	Mean	Tran. cost	Tran. costs
rd_mve	11.85***	*0.00425*	−0.00333	*0.00045*	−0.00164	*0.00027*	*0.00055*
agr	−7.27***	*−0.00278*	*−0.00012*	0.00057	0.00290	*−0.00057*	*0.00125*
sue	3.00***	*0.00051*	*0.00077*	−0.00019	−0.00254	*0.00146*	*0.00240*
turn	−3.41***	*−0.00806*	0.00502	0.00279	0.00068	*−0.00043*	*0.00177*
retvol	−1.92***	*−0.00623*	0.00148	0.00292	0.00323	*−0.00139*	*0.00468*
std_turn	1.28***	*0.00217*	−0.00433	*0.00214*	−0.00080	*0.00082*	*0.00493*
zerotrade	−1.53***	*−0.00129*	0.00284	*−0.00205*	0.00124	*−0.00075*	*0.00235*
chatoia	4.51**	*0.00029*	*0.00008*	−0.00005	−0.00077	*0.00046*	*0.00116*
chtx	1.36**	*0.00026*	−0.00022	*0.00015*	−0.00123	*0.00104*	*0.00232*
beta	3.39**	*0.01398*	−0.01829	*0.00419*	−0.00008	*0.00021*	*0.00126*
ps	4.94**	*0.00156*	−0.00027	−0.00068	−0.00127	*0.00066*	*0.00140*
gma	6.60**	*0.00278*	−0.00298	*0.00069*	−0.00066	*0.00016*	*0.00090*
herf	−5.78**	*−0.00144*	0.00061	0.00041	0.00061	*−0.00019*	*0.00077*
mom1m	−0.62**	*−0.00102*	0.00258	*−0.00109*	0.00164	*−0.00211*	*0.00857*
bm_ia	2.85**	*0.00148*	−0.00168	*0.00072*	−0.00081	*0.00029*	*0.00128*
stdcf	−5.05*	*−0.00259*	0.00101	0.00068	0.00114	*−0.00024*	*0.00067*
pchgm_pchsale	3.46*	*0.00034*	*0.00006*	−0.00003	−0.00079	*0.00042*	*0.00122*
chcsho	−3.11*	*−0.00111*	*−0.00166*	0.00092	0.00228	*−0.00044*	*0.00123*
bm	1.74*	*0.00148*	*0.00122*	−0.00082	−0.00205	*0.00017*	*0.00121*
chmom	−0.67	*−0.00065*	0.00166	*−0.00073*	0.00044	*−0.00072*	*0.00404*
baspread	0.55	*0.00240*	−0.00795	*0.00329*	*0.00279*	−0.00053	*0.00322*
ep	1.27	*0.00206*	*0.00045*	−0.00166	−0.00104	*0.00018*	*0.00125*
idiovol	−1.80	*−0.00680*	0.00194	0.00308	0.00187	*−0.00008*	*0.00109*
roaq	−0.12	*−0.00014*	0.00292	*−0.00114*	*−0.00215*	0.00051	*0.00186*
mve	−2.28	*−0.00077*	0.00092	*−0.00034*	0.00022	*−0.00003*	*0.00045*
mom12m	−0.61	*−0.00109*	0.00418	*−0.00066*	*−0.00275*	0.00031	*0.00265*

This table reports the significance and marginal contributions for the parametric portfolios in the presence of transaction costs, for risk-aversion parameter $\gamma = 5$. We run a screen-and-clean significance test. For the first (screen) stage, we calibrate the regularized parametric portfolios with fivefold cross-validation and find that the lasso threshold that maximizes investor's utility is $\delta = 25$. For the second (clean) stage, we run the bootstrap experiment for the parametric portfolios using those characteristics with nonzero θ's from the first stage. Characteristic p-values are computed using the percentile method discussed in Section 2.5. To compute the optimal parametric portfolio and marginal contributions, we include all characteristics with nonzero θ's for the screen stage plus the three characteristics considered in Brandt, Santa-Clara, and Valkanov (2009): size, book to market, and momentum. For each characteristic, the first column gives the acronym, the second the optimal value of the parameter and the significance asterisks, and the next five columns give the marginal contribution of the characteristic to (1) the characteristic own-variance, (2) the covariance of the characteristic with the other characteristics in the portfolio, (3) the covariance of the characteristic with the benchmark portfolio, (4) the characteristic mean, and (5) the transaction cost. The last column reports the marginal contribution of the characteristic to transaction costs when it is traded in isolation. Contributions that drive the characteristic to be nonzero are in blue sans serif font, and contributions that drive the characteristic toward zero are in *red italic* font (cf. Footnote 29). $^{*}p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$.

that are jointly significant at the 5% level *increases* from five in the absence of transaction costs to 15 in the presence of proportional transaction costs.[32]

Table 4 explains this result. The table gives the significance and marginal contributions of the characteristics for the parametric portfolios in the presence of transaction costs. Of particular interest are the last two columns of the table, which give (i) the marginal contribution of each characteristic to

[32] For the case with proportional transaction costs estimated from daily price data reported in Section IA.1, the number of characteristics that are jointly significant increases from five to 14, and for the case of quadratic transaction costs reported in Section IA.2, the number of characteristics that are jointly significant increases from 5 to 19.

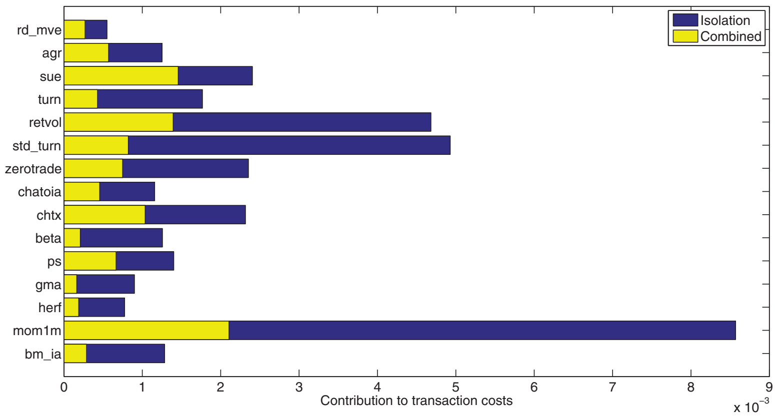

Figure 5
Marginal contribution to transaction costs of characteristics in isolation and in optimal parametric portfolio
This figure shows the marginal contribution to transaction costs when characteristics are traded in isolation and in an optimal parametric portfolio. We plot the marginal contribution to transaction costs of the 15 most significant characteristics in Table 4. The horizontal axis gives the marginal contribution to transaction costs and the vertical axis gives the acronyms of the characteristics. The blue bars represent the marginal contribution of each characteristic to transaction costs when traded in isolation (Isolation), and the yellow bars represent the marginal contribution of each characteristic to transaction costs when combined in the optimal parametric portfolio (Combined).

transaction costs when combined in the optimal parametric portfolio and (ii) the marginal contribution of each characteristic to transaction costs when traded in *isolation*; that is, independently from the benchmark portfolio and the other characteristics. Comparing these two columns reveals that the reason why the number of significant characteristics is *larger* in the presence of transaction costs is that the transaction costs associated with trading the portfolio of characteristics that maximizes the investor's utility are *substantially smaller* than those associated with trading characteristics in isolation. We find that the marginal transaction cost associated with trading the 15 significant characteristics is reduced by around 65% on average when they are combined in the optimal portfolio. This reduction is illustrated in Figure 5, which depicts the marginal contributions to transaction costs of the 15 significant characteristics for the case when the characteristics are combined in the optimal portfolio and in isolation.

