NOVA SCHOOL OF BUSINESS & ECONOMICS

Universidade Nova de Lisboa

Dissertation, presented as part of the requirements for the Degree of

Doctor of Philosophy in Economics | Finance

Essays in Empirical Asset Pricing

Fahiz Mohammed Baba Yara

Student nº 25795

A Dissertation carried out on the PhD in Economics | Finance,
under the supervision of
Professor Martijn Boons, Professor Andrea Tamoni, and Professor Melissa Prado

June 17, 2021

Abstract

The first chapter of this dissertation studies value strategies across equities, industries, commodities, currencies, global government bonds, and global stock indexes. We find that these strategies are predictable in the time series by the respective value spreads. A single component of the value spreads across asset classes capture about two-thirds of the value return predictability. The second chapter analyses returns to new and old sorts, where new (old) sorts capture the return of a characteristic-sorted portfolio immediately (longer) after portfolio formation. We find that there exist large alphas between old and new sorts. These alphas (i) translate into large improvements in Sharpe ratio, (ii) are not captured by benchmark asset pricing models, and (iii) are linked to the return differential between new and old stocks. The final chapter investigates how incorporating results from the financial-economics literature in the specification of a machine learning model can improve the resulting models' forecasts. I find that the economically motivated specification predicts better cross-sectional variation in individual stock returns and more robustly predicts time-series variation in returns to value-weighted long-short portfolios and the market portfolio.

Keywords: Value Premium, Value Spread, Machine Learning, Neural Networks, Characteristic Sorted Portfolios, Cross-sectional Return Predictability
JEL Classification: E44, G10, G11, G12, G17.

Acknowledgements

First and foremost, I would like to express my gratitude to my supervisors, Martijn Boons, Andrea Tamoni, and Melissa Prado. This thesis is the result of their guidance and supervision. I have been blessed to have them with me during this journey and will forever be indebted to them. Additionally, I am very thankful to Fernando Anjos and Miguel Ferriera for their support, comments, and suggestions throughout my Ph.D. Furthermore, I thank Irem Demirci, Emanuele Rizzo, Giorgio Otttonello, Nicholas Hirschey, Virginia Gianinazzi, Rik Frehen, Paul Karehnke, Irina Zviadadze, Hugues Langlois-Bertrand, Bradley Paye, Ngoc-Khanh Tran, Olga Obizhaeva, Tobias Sichert, Riccardo Sabbatucci, Francesco Sangiorgi, Grigory Vilkov, Chukwuma Dim, Janet Gao, Fotis Grigoris, and Craig Holden for their insightful comments, as well as conference and seminar participants at Nova SBE, VU Amsterdam, Virginia Tech, HEC Paris, Stockholm School of Economics, Frankfurt School of Finance and Management, and Indiana University.

Special thanks goes to my all friends and colleagues from Nova SBE. Namely, I thank Alfonso, Ana, Catia, Diogo, Felix, Ines, Jacob, Joana, Joao, Luis, Massimiliano, Maria, Matilde, Mattia, Robert, Sara, Stefan and Vera for making my time in Lisbon a memorable one.

This work received the support of Funda¸c˜ao para a Ci^encia e Tecnologia (FCT) under the grant PD/BD/113888/2015.

Contents

Introduction

In the first chapter, my co-authors and I show that returns to value strategies in individual equities, industries, commodities, currencies, global government bonds, and global stock indexes are predictable in the time series by their respective value spreads. In all these asset classes, expected value returns vary by at least as much as their unconditional level. A single common component of the value spreads captures about two–thirds of value return predictability and the remainder is asset class–specific. We argue that common variation in value premia is consistent with rationally time–varying expected returns, because (i) common value is closely associated with standard proxies for risk premia, such as the dividend yield, intermediary leverage, and illiquidity, and (ii) value premia are globally high in bad times.

In the second chapter, my co-authors and I study the returns to characteristic-sorted portfolios up to five years after portfolio formation. For many characteristics, like book-to-market, the persistence of return predictability does not match the persistence of the characteristic. Consequently, large alphas exist between new and old *sorts*, where new (old) sorts capture the return of a characteristic-sorted portfolio immediately (longer) after portfolio formation. These alphas (i) translate into large improvements in Sharpe ratio, (ii) are not captured by benchmark asset pricing models, and (iii) are linked to the return differential between new and old *stocks*. Since portfolios of new and old stocks are characteristic-neutral, we conclude that explanations of the cross-section based on recent observations of characteristics (and factors derived therefrom) are incomplete.

