Journal of Financial Economics 154 (2024) 103808

Contents lists available at ScienceDirect

Journal of Financial Economics

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

Persistent and transitory components of firm characteristics: Implications for asset pricing ☆

Fahiz Baba-Yara [a], Martijn Boons [b,c,*], Andrea Tamoni [d]

[a] Kelley School of Business, Indiana University, 1309 E 10th St, Bloomington Indiana, IN 47405, USA
[b] Department of Finance, Tilburg School of Economics and Management, 5000 LE Tilburg, the Netherlands
[c] Nova School of Business and Economics, Rua Holanda 1, 2775-405 Carcavelos, Portugal
[d] Rutgers Business School, Department of Finance and Economics, 1 Washington Park, Newark, NJ 07102, USA

ARTICLE INFO

Dataset link: https://data.mendeley.com/datasets/sj93bbhdcw/2

JEL classification:
G11
G12

Keywords:
Characteristics
Persistent-transitory decomposition
Cross-section
Return predictability
Discount rates

ABSTRACT

We study the horizon dimension of cross-sectional return predictability using a model where characteristics contain both persistent and transitory components. We test the implications of this model for the average returns of popular characteristic-based trading strategies at short versus long horizons after portfolio formation. Our evidence supports the claim that the relative compensation for persistent and transitory components varies across characteristics, in both magnitude and sign. Benchmark factor models cannot explain the returns of portfolios sorted on characteristics where either the persistent or transitory component is dominant. Finally, we discuss implications for the long-term discount rates of firms.

1. Introduction

In this paper, we study the relative compensation for persistent and transitory components of firm characteristics. Recent literature argues that transitory components of characteristics are the major driver of cross-sectional return predictability. For instance, Keloharju et al. (2021) decompose a large set of characteristics into their persistent and transitory components and argue that only the average transitory component predicts returns. Liu et al. (2021) further argue that the mispricing captured by characteristics is more transitory than the risk. Instead, we argue that the relative compensation for persistent and transitory components varies across characteristics, both in magnitude and sign. Since the return predictability from a characteristic such as book-to-market is quite persistent, it seems unlikely that no persistent component is priced.

To fix ideas, consider a characteristic that consists of a persistent and a transitory component: $X_t = X_t^P + X_t^T$. We assume that a firm's expected return is determined by its market beta and compensation for these components:

$$E_t(R_{t+1}) = \lambda_M \beta_M + \lambda_P X_t^P + \lambda_T X_t^T. \quad (1)$$

We further assume that the two components of X_t contribute equally to X-factor exposure. The X-factor exposure is our measure of factor risk or beta, consistent with numerous studies that construct factors from sorting stocks on not-decomposed characteristics.

☆ Nikolai Roussanov was the editor for this article and we thank him for his invaluable guidance. Further, we are grateful to an anonymous referee, Fernando Anjos, Jonathan Berk, Kent Daniel (discussant) Esther Eiling, Rik Frehen, Jens Kværner, Juhani Linnainmaa, Lars Lochstoer, Tyler Muir, Dino Palazzo (discussant), Florian Peters, Simon Rottke, Chi-Yang Tsou (discussant) and seminar participants at the 2021 American Finance Association, 2022 European Finance Association, Morgan Stanley London, 2020 Northern Finance Association, Quoniam Asset Management Frankfurt, Rutgers Business School, Tilburg University, University of Amsterdam, University of Houston, Vienna Graduate School of Finance and Vrije Unversiteit Amsterdam for helpful comments. This paper is based upon work supported by the Dutch Research Council (NWO Vidi grant 201005).

* Corresponding author at: Nova School of Business and Economics, Rua Holanda 1, 2775-405 Carcavelos, Portugal.
E-mail addresses: fababa@iu.edu (F. Baba-Yara), martijn.boons@novasbe.pt (M. Boons), atamoni@business.rutgers.edu (A. Tamoni).

https://doi.org/10.1016/j.jfineco.2024.103808
Received 29 October 2021; Received in revised form 30 January 2024; Accepted 13 February 2024
Available online 1 March 2024

We ask two questions to assess the importance of variation in the relative compensation λ_P versus λ_T. First, what are the horizon dynamics in returns of characteristic-sorted portfolios? Second, what are the long-term discount rates of firms? Although existing literature on characteristic-based return predictability focuses almost exclusively on short-term returns, studying returns at longer horizons is economically important. Characteristics that predict returns more persistently either have a larger impact on firm's discount rates (Keloharju et al., 2021) or imply more mispricing (Van Binsbergen and Opp, 2019), and the horizon of most investors is longer than a single month.[1]

To evaluate the horizon dynamics of returns, we simulate from a model that accounts for important properties of real-world data. We track the return after formation of the high-minus-low decile portfolio sorted on X_t. We denote these returns as $R_{X,(t-s),t+1}$, where $(t-s)$ refers to the sorting date and $t+1$ denotes the return observation date. The null hypothesis – of a standard characteristic-based model of expected returns – is equal compensation for persistent and transitory components, that is, $\lambda_P = \lambda_T > 0$. In this case, the alpha in a regression of the return to old sorts ($s > 0$) on the newest sort ($s = 0$) is zero, because expected returns decay exactly at the same speed as the characteristic X_t and, therefore, risk (as measured by X-factor beta).[2] Instead, this alpha is negative when only the transitory component is priced ($\lambda_P = 0$), because expected returns decay too fast relative to X_t and risk. Conversely, this alpha is positive when only the persistent component is priced ($\lambda_T = 0$). Testing the null using alphas is attractive because the regression beta controls for persistence, which varies strongly across firm characteristics in the data. Simply comparing the average returns of new and old sorts, as in Keloharju et al. (2021), does not speak to the relative compensation for persistent and transitory components in our model.

We compare model-implied distributions under the various hypotheses to the distribution obtained empirically from a large set of 56 characteristics (studied also in Freyberger et al. (2020)). For each characteristic, we construct value-weighted portfolios and track the buy-and-hold return of the high-minus-low strategy from one month up to five years after portfolio formation from 1972 through 2019. We find that old sorts provide a significantly negative alpha for 23 characteristics, whereas this alpha is positive and significant for another eight characteristics.[3] The magnitude of alphas in the left tail cannot be explained under the null of the standard characteristic-based model. In turn, the magnitude of alphas in the right tail cannot be explained under the assumption that only the transitory component of characteristics is priced (as argued in Keloharju et al., 2021). For instance, three years after portfolio formation, a high-minus-low book-to-market strategy provides a positive annualized alpha of 4.40% (t-stat = 2.28) relative to the newest book-to-market strategy. This alpha translates to a large improvement in Sharpe ratio: from 0.23 for the newest book-to-market sort to 0.45 for its optimal combination with the three-year-old book-to-market sort. Among the characteristics where old sorts provide a significant alpha, relative increases in Sharpe ratio of 100% or more are commonplace.

This evidence suggests that the compensation for persistent and transitory components varies in magnitude and sign across characteristics.[4] An advantage of our model setup is that we can easily analyze alternative characteristic-based trading strategies. One such strategy is to decompose the return of the newest sort into the return coming from new and old stocks. These stocks together make up the extreme high or low characteristic-sorted portfolio today, but only the old stocks were in (or close to) that same characteristic-sorted portfolio in the past. This decomposition is useful because real-world characteristic-based strategies have drawn inspiration from evidence based on the newest sorts. If either the persistent or transitory component is dominant, it is more profitable to trade only one of the two subsets of stocks.

We find empirically that the average return of old-minus-new stocks is highly correlated across characteristics with the alpha of old-versus-new sorts, as predicted by the model. Old-minus-new stock returns also vary strongly in magnitude and sign. For instance, a book-to-market strategy that uses only old stocks obtains an annualized return that is 7.37% (t-stat = 2.48) higher than a strategy that uses only new stocks. This result occurs even though these two sets of stocks generate the same spread in book-to-market today. Thus, firms with a persistently high book-to-market ratio capture a higher return than firms for which the same book-to-market ratio is more transitory.

We next ask whether these results are surprising from the point of view of benchmark factor models.[5] Any model that prices the newest sorts will also price the older sorts under the null of a standard characteristic-based model of expected returns. Challenging models with moments based on old-versus-new sorts is useful, because characteristic-based factors are routinely defined as the return on a new sort. Such factors have been sequentially added to the CAPM to improve explanatory power for cross-sections of new sorts. Hence, some of the improved fit at short horizons may be due to overfitting (Harvey et al., 2016; Harvey and Liu, 2021).

To reduce the dimensionality of our data, we extract the first principal component (PC1) at each horizon after portfolio formation. Aggregating over the 56 characteristics using the PC1 loadings, old-versus-new sorts generate large abnormal average returns with t-statistics well above three in all models we study.[6] The aggregated old-versus-new strategy is difficult to price, because the PC1 loadings correlate strongly with the alpha between old and new sorts.[7] We use this insight to split PC1 in two sub-components, that is, the component coming from characteristics on which PC1 loads with a positive versus negative sign. We find that the benchmark models unanimously struggle to price both sub-components. This finding represents strong, joint evidence that existing factor models cannot explain why returns decay too fast for some characteristics, but too slow for others.

Through the lens of our model, the failure of benchmark models is unsurprising. Factors based on new sorts capture the total compensation for loading on a characteristic. However, to price old-versus-new sorts and stocks, factors must also account for the relative compensation of persistent and transitory components. Independent of whether new asset pricing models describe expected returns as a function of risk or mispricing, models looking for a challenge should target the horizon dynamics presented in this paper. Indeed, any model that prices

[1] Recent work studies optimal rebalancing frequencies and finds large variation across characteristics (Novy-Marx and Velikov, 2016; Jensen et al., 2022).

[2] This null also holds approximately in popular theoretical explanations of characteristics-based return predictability, such as Gomes et al. (2003) and Zhang (2005).

[3] Our robustness checks confirm that this conclusion extends in subsamples and when estimating alternative definitions of the alpha between old and new sorts, for instance, using the decay in the characteristic spread.

[4] If the relative compensation were fixed across characteristics, the model would imply a large correlation between persistence and alphas. This prediction is also strongly rejected in the data.

[5] The complete set of models we study includes the CAPM as well as the models of Fama and French (1993), Frazzini and Pedersen (2014), Fama and French (2015), Hou et al. (2015), Stambaugh and Yuan (2016), Daniel et al. (2020a), and Daniel et al. (2020b).

[6] We find qualitatively and quantitatively similar results when we aggregate the old-minus-new stock strategies using PC1 loadings. This aggregated old-minus-new stock strategy is particularly interesting because it does not load meaningfully on any of the 56 characteristics we study (i.e., the strategy is characteristic-neutral).

[7] Note that the principal component loadings are determined only by the covariance matrix of returns, thus ignoring the relative performance of old and new sorts. Moreover, we discuss in Online Appendix A an extension of our model that can match the empirical link between principal component loadings and the performance of old-versus-new sorts using minimal additional assumptions.

returns at all horizons will get price levels right (Cho and Polk, 2020; van Binsbergen et al., 2023).

To assess the implications for long-term discount rates, we follow the approach in Keloharju et al. (2021, Section 6.3). The implied discount rate r solves the Gordon growth equation: $P = \frac{D}{r-g}$, where P follows from discounting an implied cash flow stream at rates derived from the realized average returns of a characteristic-sorted portfolio.[8] The model implies that the high-minus-low portfolio difference in implied discount rate is small when only the transitory component is priced. In contrast, the difference can be large when the persistent component is priced.

In the data, the high-minus-low difference in discount rates is large at 2.5% (value-weighted portfolios) and 3.5% (equal-weighted portfolios) when we zoom in on firms that load strongly on characteristics with returns that decay too slow, that is, the characteristics on which PC1 loads with a positive sign. These effects are at least 2.5 times larger than what is found in Keloharju et al. (2021) and are economically important.[9] The reason for this gap is that these authors focus on firms that load strongly on the average of a large set of characteristics, which overweights characteristics with a dominant transitory component. Our simulations show that large discount rate effects are most likely generated in a world where the compensation for the persistent component of the characteristic is large relative to the transitory component. Single sorts on some characteristics also imply meaningful discount rate differences. For instance, in the value-weighted (equal-weighted) case, firms in the high size and book-to-market deciles have an implied discount rate that is, respectively, 2% and 1.5% (3% and 2.1%) larger than firms in the low deciles. In contrast, for characteristics like profitability and investment, discount rate differences are smaller. We conclude that there are subsets of firms for which long-term discount rates are meaningfully affected by characteristic-based return predictability.

Overall, our evidence that the relative compensation for persistent and transitory components varies strongly across characteristics challenges popular explanations of the cross-section based solely on recent observations of firm characteristics. Our results are also important for investors, because most stock-picking applications explicitly reduce the information set to the most recent values of firm characteristics. Finally, characteristic-based return predictability should not be ignored in capital budgeting.

1.1. Literature

The literature on characteristics-based return predictability is vast, but almost exclusively studies the relation between characteristics and short-term returns. In recent machine learning literature, the goal has been to find the (potentially higher-order) functional form of a large set of characteristics that best predicts short-term returns (see, e.g., Kozak et al. (2020), Freyberger et al. (2020), and Gu et al. (2020)). Similarly, empirical asset pricing tests typically use factors and test assets derived from sorting stocks on recent observations of characteristics (see, e.g., Fama and French (2015, 2018), Hou et al. (2015, 2018)). We derive new moments from the returns at longer horizons after portfolio formation that challenge these models.

We share the objective of generating new moments to test for model misspecification with a number of recent papers. Chernov et al. (2022) show that the restrictions implied by a stochastic discount factor (SDF) that prices single period returns of popular factors, like those of Fama and French, do not hold for long-term returns of the same factors. While the focus of Chernov et al. (2022) is on multi-period compounded returns of rebalanced factors, we focus on the buy-and-hold returns of characteristic-sorted portfolios. Our test assets are motivated by a novel decomposition of characteristics in persistent and transitory components, which allows us to speak directly to the popular characteristics-based model of expected returns. Liu et al. (2021) discuss a related channel for model misspecification when mispricing is more transitory than factor risk. In this case, including factors that contain a mispricing component will distort the firm's expected return after the mispricing is corrected. Their proposed channel is one motivation for our alternative hypothesis that the compensation for the transitory component of the characteristic is relatively large. However, this channel cannot explain the evidence for the subset of characteristics with returns that decay too slow. For the same reason, our evidence contributes to Keloharju et al. (2021), who show that returns decay too fast for the average characteristic.