A stark example of the trading-diversification benefits from combining characteristics is the short-term reversal characteristic (*mom1m* in the 14th row of Table 4), which has an enormous marginal contribution to transaction costs if traded in isolation (marginal contribution = 0.00857), but a dramatically smaller marginal contribution to transaction costs when traded in the optimal portfolio (marginal contribution = 0.00211). As a result, the short-term reversal characteristic, which is significant in the absence of transaction costs as shown in Table 2, is significant even in the presence of transaction costs when combined in the optimal portfolio of characteristics.[33]

In Section 3, we showed analytically and empirically that combining characteristics in an equally weighted portfolio results in a substantial reduction in turnover compared to trading them in isolation. The results in this section confirm that combining characteristics in the optimal parametric portfolio leads to a substantial reduction in *transaction costs*, and hence, an increase in the investor's utility. The explanation is that combining a larger number of characteristics is advantageous in the presence of transaction costs because the benefits from trading diversification grow with the number of characteristics exploited, as shown in Proposition 2. The main takeaway is that transaction costs *increase* the dimension of the cross-section of stock returns and provide a rationale for nonsparse characteristic-based asset-pricing models.

6. Out-of-Sample Analysis

The previous sections studied the effect of transaction costs on the number of characteristics that are jointly significant *in sample*. In this section, to alleviate data-mining concerns, we study whether an investor can improve *out-of-sample* performance net of transaction costs by exploiting a larger set of characteristics than that considered in prominent asset-pricing models.

6.1 Methodology for out-of-sample evaluation

To evaluate the out-of-sample performance of the various portfolio strategies we use a "rolling-horizon" procedure, similar to that used in DeMiguel, Garlappi, and Uppal (2009b). First, we choose a window over which to perform the estimation. The total number of monthly observations in the data set is $T_{tot}=419$, and we choose an estimation window of $T=100$. Second, using the return data over the estimation window, we compute the various portfolios we study. Third, we repeat this "rolling-window" procedure for the next month, by including the data for the next month and dropping the data for the earliest month. We continue doing this until the end of the data set is reached. At the end of this process, we have generated $T_{tot}-T=319$ portfolio-weight vectors, w_t^j, for $t=T,\ldots,T_{tot}-1$ and for each strategy j. Holding the portfolio w_t^j for one month gives the *out-of-sample* return net of transaction costs at time $t+1$:

$$r_{t+1}^j=(w_t^j)^\top r_{t+1}-\|\Lambda_t(w_t^j-(w_{t-1}^j)^+)\|_1,$$

[33] This result contrasts with those of DeMiguel, Nogales, and Uppal (2014) and Novy-Marx and Velikov (2016), who find that the short-term reversal characteristic is not profitable after transaction costs *when traded in isolation*. DeMiguel, Nogales, and Uppal (2014) find that a short-term reversal (contrarian) strategy is not profitable in the presence of even modest proportional transaction costs of 10 basis points. Novy-Marx and Velikov (2016) find that the short-term reversal strategy does not improve the investment opportunity set of an investor with access to the Fama and French (2016) and Carhart (1997) factors, even when a buy-and-hold transaction-cost-mitigation strategy is employed.

where $(w_{t-1}^j)^+$ is the portfolio for the jth strategy before rebalancing at time t; that is,

$$(w_{t-1}^j)^+ = w_{t-1}^j \circ (e_{t-1} + r_t),$$

and Λ_t, e_{t-1}, and $x \circ y$ are as defined in Section 2.2. Then, for each portfolio we study, we compute the monthly turnover, and the out-of-sample annualized mean, standard deviation, and Sharpe ratio of returns net of transaction costs:

$$\text{turnover}^j = \frac{1}{T_{tot} - T} \sum_{t=T}^{T_{tot}-1} \| w_t^j - (w_{t-1}^j)^+ \|_1,$$

$$\hat{\mu}^j = \frac{12}{T_{tot} - T} \sum_{t=T}^{T_{tot}-1} (w_t^j)^\top r_{t+1},$$

$$\hat{\sigma}^j = \left(\frac{12}{T_{tot} - T} \sum_{t=T}^{T_{tot}-1} \left((w_t^j)^\top r_{t+1} - \hat{\mu}^j \right)^2 \right)^{1/2}, \quad \text{and}$$

$$\widehat{\text{SR}}^j = \frac{\hat{\mu}_j}{\hat{\sigma}_j}.$$

To test whether the out-of-sample performance of the regularized parametric portfolio is statistically significantly better than that of the other portfolios we consider, we use the iid bootstrap method in Ledoit and Wolf (2008), with 10,000 bootstrap samples to construct a one-sided confidence interval for the difference between Sharpe ratios. We use three/two/one asterisks (*) to indicate that the difference is significant at the 0.01/0.05/0.10 level.[34]

6.2 Out-of-sample performance

Table 5 reports the out-of-sample performance of several portfolios in the presence of transaction costs with risk-aversion parameter $\gamma = 5$. Panel A reports the performance for the portfolios that do not use any characteristics, which are the benchmark value-weighted portfolio (VW) and the equally weighted portfolio ($1/N$). Panel B reports the performance of three parametric portfolios: two portfolios that exploit a small number of characteristics and the regularized portfolio that exploits a large set of 51 characteristics.[35] The first parametric

[34] Note that to reduce computation time, we compute the optimal parameter vector θ only in January of each year and use this parameter vector to compute the parametric portfolios for every month of the year. Also, we find that the regularized parametric portfolios that solve problem (15)–(16) result in very large turnovers. Although we find that these portfolios are profitable even after transaction costs (see Section IA.12.3 of the Online Appendix), they may not be implementable for institutional investors facing turnover constraints. Therefore, we report the results for the parametric portfolios after scaling them to control for turnover. Specifically, we scale the optimal parameter vector θ so that the portfolio monthly turnover is around 100%.

[35] For the regularized parametric portfolio, we calibrate the lasso threshold to optimize mean-variance utility by using the fivefold cross-validation methodology explained in Section 2.5, but using only the 100 observations in each estimation window so that there is no look-ahead bias.

Table 5
Out-of-sample performance

Policy	Turnover	Mean	SD	SR
A: Portfolios with no characteristics				
VW	0.050	0.085	0.150	0.567***
1/N	0.134	0.085	0.177	0.482***
B: Portfolios with characteristics				
Size/val./mom.	0.754	0.145	0.215	0.675***
Size/val./inv./prof.	0.963	0.236	0.220	1.072**
Regularized	0.979	0.241	0.178	1.356

This table reports the out-of-sample performance of the different portfolios in the presence of transaction costs, for risk-aversion parameter $\gamma = 5$. Panel A reports the performance for the portfolios that do not use any characteristics, which are the benchmark value-weighted portfolio (VW) and the equally weighted portfolio ($1/N$). Panel B reports the performance of two parametric portfolios that exploit a small number of characteristics, and the regularized parametric portfolio that exploits a large set of 51 characteristics. The first parametric portfolio exploits the size, book-to-market, and momentum characteristics (Size/val./mom.). The second parametric portfolio exploits the size, book-to-market, asset growth, and gross profitability characteristics (Size/val./inv./prof.). The third portfolio is the regularized parametric portfolio that exploits all 51 characteristics (Regularized). The lasso threshold is calibrated using cross-validation over the estimation window. For each portfolio, the first column reports the monthly turnover, and the next three columns report the out-of-sample annualized mean, standard deviation, and Sharpe ratio of returns, net of transaction costs. We test the significance of the difference of the Sharpe ratio of each portfolio with that of the regularized parametric portfolio. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

portfolio exploits the *three* characteristics considered in Brandt, Santa-Clara, and Valkanov (2009): size, book to market, and momentum. The second parametric portfolio exploits four characteristics: size, book to market, asset growth, and gross profitability, which include the investment and profitability characteristics highlighted in Fama and French (2016) and Hou, Xue, and Zhang (2014).

We observe from Table 5 that the gains from exploiting a large set of characteristics are significant: the regularized parametric portfolios achieve an out-of-sample Sharpe ratio that is 100% higher than that of the parametric portfolios based on three characteristics and 25% higher than that of the parametric portfolios based on four characteristics, with the differences being statistically significant. The magnitude of the economic gains is evident also from Figure 6, which depicts the out-of-sample cumulative returns of the value-weighted portfolio and the three parametric portfolios we consider, after scaling them so that they all have the same volatility.

These out-of-sample results confirm that in the presence of transaction costs the cross-section of stock returns is not fully explained by a small number of characteristics.