In the final chapter, I study how incorporating results from the financial economics literature in the specification of a machine learning model improves the predictive ability of the model. Previous research finds that machine learning methods predict short-term return variation in the cross-section of stocks, even when these methods do not impose strict economic restrictions. However, without such restrictions, the models' predictions fail to generalize in a number of important ways, such as predicting time-series variation in returns to the market portfolio and long-short characteristic sorted portfolios. I show that this shortfall can be remedied by imposing restrictions, that reflect findings in the financial economics literature, in the architectural design of a neural network model and provide recommendations for using machine learning methods in asset pricing. Additionally, I study return predictability over multiple future horizons, thus shedding light on the dynamics of intermediate and long-run conditional expected returns.

Chapter 1

Value Return Predictability Across Asset Classes and Commonalities in Risk Premia

This chapter was written in co-authorship with Martijn Boons and Andrea Tamoni. We thank the editor Jules H. van Binsbergen, an anonymous referee, an anonymous second editor, Fernando Anjos, Matt Billett, Victor deMiguel, Miguel Ferreira, Simon Haynes, Chris Hrdlicka, Sreeni Kamma, Sonia Konstantinidi, Dong Lou, Igor Makarov, Richard Payne, Christopher Polk, Melissa Prado, Alessandro Previtero, Pedro Santa–Clara, Wenyu Wang, Michael Weber, Michela Verardo, and seminar participants at Cass Business School, Foster School of Business (University of Washington, Seattle), Kelley School of Business (Indiana University), London School of Economics, Nova School of Business and Economics, Queen Mary University, Rutgers Business school, Virginia Tech, the 2018 Conference in Frontiers of Factor Investing (Lancaster University), Unigestion and Fulcrum Asset Management for helpful comments. This work was supported by Fundação para a Ciência e Tecnologia, Portugal. The authors have no conflicts of interest to disclose.

1.1 Introduction

In this paper, we show that the expected returns of long–short value strategies in a range of asset classes increase in the value spread. The value spread is the difference between the value signal in the long versus the short portfolio, and its relation to value premia can be motivated from standard present value logic (e.g., Vuolteenaho, 2002; Froot and Ramadorai, 2005). The time variation in value premia we document is both economically and statistically large. At the one–year horizon, the R^2 in a time series predictive regression equals 14% and 6% for U.S. individual equities and industries, respectively, as well as 9%, 11%, 19% and 8% for commodities, currencies, global government bonds, and global stock indexes, respectively, and 13% in a pooled regression. In all these asset classes, a standard deviation increase in the value spread predicts an increase in the expected value return of the same order of magnitude (or more) as the unconditional value premium. Thus, expected returns on value strategies vary over time by at least as much as their already puzzling level.

Cochrane, 2011a emphasizes that the value premium continues to be one of the main "puzzles" in finance, as the long–standing debate between rational explanations and mispricing is still unresolved. To provide new insight, we analyze the economic drivers of the time variation in expected value returns. We first decompose the value spread into a common component, defined as the first principal component of value spreads, and asset class–specific components. While the common component captures about half of the variation in value spreads, it captures more — about two–thirds — of the variation in expected value returns in the pool of asset classes. The remaining one–third is asset class–specific. Quantifying the relative contribution of these two components to predictability is important, because a large and significant common component is evidence of market integration. Despite this fact, there is little evidence in the literature for common return predictability across asset classes (e.g. Cochrane, 2011a).

We argue that time–varying risk premia drive the common component of value. Indeed, we find that expected value returns are globally high in bad times and remain so for a number years. Moreover, two proxies for the risk of financial intermediaries — market leverage and funding liquidity — together with a measure of risk aversion explain the bulk of time variation in the common component. Thus, our evidence builds support for the recent theoretical literature on intermediary–based asset pricing (He and Krishnamurthy, 2012; He and Krishnamurthy, 2013; Brunnermeier and Sannikov, 2014), as well as for asset pricing models featuring time–varying risk aversion (Campbell and Cochrane, 1999; Menzly, Santos, and Veronesi, 2004; Santos and Veronesi, 2016). Thus, the quantitatively large amount of value return predictability we find in asset classes with potentially different investors and institutional settings presents a challenge for asset pricing theory. Our results suggest that a full explanation of the value premium requires a general framework, where in bad times investors shy away from holding different risky assets, so that value spreads widen simultaneously.