Our results regarding the long-term discount rate take the perspective that characteristic-based return predictability captures compensation for risk. However, our results also resonate with recent literature that takes the perspective that characteristics capture mispricing. For instance, Cho and Polk (2020) and van Binsbergen et al. (2023) use longer-horizon returns of characteristic-sorted portfolios to estimate the price wedge, that is, the difference between the market price of an asset and the rationally discounted present value of the asset's future cash flows. van Binsbergen et al. (2023) focus on the dynamics of these price wedges at the portfolio and firm level and their potential for real capital misallocations. Consistent with variation in the relative compensation for persistent and transitory components, both of these papers find that the total mispricing implied by some characteristics, such as profitability, is small, whereas it is large for others, such as book-to-market. In particular, Cho and Polk (2020) focus on the interaction between value and quality as the main determinant of price wedges in the cross-section.

Daniel et al. (2020b) argue that factors can be traded more profitably by combining a factor, like the high-minus-low book-to-market portfolio, with an offsetting position in a hedge portfolio that has a zero loading on the characteristic (book-to-market) and a maximum loading on the factor (see, also, Daniel and Titman, 1997; Herskovic et al., 2019). We argue that combinations of newer and older sorts are attractive investments and show that these combinations provide returns that are not captured by popular factors, including the optimally hedged factors of Daniel et al. (2020b). In fact, we reject the common assumption that firms' loadings on the SDF are a function of current values of characteristics, such as size, book-to-market, profitability, investment, and momentum. The reason is that our aggregate old-minus-new stock strategy is approximately neutral with respect to these characteristics, but has a non-zero average excess return. This return originates from the variation in the relative compensation for persistent and transitory components of these characteristics, a variation which has been largely overlooked in the literature.

2. Data

In this paper we study 56 characteristics that are similar to those in Freyberger et al. (2020) and described in detail in Table OA.1 of the Online Appendix. For all U.S. common stocks traded on the NYSE, AMEX or NASDAQ from July 1964 through December 2019, we collect monthly and daily stock market data from the Center for Research in Security Prices (CRSP) and annual balance-sheet data from Compustat. Following Green et al. (2017) and Gu et al. (2020), we delay monthly variables by one month and annual variables by six months. We construct value-weighted decile portfolios for each characteristic, splitting each portfolio at NYSE breakpoints to reduce the influence of micro-cap stocks. We track the buy-and-hold returns of these decile portfolios. When a stock delists, we reallocate the investment in this stock (net

[8] We assume a growth rate $g = 1\%$ in our main analyses, but present qualitatively similar results for $g = 5\%$ in a robustness check. We track realized average returns up to 10 years after portfolio formation. After year 10, we assume the per period discount rate has converged back to the expected market return of 8% (i.e., a 2% risk-free rate plus a 6% market risk premium).

[9] For instance, if the discount rate for an asset is 9.5% and using $D = 1\$$ and $g = 1\%$, the asset's price would be 11.8. If the discount rate is only 6.5%, the price would be almost 55% larger at 18.2.

Table 1

Old and new sorts on popular characteristics. This table reports summary statistics for old and new sorts on size, book-to-market, profitability, and investment. We track the returns of long-short decile portfolios (value-weighted and split at NYSE breakpoints) for each characteristic from one month to five years after portfolio formation. Panel A shows the average high-minus-low return. Panel B reports the alpha of old with respect to new sorts. The unconditional alpha, α^u, is the intercept from a regression of the return of an old sort on the contemporaneous return of the newest sort: $R_{X,(t-s),t+1} = \alpha^u_s + \beta^u_s R_{X,(t),t+1} + \epsilon_{X,(t-s),t+1}$ (see Eq. (2)). The conditional alpha, α^c, is calculated as the average return of a strategy that invests in $R_{X,(t-s),t+1}$ and hedges in each month t the conditional exposure to $R_{X,(t),t+1}$. Following Eq. (3), we estimate this exposure over a 36 month historical rolling window. The t-statistics are calculated using White (1980) heteroskedasticity-consistent standard errors. Panel C reports the Sharpe ratio of the newest sort, Sharpe($R_{X,(t),t+1}$), and the maximum increase in Sharpe ratio achievable from combining the newest sort with the older sorts (based on either the unconditional or conditional alpha). The sample period runs from July 1972 through December 2019.

Horizon s	Size			Book-to-market			Profitability			Investment		
	Panel A: Average return of new and old sorts ($R_{X,(t-s),t+1}$)											
	Avg.	t-stat		Avg.	t-stat		Avg.	t-stat		Avg.	(t-stat)	
0	2.01	(0.86)		5.24	(1.61)		5.09	(3.32)		6.06	(3.62)	
12	4.45	(1.91)		7.15	(3.14)		2.94	(2.05)		2.86	(1.99)	
24	3.48	(1.50)		5.49	(2.62)		1.50	(1.05)		0.47	(0.31)	
36	2.44	(1.13)		5.54	(2.75)		-0.02	(-0.01)		0.83	(0.53)	
48	3.08	(1.48)		5.97	(3.03)		0.16	(0.11)		1.02	(0.62)	
60	2.54	(1.17)		4.44	(2.23)		-0.35	(-0.23)		-0.78	(-0.47)	
	Panel B: Alphas of old-versus-new sorts											
	α^u_s	β^u_s	α^c_s	α^u_s	β^u_s	α^c_s	α^u_s	β^u_s	α^c_s	α^u_s	β^u_s	α^c_s
12	2.63	0.90	3.27	4.66	0.48	4.78	-1.09	0.79	-1.02	0.49	0.39	0.65
	(2.74)	(20.78)	(3.03)	(2.95)	(8.31)	(2.79)	(-1.42)	(15.73)	(-1.42)	(0.38)	(10.82)	(0.49)
24	1.77	0.85	2.65	3.66	0.35	4.17	-1.84	0.66	-1.67	-1.15	0.27	-1.43
	(1.53)	(18.57)	(2.14)	(2.14)	(7.07)	(2.23)	(-1.83)	(10.67)	(-1.75)	(-0.79)	(6.06)	(-0.98)
36	0.83	0.80	1.71	4.15	0.26	4.40	-3.11	0.61	-3.25	-0.79	0.27	-1.59
	(0.78)	(21.11)	(1.54)	(2.35)	(5.31)	(2.28)	(-2.72)	(11.02)	(-3.02)	(-0.52)	(6.70)	(-1.04)
48	1.56	0.76	2.35	4.65	0.25	4.64	-2.81	0.58	-2.83	-1.30	0.38	-1.77
	(1.43)	(20.03)	(2.10)	(2.66)	(5.43)	(2.41)	(-2.46)	(10.95)	(-2.69)	(-0.85)	(7.45)	(-1.16)
60	1.03	0.75	1.84	3.18	0.24	2.71	-3.18	0.56	-3.16	-2.85	0.34	-3.53
	(0.83)	(15.84)	(1.40)	(1.78)	(5.59)	(1.37)	(-2.64)	(8.48)	(-2.81)	(-1.79)	(5.65)	(-2.30)
	Panel C: Improvements in Sharpe ratio											
	Sharpe ratio of the newest sort ($R_{X,(t),t+1}$)											
0	0.12			0.23			0.48			0.52		
	Max. Sharpe($R_{X,(t-s),t+1}$, $R_{X,(t),t+1}$) - Sharpe($R_{X,(t),t+1}$)											
	u		c	u		c	u		c	u		c
12	0.28		0.34	0.23		0.25	0.04		0.03	0.00		0.00
24	0.12		0.22	0.15		0.21	0.07		0.04	0.01		0.01
36	0.04		0.14	0.17		0.22	0.14		0.13	0.01		0.01
48	0.11		0.21	0.21		0.22	0.12		0.11	0.01		0.02
60	0.05		0.12	0.11		0.10	0.13		0.12	0.06		0.09

of the delisting return) to the non-missing stocks in the portfolio using value-weights.

The return to a characteristic-sorted portfolio is defined as the return of the zero-cost, long-short portfolio formed from buying the high portfolio and selling the low portfolio[10]:

$$R_{X,(t-s),t+1} = R^{High}_{X,(t-s),t+1} - R^{Low}_{X,(t-s),t+1}.$$

In this definition, the first subscript refers to the characteristic, $X = 1, \ldots, 56$; the second subscript refers to the date of portfolio formation or sorting date, $(t-s)$ where $s \geq 0$; and the third subscript refers to the return realization date, $t+1$. By varying the sorting date, we observe contemporaneous returns to the newest sort ($s = 0$), which is the focus in most of the literature, and older sorts ($s > 0$). Similar to Jegadeesh and Titman (1993), we combine three sorts for each horizon $s > 0$ to reduce noise. For brevity and because some characteristics are updated only once per year, we initially focus on $s = 0, 12, 24, \ldots, 60$.[11] Dictated by data availability and a burn-in period for some of our estimates, the sample period for all our results is July 1972 through December 2019.

3. Motivating evidence

To motivate our study of the persistent and transitory components of characteristics and their impact on the relative performance of new and old sorts, we focus on the characteristics in the Fama and French (2015) model: size, book-to-market, profitability and, investment.

Panel A of Table 1 presents summary statistics for the characteristic-sorted portfolios. We see that each portfolio obtains a positive average return one month after formation, ranging from 2.01% for size to 6.06% for investment. The persistence of return predictability varies considerably across these popular characteristics, however. The book-to-market effect is large and at least marginally significant at all horizons up to five years after portfolio formation. In fact, the book-to-market effect is largest one year after portfolio formation at 7.15%, after which it slowly decreases to 4.44% at the five-year mark. Similarly, the size ef-

[10] For a characteristic X that predicts returns with a negative sign, such as size, we sort on $-1 \times X$. Signing the characteristics in this way makes our results more comparable to previous work (e.g., Freyberger et al. (2020) and Haddad et al. (2020)), but leaves our main conclusions unchanged. If some return R_X expands the mean-variance frontier, R_{-X} will do so as well.

[11] Only for a handful of characteristics do returns beyond the five-year horizon provide an abnormal return relative to both the newest sort ($s = 0$) and the older sorts with $s \leq 60$.

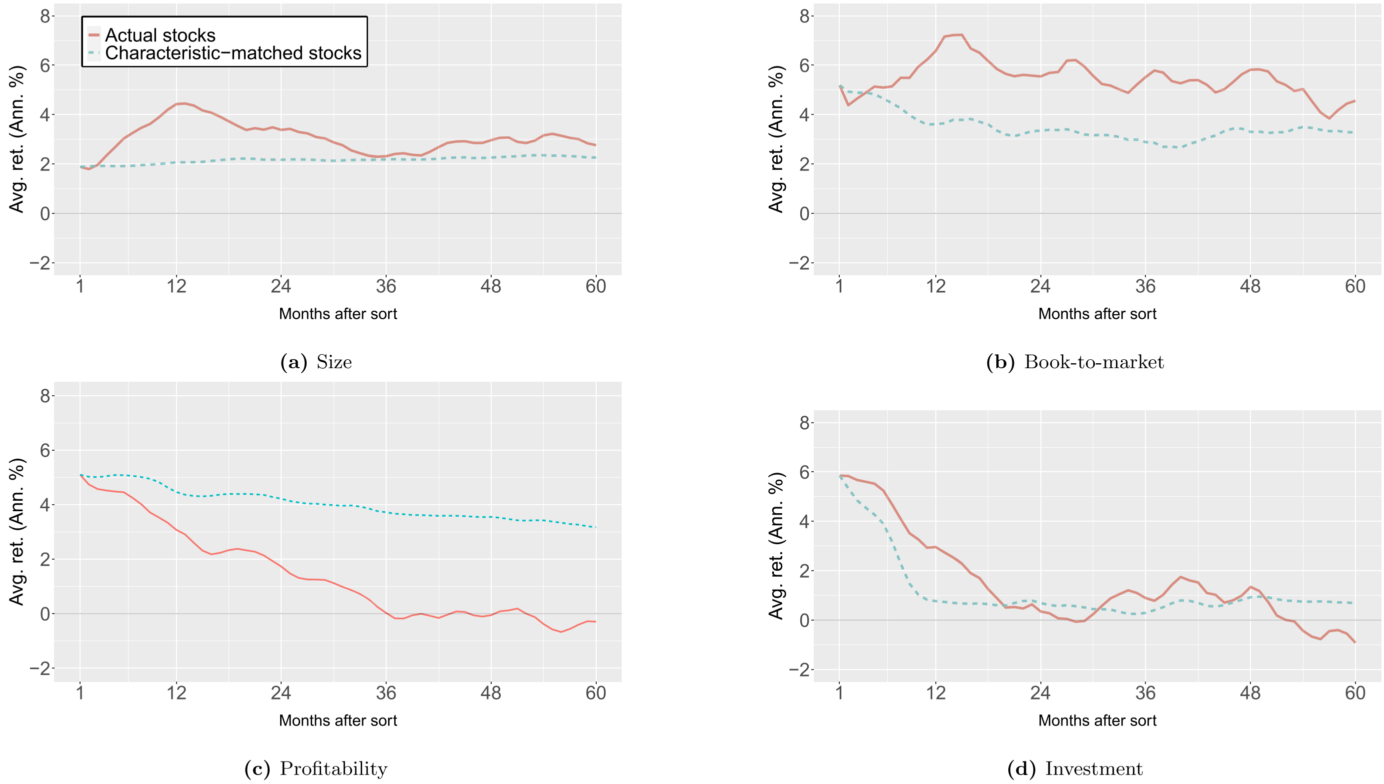


Fig. 1. Returns and characteristic-matched returns We plot annualized average returns and characteristic-matched average returns of the high-minus-low decile portfolios sorted on size, book-to-market, profitability, and investment. We consider horizons from one month up to five years after portfolio formation. The sample period runs from July 1972 through December 2019.

fect is substantial at all horizons, but largest one year after portfolio formation at 4.45%. In contrast, the effects for profitability and investment are small and insignificant at horizons beyond two years after portfolio formation.

Under the null of a standard characteristic-based model of expected returns, the persistence of return predictability matches the persistence of the characteristic. In Fig. 1, we analyze this implication using the approach of Keloharju et al. (2021, Section 7.2). In each month up to five years after portfolio formation, we plot the return of the long-short portfolio as well as the return of a characteristic-matched portfolio.[12] First, we see that persistence – proxied by the matched portfolio return – varies importantly across characteristics, with book-to-market (investment) being the most (least) persistent. Second, we see that returns are too persistent (i.e., returns decay slowly relative to the matched return) for book-to-market and size. In contrast, returns are too transitory (i.e., returns decay relatively fast) for profitability. For investment, returns decay initially too slow, but later too fast.