6.3 Can factor models explain regularized portfolio returns?

The previous section demonstrates that the regularized parametric portfolios that exploit a large set of 51 characteristics significantly outperform the two parametric portfolios that exploit only small sets of characteristics. To check the robustness of this result, we run a time-series regression of the out-of-sample returns of the regularized parametric portfolio onto three sparse factor models from the literature: the Fama and French (1993) and Carhart (1997) four-factor

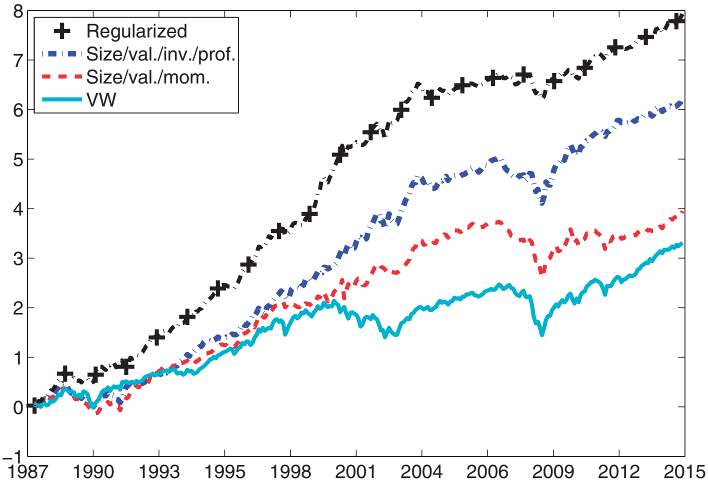

Figure 6
Out-of-sample cumulative returns
This figure shows the out-of-sample cumulative returns of the value-weighted portfolio (VW) and three different parametric portfolios in the presence of transaction costs, for risk-aversion parameter $\gamma=5$. Two of the parametric portfolios exploit a small number of characteristics. The first parametric portfolio exploits the size, book-to-market, and momentum characteristics (Size/val./mom.). The second parametric portfolio exploits the size, book-to-market, asset growth, and gross profitability characteristics (Size/val./inv./prof.). The third parametric portfolio is the regularized parametric portfolio that exploits all 51 characteristics (Regularized). The lasso threshold is calibrated using cross-validation over only the estimation window. For comparison purposes we normalize all portfolio returns so that they have the same volatility.

model (FFC), the Fama and French (2016) five-factor model (FF5), and the Hou, Xue, and Zhang (2014) four-factor model (HXZ). All factors are obtained from Kenneth French's and Lu Zhang's Web sites. Table 6 shows that none of these three sparse factor models fully explain the returns of the regularized parametric portfolios, which achieve an economically and statistically significant abnormal average monthly return of about 1% for each of the three models.[36]

This analysis, however, does not account for the transaction costs that an investor would incur to exploit the characteristics underlying the sparse factor models. To check whether transaction costs affect the abnormal out-of-sample returns delivered by the regularized portfolio, we compute the generalized alpha in Novy-Marx and Velikov (2016). Table 7 reports the intercept, slope, and t-statistic (in brackets) from regressing the out-of-sample regularized portfolio returns net of transaction costs onto the out-of-sample returns net of transaction costs of the parametric portfolio that exploits: (1) the size, book-to-market,

[36] The table also shows that the regularized parametric-portfolio returns load significantly on the market, value (HML), and momentum (UMD) factors for the FFC model, on the market, value, and investment (CMA) factors for the FF5 model, and on the market, investment (I/A), and profitability (ROE) factors for the HXZ model. Finally, the loading of the regularized parametric-portfolio returns on the market factor is close to one because, following Brandt, Santa-Clara, and Valkanov (2009), we use the value-weighted portfolio as the benchmark for the parametric portfolios.

Table 6
Factor loadings of regularized parametric portfolios

FFC	Coefficient	FF5	Coefficient	HXZ	Coefficient
α	0.0115	α	0.0102	α	0.0095
	[4.12]		[3.59]		[2.89]
Market	0.8898	Market	0.9747	Market	0.9147
	[15.29]		[15.35]		[11.90]
SMB	0.0745	SMB	0.1212	SMB	0.2547
	[0.49]		[0.84]		[1.37]
HML	0.3697	HML	−0.2640	I/A	0.7491
	[1.84]		[−1.71]		[2.65]
UMD	0.3249	RMW	0.2554	ROE	0.3316
	[2.46]		[1.31]		[1.69]
		CMA	1.0852		
			[3.64]		

This table reports the intercept, slopes, and t-statistics (in brackets) from regressing the out-of-sample regularized portfolio returns onto three sparse factor models: (1) the Fama and French (1993) and Carhart (1997) four-factor model (FFC) that includes the market, size (SMB), value (HML), and momentum (UMD) factors; (2) the Fama and French (2016) five-factor model (FF5) that includes the market, size, value, profitability (RMW), and investment (CMA) factors; and (3) the Hou, Xue, and Zhang (2014) four-factor model (HXZ) that includes the market, size, investment (I/A), and profitability (ROE) factors. We report t-statistics with Newey-West adjustments of 12 lags. Factors are obtained from Kenneth French's and Lu Zhang's Web sites.

Table 7
Generalized alpha of regularized parametric portfolios

	Size/val./mom.	Size/val./inv./prof.
Generalized α	0.0132	0.0090
	[5.56]	[4.21]
Slope	0.5719	0.5647
	[13.54]	[11.83]

This table reports the intercept, slope, and t-statistic (in brackets) from regressing the out-of-sample regularized portfolio returns onto the out-of-sample returns net of transaction costs of the parametric portfolio that exploits: (1) the size, book-to-market, and momentum characteristics (Size/val./mom.); and (2) the size, book-to-market, investment, and profitability characteristics (Size/val./inv./prof.). We report t-statistics with Newey-West adjustments of 12 lags.

and momentum characteristics (Size/val./mom.); and (2) the size, book-to-market, investment, and profitability characteristics (Size/val./inv./prof.). We observe that the regularized parametric portfolio has an economically and statistically significant generalized alpha of about 1% with respect to these two portfolios.

The results in Tables 6 and 7 confirm that sparse factor models cannot fully explain the out-of-sample performance of the regularized parametric portfolios.

7. Conclusion

A multitude of variables have been proposed to explain the cross-section of stock returns. When addressing the challenge posed in Cochrane (2011), which we highlighted in the introduction, the existing literature either ignores transaction costs or considers one characteristic at a time. We, in contrast, study the impact of *transaction costs* on the number of characteristics that are *jointly* significant for an investor's portfolio. We show analytically that combining characteristics *always* reduces turnover, and thus, transaction costs. The ability to reduce transaction costs by investing in a larger number of characteristics changes the optimal portfolio of a utility-maximizing investor, and hence, increases the dimension of the cross-section. Our empirical work establishes that the magnitude of this effect is substantial: transaction costs roughly double the number of jointly significant characteristics. Our findings have implications for asset-pricing theories based on SDFs because the investor's optimality condition determines not only her optimal portfolio but also the associated SDF. In particular, our work shows that transaction costs provide a rationale for considering a larger number of characteristics than that in prominent asset-pricing models.

Appendix A. Relation to Regression Approaches

In this appendix, we study the relation of our approach based on parametric portfolios to the regression approaches frequently used in the literature. Section A.1 studies the relation to Fama-MacBeth cross-sectional regressions, Section A.2 to time-series regressions, and Section A.3 to the generalized-alpha approach developed in Novy-Marx and Velikov (2016). Proofs for the propositions and corollary that appear in this section are given in Appendix C.