Another challenge to asset pricing models follows from the asset class–specific components of the value spread, which point to a mix of risk and mispricing. Although these components load on risk proxies, such as leverage and uncertainty, we find that the loadings vary considerably across asset classes. In addition, the risk proxies leave a large share of asset class–specific value return predictability unexplained, which points to mispricing. Consistent with these findings, we show that the common component of the value spread contributes relatively more than the asset class–specific components to value return predictability in the recent subsample. We find that common value is strongly associated with proxies for the risk of financial intermediaries, and financial intermediation has become more important over time. Moreover, if limits to arbitrage partially drive the asset class–specific components of value return predictability, one would expect these components to become less important over time.

Our results contribute to the asset pricing literature in various ways. Unconditional value premia are documented for U.S. individual equities (Fama and French, 1992a), international equities (e.g., Fama and French, 1998; Liew and Vassalou, 2000),

A contemporaneous paper, Asness et al., 2018, independently reaches the same conclusion that value returns are predictable in different asset classes. The key difference from their paper is that we use the value spread as a simple measure of the expected return to a value strategy and analyze its variation over time in a pool of asset classes. This setup allows us to decompose value into common and asset class–specific components, thus enabling us to highlight the close association between common value and aggregate risk premia. Asness et al., 2018 focus on "deep" value events. They have more extensive data for equities, which enables them to highlight the fundamentals of low and high value stocks and to test more rigorously alternative behavioral theories for the value effect.

and alternative asset classes (Asness, Moskowitz, and Pedersen, 2013a). Whereas our paper is about comovement in *expected* value returns, 2013a show that *realized* value returns comove across asset classes. Conditional tests are relatively powerful in distinguishing between competing asset pricing models Campbell and Cochrane, 2000; Cochrane, 2001; Nagel and Singleton, 2011. Therefore, the large amount of common variation in expected value returns that we document sets a higher hurdle for rational, risk–based models than what 2013a discuss.

We analyze the ability of the value spread to predict a value–minus–growth portfolio return over time, whereas many studies attempt to forecast (long–only) returns using valuation ratios. Lewellen, 1999 and Cochrane (2011a) predict the returns of diversified equity portfolios with their book–to–market ratio. Cochrane (2011a, p. 1099) concludes that "variation over time in a given portfolio's book–to–market ratio is a much stronger signal of return variation than the same variation across portfolios in average book–to–market ratio." Kelly and Pruitt, 2013 conclude that the expansion and compression of the cross–section of value characteristics contains information about expected stock market returns. We show that this conclusion applies equally to expected value returns in all the asset classes we study.

Our findings for the value spread in individual equities are consistent with those of Asness et al., 2000a. Using data for large U.S. individual equities from 1982 to 1999, they find that industry–adjusted value spreads have predictive power for value–minus–growth returns. We contribute to this literature by studying (i) the value spread in other asset classes; (ii) the relative contribution of common and specific components of the value spread to predictability, as well as their economic drivers; and (iii) the potential for timing and rotation using the value spread in an out–of–sample setting. In particular, we find that value returns are predictable in real time, which alleviates concerns that our in–sample evidence is spurious.

Our multi–asset approach is uniquely suited to answer some of the central questions in asset pricing: Do expected returns vary over time and across assets? If so, by how much? And is this time variation driven by risk or mispricing? Our risky common component of expected value returns cannot be identified by analyzing a

Similarly, 2003a show that the return of the Fama and French, 1993a HML factor is predictable by the HML value spread. Asness et al., 2017 study strategies that time and rotate value, momentum, and betting–against–beta in equities using their respective value spreads.

single value strategy in isolation. This fact helps to explain recent mixed evidence on the question of whether the equity value premium is driven by risk or mispricing (Golubov and Konstantinidi, 2016; Gerakos and Linnainmaa, 2018a). Our work also contributes to the literature on global asset pricing, where "betting against beta" (Frazzini and Pedersen, 2014a), "carry" (Koijen et al., 2018), and downside risk (Lettau, Maggiori, and Weber, 2014) are shown to be factors in U.S. individual equities, as well as a host of other asset classes. In contrast to us, these papers mostly characterize unconditional premia. 2017 characterize conditional return variation in stocks, currencies, and bonds and argue, just like us, that long–short returns are more predictable than long–only market returns. Haddad, Kozak, and Santosh (2017) analyze a different strategy and a different predictor in each asset class. We analyze a single strategy (value) and a single predictor (the value spread) in all asset classes, and we extract a single common component.