What do these results imply for the relative performance of new and old sorts? To answer this question, we regress the returns of older sorts on the newest sort:

$$R_{X,(t-s),t+1} = \alpha_s^u + \beta_s^u R_{X,(t),t+1} + \varepsilon_{X,(t-s),t+1}. \quad (2)$$

The model presented in the next section formally motivates why the beta in this regression is a suitable control for characteristic-persistence. Our main interest is in the alpha, measuring the unconditional abnormal performance of old sorts relative to the newest sort. Because persistence may vary over time, we also calculate a conditional alpha using exposures estimated over a 36-month rolling window:

$$\alpha_s^c = E(R_{X,(t-s),t+1} - \beta_{s,t}^c R_{X,(t),t+1}), \text{ with } \beta_{s,t}^c \text{ from:} \quad (3)$$

$$R_{X,(\tau-s),\tau+1} = \alpha_s + \beta_{s,t}^c R_{X,(\tau),\tau+1} + \varepsilon_{X,(\tau-s),\tau+1}, \ \tau = t-36 : t-1. \quad (4)$$

In Panel B of Table 1 we see that a large number of these alphas are economically and statistically significant. Let us focus on the conditional alpha (α_s^c) from Eq. (3). This alpha is large and positive at all horizons after portfolio formation for book-to-market and size, consistent with the fact that returns decay too slow for these characteristics.[13] For book-to-market, the alpha is significant up to four years out at about 4.50%; for size, the alpha is significant at the one-, two-, and four-year mark at about 2.75%. Consistent with the fact that returns decay too fast for profitability, we see negative alphas for this characteristic. These alphas are significant at about −3.00% from three to five years after portfolio formation.

A significant alpha implies that the maximum Sharpe ratio from investing in the optimal portfolio of the older and the newest sort is significantly larger than the Sharpe ratio of the newest sort. We present these improvements in Sharpe ratio in Panel C. For book-to-market, the Sharpe ratio from investing in the newest sort, $R_{BM,(t),t+1}$, equals 0.23. The Sharpe ratio more than doubles to 0.48 (= 0.23 + 0.25) when an investment in the older sort, $R_{BM,(t-12),t+1}$, is added.[14] Similarly, for

[12] The matched portfolio return is calculated by tracking the portfolio rank of each stock that is assigned to the high and low decile at the sorting date. In each month after portfolio formation, we replace the stock's actual return with the value-weighted return of the decile to which the stock belongs at that point in time. We finally take the value-weighted average within each decile of these surrogate returns.

[13] Although variation in past returns contributes to the alphas for size and book-to-market, we show below that past returns-related variables, such as momentum, cannot explain the alphas we find in the larger cross-section of 56 characteristics.

[14] The optimal portfolio invests 1.22 and −0.22 in $R_{BM,(t-12),t+1}$ and $R_{BM,(t),t+1}$, respectively, in the unconditional specification (1.38 and −0.38 in the conditional specification). Although optimal weights are more extreme for

size, the optimal combination of $R_{Size,(t-12),t+1}$ and $R_{Size,(t),t+1}$ obtains a Sharpe ratio that is more than triple the Sharpe ratio of an investment in $R_{Size,(t),t+1}$ alone: 0.46 versus 0.12. For both book-to-market and size, the increase in Sharpe ratio falls gradually as time passes after portfolio formation, although it remains economically large at over 0.10 for all sorting dates in the conditional specification. For profitability and investment, the largest increases in Sharpe ratio are observed when the return three and five years after portfolio formation is combined with the return one month after portfolio formation, at 0.13 and 0.09, respectively. These absolute increases translate to relative increases in Sharpe ratio of about 20%, which is non-negligible economically.

We conclude that there are significant alphas between new and old sorts and the sign and magnitude of these alphas vary substantially even among the most popular characteristics in the literature. To the best of our knowledge, we are the first to estimate these alphas and show that they vary in sign across characteristics. If these alphas represent a rejection of the null of a standard characteristic-based model of expected returns, what is a suitable alternative?

4. Model

We start with a simple model that outlines the main predictions for new and old sorts that we test in this paper. We discuss the key trade-offs between characteristic-persistence and the persistence of return predictability under alternative hypotheses motivated by previous literature. Towards the end of this section, we turn to a richer specification that accounts for more properties of the data.

4.1. A simple data generating process

Following Keloharju et al. (2021), we assume that firm characteristics consist of a persistent and a transitory component. For a generic characteristic X, with an unconditional mean equal to zero and variance equal to one, we assume:

$$X_{i,t} = X^P_{i,t} + X^T_{i,t}, \quad (5)$$

$$X^P_{i,t} = \rho X^P_{i,t-1} + \epsilon^P_{i,t}, \text{ and} \quad (6)$$

$$X^T_{i,t} = \epsilon^T_{i,t}, \quad (7)$$

such that $X^P_{i,t}$ follows an AR(1)-process with persistence ρ. We initially draw the random normal shocks $\epsilon^P_{i,t}$ and $\epsilon^T_{i,t}$ at the annual frequency. In Section 4.5, we map our model to the monthly frequency usually studied in the literature.

We define a stock's expected return to be a function of its market beta and the components of X:

$$E_t(R_{i,t+1}) = \lambda_M \beta_{i,M} + \lambda_P X^P_{i,t} + \lambda_T X^T_{i,t} \,. \quad (8)$$

We allow the relative compensation for the two components to differ. Differential compensation is key to understanding the relative performance of old and new sorts in the data. Although our main focus is on a two-factor model, our arguments extend to a model with multiple characteristics (see Online Appendix A, for instance). Moreover, we control for a large set of benchmark factor models when we analyze the implications of our results for asset pricing models in Section 6.

Realized returns are defined as follows:

$$R_{i,t+1} = E_t(R_{i,t+1}) + \beta_{i,M}(R_{M,t+1} - E_t(R_{M,t+1})) + \beta_{i,F_X,t}(F_{X,t+1} - E_t(F_{X,t+1})) + \epsilon_{i,t+1}, \quad (9)$$

where the exposure to the second factor $F_{X,t+1}$ – referred to as the "X-factor" – is a function of the characteristic:

$$\beta_{i,F_X} = c_{F_X} X_{i,t} + \epsilon_{i,F_X}. \quad (10)$$

We set the loading c_{F_X} such that the long-short portfolio resulting from a sort on the characteristic X has a unit X-factor beta. We start out assuming that $\epsilon_{i,t+1}$ is a random normal idiosyncratic return shock. Then, as the number of stocks $i = 1, \ldots, N$ grows larger, the return of the long-short decile portfolio converges to $F_{X,t+1}$. This setup is consistent with numerous studies that construct factors as long-short characteristic-sorted portfolios. Indeed, existing factors do not distinguish between the persistent and transitory components of firm characteristics.

Our model allows us to study the relative compensation for these components through the returns of old versus new sorts. We prefer to analyze these returns rather than decomposing firm characteristics empirically, as for instance Keloharju et al. (2021) do, for two reasons. First, portfolios sorted on not-decomposed characteristics feature widely in academic and practitioner literature. Second, we would need to make additional assumptions about the dynamics of firm characteristics to perform the decomposition empirically. That said, we discuss below a robustness check using a simple empirical decomposition of characteristics.

Furthermore, our model yields interesting and testable implications without having to take a stand on how much risk and mispricing contribute to the components of X. Previous literature provides suggestive evidence for this issue, however. For instance, Kothari et al. (1995) and Cohen et al. (2009) find that the size and book-to-market effects align better with risk measured over horizons considerably longer than a single month. In the context of our model, this would imply that risk lines up better with the persistent component of characteristics. In contrast, the evidence in Liu et al. (2021) suggests that the mispricing captured by performance-related characteristics (such as profitability) is more in line with the transitory component of these characteristics.

4.2. Hypotheses

Our null hypothesis is that $\lambda_P = \lambda_T$ in Eq. (8), consistent with the standard characteristic-based model of expected returns studied in the literature. Under this null, a unit increase in either X^P or X^T yields the same increase in expected return, just as it yields the same increase in exposure to the X-factor. Moreover, expected returns of a long-short portfolio sorted on X decay at exactly the same speed as the characteristic X itself.

Our first alternative hypothesis, $\lambda_P = 0$ and $\lambda_T > 0$, follows Keloharju et al. (2021) who argue that return predictability is driven by the transitory components of characteristics. Under this alternative, a unit increase in X^T yields a larger increase in expected return than a unit increase in X^P, even though they yield the same increase in exposure to the X-factor. Thus, the return compensation for X^T is relatively too high when compared to its risk. This assumption also implies that expected returns decay faster than the characteristic X, because returns decay at the speed of the transitory component. A good example of such a characteristic is profitability in Fig. 1.

Our second alternative hypothesis is that $\lambda_P > 0$ and $\lambda_T = 0$. Under this alternative, the return compensation for X^T is relatively too low when compared to its risk. Moreover, in this case the decay in expected returns is slower than the characteristic X, because now returns decay at the speed of the persistent component. A good example of such a characteristic is book-to-market in Fig. 1. Indeed, existing work shows that some characteristics, including book-to-market, predict returns very persistently (see Cho and Polk, 2020; van Binsbergen et al., 2023). We will argue that a relatively large compensation for the persistent component is necessary to fit the data for such characteristics.

4.3. Calibration

Through the lens of our model, the fact that book-to-market returns decay relatively slowly (see Fig. 1) indicates the presence of a component that is more persistent than the book-to-market ratio itself. To

a few of the 56 characteristics studied below, we find in those cases that the improvement in Sharpe ratio is only slightly smaller when we restrict the weights to be in the interval $[-2, +2]$.

calibrate the model, we rely on the book-to-market ratio and set the persistence of the component X^P to $\rho = 0.85$. This number is only slightly above the persistence of the book-to-market ratio in the data (= 0.75) and our results are qualitatively insensitive to this choice.

Following Keloharju et al. (2021) and to be conservative, we calibrate λ_P and λ_T such that the cross-sectional standard deviation of $E_t(R_{i,t+1})$ is equal to 4.8% per year. Given a cross-sectional standard deviation of realized annual returns of about 50%, this number would imply an R^2 of less than 1% in a cross-sectional regression of realized returns on expected returns. We draw market betas randomly from $\mathcal{N}(1, 0.6^2)$, consistent with Welch (2022). Similarly, and consistent with stocks' exposures to a long-short book-to-market factor in the data, we set the cross-sectional standard deviation of X-factor betas equal to 0.6. For the high-minus-low decile portfolio from a sort on X to have a unit beta with respect to the X-factor, we set $c_{F_X} = 0.285$.[15] This number follows from the fact that the high-minus-low spread in X is equal to 2×1.755 based on the properties of a truncated normal distribution and the uncorrelatedness of the two components of X. We set the market risk premium $\lambda_M = 6\%$. Thus, in the standard characteristic-based model of expected returns we have $\lambda_P = \lambda_T = 3.17\%$. Under the two alternatives, the compensation for the two components of the characteristic depends on $Var\left(X_{i,t}^P\right)$. For instance, for $Var\left(X_{i,t}^P\right) = 0.5$, we have $\lambda_T = 4.49\%$ (and $\lambda_P = 0$) under the first alternative inspired by Keloharju et al. (2021) and $\lambda_P = 4.49\%$ (and $\lambda_T = 0$) under the second alternative.[16] With these parameter values, about 45% of expected return variation is coming from variation in the characteristic and the rest from variation in market beta.

4.4. Predictions

With the above setup, we can calculate the alpha of old to new sorts (as well as other characteristic-based strategies; see Section 4.6) in closed form. It turns out that this alpha clearly distinguishes the alternative hypotheses.

As mentioned above, the spread in X between the top and bottom decile for the newest sort on X equals 2×1.755. Since the persistent and transitory components are uncorrelated, this spread derives for a proportion $Var(X_i^P)$ from the persistent component and (1-$Var(X_i^P)$) from the transitory component. Thus, the expected return of the newest sort equals:

$$E(R_{X,(t),t+1}) = 2 \times 1.755 \times (Var(X_{i,t}^P)\lambda_P + (1 - Var(X_{i,t}^P))\lambda_T). \tag{11}$$

Since the persistence of the characteristic $Corr(X_{i,t}, X_{i,t-s})$, and thus also the persistence of X-factor betas, equals $Var\left(X_{i,t}^P\right)\rho^s$, the expected return of older sorts equals[17]:

$$E(R_{X,(t-s),t+1}) = 2 \times 1.755 \times Var(X_{i,t}^P) \times \rho^s \times \lambda_P. \tag{12}$$

Given that the long-short portfolio does not load on the market, the realized returns on the newest and old sorts equal:

$$R_{X,(t),t+1} = E(R_{X,(t),t+1}) + F_{X,t+1} + \epsilon_{X,(t),t+1} \text{ and} \tag{13}$$

$$R_{X,(t-s),t+1} = E(R_{X,(t-s),t+1}) + Var(X_{i,t}^P) \times \rho^s \times F_{X,t+1} + \epsilon_{X,(t-s),t+1}. \tag{14}$$

[15] This choice of c_{F_X} pins down the R^2 in the regression of Eq. (10): $0.285^2/0.6^2 = 0.23$.

[16] These numbers follow from solving $0.048 = \sqrt{\lambda_M^2 \times Var(\beta_{i,M}) + \lambda_P^2 \times Var\left(X_{i,t}^P\right) + \lambda_T^2 \times Var\left(X_{i,t}^T\right)}$.

[17] Time-variation in risk premia should not change our results because we are comparing the return of old and new sorts at the same point in time. As it is standard in the literature, we assume that the compensation for a unit loading on a characteristic (or one of its components) is independent of the time of portfolio formation.