A.1 Relation to Fama-MacBeth Regressions

In this section, we study analytically and empirically the relation between our approach and the Fama-MacBeth regressions in the absence of transaction costs. The Fama-MacBeth procedure can be described as running cross-sectional regressions of stock returns, r_t, onto firm-specific characteristics at each date t:

$$r_t = X_{t-1}\lambda_t + \epsilon_t, \tag{A.1}$$

where $X_{t-1} \in \mathbb{R}^{N_{t-1}\times K}$ is the matrix of firm-specific characteristics at time $t-1$,[37] $\lambda_t \in \mathbb{R}^K$ is the vector of slopes at time t, and $\epsilon_t \in \mathbb{R}^{N_{t-1}}$ is the vector of pricing errors at time t. The Fama-MacBeth approach then tests the significance of the average of the slopes over time, $\overline{\lambda}$.

Most of the existing literature estimates the Fama-MacBeth cross-sectional regressions using ordinary least squares (OLS). Lewellen, Nagel, and Shanken (2010), however, recommend using generalized least squares (GLS) cross-sectional regressions because their goodness-of-fit metric has a clear economic interpretation. In particular, Lewellen, Nagel, and Shanken (2010) extend a result in Kandel and Stambaugh (1995) to show that the GLS R^2 measures the mean-variance efficiency of the model's factor-mimicking portfolios.[38] The following proposition clarifies the relation between our portfolio approach and the Fama-MacBeth OLS and GLS regressions.

[37] For the sake of simplicity and without loss of generality, we assume that X_{t-1} is divided by the number of firms at time $t-1$, as we do for parametric portfolios.

[38] Lewellen, Nagel, and Shanken (2010) study two-pass cross-sectional regressions, rather than Fama-MacBeth regressions (see Cochrane 2009, sections 12.2 and 12.3). For our theoretical analysis, we make the simplifying assumption that the characteristics are time invariant, and in this case the cross-sectional regressions coincide with the Fama-MacBeth regressions. In addition, we use firm-specific characteristic data, rather than factor data, and thus all of our analysis is based on a single pass regression of stock returns onto characteristics.

Proposition A.1. Assume that the standardized firm characteristics are constant through time so that $X_t = X$. Then, the OLS and GLS Fama-MacBeth average slopes are

$$\overline{\lambda}_{OLS} = (X^\top X)^{-1} X^\top \widehat{\mu}_r, \quad \text{and} \tag{A.2}$$

$$\overline{\lambda}_{GLS} = (X^\top \widehat{\Sigma}_r^{-1} X)^{-1} X^\top \widehat{\Sigma}_r^{-1} \widehat{\mu}_r, \tag{A.3}$$

where $\widehat{\mu}_r \in \mathbb{R}^N$ is the sample mean of stock returns and $\widehat{\Sigma}_r \in \mathbb{R}^{N\times N}$ is the sample covariance matrix of stock returns. Assume also that the sample vector of covariances between the benchmark portfolio return and the characteristic portfolio return vector is zero ($\sigma_{bc} = 0$). Then the optimal mean-variance parametric portfolio is

$$\theta^* = \frac{1}{\gamma} (X^\top \widehat{\Sigma}_r X)^{-1} X^\top \widehat{\mu}_r. \tag{A.4}$$

Proposition A.1 shows that the OLS and GLS Fama-MacBeth slopes differ in general from the mean-variance parametric-portfolio weights; that is, testing the significance of Fama-MacBeth slopes is different from testing the significance of the weights a mean-variance investor assigns to each characteristic. Note, in particular, that the OLS and GLS Fama-MacBeth slopes are different in general from the mean-variance parametric-portfolio weights *unless* the sample covariance matrix of asset returns is equal to the identity matrix ($\Sigma_r = I$).

The following corollary provides further insight into the difference between the parametric-portfolio weights and the OLS Fama-MacBeth slopes.

Corollary A.1. Let the assumptions in Proposition A.1 hold, and assume in addition that the columns of the firm-specific characteristic matrix X are orthonormal; that is, $X^\top X = I$. Then, the optimal mean-variance parametric portfolio is

$$\theta^* = \frac{1}{\gamma} \widehat{\Sigma}_c^{-1} \overline{\lambda}_{OLS}, \tag{A.5}$$

where $\widehat{\Sigma}_c$ is the sample covariance matrix of characteristic returns and γ is the risk-aversion parameter.

Corollary A.1 shows that, for the particular case in which the columns of the firm-specific characteristic matrix are orthonormal, there is a componentwise one-to-one relation between mean-variance parametric-portfolio weights and OLS Fama-MacBeth slopes *only if* the sample covariance matrix of characteristic returns, $\widehat{\Sigma}_c$, is diagonal.[39] If, on the other hand, characteristic returns are correlated, then a given characteristic k could have a zero OLS Fama-MacBeth slope ($\overline{\lambda}_k = 0$), and yet have a nonzero parametric-portfolio weight ($\theta_k^* \neq 0$). This is the case, for instance, when the correlation of the kth characteristic return with the returns on the other characteristics can be exploited by the investor to reduce risk, and thus, improve her overall mean-variance utility.

The above theoretical results demonstrate that testing the significance of Fama-MacBeth slopes will, in general, produce results that are different from those of testing the significance of the weights that a mean-variance investor assigns to each characteristic. We now compare empirically the significance results from OLS Fama-MacBeth regressions with those of our approach.[40] Table A1 reports the significance of the Fama-MacBeth slopes for the 6 characteristics we found to be

[39] To see this, note that if $\widehat{\Sigma}_c$ is diagonal, then $\theta_k^* = (\overline{\lambda}_{OLS})_k / (\gamma (\widehat{\Sigma}_c)_{kk})$, where $(\widehat{\Sigma}_c)_{kk}$ is the kth element of the diagonal of $\widehat{\Sigma}_c$, and thus there is a one-to-one correspondence between the kth component of θ^* and the kth component of $\overline{\lambda}_{OLS}$.

[40] We do not run GLS Fama-MacBeth regressions because the sample covariance matrix of stock returns is singular for our case with thousands of stocks and only hundreds of monthly dates.

Table A1
Fama-MacBeth regressions for significant characteristics

Characteristic	Multiple	Individual
Unexpected quarterly earnings (sue)	0.0019	0.0027
	[7.38]	[7.10]
Return volatility (retvol)	−0.0037	−0.0032
	[−4.42]	[−2.22]
Asset growth (agr)	−0.0026	−0.0031
	[−5.39]	[−5.09]
1-month momentum (mom1m)	−0.0033	−0.0017
	[−4.67]	[−2.13]
Gross profitability (gma)	0.0020	0.0007
	[3.80]	[1.34]
Beta (beta)	0.0013	0.0001
	[0.99]	[0.04]
Book to market (bm)	0.0016	0.0021
	[2.11]	[2.17]
Size (mve)	−0.0007	−0.0002
	[−1.76]	[−0.40]
12-month momentum (mom12m)	0.0026	0.0030
	[2.43]	[2.45]

This table reports the slope coefficients from Fama-MacBeth regressions and the corresponding t-statistics (in brackets) with Newey-West adjustments of 12 lags. We report the results for multiple and individual regressions for the 6 most significant characteristics in the absence of transaction costs, and the three characteristics considered in Brandt, Santa-Clara, and Valkanov (2009): book to market (bm), size (mve), and momentum ($mom12m$).

significant in Section 4 plus size, book to market, and momentum. The first column lists the name of the characteristics, the second column reports the multiple regression slopes and Newey-West t-statistics (in brackets),[41] and the third column reports the individual regression slopes and Newey-West t-statistics.

We see from Table A1 that the five characteristics that are significant at the 5% level in Section 4 are also jointly significant for cross-sectional regressions. However, in contrast to the finding in Section 4, beta is not significant in the Fama-MacBeth regressions even at the 10% level. This is because, as shown in Proposition A.1, Fama-MacBeth slopes differ in general from parametric-portfolio weights when the returns on the characteristics are correlated over time and the investor can exploit this to reduce the risk of the mean-variance portfolio. Regarding the book-to-market and momentum characteristics, we see from Table A1 that both book to market (*bm*) and 12-month momentum (*mom12m*) are significant for multiple cross-sectional regressions, whereas they were not significant from a portfolio perspective. Intuitively, these characteristics are significant in multiple cross-sectional regressions because these regressions ignore the large contribution of these characteristics to the risk of the overall portfolio of characteristics, which reduces their appeal from a mean-variance portfolio perspective.