1.2 Data and Methodology

In this section, we describe the construction of our value measures and value returns in different asset classes. The sources and procedures to clean the data are in the Internet Appendix 3.C. There, we also validate our key result using the value returns of 2013a. As is common in the literature, we use the book–to–market ratio as our measure of value for individual equities, industries, and global stock indexes. For the remaining asset classes, we follow 2013a and measure value using long–term past returns. This choice is inspired by the literature documenting a direct link between past returns and book–to–market ratios, both empirically (DeBondt and Thaler, 1985; Fama and French, 1996a; Gerakos and Linnainmaa, 2018a) and theoretically (Daniel, Hirshleifer, and Subrahmanyam, 1998; Hong and Stein, 1999; Vayanos and Woolley, 2013).

2012, Neuhierl and Weber, 2017, and Moreira and Muir, 2017 also present global evidence for return predictability, respectively, due to time series momentum, monetary momentum, and volatility timing.

1.2.1 Value in Different Asset Classes

U.S. individual equities and industries

The U.S. individual equities data are from the Center for Research in Security Prices (CRSP) and Compustat. Following 2013a, we limit the analysis to a sample from January 1972 to December 2017 and a universe of stocks that is liquid and can be traded at reasonably low cost in sizable trading volume. To be precise, we include in our value strategies only those stocks that cumulatively account for 90% of the total market capitalization in CRSP, which cutoff yields an average of 495 stocks for our portfolios. The idea is twofold. This allows us to provide conservative estimates for an implementable set of trading strategies. The cutoff also allows for a better comparison with the value strategies in alternative asset classes, where the securities are relatively liquid.

To measure value for each firm i, we use the ratio of the book value to the market value of equity, or the book–to–market ratio, $BM_{i,t}$, as in Fama and French, 1992a. Book values are observed each June and refer to the previous fiscal year–end. Market values are updated monthly as in Asness and Frazzini, 2013, but we also consider annually updated market values in a robustness check. Consistent with the literature, we exclude financial firms: a given book–to–market ratio might indicate distress for a non–financial firm, but not for a financial firm (Fama and French, 1995). We denote this measure $BM_{i,t,ExFin}$. Because many financial firms are large and in the investment opportunity set of most investors, we also consider a second set of industry–adjusted book–to–market ratios. To find the industry–adjusted book–to–market ratio for stock i, $BM_{i,t,IndAdj}$, we subtract from its book–to–market ratio the value–weighted average book–to–market ratio of the industry to which stock i belongs. 2000b and Cohen and Polk, 1998 find that industry–adjusted value strategies are relatively attractive. They argue that there is no unconditional value effect across industries. To determine whether there is a conditional value effect, we sort seventeen industry portfolios on the value–weighted average book–to–market ratio within each industry. To be consistent with our analysis of individual stocks, we construct the seventeen industry portfolios using only the restricted set of relatively large stocks.

Commodity futures

We obtain commodity futures price data for Crude Oil, Gasoline, Heating Oil, Natural Gas, Gas–Oil Petroleum, Coffee, Rough Rice, Orange Juice, Cocoa, Soybean Oil, Soybean Meal, Soybeans, Corn, Oats, Wheat, Cotton, Gold, Silver, Platinum, Feeder Cattle, Live Cattle, Lean Hogs (from the Commodity Research Bureau) and Aluminum, Nickel, Tin, Lead, Zinc, and Copper (from Datastream). We define value for commodities as the negative of the five–year spot return. As is common in the literature, we use the more liquid first–nearby futures price to proxy for the spot price. The sample period for commodities runs from January 1972 to December 2017.

Currencies

We obtain spot and forward currency exchange rates for Australia, Canada, Germany (spliced with the euro), Japan, New Zealand, Norway, Sweden, Switzerland, United Kingdom, and the United States. To measure value, we use the five–year change in relative purchasing power parity, which is calculated as the negative of the five–year spot return adjusted by the five–year foreign˘U.S. inflation difference. Currency value is large when the foreign currency has weakened relative to the dollar. As noted in Menkhoff et al., 2016, using five–year changes avoids potential problems that may arise from non–stationarity and base–year effects. The sample period for currencies runs from February 1976 to December 2017.