Since $\epsilon_{X,(t),t+1}$ converges to zero as the number of stocks grows large, we find that the newest sort is the X-factor. In this case, the beta in a regression of the older sorts on the newest sort (see Eq. (2)) equals the characteristic spread that remains s years after portfolio formation as a fraction of the characteristic spread at portfolio formation: $\frac{2\times 1.755 \times Var(X_{i,t}^P)\times\rho^s}{2\times 1.755} = Var(X_{i,t}^P)\rho^s$. The alpha in this regression equals:

$$\begin{aligned}\alpha_s^u &= E(R_{X,(t-s),t+1}) - \beta_s^u E(R_{X,(t),t+1}) \\ &= 2 \times 1.755 \times \left(Var(X_{i,t}^P) - Var(X_{i,t}^P)^2\right) \times \rho^s \times (\lambda_P - \lambda_T).\end{aligned} \tag{15}$$

This alpha is zero under the null of the standard characteristic-based model of expected returns ($\lambda_P = \lambda_T = 0$).[18] Old sorts are not priced by the newest sort under the two alternatives. Under the first alternative ($\lambda_P = 0,\ \lambda_T > 0$), we find that $\alpha_s^u < 0$, because old sorts contain relatively more unpriced exposure to the X-factor than the newest sort. This follows from the fact that exposure decays from time $t - s$ to t by an amount $Var(X_{i,t}^P)\rho^s$, whereas expected returns decay immediately with the transitory component. This implication is consistent with Liu et al. (2021), who argue that mispricing is more transitory than risk exposure. Under the second alternative ($\lambda_P > 0,\ \lambda_T = 0$), we find that $\alpha_s^u > 0$, because old sorts contain relatively less unpriced exposure to the X-factor than the newest sort. In this case, exposure still decays by an amount $Var(X_{i,t}^P)\rho^s$, but expected returns decay more slowly with the persistent component, that is, from $t - s$ to t by an amount ρ^s. Quantitatively, assuming $Var(X_{i,t}^P) = 0.5$ and focusing on the three-year horizon, we show that the model generates alphas that are economically meaningful at −2.42% (first alternative) and 2.42% (second alternative).

Finally, our model implies a simple test of the idea that all characteristics are created equal except for their persistence. In particular, given a choice for $Var(X_{i,t}^P)$, the correlation between alpha and characteristic-persistence ρ is 1 if $\lambda_P > \lambda_T$ and −1 if $\lambda_P < \lambda_T$. In other words, a low (in absolute value) correlation between alpha and persistence in the data is evidence against the idea that the relative compensation for the persistent and transitory component is the same for all characteristics.

4.5. Richer specification

We now turn to a richer specification that accounts for important properties of empirical data, closely following Keloharju et al. (2021). We simulate $B = 10{,}000$ samples of 570 months. In each month t, the size of the simulated cross-section is equal to the data. We continue to generate X^T and X^P annually, but we repeat the annual observations for 12 continuous months, similar to the approach of Fama and French (1992). We draw the relative contribution of X^T and X^P to the variance of X from a uniform distribution ($U[0.25, 75]$). We do so because this relative contribution is likely to vary across characteristics in the data. To capture the conditional covariance structure of the market ($R_{M,t}$) and the characteristic-based factor ($F_{X,t}$) returns, we estimate their empirical covariance matrix Σ_{F_t} $(= Var([R_{M,t}, F_{X,t}]))$ using a rolling window centered at month t (from three years before to three years after). We then draw the two de-meaned factor returns in each month t from the multivariate normal distribution $\mathcal{N}(0_2, \Sigma_{F_t})$.

We specify the idiosyncratic component of realized returns, $\epsilon_{i,t}$ in Eq. (9), to have the same factor structure as idiosyncratic returns in the data (see Section 4.2 in Keloharju et al., 2021). In each month t, we use a rolling window centered at t to run a time-series regression for all N_t stocks on the market and a long-short book-to-market portfolio. We standardize the month t residuals cross-sectionally and denote this N_t-vector ε_t. Over the same rolling window, we estimate the $N_t \times N_t$ conditional covariance matrix of stock returns, Σ_{N_t}. We follow Higham (2002) and replace Σ_{N_t} with its nearest positive definite

[18] We show in Online Appendix B that popular structural explanations of characteristics-based return predictability, such as Gomes et al. (2003) and Zhang (2005), are consistent with this null.

Table 2

Old-versus-new sorts and stocks in model simulations. This table reports results from 10,000 simulations of the model presented in Section 4. In Panel A, we present the 5th, 50th, and 95th percentiles of the simulated distribution of the alpha from a regression of a three-year old sort on the newest sort (see Eq. (15)). In Panel B, we present the same distribution for the average return of old-minus-new stocks (as defined in Section 4.6). We also report in each panel (i) the analytical estimate, derived from a setting with IID idiosyncratic return shocks, and (ii) the distribution of the estimate in the data, derived from the large set of 56 characteristics. For the model simulations, we consider three cases. Under the null of a standard characteristic-based model of expected returns, the compensation for the transitory and persistent component of the characteristic is equal: $\lambda_P = \lambda_T = 3.17\%$. Under the first alternative, only the transitory component is compensated in expected returns: $\lambda_P = 0,\ \lambda_T = -4.49\%$. Under the second alternative, only the persistent component is compensated in expected returns: $\lambda_P = 4.49\%,\ \lambda_T = 0$.

		5^{th}	50^{th}	95^{th}	Analytical
		Panel A: Alpha of old-versus-new sort			
Null:	$\lambda_P = \lambda_T > 0$	-2.46%	0.02%	2.53%	0.00%
Only transitory:	$\lambda_T > \lambda_P = 0$	-4.73%	-2.25%	0.31%	-2.42%
Only persistent:	$\lambda_P > \lambda_T = 0$	-0.27%	2.24%	4.79%	2.42%
Empirical		-4.67%	-0.76%	3.02%	
		Panel B: Average return of old-minus-new stocks			
Null:	$\lambda_P = \lambda_T > 0$	-1.58%	0.52%	2.57%	0.54%
Only transitory:	$\lambda_T > \lambda_P = 0$	-3.89%	-1.79%	0.34%	-1.89%
Only persistent:	$\lambda_P > \lambda_T = 0$	0.42%	2.50%	4.57%	2.65%
Empirical		-4.86%	-0.22%	5.90%	

matrix, $\hat{\Sigma}_{N_t}$. Finally, we generate the vector of month-t idiosyncratic returns, ϵ_t, by randomizing the elements of ε_t and post-multiplying it with the Cholesky factor of $\hat{\Sigma}_{N_t}$.

In Panel A of Table 2, we report the 5th, 50th, and 95th percentiles of the simulated distribution of old-versus-new alphas at the three-year horizon (see Eq. (2)). Under the null and the two alternatives, the center of the alpha-distribution is close to the analytical value from our simple model, which is reported in the last column. The width of the alpha-distribution also contains important information. For instance, under the null that the persistent and transitory components capture the same expected return compensation, one is unlikely to find alphas larger than about 2.5% per year in absolute magnitude (the 5th and 95th percentile equal −2.46% and 2.53%, respectively). In contrast, under the alternative that only the transitory component is compensated in expected returns, one is unlikely to find large positive alphas (the 95th percentile equals 0.31%). In turn, under the alternative that only the persistent component is compensated, one is unlikely to find large negative alphas (the 5th percentile equals −0.27%). In Section 5, we analyze how each of these distributions compares to the empirical distribution of alphas, which we obtain by estimating alphas for a large set of characteristics.

Note that using our approach the attrition rate of the simulated portfolios after formation equals the rate at which stocks exit the CRSP file in the data. As a result, approximately one-third of the stocks have been dropped from the simulated high and low portfolios five years after portfolio formation. We show in Table OA.2 that the simulated distribution (of old-versus-new alphas as well as other metrics introduced below) is robust to imposing a higher attrition rate of 50% for the high portfolio. This attrition rate is consistent with the stocks in the highest book-to-market decile in the data.[19] Furthermore, even though characteristics predict returns quite persistently in our simulations, we show in Online Appendix C that our rich specification does not generate counterfactually strong cross-sectional return predictability.

[19] Attrition is imposed randomly in our simulation. While outside the scope of our model, it is an interesting question what drives attrition in the data (bankruptcy, M&A, going private or some other reason for delisting) and how each of these drivers impacts characteristic-based return predictability. We thank the Associate Editor for pointing this out.

4.6. Old-minus-new stocks

Before turning to the data, we note that an advantage of our model setup is that we can easily analyze alternative characteristic-based trading strategies. One strategy of particular interest decomposes the return of the newest sort into the return coming from new and old stocks. This decomposition is of interest because real-world characteristic-based investment strategies have drawn inspiration from evidence based on the newest sorts but, as we will see below, it is more profitable to trade only one of these two subsets of stocks if one of our alternative hypotheses holds. We define old stocks as those stocks in the high (low) portfolio at time t that already had a relatively high (low) value of the characteristic three years ago. New stocks in the high (low) portfolio, instead, have seen a large increase (decrease) in the value of the characteristic. Formally, our approach entails a dependent double sort into deciles sorted on X_t, and within the high and low decile, into two portfolios split at the (within-portfolio) median of X_{t-36}.

We see in Panel B of Table 2 that the old-minus-new stock difference in expected returns behaves similarly to the alpha of old-versus-new stocks. The median old-minus-new difference is small at 0.52% under the null, is large and negative at -1.79% assuming only the transitory component is priced, and is large and positive at 2.50% assuming only the persistent component is priced. These medians are close to the analytical estimates (reported in the last column) that follow from combining the relative loadings of old and new stocks on X^P and X^T with the compensation for the two components. Assuming $Var(X^P_{i,t}) = 0.5$, these loadings equal 0.59 and −0.42, respectively. Given that old (new) stocks load relatively more strongly on X^P (X^T), it follows naturally that the old-minus-new stock difference changes sign between the two alternative hypotheses. Note also that these relative loadings imply that the old-minus-new stock strategy does not load strongly on the total characteristic X, which sums X^P and X^T. Thus, under the null that expected return variation is driven solely by X, it would be surprising if such strategies obtain a large and significant return in the data.

5. Testing the model

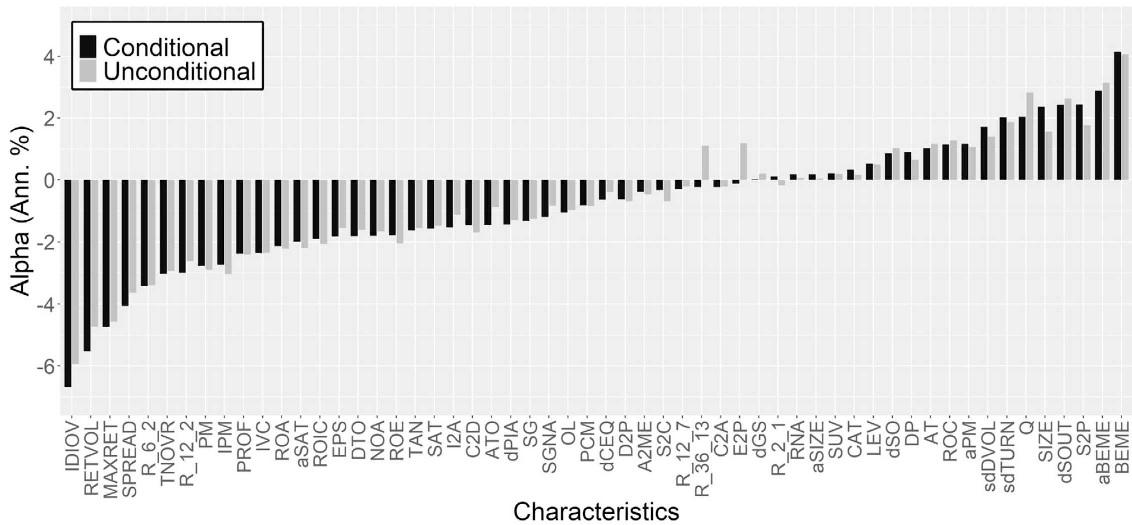
In this section, we generalize the results from Section 3 to the full set of 56 characteristics and interpret these results through the lens of our model.

5.1. Alphas of old-versus-new sorts

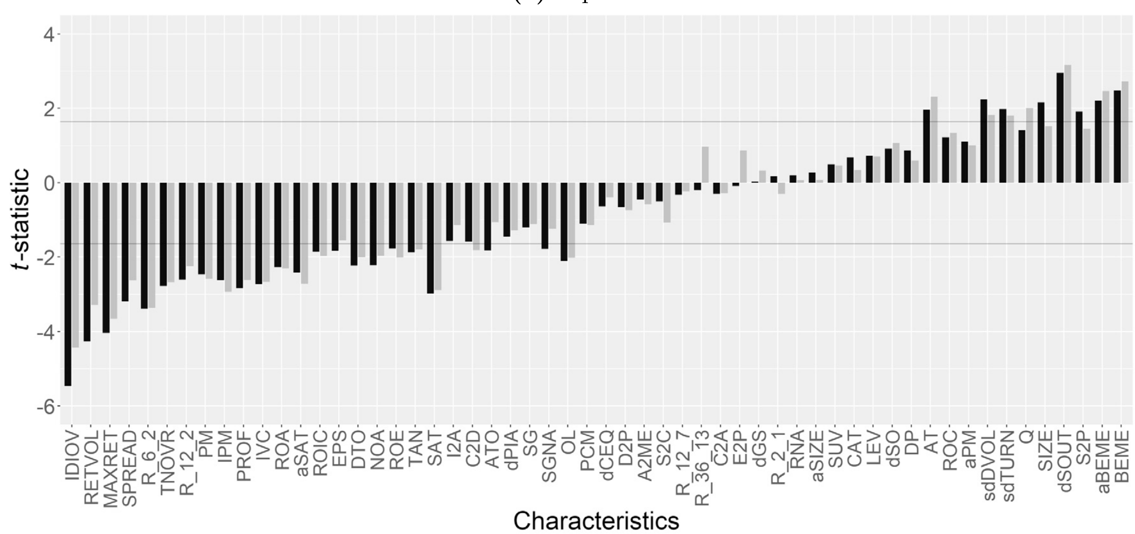
To start, we present both unconditional and conditional alphas (see Eqs. (2) and (3)) in Fig. 2. To facilitate interpretation, we sort the characteristics from left to right on the conditional alphas and, to see the big picture, we focus on a strategy that averages the returns from one to five years after portfolio formation, denoted $R_{X,(t-60:t-12),t+1}$. We present conditional alphas for each individual horizon $s = 12, 24, 36, 48, 60$ in Figure OA.1 of the Online Appendix.[20]

In Panel A of Fig. 2, we see that the empirical distribution of alphas is wide, with conditional (unconditional) alphas ranging from −6.69% to 4.14% (−5.94% to 4.06%). For at least half of the characteristics, these alphas are significant at the 10%-level. In the conditional specification, for instance, the alpha is negative and significant at the 10%-level for 23 characteristics. Among the largest negative alphas, we find a number of characteristics related to (idiosyncratic) return volatility, momentum, and profitability. The alpha in this specification is positive and significant for eight characteristics, such as share issuance, value (broadly defined), and illiquidity. In Figure OA.1, we see that there is some

[20] Focusing on the average is conservative: we find even larger increases in Sharpe ratio when we optimally choose one of the five older sorts to be included in a portfolio with the newest sort. The horizon s for which this maximum Sharpe ratio is obtained varies across characteristics.