A.2 Relation to Time-Series Regressions

In this section, we study analytically and empirically the relation of our portfolio approach to the time-series regression approach in the absence of transaction costs. The time-series approach may be described as regressing the return of a *new* characteristic long-short portfolio onto the returns of K_c *commonly* accepted characteristic long-short portfolios; that is,

$$r_{n,t} = \alpha_{TS} + \beta_{TS}^{\top} r_{c,t} + \epsilon_t, \tag{A.6}$$

where $r_{n,t} \in \mathbb{R}$ is the return of the *new* characteristic long-short portfolio at time t, $r_{c,t} \in \mathbb{R}^{K_c}$ is the return of the commonly accepted characteristic long-short portfolios at time t, the error term $\epsilon_t \in \mathbb{R}$

41 We compute t-statistics with Newey-West adjustments of 12 lags, as in Green, Hand, and Zhang (2017).

follows a Normal distribution with zero mean and standard deviation σ_ϵ, $\alpha_{TS} \in \mathbb{R}$ is the intercept of the regression, and $\beta_{TS} \in \mathbb{R}^{K_c}$ is the slope vector. If the intercept in this regression is significant, the return on the new characteristic is not fully explained by the return of the commonly accepted characteristics. Gibbons, Ross, and Shanken (1989) show that a significant intercept implies that the new characteristic-based long-short portfolio improves the investment opportunity set of a mean-variance investor who already has access to the returns of the set of commonly accepted characteristics.

As explained above, the time-series regression approach tests the significance of the intercept. In contrast, the following proposition shows that, in the absence of transaction costs, our approach is equivalent to testing the significance of the *slopes* of a particular constrained time-series multiple regression. Britten-Jones (1999) shows that the tangency mean-variance portfolio can be identified by solving a linear regression. We extend this result to the context of *any* parametric portfolio on the mean-variance efficient frontier by introducing a constraint on the mean return of the portfolio.

Proposition A.2. For a given risk-aversion parameter γ, the optimal parameter θ^* for the mean-variance parametric-portfolio problem without transaction costs (5) is equal to the ordinary least square (OLS) estimate of the slope vector in the following time-series regression model:

$$r_{b,t} = \alpha - \beta^\top r_{c,t} + \epsilon_t, \tag{A.7}$$

subject to the constraint that

$$\beta^\top \mu_c = (\theta^*)^\top \mu_c, \tag{A.8}$$

where $r_{b,t} \in \mathbb{R}$ is the return of the benchmark portfolio, $r_{c,t} \in \mathbb{R}^K$ is the return on the characteristics, $\alpha \in \mathbb{R}$ is the intercept, $\beta \in \mathbb{R}^K$ is the slope vector, μ_c is the mean characteristic return vector, and $(\theta^*)^\top \mu_c$ is the average return of the mean-variance parametric portfolio.

The advantage of the parametric-portfolio approach is that by focusing on the slopes, it allows one to test the significance of the different characteristics when they are considered *jointly*. The traditional time-series approach, on the other hand, is designed to test the significance of a single characteristic when it is added to a set of commonly accepted characteristics. This is a limitation of the time-series regression approach because the result of the statistical inference depends on the sequence in which variables are selected. For instance, when regressing the return of each characteristic in our data set onto the returns of the four Fama and French (1993) and Carhart (1997) factors downloaded from Kenneth French's Web site, we find that eight characteristics are significant in the absence of transaction costs, but beta *is not* significant.[42] Beta, however, *is* significant when its returns are regressed onto the four Fama and French (1993) and Carhart (1997) factors *plus* the return of the return-volatility long-short portfolio, because beta helps to hedge the return-volatility characteristic.[43] Accordingly, beta *matters* if one controls for return volatility.[44] Our portfolio approach considers all characteristics simultaneously and finds that return volatility and beta are jointly significant together with four other characteristics. These empirical results highlight the importance of considering all characteristics simultaneously. Other advantages of our portfolio approach are that it allows one to consider transaction costs in a straightforward manner and to identify the marginal contribution of each characteristic to the investor's utility.

42 We run 48 significance tests corresponding to the 51 characteristics except size, value, and momentum and thus, following Harvey, Liu, and Zhu (2015) we apply Bonferroni's adjustment.

43 We again apply Bonferroni's adjustment.

44 This result is analogous to that in Asness et al. (2018), which finds that despite the weak performance of the *size* characteristic when evaluated in isolation, it becomes significant once it is considered in combination with a quality characteristic.

A.3 Relation to Generalized Alpha

In this section, we compare empirically the results from our portfolio approach in the presence of transaction costs with those from using the generalized alpha developed by Novy-Marx and Velikov (2016), who extend the traditional time-series regression framework to take transaction costs into account. Novy-Marx and Velikov (2016) propose computing the returns of the mean-variance portfolio in the presence of transaction costs for the commonly accepted characteristics, MVE_X, and the returns of the mean-variance portfolio in the presence of transaction costs for the commonly accepted characteristics plus the new characteristic, $\text{MVE}_{X,y}$. Then it runs the following regression:

$$\text{MVE}_{X,y}/w_y = \alpha + \beta \text{MVE}_X + \epsilon, \tag{A.9}$$

where w_y is the weight of the mean-variance portfolio on the new characteristic. Novy-Marx and Velikov (2016) show that in the absence of transaction costs, the generalized alpha in (A.9) equals the alpha from the traditional time-series approach. In the presence of transaction costs, this approach tests the significance of adding the new characteristic to a set of commonly accepted characteristics taking transaction costs into account.[45]

As discussed in Section A.2, the main advantage of our portfolio approach with respect to the time-series approach is that it considers all characteristics simultaneously and tests their significance when considered jointly, whereas the time-series regressions are designed to consider one characteristic at a time. To illustrate this, we compute the generalized alpha for each of our characteristics with respect to the four Fama and French (1993) and Carhart (1997) factors downloaded from Kenneth French's Web site. We find that, in the presence of transaction costs, *none* of the characteristic portfolios has a significant generalized alpha with respect to the four factors.[46] However, in the absence of transaction costs, Section A.2 showed that eight characteristics were significant with respect to the four factors. That is, the number of characteristics that are significant with respect to the four factors for the time-series approach *decreases* in the presence of transaction costs when the characteristics are considered in isolation. In contrast, our portfolio approach shows that the number of significant characteristics *increases* in the presence of transaction costs. This is because our approach allows one to consider all characteristics simultaneously and identify the optimal combination of characteristics that results in substantial trading diversification.

Appendix B. Optimal Portfolio and SDF

In this appendix, we establish the relation between the investor's optimal portfolio and its associated SDF in the presence of transaction costs. For exposition purposes, we first derive the relation for the case where the transaction-cost function is differentiable, as is the case for the quadratic transaction costs that we consider in Section IA.2 of the Online Appendix. Then, we show how the relation can be extended to the case where the transaction-cost function is convex, but not differentiable, as is the case for the proportional transaction costs considered in the main body of the manuscript.

B.1 Differentiable Transaction-Cost Function

Consider an investor who holds a parametric portfolio with return $r_{pt} = r_{bt} + \theta^\top r_{ct}$, where r_{bt} is the benchmark portfolio return, r_{ct} is the characteristic return vector, and θ is the parameter vector.

[45] Although the implementation in Novy-Marx and Velikov (2016) considers the transaction cost associated with each characteristic *independently*, here we extend the approach in Novy-Marx and Velikov (2016) to capture trading diversification.

[46] To address the multiple testing problem, we again apply Bonferroni's adjustment because we carry out 48 significance tests corresponding to our 51 characteristics except size, value, and momentum.

The investor selects the parametric portfolio that maximizes her expected utility of returns net of transaction costs:

$$\max_{\theta} \quad E\left[u\left(r_{bt}+\theta^{\top} r_{ct}-\mathrm{TC}(\theta)\right)\right], \tag{B.1}$$

where $\mathrm{TC}(\theta)$ is the transaction cost of holding the parametric portfolio.