Global government bonds

We obtain global government bond data for Australia, Canada, New Zealand, Germany, Japan, Norway, Sweden, Switzerland, the United Kingdom, and the United States. We consider two sets of returns. Synthetic prices and returns for a one–month futures contract on a ten–year bond are derived for all countries from zero coupon, government bond yields. Traded bond index futures returns are available for six countries only (Australia, Canada, Germany, Japan, U.K. and U.S.). We define two measures of value for bonds using synthetic prices and yields, because the cheapest–to–deliver feature of traded bond futures makes it hard to compare yields over time and across countries. The first measure is the negative of the five–year futures return (˘5–year return). The second is the five–year change in the ten–year yield (5–year Δy). Using five–year changes in yields avoids potential problems that may arise from trends and unconditional differences across bond markets in default risk, for instance. Throughout the paper, our main focus is on strategies that use the first value measure to invest in the traded bond futures, but we present a number of robustness checks for the second value measure and synthetic bond returns. The sample period for global government bonds runs from January 1991 to December 2017.

Global stock indexes

The universe of developed country stock index futures consists of Australia, Canada, France, Germany, Hong Kong, Italy, Japan, Netherlands, Spain, Sweden, Switzerland, the United Kingdom, and the United States. To measure value for global stock indexes, we use the inverse of the MSCI price–to–book ratio ($MSCI_{BP}$). Dictated by data availability, the sample period for these stock indexes runs from January 1994 to December 2017.

1.2.2 Value Returns and Value Spreads

To construct value returns, we sort securities within each asset class into P groups based on the cross–section of value measures, $V_{i,t}$. For individual stocks, we form market value–weighted decile portfolios ($P = 10$) each month and define the value portfolio as decile 10 (High) and the growth portfolio as decile 1 (Low). For all other classes, we set $P = 2$ and form an equal–weighted High and Low portfolio by splitting the securities at the median of ranked values. We denote with $R_{t+1}^{H\text{˘}L}$ the return of the High–minus–Low value portfolio in the month after sorting. We also report results from an alternative rank–weighting procedure that weights each security $i = 1, \ldots, N_t$ at time t according to its rank in the cross–section:

$$w_{i,t}^{Rank} = q_t \left(\text{Rank}(V_{i,t}) \text{˘} \frac{\sum_i^{N_t} \text{Rank}(V_{i,t})}{N_t} \right).$$

The weights sum to zero, thus representing a dollar–neutral long–short portfolio. The scaling factor q_t ensures that we are one dollar long and one dollar short. The return of this rank–weighted strategy is calculated as $R_{t+1}^{Rank} = \sum_i w_{i,t}^{Rank} R_{i,t+1}$.

Throughout the paper, whenever we are predicting returns over horizons longer than one month, we separately compound total returns on the long and short position of these value strategies and then take the difference. These long and short positions are rebalanced for every month.

1.2.3 Predicting Value Returns with the Value Spread

The signal of interest is the value spread, which is defined as the difference between the average value signal in the High and Low portfolio, $VS_t^{H˘L} = V_t^{H}˘V_t^{L}$, or the rank–weighted average value signal, $VS_t^{Rank} = \sum_i w_{i,t}^{Rank} V_{i,t}$. We conduct predictive regressions of value returns (compounded over horizon h) on the lagged value spread:

$$R_{t+1:t+h}^{x} = a_h + b_h VS_t^{x} + \varepsilon_{t+1:t+h} \text{ for } x = H˘L, Rank. \tag{1.1}$$

This regression is easily motivated economically. For equities, consider the log–linear present value model employed in Vuolteenaho, 2002. If the book–to–market ratio is well–behaved, then:

$$\theta_t = \sum_{j=0}^{\infty} \rho^j r_{t+1+j} + \sum_{j=0}^{\infty} \rho^j (˘e_{t+1+j}) + \sum_{j=0}^{\infty} \rho^j k_{t+1+j}, \tag{1.2}$$

where θ_t is the log book–to–market ratio, $r_{t+1} \equiv \log\left(1 + \frac{\Delta ME_{t+1} + D_{t+1}}{ME_t}\right)$ denotes the log stock return, and $e_{t+1} \equiv \log\left(1 + \frac{\Delta BE_{t+1} + D_{t+1}}{BE_t}\right)$ is the log clean–surplus accounting return on equity. Next, consider a portfolio that is long high book–to–market stocks and short low book–to–market stocks. We apply Equation (1.2) to both portfolios, take conditional expectations, difference, and reorganize, to get:

$$E_t\left[\sum_{j=0}^{\infty} \rho^j r_{t+1+j}^{H˘L}\right] = \theta_t^{H}˘\theta_t^{L} + E_t\left[\sum_{j=0}^{\infty} \rho^j (e_{t+1+j}^{H}˘e_{t+1+j}^{L})\right]. \tag{1.3}$$

Empirically, we abstract from the correction for the spread in discounted future expected profitability. Thus, the specification in Equation (1.1) provides a lower bound on the predictability of value returns (Asness et al., 2000a).