(a) Alphas



(b) t-statistics

Fig. 2. Alphas of old-versus-new sorts This figure presents the unconditional and conditional alphas (α^u and α^c in Panel A and associated White (1980) heteroskedasticity-consistent t-statistics in Panel B) of the old sorts with respect to the newest sort for 56 characteristics. We report this alpha for a single combination of five old sorts: $R_{X,(t-60:t-12),t+1} = 1/5(R_{X,(t-12),t+1} + R_{X,(t-24),t+1} + \ldots + R_{X,(t-60),t+1})$, such that it represents the abnormal average return from one to five years after portfolio formation. To facilitate interpretation, we sort the characteristics from low to high α^c. The sample period runs from July 1972 through December 2019.

within-characteristic variation across horizons, but alphas are roughly increasing from left to right at all horizons.

In short, we find both negative and positive old-versus-new alphas that are large and significant. This finding is robust along various dimensions. First, we show in Figure OA.2 that alphas are largely unaffected when we additionally control for exposure to the market. We study a broader set of benchmark factors in the next section. Figure OA.3 shows similar alphas when we split our sample into two halves. Figure OA.4 and OA.5, respectively, show that alphas are not driven by a small set of extreme returns in NBER recessions nor exclusively by periods of high sentiment. Figure OA.6 shows virtually identical alphas when we correct for survivorship bias. While the return of the old sort, $R_{X,(t-s),t+1}$, conditions on firm survival from $t-s$ to t, the return of the newest sort, $R_{X,(t),t+1}$, does not. For these survivorship bias-corrected alphas, we calculate the return of the newest sort using only those stocks that were already in the CRSP file at $t-s$.

Further, we see in Fig. 3 that these alphas translate to large increases in Sharpe ratio from optimally combining the older sort ($R_{X,(t-60:t-12),t+1}$) with the newest sort ($R_{X,(t),t+1}$). The V-shaped pattern indicates that for characteristics where the alpha is large in absolute magnitude, the increase in Sharpe ratio also tends to be large. For 32 (15) out of 56 characteristics, the absolute increase in Sharpe ratio is over 0.10 (0.20) in the conditional specification. For 18 (13) of these characteristics, this number translates to a relative increase in Sharpe ratio of over 100%. Thus, combining new and old sorts considerably improves investment opportunities. This insight is useful for investors, like mutual funds, that for various reasons (e.g., clientele and menu effects or specialization) may target only one (or a few) characteristic themes, such as value. Since these themes are usually defined broadly, how best to construct the characteristic has remained an open question.

5.2. *Alphas in the data versus the model*

Benchmarking these results to our model (cf., Panel A of Table 2), the following stands out. First, under the null of a standard characteristic-based model of expected returns ($\lambda_P = \lambda_T$), one is un-

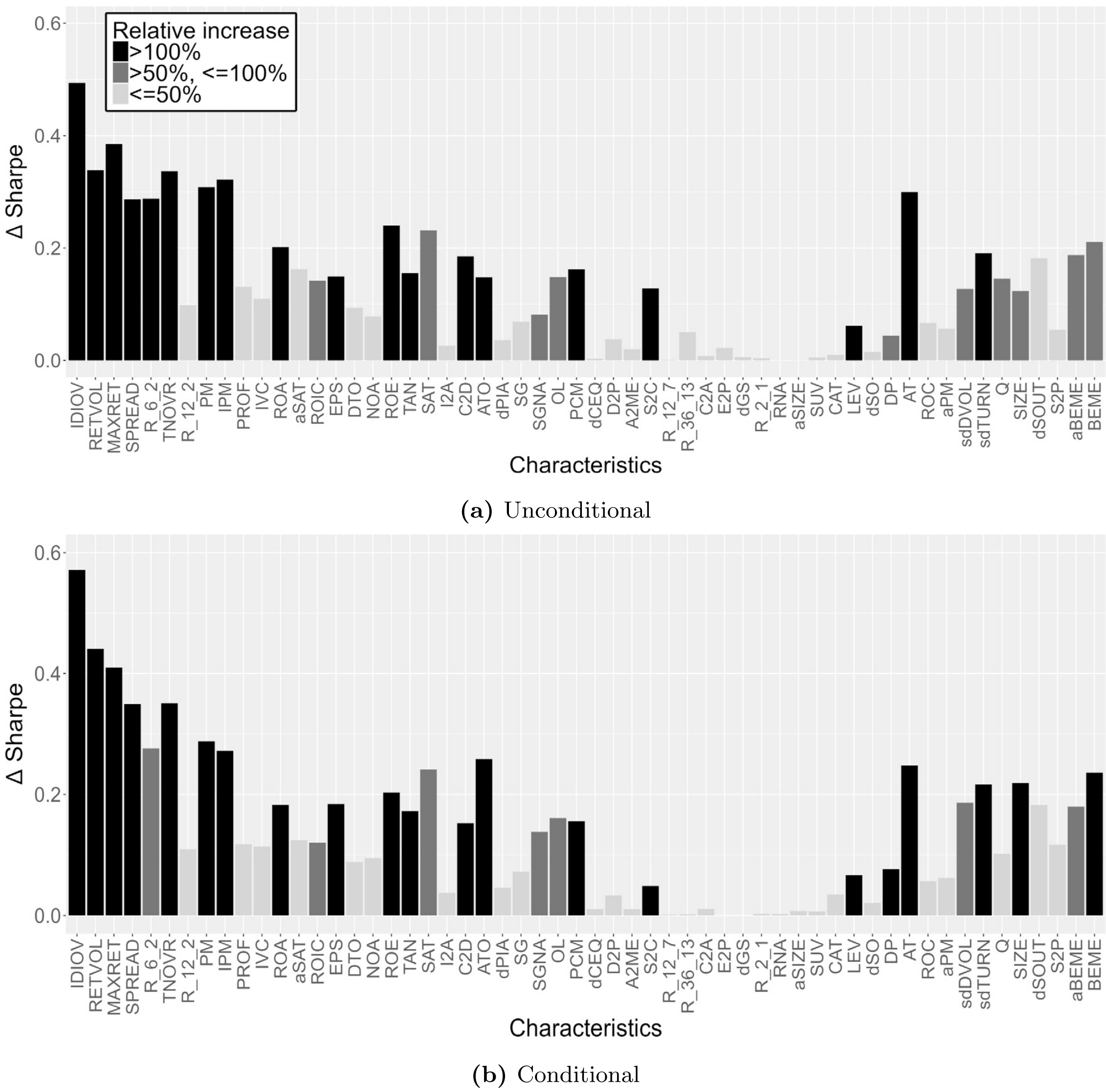


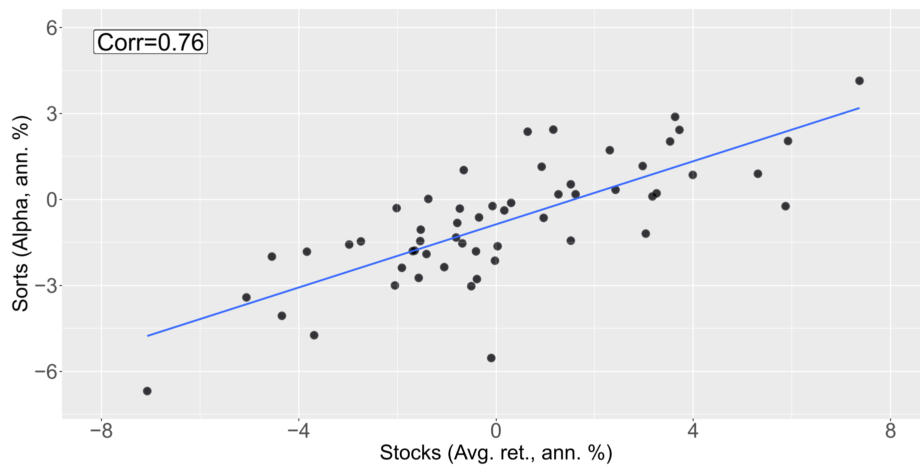
Fig. 3. Increases in Sharpe ratio. This figure presents the maximum improvement in Sharpe ratio from combining the newest sort, $R_{X,t,t+1}$, with a single combination of five old sorts, $R_{X,(t-60:t-12),t+1}$. The improvement Δ Sharpe = Max. Sharpe($R_{X,(t-60:t-12),t+1}$, $R_{X,(t),t+1}$) - Sharpe($R_{X,(t),t+1}$). We color code these improvements to highlight the relative increase in Sharpe ratio: Δ Sharpe / Sharpe($R_{X,(t),t+1}$)-1. To calculate the maximum Sharpe ratio, we optimally combine the newest sort with either the old sort (Panel A) or the conditionally hedged return of the old sort (Panel B, see Eq. (3)). The sample period runs from July 1972 through December 2019.

likely to find negative alphas as large and as plentiful as what we find in the data. For instance, the 5th percentile of alphas in the data (−4.67%) is well below the 5th percentile under the null (−2.46%), but close to the 5th percentile under the first alternative (−4.73%, when $\lambda_P = 0$ and $\lambda_T > 0$). This evidence is consistent with Keloharju et al. (2021) and the idea that the transitory component drives return predictability for the average characteristic. However, under this alternative, one is unlikely to find positive alphas as large and as plentiful as what we find in the data. For instance, the 95th percentile of alphas in the data (3.02%) is well above the 95th percentile (0.31%) under the alternative that only the transitory component is priced. Our evidence thus suggests that there is an important subset of characteristics for which the persistent component must be priced. In fact, given that the 95th percentile under the null (2.53%) is smaller than what we find in the data, we conclude that it is most likely that the right tail of alphas in the data is generated by characteristics where the compensation for the persistent component is larger than the compensation for the transitory component. We conclude that the compensation for the persistent and transitory components must vary significantly across characteristics.

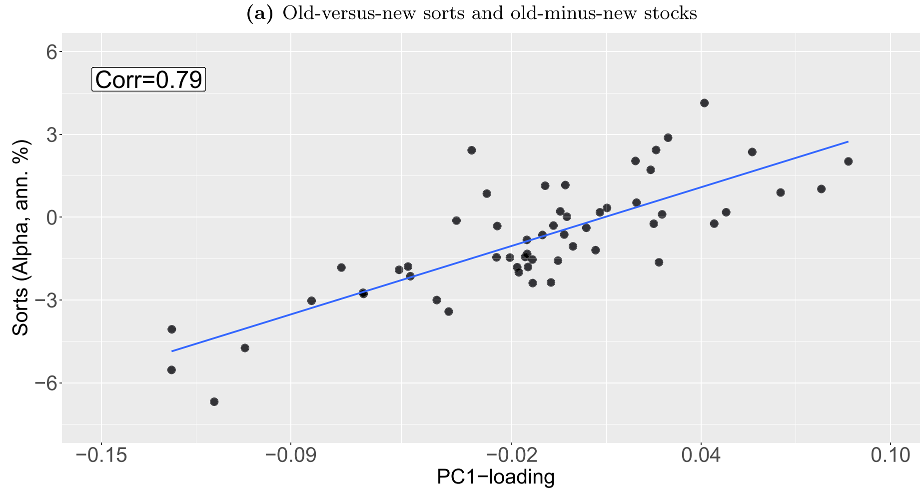
Consistent with this conclusion, Figure OA.7 shows that the correlation across characteristics between alphas and persistence (measured as the beta in the regression of the old sort on the newest sort) is small. As discussed in Section 4.4, if the relative compensation of persistent and transitory components were fixed across characteristics, the model (cf., Eq. (15)) would predict a large correlation between persistence and alphas. Finally, in Panel A of Figure OA.8 we show that, consistent with the model, there is a large correlation of 0.89 between the characteristic spread that remains three years after portfolio formation ($\frac{X_{H-L,(t-36),t}}{X_{H-L,(t),t}}$) and the beta obtained from regressing $R_{X,(t-36),t+1}$ on $R_{X,(t),t+1}$. As a result, Panel B of this figure shows that alphas at the three-year horizon are largely unchanged when we use this characteristic spread to define an alternative old-versus-new strategy.

5.3. Old-minus-new stocks

Last, we study the decomposition in old and new stocks introduced in Section 4.6:



(a) Old-versus-new sorts and old-minus-new stocks



(b) Old-versus-new sorts and loadings of PC1

Fig. 4. Old-versus-new sorts, old-minus-new stocks, and PC1 loadings. Panel A of the figure presents a scatter plot showing the correlation across characteristics between the conditional alpha of old-versus-new sorts (see also Fig. 2) and the average return of the old-minus-new stock strategy (defined in Section 5.3). Panel B presents a scatter plot showing the correlation across characteristics between the conditional alpha of old-versus-new sorts and the loadings of the first principal component of the newest sorts (PC1, see Section 6.1). The sample period runs from July 1972 through December 2019.

$$R^{New}_{X,(t),t+1} = R^{High,New}_{X,(t),t+1} - R^{Low,New}_{X,(t),t+1} \text{ and} \quad (16)$$

$$R^{Old}_{X,(t),t+1} = R^{High,Old}_{X,(t),t+1} - R^{Low,Old}_{X,(t),t+1}. \quad (17)$$

In particular, $R^{High,Old}_{X,(t),t+1}$ is the return of a strategy that is long a value-weighted portfolio of the subset of stocks in the high decile portfolio at time t with lagged characteristic X_{t-36} above the within-portfolio median of X_{t-36}. $R^{Low,Old}_{X,(t),t+1}$ is defined analogously. The return for new stocks, $R^{New}_{X,(t),t+1}$, uses all remaining stocks in the high and low portfolios at time t. This decomposition ensures that the new and old portfolios contain (roughly) the same number of stocks for all characteristics.