Assuming that the transaction-cost function, $\mathrm{TC}(\theta)$, is differentiable, the investor's first-order optimality condition is

$$E\left[u'(r_{bt}+\theta^{\top} r_{ct}-\mathrm{TC}(\theta))(r_{ct}-\mathrm{TC}'(\theta))\right]=0, \tag{B.2}$$

where $u'(\cdot)$ and $\mathrm{TC}'(\cdot)$ are the first derivatives of the utility and transaction cost functions, respectively. From Equation (B.2) it is apparent that the investor's optimal portfolio, θ, will depend on transaction costs.

To derive the SDF associated with the investor's optimal portfolio, one can rewrite Equation (B.2) as

$$E\left[M_t(r_{ct}-\mathrm{TC}'(\theta))\right]=0, \tag{B.3}$$

where the SDF is

$$M_t=u'(r_{bt}+\theta^{\top} r_{ct}-\mathrm{TC}(\theta)). \tag{B.4}$$

Thus, the SDF is the marginal utility of the returns net of transaction costs of the optimal parametric portfolio. This demonstrates that our finding that transaction costs increase the number of characteristics that are significant for the investor's optimal portfolio applies also to the associated SDF.[47] Note that transaction costs affect asset prices through *three* channels. First, Equation (B.3) shows that the SDF prices returns *net* of marginal transaction costs. Second, Equation (B.4) shows that the SDF depends on the investor's optimal portfolio, θ, which itself depends on transaction costs. Third, the SDF is the investor's marginal utility evaluated using returns *net* of transaction costs.

B.2 Convex Transaction-Cost Function

We now establish the relation for the case where the transaction-cost function is convex, but not differentiable, as in the case with proportional transaction costs considered in the main body of the manuscript.

Most popular utility functions are differentiable. This is the case for power utility or for the quadratic utility that underlies mean-variance preferences. Also, it is straightforward to show that the proportional transaction-cost function in Equation (7) of the manuscript is convex and Lipschitz continuous. Therefore, Clarke (1990, theorem 2.6.6) implies that the chain rule can be applied to obtain the subdifferential of the investor's utility. Thus, the investor's first-order optimality condition is

$$0 \in E\left[u'(r_{bt}+\theta^{\top} r_{ct}-\mathrm{TC}(\theta))(r_{ct}-\partial\mathrm{TC}(\theta))\right], \tag{B.5}$$

where $\partial TC(\cdot)$ is the subdifferential of the transaction-cost function. Note that, unlike a differential, the subdifferential is a set-valued function, and thus, the optimality conditions state that zero must be an element in the set defined by the expectation of the subdifferential of the investor's utility function.

47 For the case with quadratic utility, which underlies mean-variance preferences, the SDF is an affine function of the parametric portfolio returns net of transaction costs, and thus, the optimal portfolio and the SDF are particularly closely related.

To derive the SDF, note that the first-order optimality condition can be rewritten as

$$0 \in E[M_t(r_{ct} - \partial \text{TC}(\theta))], \tag{B.6}$$

where the SDF is

$$M_t = u'(r_{bt} + \theta^\top r_{ct} - \text{TC}(\theta)). \tag{B.7}$$

The pricing condition (B.6) states that zero must be an element of the expectation of the SDF multiplied by the return vector minus the subdifferential of the transaction-cost function. For the case with proportional transaction-cost function, the subdifferential is a polyhedral set, and thus, the pricing condition can be rewritten as a system of pricing inequalities. This is consistent with Luttmer (1996) and De Roon, Nijman, and Werker (2001). In particular, Luttmer (1996) states that "In an economy with proportional transaction costs consumer intertemporal marginal rates of substitution have to satisfy a set or Euler *inequalities*."

Finally, it is clear from Equations (B.6) and (B.7) that transaction costs affect the SDF for the case with convex nondifferentiable transaction-cost function through the same three channels discussed for the case with differentiable transaction-cost function.

Appendix C. Proofs for All Propositions

Proof of Proposition 1

Equation (3) shows that the parametric portfolio is a combination of the benchmark portfolio and the K standardized firm-specific characteristics, scaled by the number of firms N_t. Therefore, we can define this combination as $w = [1, \theta] \in \mathbb{R}^{K+1}$ and the vector of benchmark and characteristic returns as $R_t = [r_{b,t}, r_{c,t+1}/N_t]$. Under this specification, the mean-variance parametric-portfolio problem takes the familiar form:

$$\min_{w} \quad \frac{\gamma}{2} w^\top \widehat{\Sigma} w - w^\top \widehat{\mu}, \tag{C.1}$$

$$\text{s.t.} \quad w_1 = 1, \tag{C.2}$$

where $w = [w_1, \theta] \in \mathbb{R}^{K+1}$ and $\widehat{\Sigma}$ and $\widehat{\mu}$ are the sample covariance matrix and mean of $R_t = [r_{b,t}, r_{c,t+1}]$. The result follows by using straightforward algebra to eliminate the decision variable w_1 and the constraint, and then removing terms in the objective function that do not depend on the parameter-vector θ.

Proof of Proposition 2

The marginal contributions of the characteristics are given by the subdifferential of the objective function in (10) with respect to θ. Note that the first four terms in (10) are differentiable with respect to θ and thus their subdifferentials coincide with their gradient. It is straightforward to show that the gradients of these four terms are given by the first four terms in the right-hand side of (11).

The only term that is not differentiable is the transaction cost from trading asset i at time $t+1$. From expression (7), we can define the transaction-cost term for asset i at time $t+1$ as

$$u_{i,t+1} = |\Lambda_{ii,t}\left(w_{i,t+1}(\theta) - w_{i,t}^+(\theta)\right)|, \tag{C.3}$$

where $\Lambda_{ii,t}$ is the associated transaction-cost parameter for asset i at time t. Therefore, it suffices to characterize the subdifferential of expression (C.3). Note that the function inside the absolute value is differentiable with respect to θ. Thus, applying the chain rule for subdifferentials, we have that the subdifferential of $u_{i,t+1}$ with respect to the kth parametric-portfolio weight θ_k is equal to the subdifferential of the absolute value function times the differential of $\Lambda_{ii,t}\left(w_{i,t+1}(\theta) - w_{i,t}^+(\theta)\right)$.

Note that $\Lambda_{ii,t} > 0$ and thus, the subdifferential of the absolute-value function is given by the sign function as precisely defined in (13). Finally, the differential of the term $\Lambda_{ii,t}\left(w_{i,t+1}(\theta) - w^+_{i,t}(\theta)\right)$ is

$$\frac{d[\Lambda_{ii,t}\left(w_{i,t+1}(\theta) - w^+_{i,t}(\theta)\right)]}{d\theta_k} = \Lambda_{ii,t}[(X_{t+1})_{ik} - (X_t)_{ik}(1+r_{i,t+1})].$$

The result follows by adding the subdifferentials of $u_{i,t+1}$ for $i = 1,2,\ldots,N_t$, and then combining the subdifferentials with respect to θ_k for $k = 1,2,\ldots,K$ into a single vector.

Proof of Proposition 3

Part 1. The trade in the ith stock required to rebalance an equally weighted portfolio of K characteristics is:

$$\text{trade}_i^{ew} = \frac{1}{K}\sum_{k=1}^{K}\text{trade}_{i,k} = \frac{1}{K}\sum_{k=1}^{K}[(X_{t+1})_{i,k} - (X_t)_{i,k}(1+r_{i,t+1})]. \tag{C.4}$$

Because $\text{trade}_{i,k}$ for $k = 1,2,\ldots,K$ are jointly distributed as a multivariate Normal distribution with zero mean and and covariance matrix Ω, we have that trade_i^{ew} is distributed as a Normal distribution with zero mean and standard deviation $\sqrt{e^\top \Omega e}/K$.

By definition, the average trading volume (turnover) in the ith stock required to rebalance an equally weighted portfolio of the K characteristics is the average of the absolute value of trade_i^{ew}. Geary (1935) shows that the mean absolute deviation of a Normally distributed random variable is $\sqrt{2/\pi}$ times its standard deviation. Therefore, the average turnover in the ith stock required to rebalance an equally weighted portfolio of K characteristics is

$$\text{turnover}(\text{trade}_i^{ew}) = \sqrt{2/\pi} \times \sqrt{e^\top \Omega e}/K. \tag{C.5}$$

Following a similar argument, the average cost of the trade in the ith stock required to rebalance a quantity $1/K$ of each of the K characteristics in isolation is

$$\text{turnover}(\text{trade}_i^{iso}) = \sqrt{2/\pi} \times \sum_{k=1}^{K}\frac{\sqrt{\Omega_{kk}}}{K}. \tag{C.6}$$

Taking the ratio of (C.5) to (C.6), we get

$$\frac{\text{turnover}(\text{trade}_i^{ew})}{\text{turnover}(\text{trade}_i^{iso})} = \frac{\sqrt{e^\top \Omega e}}{\sum_{k=1}^{K}\sqrt{\Omega_{kk}}}. \tag{C.7}$$

To show that this ratio is strictly smaller than one, we note that the square of the ratio in (C.7) is

$$\frac{e^\top \Omega e}{(\sum_{k=1}^{K}\sqrt{\Omega_{kk}})^2} = \frac{\sum_{k=1}^{K}\Omega_{kk} + \sum_{l=1}^{K}\sum_{m\neq l}\rho_{lm}\sqrt{\Omega_{ll}}\sqrt{\Omega_{mm}}}{\sum_{k=1}^{K}\Omega_{kk} + \sum_{l=1}^{K}\sum_{m\neq l}\sqrt{\Omega_{ll}}\sqrt{\Omega_{mm}}}, \tag{C.8}$$

where ρ_{lm} is the correlation between the rebalancing trade in the ith stock for the lth and mth characteristics. The ratio in (C.8) is smaller than one because $\rho_{lm} < 1$ by the assumption that Ω is positive definite.

Part 2. Because Ω is symmetric with respect to the K characteristics, we have that trade_i^{ew} is distributed as a Normal distribution with zero mean and standard deviation $\sqrt{e^\top \Omega e}/K = \sigma(1 + \rho(K-1))/K$. The result follows using arguments identical to those in the proof of Part 1.

Part 3. Because $\rho = 0$, we have that trade_i^{ew} is distributed as a Normal distribution with zero mean and standard deviation $\sqrt{e^\top \Omega e}/K = \sigma/K$. The result follows using arguments identical to those in the proof of Part 1.

Proof of Proposition A.1
Let us consider the following cross-sectional regression model:

$$r_t = X\lambda_t + \epsilon_t, \tag{C.9}$$

where $r_t \in \mathbb{R}^N$ is the vector of stock returns at time t, $X \in \mathbb{R}^{N\times K}$ is the matrix of standardized firm characteristics, $\lambda_t \in \mathbb{R}^K$ is the vector of slopes at time t, and $\epsilon_t \in \mathbb{R}^N$ is the vector of pricing errors at time t.[48] The OLS and GLS Fama-MacBeth slopes of model (C.9) are

$$\overline{\lambda}_{OLS} = (X^\top X)^{-1} X^\top \widehat{\mu}_r \tag{C.10}$$

$$\overline{\lambda}_{GLS} = (X^\top \widehat{\Sigma}_r^{-1} X)^{-1} X^\top \widehat{\Sigma}_r^{-1} \widehat{\mu}_r, \tag{C.11}$$

where $\widehat{\mu}_r$ is the vector of sample mean returns. It is straightforward to see that $\overline{\lambda}_{OLS}$ and $\overline{\lambda}_{GLS}$ are identical when $\widehat{\Sigma}_r$ is the identity matrix. On the other hand, we know that the solution of a mean-variance parametric portfolio is

$$\theta^* = \frac{1}{\gamma} \widehat{\Sigma}_c^{-1} \widehat{\mu}_c - \widehat{\Sigma}_c^{-1} \widehat{\sigma}_{bc}. \tag{C.12}$$

Now, given the assumption that firm characteristics are constant, we can define the vector of mean characteristic-portfolio returns and the covariance matrix of characteristic-portfolio returns as $\widehat{\mu}_c = X^\top \widehat{\mu}_r$ and $\widehat{\Sigma}_c = X^\top \widehat{\Sigma}_r X$, respectively. Assuming that the covariance between characteristic portfolio returns and the benchmark portfolio is zero, expression (C.12) can be then expressed as

$$\theta^* = \frac{1}{\gamma} (X^\top \widehat{\Sigma}_r X)^{-1} X^\top \widehat{\mu}_r. \tag{C.13}$$

Therefore, one can see that $\overline{\lambda}_{OLS}$, $\overline{\lambda}_{GLS}$, and θ^* will be equivalent when $\widehat{\Sigma}_r$ is the identity matrix of dimension N and the covariance between characteristic portfolio returns and the benchmark portfolio is zero.

Proof of Corollary A.1
The result in Corollary A.1 follows from the assumption that $X^\top X = I$, which implies that $\overline{\lambda}_{OLS} = X^\top \widehat{\mu}_r = \widehat{\mu}_c$. Then, if the covariance between characteristic-portfolio returns and the benchmark portfolio is zero, we can define the solution to the mean-variance parametric portfolio as

$$\theta^* = \frac{1}{\gamma} \widehat{\Sigma}_c^{-1} \overline{\lambda}_{OLS}. \tag{C.14}$$

Proof of Proposition A.2
We can estimate model (A.7) with OLS. The corresponding optimization problem, in matrix form, is

$$\begin{aligned} \min_{\alpha,\beta} \quad & r_b^\top r_b + \alpha^2 T + \beta^\top r_c^\top r_c \beta - 2\alpha r_b^\top e_T + 2 r_b^\top r_c \beta - 2\alpha e_T^\top r_c \beta \\ \text{s.t.} \quad & \widehat{\mu}_c^\top \beta = \mu_0, \end{aligned}$$

where e_T is a T-dimensional vector of ones. Now, given that $\widehat{\Sigma}_c = r_c^\top r_c - \widehat{\mu}_c \widehat{\mu}_c^\top$, $\widehat{\sigma}_{bc} = r_b^\top r_c - \widehat{\mu}_b \widehat{\mu}_c^\top$ and $e_T^\top r_c = T\widehat{\mu}_c$, we can write the above problem as

$$\begin{aligned} \min_{\alpha,\beta} \quad & r_b^\top r_b + \alpha^2 T + \beta^\top \widehat{\Sigma}_c \beta + \beta^\top \widehat{\mu}_c \widehat{\mu}_c^\top \beta - 2\alpha r_b^\top e_T + 2(\widehat{\sigma}_{bc} + \widehat{\mu}_b \widehat{\mu}_c)^\top \beta - 2\alpha T \widehat{\mu}_c^\top \beta \\ \text{s.t.} \quad & \widehat{\mu}_c^\top \beta = \mu_0. \end{aligned}$$

48 Note that we now assume that characteristics X_t and the number of firms N_t are constant through time and therefore we drop the subscript t.

Because $\widehat{\mu}_c^\top \beta$ is constant in the feasible region, we can obtain the OLS slopes of (A.7) as the solution to the following problem:

$$\min_{\beta} \quad \beta^\top \widehat{\Sigma}_c \beta + 2\widehat{\sigma}_{bc}\beta$$

$$\text{s.t.} \quad \widehat{\mu}_c^\top \beta = \mu_0,$$

which is a quadratic mean-variance optimization problem. If we set μ_0 equal to the solution of the mean-variance parametric-portfolio problem times the vector of mean characteristic portfolio returns (that is, $\mu_0 = \theta^{*\top}\widehat{\mu}_c$), the OLS slopes of the time-series model in (A.7) coincide with the solution of the mean-variance parametric-portfolio problem in (5).

References

Ang, A., R. J. Hodrick, Y. Xing, and X. Zhang. 2006. The cross-section of volatility and expected returns. *Journal of Finance* 61:259–99.

Asness, C. S., A. Frazzini, R. Israel, T. J. Moskowitz, and L. H. Pedersen. 2018. Size matters, if you control your junk. *Journal of Financial Economics* 129:479–509.