Note that Keloharju et al. (2021) decompose a firm characteristic into its historical average (the permanent component) and a residual (the transitory component). The key difference is that we sort the stocks within the high and low decile portfolio into a new and old group, whereas Keloharju et al. (2021) sort the whole cross-section of stocks using their two components. Thus, our decomposition of stocks within the high and low decile portfolio has the potential to uncover new information about how changes in characteristics predict returns. To see this by example, consider book-to-market. This characteristic generates a positive old-minus-new stock return difference, $R^{Old}_{BM,(t),t+1} - R^{New}_{BM,(t),t+1}$, of 7.37% ($t$-stat = 2.48). This difference is large even over the last 15 years of our sample at 8.52%, which is a period when the newest sort on book-to-market generates a negative return of −7.80%. This finding indicates that past changes in book-to-market predict returns with a negative sign among stocks that are in the extreme book-to-market portfolios today.[21]

We plot the average return of the old-minus-new stock strategy against the old-versus-new alpha in Panel A of Fig. 4. We see that the two are highly correlated across characteristics at 0.76, as our model predicts. In Panel B of Table 2, we see that both the left and right tail of the empirical distribution of old-minus-new stock returns is much wider than what is expected under the null ($\lambda_P = \lambda_T$). The left tail of these average returns is generated by characteristics like idiosyncratic volatility and momentum, whereas the right tail is generated by a large variety of value characteristics. Looking at the distribution of the average returns under the two alternatives, we conclude that the characteristics in the left tail are most consistent with the assumption that $\lambda_P = 0, \lambda_T > 0$, whereas the right tail is most consistent with the opposite assumption

[21] In contrast, Gerakos and Linnainmaa (2018) find that changes in book-to-market predict returns with a positive sign in the full cross-section of stocks, a result that we replicate in our data. Our finding is consistent with the idea that new stocks in the high (low) book-to-market portfolio have experienced relatively low (high) returns, a trend that continues in the next month. Thus, one can think of a book-to-market strategy using old stocks as a simple, alternative way to profitably combine book-to-market and momentum signals (see, also, Asness et al., 2013).

Table 3

Principal components of new and old sorts. Panel A of this table reports the intercept (α) and associated t-statistic (based on White (1980) heteroskedasticity-consistent standard errors) from regressing the first three principal components of old sorts ($\gamma'_{(t-s),z} R_{X,(t-s),t+1}$, $z = 1,2,3$ and $s > 0$) on a statistical factor model containing the first three principal components of new sorts ($R_{PC3,(t),t+1} = [\gamma_{(t),1}, \gamma_{(t),2}, \gamma_{(t),3}]' R_{X,(t),t+1}$). The last column of this panel reports the associated F-statistics and p-values from a (conditional heteroskedasticity-consistent) GRS test. Panel B is identical to Panel A except that we now apply the loadings of the principal components of the newest sorts to each of the old sorts. Hence, the test asset returns are defined as: $\gamma'_{(t),z} R_{X,(t-s),t+1}$. Panel C reports summary statistics for the first principal component of new and old sorts reported in Panel A. The sample period runs from July 1972 through December 2019.

	PC1 ($z=1$)		PC2 ($z=2$)		PC3 ($z=3$)		GRS test	
Horizon s	α	t-stat	α	t-stat	α	t-stat	F-stat	p-val
	Panel A: PCs of old sorts on statistical factor model ($\gamma'_{(t-s),z} R_{X,(t-s),t+1}$ on $R_{PC3,(t),t+1}$)							
12	3.11	(3.61)	0.22	(0.31)	1.29	(0.95)	4.42	0.00
24	5.03	(4.89)	0.44	(0.40)	0.36	(0.27)	8.41	0.00
36	5.27	(5.08)	-0.58	(-0.46)	0.62	(0.45)	8.92	0.00
48	5.72	(5.12)	-1.31	(-1.02)	0.95	(0.71)	9.87	0.00
60	5.45	(4.84)	-1.08	(-0.79)	0.14	(0.11)	8.92	0.00
	Panel B: PCs of old sorts with loadings fixed ($\gamma'_{(t),z} R_{X,(t-s),t+1}$ on $R_{PC3,(t),t+1}$)							
12	2.89	(3.49)	0.17	(0.25)	-2.43	(-2.00)	5.50	0.00
24	4.45	(4.41)	1.51	(1.54)	-1.35	(-1.10)	8.05	0.00
36	4.84	(4.41)	1.57	(1.30)	0.18	(0.16)	8.12	0.00
48	5.75	(5.18)	0.79	(0.65)	-1.19	(-1.10)	10.22	0.00
60	5.36	(4.72)	1.31	(1.02)	-1.93	(-1.64)	9.41	0.00

	Panel C: Summary statistics for first principal components ($\gamma'_{(t-s),z} R_{X,(t-s),t+1}$)					
	Avg. Ret.	Correlations				
		0	12	24	36	48
0	-3.76					
12	0.11	0.91				
24	1.90	0.86	0.98			
36	1.94	0.81	0.95	0.98		
48	2.10	0.78	0.92	0.95	0.98	
60	1.70	0.75	0.90	0.93	0.96	0.98

that $\lambda_P > 0, \lambda_T = 0$. Once more, we conclude that the relative compensation for persistent and transitory components must vary in both magnitude and sign across characteristics.

6. Benchmark factor models and old-versus-new strategies

Our results so far are surprising from the point of view that persistent and transitory components of characteristics contribute equally to expected returns. Whether our results are surprising from the standpoint of benchmark factor models is a question we answer in this section. In particular, are old and new sorts or stocks exposed differently to the factors in benchmark models (beyond the single characteristic-based factor featured in the model of Section 4.1)? Can this exposure explain the average return of the old-versus-new strategies?

6.1. Principal components of old and new sorts

To answer these questions, we start by extracting three principal components (PC) at each horizon $s = 0, 12, \ldots, 60$ after portfolio formation. In this way, we focus on the dominant sources of variation in our panel of characteristic-sorted portfolio returns and increase the power of our tests.[22] At each horizon, the three PCs explain about 60% of the total variation in returns. We rescale each PC so that the sum of absolute weights equals two. In the spirit of previous literature, we treat the PCs extracted from the newest sorts (with returns $R_{X,(t),t+1}$) as a statistical factor model.

In Table 3 we ask whether this statistical factor model can price the PCs extracted from older sorts using regressions of the form:

$$\gamma'_{(t-s),z} R_{X,(t-s),t+1} = \alpha_{s,z} + \beta'_{s,z} R_{PC3,(t),t+1} + \epsilon_{(t-s),z,t+1}, \tag{18}$$

where $z = 1,2,3$ and $R_{PC3,(t),t+1} = [\gamma_{(t),1}, \gamma_{(t),2}, \gamma_{(t),3}]' R_{X,(t),t+1}$. The Gibbons et al. (1989, GRS) tests presented in the last column of Panel A strongly reject the statistical factor model with p-values below 0.0012 at all horizons. Looking at the alphas of the individual PCs, we see that the rejection is driven by the first PC (PC1). This portfolio provides a large and significant alpha ranging from 3.11 (t-stat = 3.61) at the one-year horizon to 5.72% (t-stat = 5.12) at the four-year horizon. Neither PC2 nor PC3 generates a significant alpha at any of the horizons we consider.

These conclusions do not result from the loadings of PC1 changing across horizons s. Indeed, in Panel B we show similarly large alphas when we apply the loadings of PC1 extracted from the newest sorts, $\gamma_{(t),1}$, to the older sorts at each horizon s (instead of $\gamma_{(t-s),1}$). This result suggests that returns are highly correlated in the time-series for the average characteristic. The summary statistics presented in Panel C confirm this intuition. For instance, the correlation between PC1 of the newest sorts and PC1 of the three-year-old sorts is high at 0.81. At the same time, the average returns of these two strategies are wildly different at −3.76% versus 1.94%.

[22] The SDF can be suitably approximated using only a few PC factors when test assets do not each represent an independent source of priced risk (e.g., Kozak et al., 2020; Kelly et al., 2019; Haddad et al., 2020; Lettau and Pelger, 2020). The main conclusions of this section are unchanged when we use the Lettau and Pelger (2020) risk premium PCA approach (see Table OA.7).

Table 4
Old and new sorts in benchmark factor models. This table presents average returns as well as the intercept (α), the associated t-statistic (based on White (1980) heteroskedasticity-consistent standard errors), and the factor contribution ($\beta'_{s,1}\mu_F$) from regressing the first principal component (PC1) of new and old sorts on benchmark factor models. In Panel A, we consider each horizon in isolation and define the return of the test asset as $\gamma'_{(t),1}R_{X,(t-s),t+1}$ for $s \geq 0$. In Panel B, we look at old-minus-new strategies where we define the return of the test asset as $\gamma'_{(t),1}(R_{X,(t-s),t+1} - R_{X,(t),t+1})$ for $s > 0$. The sample period runs from July 1972 through December 2019.

			CAPM			FF3M			FF5M			FF5M+MOM		
Horizon s	Avg. ret.	t-stat	α	t-stat	$\beta'_{s,1}\mu_F$	α	t-stat	$\beta'_{s,1}\mu_F$	α	t-stat	$\beta'_{s,1}\mu_F$	α	t-stat	$\beta'_{s,1}\mu_F$
	Panel A: PC1 of new and old sorts ($\gamma'_{(t),1}R_{X,(t-s),t+1}$)													
0	-3.76	(-1.06)	-10.15	(-3.55)	6.39	-9.43	(-4.82)	5.67	-2.72	(-1.68)	-1.03	0.20	(0.13)	-3.96
12	1.03	(0.33)	-4.32	(-1.67)	5.35	-3.33	(-2.05)	4.36	1.78	(1.33)	-0.75	2.14	(1.59)	-1.11
24	2.84	(0.96)	-1.95	(-0.78)	4.79	-0.55	(-0.35)	3.39	3.90	(2.81)	-1.06	3.86	(2.72)	-1.01
36	2.98	(1.07)	-1.54	(-0.64)	4.53	0.14	(0.09)	2.85	4.64	(3.45)	-1.65	4.54	(3.27)	-1.55
48	3.87	(1.44)	-0.54	(-0.23)	4.41	1.16	(0.80)	2.70	5.59	(4.33)	-1.72	5.58	(4.22)	-1.71
60	3.46	(1.34)	-0.80	(-0.36)	4.26	0.66	(0.46)	2.80	4.83	(3.75)	-1.36	4.93	(3.65)	-1.47
	Panel B: PC1 of old-minus-new sorts ($\gamma'_{(t),1}(R_{X,(t-s),t+1} - R_{X,(t),t+1})$)													
12	4.79	(4.11)	5.83	(5.20)	-1.05	6.10	(5.67)	-1.31	4.50	(3.83)	0.28	1.93	(1.85)	2.85
24	6.60	(4.51)	8.20	(6.15)	-1.60	8.88	(6.91)	-2.28	6.62	(4.67)	-0.02	3.65	(2.92)	2.95
36	6.74	(4.11)	8.60	(5.74)	-1.86	9.57	(6.60)	-2.83	7.36	(4.80)	-0.62	4.33	(3.13)	2.41
48	7.62	(4.50)	9.61	(6.28)	-1.99	10.59	(7.35)	-2.97	8.31	(5.45)	-0.69	5.38	(3.73)	2.25
60	7.22	(4.10)	9.35	(5.96)	-2.13	10.09	(6.85)	-2.87	7.55	(5.04)	-0.33	4.73	(3.24)	2.49

These results nicely summarize the challenge that any asset pricing model will face, which is to price two highly correlated returns that are separated by a large difference in average return. This challenge is particularly hard when we aggregate characteristics using the loadings of PC1, because these loadings are highly correlated with the alpha between old and new sorts ($corr = 0.79$; see Panel B of Fig. 4). Interestingly, this is the case even though the loadings are determined only by the (co-)variances of sorts at a single horizon.

6.2. Do benchmark factor models price old-versus-new sorts?

To answer this question, we consider the single-factor CAPM (Sharpe (1964), Lintner (1965), Mossin (1966)); the three-factor model of Fama and French (1993, FF3M); the five-factor model of Fama and French (2015, FF5M); and, a six-factor model including the factors in the FF5M and momentum (FF5M+MOM).[23] We substitute the factors in each benchmark model on the right-hand side of Eq. (18). For consistency, we henceforth apply the same loadings (of PC1 extracted from the newest sorts, $\gamma_{(t),1}$) at each horizon after portfolio formation.[24]

In Panel A of Table 4, we see that the benchmark models unanimously struggle to jointly price PC1 of new and old sorts. There is a clear trade-off between small and big models, however. The larger benchmark models do not price the return of PC1 at longer horizons after portfolio formation. For instance, the annualized alpha in the FF5M+MOM is larger than 3.86% at all horizons $s \geq 24$ (with t-stat ≥ 2.7). These larger models perform better on the PC1 of the newest sorts ($s = 0$). For instance, the FF5M+MOM alpha for this test asset is small and insignificant at 0.20 (t-stat $= 0.13$). In contrast, the smaller benchmark models price PC1 of old sorts at most horizons after portfolio formation $s \geq 24$, but fail completely for the newest sorts. For instance, the alpha at $s = 0$ is statistically and economically large at about -10.00% in the CAPM and FF3M.

In Panel A, we also report the factor contributions ($\beta'_{s,1}\mu_F$) and show that all models imply that returns decrease as time passes after portfolio formation.[25] In contrast, realized average returns increase as time passes after portfolio formation. Thus, differential exposure to benchmark factors cannot explain the difference in average returns between new and old sorts. This conclusion is easily confirmed in Panel B, where we present the alpha of an old-versus-new PC1 strategy ($\gamma'_{(t),1}(R_{X,(t-s),t+1} - R_{X,(t),t+1})$). For all models and at all horizons s, this strategy provides a large and significant alpha (with t-statistic well above 3 in all but two cases).

In Panel A of Table 5, we show similarly large and significant alphas when we apply the PC1 loadings to the old-versus-new strategies that control for persistence at the characteristic level. Focusing on the three-year horizon, we thus define the return of the test asset as $\gamma'_{(t),1}(R_{X,(t-36),t+1} - \beta_{36}R_{X,(t),t+1})$. Under the null of our model, returns decay at the appropriate speed and alphas will be zero relative to a factor model that accurately prices the newest sorts (like FF5M and FF5M+MOM do empirically). Given that we do find large alphas in all benchmark models, it is useful to analyze the contribution of characteristics with returns that decay too fast versus too slow.

For this analysis, we split the PC1 loadings, $\gamma_{(t),1}$, into those that are positive versus negative and define the following two returns:

$$R^{Slow}_{(t-36),t+1} = \gamma^{>0,\prime}_{(t),1}(R_{X,(t-36),t+1} - \beta_{36}R_{X,(t),t+1}) \text{ and} \tag{19}$$

$$R^{Fast}_{(t-36),t+1} = \gamma^{\leq 0,\prime}_{(t),1}(R_{X,(t-36),t+1} - \beta_{36}R_{X,(t),t+1}). \tag{20}$$

Because the PC1 loadings line up quite well with alphas between old and new sorts (see Panel B of Fig. 4), $R^{Slow}_{(t-36),t+1}$ and $R^{Fast}_{(t-36),t+1}$ are driven mostly by characteristics where returns decay too slow and too fast, respectively. These aggregated returns are particularly suited to test the null that returns decay at the same speed as characteristics and are captured by factor models that accurately price the newest sorts.

Panels B and C of Table 5 present alphas from regressing these test asset returns on the benchmark factor models. In short, we see that all models struggle to price both subsets of characteristics. For instance, for the characteristics with returns that decay too slow, the alpha (in

[23] We present similar results for the models of Hou et al. (2015), Frazzini and Pedersen (2014), Daniel et al. (2020b), Stambaugh and Yuan (2016), and Daniel et al. (2020a) in Table OA.8. In short, none of these models prices both old and new sorts.

[24] In Figure OA.9, we show that these loadings are robust over time. The correlation, across characteristics, between PC1 loadings extracted from returns over the first and second half of our sample period equals 0.70.

[25] We also note that the total factor contribution is relatively large for the CAPM. Specifically, for the newest sort, the correlation across characteristics between the unconditional market beta and the loading of PC1 is large at 0.96 (see Figure OA.10). We show in Online Appendix A how a large market loading can be generated in the model making minimal assumptions inspired by the low beta anomaly.

Table 5

Old-versus-new sorts when returns decay fast versus slow. This table presents average returns and the intercept (α) with associated t-statistic (based on White (1980) heteroskedasticity-consistent standard errors) from standard factor regressions. In Panel A, the test asset is based on the return of the old-versus-new sort strategy that is the focus in our model and aggregates across characteristics using the PC1 loadings: $\gamma'_{(t),1}(R_{X,(t-36),t+1} - \beta_{36} R_{X,(t),t+1})$. We focus on the three-year horizon and present results for both an unconditional and conditional specification of β_{36} (β^u_{36} from Eq. (2) and $\beta^c_{36,t}$ from Eq. (3)). In Panel B and C, we decompose the test asset returns into the part coming from the characteristics with returns that decay too fast versus too slow, which are those characteristics on which PC1 loads with a negative versus positive sign (see Eqs. (19) and (20)). Because the absolute sum of negative loadings is about twice as large as of positive loadings, we rescale the latter such that the two test asset returns have the same volatility. The sample period runs from July 1972 through December 2019.

				CAPM		FF3M		FF5M		FF5M+MOM	
		Avg.Ret.	t-stat	α	t-stat	α	t-stat	α	t-stat	α	t-stat
		Panel A: PC1 of old-versus-new strategies ($\gamma'_{(t),1}(R_{X,(t-36),t+1} - \beta_{36} R_{X,(t),t+1})$)									
Unconditional	β^u_{36}	4.54	(4.00)	4.02	(3.57)	4.73	(4.64)	5.13	(4.70)	4.04	(3.62)
Conditional	$\beta^c_{36,t}$	4.92	(4.50)	4.46	(4.05)	4.85	(4.57)	4.63	(4.09)	3.76	(3.30)
		Panel B: Characteristics with returns that decay too fast ($R^{Fast}_{(t-36),t+1} = \gamma^{\le 0,\prime}_{(t),1}(R_{X,(t-36),t+1} - \beta_{36} R_{X,(t),t+1})$)									
Unconditional	β^u_{36}	-3.69	(-3.79)	-3.30	(-3.40)	-4.06	(-4.55)	-4.37	(-4.57)	-3.55	(-3.57)
Conditional	$\beta^c_{36,t}$	-3.99	(-4.19)	-3.67	(-3.82)	-4.07	(-4.32)	-3.86	(-3.83)	-3.29	(-3.15)
		Panel C: Characteristics with returns that decay too slow ($R^{Slow}_{(t-36),t+1} = \gamma^{>0,\prime}_{(t),1}(R_{X,(t-36),t+1} - \beta_{36} R_{X,(t),t+1})$)									
Unconditional	β^u_{36}	3.18	(3.27)	2.71	(2.77)	2.52	(2.74)	2.84	(2.96)	1.86	(1.96)
Conditional	$\beta^c_{36,t}$	3.47	(3.64)	2.93	(3.08)	2.90	(3.14)	2.91	(2.90)	1.75	(1.84)

the specification with conditional betas) is large and significant ranging from 1.75% (t = 1.84) in the FF5M+MOM to 2.93% (t = 3.08) in the CAPM. These alphas are even larger (in absolute magnitude) for the characteristics with returns that decay too fast, ranging from −4.07% (t = 4.32) in the FF3M to −3.29% (t = 3.15) in the FF5M+MOM.

From the point of view of our model, it is perhaps unsurprising that benchmark factors fail to price these two subsets of characteristics. Indeed, these factors have been added to the CAPM to help explain the cross-section of returns to new sorts. These new sorts capture the total compensation derived from loading on a characteristic. However, to price old-versus-new sorts, factor models must capture the variation in the relative compensation of persistent and transitory components. Benchmark factor models fail to do so. We show in Online Appendix D that this same conclusion follows from studying a simple empirical decomposition of firm characteristics into persistent and transitory components. We leave for future work the task of formally optimizing factor models to explain new and old sorts in more general settings. This is an important agenda because any model that prices returns at all horizons will get price levels right (Cho and Polk, 2020; van Binsbergen et al., 2023).

6.3. Do benchmark factor models price old-minus-new stocks?

We finally turn to the old and new stock decomposition. Continuing to weight the returns of the old-minus-new stock strategies using the PC1 loadings, we have the following test asset returns for the case of old stocks:

$$R^{Old}_{(t),t+1} = \gamma'_{(t),1} R^{Old}_{X,(t),t+1}, \tag{21}$$

$$R^{Old,Slow}_{(t),t+1} = \gamma^{>0,\prime}_{(t),1} R^{Old}_{X,(t),t+1}, \text{ and} \tag{22}$$

$$R^{Old,Fast}_{(t),t+1} = \gamma^{\le 0,\prime}_{(t),1} R^{Old}_{X,(t),t+1}. \tag{23}$$

We analogously define the test asset returns for new stocks.

We present average returns and alphas with respect to the benchmark factor models in Table 6. In Panel A, we see that the aggregated difference in average returns between old and new stocks is large and significant at 4.33% (t-stat = 2.74). These old-minus-new differences are similarly large and significant when we control for exposure to the benchmark factors, as alphas range from 3.18% (t-stat = 2.04) in the CAPM to 4.91 (t-stat = 3.54) in the FF5M. These results are consistent with the relative performance of old-versus-new sorts (see Table 4), both qualitatively and quantitatively, and provide additional evidence against the null that $\lambda_P = \lambda_T$.

In Panels B and C, we study the subsets of characteristics with returns that decay too slow versus too fast (Eqs. (22) and (23)). We see that the models struggle for both subsets. For the characteristics with returns that decay too fast, the alpha is negative in all models and significant in the FF3M, FF5M and FF5M+MOM (at values below −2.44%, $t < -2.07$). The alphas are similarly large (in absolute magnitude) and significant in all models for the characteristics with returns that decay too slow. In this case, the alpha ranges from 1.92% (t = 2.35) in the FF5M+MOM to 2.96% (t = 4.08) in the FF3M. In all, the relative performance of old and new stocks varies in sign across characteristics and this variation is not explained by benchmark factor models, just like we saw for the relative performance of old and new sorts.

It is interesting to note that these large old-minus-new stock differences occur even though the aggregated strategies are roughly characteristic-neutral. In Figure OA.11, we show the difference between old and new stocks in the loading on each of the 56 characteristics for the three PC1-weighted strategies. These loadings are presented as a fraction of the loading on each characteristic from a single sort.[26] Overall, we see that the old-minus-new differences in the loadings are small, both when aggregating by using all PC1 loadings and when aggregating by using the positive and negative subsets of these loadings.

[26] Taking book-to-market as an example, these fractions are calculated as follows. For each of the 56 characteristics denoted X, we calculate the time-series average of the median book-to-market ratio in the high and low portfolio among new and old stocks, denoted, for instance, $BM^{New}_{X,H}$. We then take the difference between the high and low portfolio, denoted $BM^{New}_{X,H-L}$ and $BM^{Old}_{X,H-L}$. Finally, we weight the difference between new and old stocks using the PC1 loadings: $\gamma'_{(t),1}(BM^{Old}_{X,H-L} - BM^{New}_{X,H-L})$. This result tells us how much the PC1 strategy loads on book-to-market, and we compare this loading to the high-minus-low book-to-market spread from a single sort on book-to-market. We analogously calculate the loading on all other characteristics.

Table 6

Old and new stocks in benchmark factor models. This table presents average returns for old and new stock strategies as well as the intercept (with corresponding t-statistic in parentheses) from regressing these returns on benchmark factor models. As discussed in Section 5.3, the returns of new, $R^{New}_{X,(t),t+1}$, and old, $R^{Old}_{X,(t),t+1}$, stocks together make up the return of the newest sort, $R_{X,(t),t+1}$. As in the previous tables, we aggregate the new and old stock returns across the 56 characteristics using the PC1 loadings. For instance, the aggregated return of old stocks presented in the first row of Panel A is defined as $\gamma'_{(t),1} R^{Old}_{X,(t),t+1}$. In Panel B and C, we decompose these test asset returns into the part coming from the characteristics with returns that decay too fast versus too slow, which are those characteristics on which PC1 loads with a negative versus positive sign (see Eqs. (22) and (23)). Because the absolute sum of negative loadings is about twice as large as positive loadings, we rescale the latter such that the old-minus-new stock return in Panel C has the same volatility as the old-minus-new stock return in Panel B. The sample period runs from July 1972 through December 2019.

			CAPM		FF3M		FF5M		FF5M+MOM	
	Avg. ret.	t-stat	α	t-stat	α	t-stat	α	t-stat	α	t-stat
	Panel A: PC1 of old and new stock strategies (e.g., $\gamma'_{(t),1} R^{Old}_{X,(t),t+1}$)									
Old	-1.41	(-0.35)	-8.22	(-2.49)	-6.85	(-3.27)	-0.04	(-0.02)	2.59	(1.54)
New	-5.74	(-1.78)	-11.40	(-4.33)	-11.09	(-5.66)	-4.95	(-2.97)	-1.87	(-1.14)
Old-minus-new	4.33	(2.74)	3.18	(2.04)	4.25	(3.15)	4.91	(3.54)	4.46	(3.05)
	Panel B: Characteristics with returns that decay too fast (e.g., $R^{Old,Fast}_{(t),t+1} = \gamma^{\le 0,\prime}_{(t),1} R^{Old}_{X,(t),t+1}$)									
Old	3.67	(1.17)	9.15	(3.57)	7.76	(4.40)	2.22	(1.58)	-0.09	(-0.06)
New	6.36	(2.61)	10.79	(5.49)	10.20	(6.47)	5.37	(3.95)	3.20	(2.29)
Old-minus-new	-2.69	(-1.96)	-1.64	(-1.22)	-2.44	(-2.07)	-3.15	(-2.62)	-3.29	(-2.68)
	Panel C: Characteristics with returns that decay too slow (e.g., $R^{Old,Slow}_{(t),t+1} = \gamma^{>0,\prime}_{(t),1} R^{Old}_{X,(t),t+1}$)									
Old	3.69	(2.37)	1.53	(1.11)	1.50	(2.05)	3.57	(5.01)	4.11	(5.51)
New	1.01	(0.68)	-0.99	(-0.77)	-1.46	(-1.75)	0.69	(0.90)	2.19	(3.29)
Old-minus-new	2.69	(3.60)	2.53	(3.35)	2.96	(4.08)	2.88	(3.84)	1.92	(2.35)

Thus, our evidence indicates that return spreads from stocks that have been in the extreme portfolios for a longer period are not the same as return spreads from stocks that are new to the extreme portfolios, even when these old and new stocks have the same current level of the characteristic. This fact represents a firm rejection of the standard characteristic-based model of expected returns, but it is consistent with variation in the compensation for persistent and transitory components of characteristics under our alternative hypotheses. This finding extends the work of Daniel and Titman (1997), who show that returns can vary with a characteristic, even holding risk exposure fixed. We show that returns can vary even holding the characteristic itself fixed. Furthermore, investors trading these characteristics should carefully consider the distinction between new and old stocks. Our new and old stock portfolios are tradable and require a position in fewer stocks than the original strategies. In addition, old stock portfolios will require less rebalancing.

7. Long-term discount rates

Keloharju et al. (2021) argue that characteristic-based return predictability has little impact on the long-term discount rates of firms, because it is transitory. However, our results so far suggest that there are characteristics for which returns decay quite slowly. If a firm loads strongly on these characteristics, we expect its long-term discount rate to be affected in a meaningful way.

We calculate long-term discount rates using a discounted cash flow approach as in Keloharju et al. (2021, Section 6.3). The implied discount rate r of an asset solves the Gordon growth equation: $P = \frac{D}{r-g}$, where P follows from discounting the cash flow stream implied by $D = 1\$$ and some annual growth rate g at a hypothetical term structure of per-period discount rates. This term structure is derived from average annual returns up to 10 years after formation of the high- or low-decile portfolio sorted on some characteristic. We center these realized returns at the 8% expected return of the market (2% risk-free rate + 6% market risk premium). After year 10, we assume the per-period discount rate has converged back to the expected market return. We simulate these long-term discount rates from our model (using the rich specification from Section 4) and compare them to a variety of portfolios in the data.