Asness, C. S., T. J. Moskowitz, and L. H. Pedersen. 2013. Value and momentum everywhere. *Journal of Finance* 68:929–85.

Back, K., N. Kapadia, and B. Ostdiek. 2015. Testing factor models on characteristic and covariance pure plays. Working Paper.

Baker, M., P. Luo, and R. Taliaferro. 2017. Detecting anomalies: The relevance and power of standard asset pricing tests. Working Paper, HBS.

Bali, T. G., R. F. Engle, and S. Murray. 2016. *Empirical asset pricing: The cross section of stock returns*. Hoboken, NJ: John Wiley & Sons.

Barroso, P., and P. Santa-Clara. 2015. Beyond the carry trade: Optimal currency portfolios. *Journal of Financial and Quantitative Analysis* 50:1037–56.

Black, F. 1993. Beta and return. *Journal of Portfolio Management* 20:8–18.

Brandt, M. W., P. Santa-Clara, and R. Valkanov. 2009. Parametric portfolio policies: Exploiting characteristics in the cross-section of equity returns. *Review of Financial Studies* 22:3411–47.

Britten-Jones, M. 1999. The sampling error in estimates of mean-variance efficient portfolio weights. *Journal of Finance* 54:655–71.

Bryzgalova, S. 2015. Spurious factors in linear asset pricing models. Working Paper, Stanford GSB.

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

Chatterjee, A., and S. N. Lahiri. 2011. Bootstrapping lasso estimators. *Journal of the American Statistical Association* 106:608–25.

Chen, A., and M. Velikov. 2017. Accounting for the anomaly zoo: A trading cost perspective. Working Paper, Federal Reserve Board.

Chordia, T., A. Goyal, and A. Saretto. 2017. p-hacking: Evidence from two million trading strategies. Working Paper, HEC Lausanne.

Clarke, F. H. 1990. *Optimization and nonsmooth analysis*. SIAM, Philadelphia.

Cochrane, J. H. 2009. *Asset pricing, revised edition*. Princeton, NJ: Princeton University Press.

Cochrane, J. H. 2011. Presidential address: Discount rates. *Journal of Finance* 66:1047–108.

De Roon, F. A., T. E. Nijman, and B. J. Werker. 2001. Testing for mean-variance spanning with short sales constraints and transaction costs: The case of emerging markets. *Journal of Finance* 56:721–42.

DeMiguel, V., L. Garlappi, F. J. Nogales, and R. Uppal. 2009a. A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms. *Management Science* 55:798–812.

DeMiguel, V., L. Garlappi, and R. Uppal. 2009b. Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? *Review of Financial Studies* 22:1915–53.

DeMiguel, V., F. J. Nogales, and R. Uppal. 2014. Stock return serial dependence and out-of-sample portfolio performance. *Review of Financial Studies* 27:1031–73.

Efron, B., and T. Hastie. 2016. *Computer age statistical inference*. Cambridge: Cambridge University Press.

Fama, E. F. 1991. Efficient capital markets: II. *Journal of Finance* 47:427–65.

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

———. 2016. Dissecting anomalies with a five-factor model. *Review of Financial Studies* 29:69–103.

Feng, G., S. Giglio, and D. Xiu. 2017. Taming the factor zoo. Working Paper, Chicago Booth.

Frazzini, A., R. Israel, and T. J. Moskowitz. 2015. Trading costs of asset pricing anomalies. Working Paper, Fama-Miller Center.

Freyberger, J., A. Neuhierl, and M. Weber. 2018. Dissecting characteristics nonparametrically. Working Paper, Chicago Booth.

Geary, R. C. 1935. The ratio of the mean deviation to the standard deviation as a test of normality. *Biometrika* 27:310–32.

Gibbons, M. R., S. A. Ross, and J. Shanken. 1989. A test of the efficiency of a given portfolio. *Econometrica* 57:1121–52.

Goyal, A. 2012. Empirical cross-sectional asset pricing: A survey. *Financial Markets and Portfolio Management* 26:3–38.

Green, J., J. R. Hand, and X. F. Zhang. 2017. The characteristics that provide independent information about average U.S. monthly stock returns. *Review of Financial Studies* 30:4389–436.

Hand, J. R., and J. Green. 2011. The importance of accounting information in portfolio optimization. *Journal of Accounting, Auditing & Finance* 26:1–34.

Harvey, C. R., and Y. Liu. 2018. Lucky factors. Working Paper.

Harvey, C. R., Y. Liu, and H. Zhu. 2015. …and the cross-section of expected returns. *Review of Financial Studies* 29:5–68.

Hastie, T., R. Tibshirani, and M. Wainwright. 2015. *Statistical learning with sparsity: The lasso and generalizations*. Boca Raton, FL: CRC Press.

Hou, K., C. Xue, and L. Zhang. 2014. Digesting anomalies: An investment approach. *Review of Financial Studies* 28:650–705.

———. 2017. Replicating anomalies. Working Paper.

Kandel, S., and R. F. Stambaugh. 1995. Portfolio inefficiency and the cross-section of expected returns. *Journal of Finance* 50:157–84.

Kelly, B., and S. Pruitt. 2015. The three-pass regression filter: A new approach to forecasting using many predictors. *Journal of Econometrics* 186:294–316.

Kelly, B., S. Pruitt, and Y. Su. 2018. Characteristics *are* covariances: A unified model of risk and return. Working Paper, Yale University.

Kogan, L., and M. H. Tian. 2013. Firm characteristics and empirical factor models: A data-mining experiment. Discussion Paper, FRB International Finance.

Korajczyk, R. A., and R. Sadka. 2004. Are momentum profits robust to trading costs? *Journal of Finance* 59:1039–82.

Kozak, S., S. Nagel, and S. Santosh. Forthcoming. Shrinking the cross section. *Journal of Financial Economics*.

Ledoit, O., and M. Wolf. 2008. Robust performance hypothesis resting with the Sharpe ratio. *Journal of Empirical Finance* 15:850–59.

Lewellen, J., S. Nagel, and J. Shanken. 2010. A skeptical appraisal of asset pricing tests. *Journal of Financial Economics* 96:175–94.

Light, N., D. Maslov, and O. Rytchkov. 2017. Aggregation of information about the cross section of stock returns: A latent variable approach. *Review of Financial Studies* 30:1339–81.

Linnainmaa, J. T., and M. R. Roberts. 2018. The history of the cross-section of stock returns. *Review of Financial Studies* 31:2606–49.

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

Luttmer, E. G. 1996. Asset pricing in economies with frictions. *Econometrica* 64:1439–67.

McLean, R. D., and J. Pontiff. 2016. Does academic research destroy stock return predictability? *Journal of Finance* 71:5–32.

Meinshausen, N., L. Meier, and P. Buhlmann. 2009. p-values for high-dimensional regression. *Journal of the American Statistical Association* 104:1671–81.

Meinshausen, N., and B. Yu. 2009. Lasso-type recovery of sparse representations for high-dimensional data. *Annals of Statistics* 37:246–70.

Messmer, M., and F. Audrino. 2017. The (adaptive) lasso in the zoo-firm characteristic selection in the cross-section of expected returns. Working Paper, University of St. Gallen.

Novy-Marx, R. 2016. Backtesting strategies based on multiple signals. Working Paper, Simon Graduate School of Business.

Novy-Marx, R., and M. Velikov. 2016. A taxonomy of anomalies and their trading costs. *Review of Financial Studies* 29:104–47.

Politis, D. N., and J. P. Romano. 1994. The stationary bootstrap. *Journal of the American Statistical Association* 89:1303–13.

Rockafellar, R. T. 2015. *Convex analysis*. Princeton, NJ: Princeton University Press.

Rytchkov, O., and X. Zhong. 2017. Information aggregation and data snooping. Working Paper, Fox School of Business and Management.

Tibshirani, R. 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)* 58:267–88.

Wasserman, L., and K. Roeder. 2009. High-dimensional variable selection. *Annals of Statistics* 37:2178–201.