In Panel A of Table 7, we present the distribution of long-term discount rates from the model for the high and low decile portfolio sorted on the characteristic X using $g = 1\%$. Consistent with Keloharju et al. (2021), the high-minus-low difference in discount rate is small under the alternative that $\lambda_P = 0$ and $\lambda_T > 0$. The median difference equals 0.26% and the 95th percentile difference is 1.2%. There is considerably more discount rate variation when the persistent component of the characteristic is priced. Under the null ($\lambda_P = \lambda_T > 0$), the median high-low difference equals 1.57%. Under the alternative that $\lambda_P > 0$ and $\lambda_T = 0$, the median difference equals 1.97%. If, as we have argued above, there are characteristics with horizon dynamics that are more consistent with the latter alternative (or even just the null), we would expect that discount rates of firms loading strongly on these characteristics will be affected more than what is suggested in Keloharju et al. (2021).

To see whether this is the case, we present in Panel B the discount rate for the high and low portfolios sorted on two composite characteristics. These two characteristics average at the firm level over the subset of characteristics on which PC1 loads with a negative versus positive sign.[27] Thus, we compare firms that load strongly on characteristics with returns that decay fast versus slow relative to the characteristic itself. Because we are not comparing characteristics with returns that decay fast or slow in an absolute sense, these results likely represent a lower bound on the discount rate implications of characteristic-based return predictability. We report results for both value- and equal-weighted portfolio returns. While all of our results so far are based on value-weighting, equal-weighting is arguably more representative

[27] We rank-normalize the characteristics and average them for all firms with at least half of the respective subset of characteristics available. Our results are similar if we weight the characteristics using the relative magnitude of the loadings, but this approach is more sensitive to missing characteristics at the firm level.

Table 7

Long-term discount rates in the model and data. In this table, we compare implied discount rates r from our model to the data (see Section 7 for more detail). The r is found solving: $P = \frac{D}{r-g}$, where we set $D = 1$, $g = 1\%$, and the fundamental price P is calculated by discounting the implied cash flow stream using a hypothetical term structure of per-period discount rates derived from realized annual returns of a characteristic-sorted portfolio. We track realized annual returns up to 10 years after portfolio formation and assume that the per-period discount rate has converged back to the expected market return of 8% after year 10. In Panel A, we report the simulated distribution of implied discount rates for the high- and low-decile portfolio as well as the high-minus-low difference for the three cases in the model of Section 4. Under the null of a standard characteristic-based model of expected returns, the compensation for the transitory and persistent component of the characteristic is equal: $\lambda_P = \lambda_T = 3.17\%$. Under the first alternative, only the transitory component is compensated in expected returns: $\lambda_P = 0,\ \lambda_T = 4.49\%$. Under the second alternative, only the persistent component is compensated in expected returns: $\lambda_P = 4.49\%,\ \lambda_T = 0$. In Panel B, we report the implied discount rate from the data when we sort stocks on two composite characteristics that average at the firm level over the subset of characteristics on which PC1 loads with a positive versus negative sign (and the decay in returns is fast versus slow). In Panel C, we report the implied discount rate from the data when we sort stocks on a single characteristic (size, book-to-market, profitability, and investment). The sample period runs from July 1972 to December 2019. In Panels B and C, we report results for both value- and equal-weighted portfolios.

		High			Low			High-minus-low		
		Panel A: Model simulations								
		5^{th}	50^{th}	95^{th}	5^{th}	50^{th}	95^{th}	5^{th}	50^{th}	95^{th}
Null:	$\lambda_P = \lambda_T > 0$	6.85%	8.86%	11.30%	5.65%	7.21%	9.08%	0.65%	1.57%	2.99%
Only transitory:	$\lambda_T > \lambda_P = 0$	6.28%	8.12%	10.38%	6.20%	7.85%	9.89%	-0.63%	0.26%	1.20%
Only persistent:	$\lambda_P > \lambda_T = 0$	7.02%	9.08%	11.56%	5.52%	7.03%	8.87%	1.05%	1.97%	3.35%
		Panel B: Composite characteristics (Data)								
		Value-weighted								
Fast		7.86%				8.13%				-0.27%
Slow		9.33%				6.83%				2.51%
		Equal-weighted								
Fast		7.88%				8.11%				-0.23%
Slow		9.93%				6.39%				3.53%
		Panel C: Individual characteristics (Data)								
		Value-weighted								
Size		8.79%				7.27%				1.52%
Book-to-market		9.04%				7.06%				1.97%
Profitability		8.17%				7.82%				0.35%
Investment		8.08%				7.92%				0.15%
		Equal-weighted								
Size		9.11%				7.00%				2.11%
Book-to-market		9.60%				6.63%				2.98%
Profitability		8.33%				7.68%				0.65%
Investment		8.90%				7.17%				1.73%

of the typical firm in the cross-section. Moreover, there is no notion of firm size in our model.

We observe that the discount rate is meaningfully different for firms that load strongly on characteristics with returns that decay too slow. Firms that load positively on these characteristics have an implied discount rate of about 9.5%, which is relative to about 6.5% for the firms that load negatively. The difference of 2.51% (value-weighted) and 3.53% (equal-weighted) is economically large and about 2.5 to 3.5 times larger than what Keloharju et al. (2021) report for the average characteristic (1.04%, see their Table 5). If we assume a higher growth rate $g = 5\%$, the relative difference with Keloharju et al. (2021) is even larger. As reported in Table OA.9, we estimate a discount rate difference of 1.25% (value-weighted) and 1.73% (equal-weighted) for this high-duration case. This estimate is five to seven times larger than what Keloharju et al. (2021) report for the average characteristic (0.24%).

Comparing these high-minus-low differences to the percentiles of the simulated distribution, we see that such large estimates are unlikely to be generated in a world where there is no compensation for the persistent component. Instead, because the value-weighted estimates fall just short of the 95th percentile under the null and the equal-weighted estimates fall above the 95th percentile under the alternative ($\lambda_P > 0$ and $\lambda_T = 0$), we conclude that the compensation for the persistent component is likely to be relatively large (meaning $\lambda_P > \lambda_T$) among the subset of characteristics with returns that decay too slow. In stark contrast, we see in Panel B of Table 7 that the difference in implied discount rates is small for firms that load strongly on characteristics with returns that decay too fast. The fact that this difference is negative in the data is potentially consistent with long-term reversal and not something that can be generated from our model.

In Panel C of Table 7, we show that there is meaningful discount rate variation even in single sorts on popular characteristics. Consistent with the slow decay of their returns in Fig. 1, we find in the value-weighted case that firms in the high book-to-market and size decile have an implied discount rate that is 2% and 1.5% larger, respectively, than firms in the low deciles. For the equal-weighted case, these differences are again larger at 3% and 2.1%. Thus, book to market and size are characteristics for which it is highly unlikely that only the transitory component is priced. In contrast, and consistent with the fast decay of their returns in Fig. 1, discount rate differences are smaller for firms that load strongly on profitability and investment.

We conclude that there are subsets of firms for which characteristic-based return predictability impacts their discount rate by substantially more than what is suggested in Keloharju et al. (2021). That said, we agree with these authors that returns at short horizons after portfolio formation are typically larger than returns at longer horizons and, consequently, not representative of the discount rate implications of characteristic-based return predictability.

8. Conclusion

In this paper, we study the long-term returns to characteristic-based strategies to shed light on the relative compensation for persistent and transitory components of characteristics. We uncover large abnormal returns between old and new sorts as well as for a closely related strategy that builds on a novel decomposition into old and new stocks. These abnormal returns translate to large improvements in Sharpe ratio. What is most surprising is that these abnormal returns vary in sign across characteristics.

To assess the economic importance of this result, we develop a simple model of characteristics containing a persistent and a transitory component. We simulate from this model using a rich specification that accounts for important features of empirical (return and characteristics) data. We show that sign-variation in abnormal returns between old and new sorts (and stocks) provides strong evidence against the null of the standard characteristic-based model of expected returns. Under this null, the compensation for the persistent and transitory component of the characteristic is equal. The sign-variation is even harder to explain from the perspective that only the transitory component of characteristics is priced, a perspective that has been endorsed in recent work. Rather, we conclude that all characteristics are not created equal: the relative compensation for persistent and transitory components varies strongly across characteristics.

If the persistent components of some characteristics are priced, the return predictability generated by these characteristics should meaningfully affect the long-term discount rates of firms. We confirm this insight and argue that there are subsets of firms for which the long-term discount rate impact of characteristic-based return predictability is substantially larger than what recent work suggests.

In all, our evidence has implications for investors trading characteristics, managers estimating discount rates, and academics testing asset pricing models. We leave for future work the examination of the risk and mispricing drivers of the relative compensation for persistent and transitory components of characteristics.

CRediT authorship contribution statement

Fahiz Baba-Yara: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Martijn Boons:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Andrea Tamoni:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The codes and data of the article can be found at https://data.mendeley.com/datasets/sj93bbhdcw/2.

Appendix A. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jfineco.2024.103808.

References

Asness, C.S., Moskowitz, T.J., Pedersen, L.H., 2013. Value and momentum everywhere. J. Finance 68, 929–985.
Chernov, M., Lochstoer, L.A., Lundeby, S.R., 2022. Conditional dynamics and the multi-horizon risk-return trade-off. Rev. Financ. Stud. 35, 1310–1347.
Cho, T., Polk, C., 2020. Asset pricing with price levels. Available at SSRN 3499681.
Cohen, R.B., Polk, C., Vuolteenaho, T., 2009. The price is (almost) right. J. Finance 64, 2739–2782.
Daniel, K., Hirshleifer, D., Sun, L., 2020a. Short- and long-horizon behavioral factors. Rev. Financ. Stud. 33, 1673–1736.
Daniel, K., Mota, L., Rottke, S., Santos, T., 2020b. The cross-section of risk and returns. Rev. Financ. Stud. 33, 1927–1979.
Daniel, K., Titman, S., 1997. Evidence on the characteristics of cross sectional variation in stock returns. J. Finance 52, 1–33.
Fama, E.F., French, K.R., 1992. The cross-section of expected stock returns. J. Finance 47, 427–465.
Fama, E.F., French, K.R., 1993. Common risk factors in the returns on stocks and bonds. J. Financ. Econ. 33, 3–56.
Fama, E.F., French, K.R., 2015. A five-factor asset pricing model. J. Financ. Econ. 116, 1–22.
Fama, E.F., French, K.R., 2018. Choosing factors. J. Financ. Econ. 128, 234–252.
Frazzini, A., Pedersen, L.H., 2014. Betting against beta. J. Financ. Econ. 111, 1–25.
Freyberger, J., Neuhierl, A., Weber, M., 2020. Dissecting characteristics nonparametrically. Rev. Financ. Stud. 33, 2326–2377.
Gerakos, J., Linnainmaa, J.T., 2018. Decomposing value. Rev. Financ. Stud. 31, 1825–1854.
Gibbons, M.R., Ross, S.A., Shanken, J., 1989. A test of the efficiency of a given portfolio. Econometrica, 1121–1152.
Gomes, J., Kogan, L., Zhang, L., 2003. Equilibrium cross section of returns. J. Polit. Econ. 111, 693–732.
Green, J., Hand, J.R., Zhang, X.F., 2017. The characteristics that provide independent information about average us monthly stock returns. Rev. Financ. Stud. 30, 4389–4436.
Gu, S., Kelly, B., Xiu, D., 2020. Empirical asset pricing via machine learning. Rev. Financ. Stud. 33, 2223–2273.
Haddad, V., Kozak, S., Santosh, S., 2020. Factor timing. Rev. Financ. Stud. 33, 1980–2018.
Harvey, C.R., Liu, Y., 2021. Lucky factors. J. Financ. Econ. 141, 413–435.
Harvey, C.R., Liu, Y., Zhu, H., 2016. … and the cross-section of expected returns. Rev. Financ. Stud. 29, 5–68.
Herskovic, B., Moreira, A., Muir, T., 2019. Hedging risk factors. Working Paper.
Higham, N.J., 2002. Computing the nearest correlation matrix—a problem from finance. IMA J. Numer. Anal. 22, 329–343.
Hou, K., Mo, H., Xue, C., Zhang, L., 2018. Which factors? Rev. Finance 23, 1–35.
Hou, K., Xue, C., Zhang, L., 2015. Digesting anomalies: an investment approach. Rev. Financ. Stud. 28, 650–705.
Jegadeesh, N., Titman, S., 1993. Returns to buying winners and selling losers: implications for stock market efficiency. J. Finance 48, 65–91.
Jensen, T.I., Kelly, B.T., Malamud, S., Pedersen, L.H., 2022. Machine learning and the implementable efficient frontier. Available at SSRN 4187217.
Kelly, B.T., Pruitt, S., Su, Y., 2019. Characteristics are covariances: a unified model of risk and return. J. Financ. Econ. 134, 501–524.
Keloharju, M., Linnainmaa, J.T., Nyberg, P., 2021. Long-term discount rates do not vary across firms. J. Financ. Econ. 141, 946–967.
Kothari, S.P., Shanken, J., Sloan, R.G., 1995. Another look at the cross-section of expected stock returns. J. Finance 50, 185–224.
Kozak, S., Nagel, S., Santosh, S., 2020. Shrinking the cross-section. J. Financ. Econ. 135, 271–292.
Lettau, M., Pelger, M., 2020. Factors that fit the time series and cross-section of stock returns. Rev. Financ. Stud. 33, 2274–2325.
Lintner, J., 1965. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. Rev. Econ. Stat., 13–37.
Liu, J., Moskowitz, T.J., Stambaugh, R.F., 2021. Pricing without mispricing. Tech. Rep. National Bureau of Economic Research.
Mossin, J., 1966. Equilibrium in a capital asset market. Econometrica, 768–783.
Novy-Marx, R., Velikov, M., 2016. A taxonomy of anomalies and their trading costs. Rev. Financ. Stud. 29, 104–147.
Sharpe, W.F., 1964. Capital asset prices: a theory of market equilibrium under conditions of risk. J. Finance 19, 425–442.

Stambaugh, R.F., Yuan, Y., 2016. Mispricing factors. Rev. Financ. Stud. 30, 1270–1315.
van Binsbergen, J., Boons, M., Opp, C., Tamoni, A., 2023. Dynamic asset (mis)pricing: build-up vs. resolution anomalies. J. Financ. Econ. 147, 406–431.
Van Binsbergen, J.H., Opp, C.C., 2019. Real anomalies. J. Finance 74, 1659–1706.
Welch, I., 2022. Simply better market betas. Crit. Finance Rev. 11, 37–64.
White, H., 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica 48, 817–838.
Zhang, L., 2005. The value premium. J. Finance 60, 67–103.