As an alternative motivation, consider the investment–based asset pricing model of Zhang, 2005a. In this model, the value spread predicts value returns in the time series because it signals time variation in the risk premia of value versus growth stocks. In bad times, the market value of value firms decreases (relative to growth firms) as they are burdened with more unproductive capital and face large adjustment costs. Consequently, value is more risky exactly when risk premia are high. Finally, the value spread can be motivated on purely statistical grounds. In Section 1.B.7 of the Internet Appendix, we show that the partial least squares method of Kelly and Pruitt, 2015 selects the High–minus–Low value spread as the optimal forecasting factor derived from the cross–section of portfolio–level book–to–market ratios.

Similar to Equation (1.2), the present value formulation of Froot and Ramadorai, 2005 shows that expected currency returns are a key driver of real exchange rates. This motivates using real exchange rates as a measure of value for currencies. For bonds, the yield is a natural value metric, where a high yield indicates that the bond is relatively cheap. As for the case of equities, our regressions for currencies and bonds provide a lower bound on the predictability of value returns, since one can likely improve on our results by controlling for expected real interest rate differentials, in the case of currencies (Menkhoff et al., 2016), and differences in expected long–term inflation, in the case of bonds (Asness et al., 2018). Because these adjustments need to be estimated and are different across asset classes, we prefer the simpler, directly observable, measures of value that are used in 2013a.

In the regressions of value returns on the value spread, we consider forecasting horizon h up to four years. Horizons longer than one month help to mitigate the countervailing momentum effect Asness and Frazzini, 2013 and better resemble

---

Indeed, the predictive ability of the value spread in U.S. individual equities improves when incorporating the restrictions in Equation (1.3) in a filtering approach (Rytchkov, 2010) or by using the implied costs of capital to control for differences in earnings growth rates and payout ratios (Li, Ng, and Swaminathan, 2014).

Similarly, one can strengthen the results by combining different measures of value in a single asset class. For instance, larger unconditional value effects are found for equities when combining earnings–to–price, sales–to–price, and book–to–price (Asness et al., 2000a; Israel and Moskowitz, 2013).

the experience of actual value investors. It is important to note that long–horizon regressions of value returns on the value spread are relatively less affected by the inferential problems that are commonly associated with predictability. High first–order autocorrelation of the predictor and Stambaugh, 1999 bias have been put forward as leading causes of inaccurate inference when predicting aggregate stock market returns (e.g., Valkanov, 2003; Lewellen, 2004; Boudoukh, Richardson, and Whitelaw, 2006). However, the monthly autocorrelation of value spreads in the different asset classes ranges from 0.95 to 0.98 (see Panel A of Table 1.A.1), which is small relative to an autocorrelation of 0.993 for the dividend yield over our sample period from 1972 to 2017. Moreover, as we show in Table 1.C.1 in the Internet Appendix, the Stambaugh bias is small when predicting value returns with the value spread in individual equities. The intuition for this result is that the left–hand side in Equation (1.1) is a difference in returns between two portfolios, which we regress on the corresponding difference in valuation ratios. This setup in differences largely breaks the mechanical relation that exists in regressions of a single return on a price–based valuation ratio.

Thus, our setting is different from the usual setting in the predictability literature. Figure 1.C.1 in the Internet Appendix, presents the coefficient estimates, *t*–statistics, and R^2s from predictive regressions of non–overlapping value returns on the value spread, as well as market returns on the dividend yield. The figure shows that the value spread predicts value returns more strongly and farther into the future than the dividend yield predicts aggregate stock market returns.

1.2.4 Time Variation in Value Spreads

To accommodate comparison across asset classes, we standardize each value spread so that its time series average equals zero and standard deviation equals one. We present the time series of (High–minus–Low) value spreads in all seven asset classes in Figure 1.A.1.

[Insert Figure 1.A.1 about here]

Pooling also helps to alleviate concerns about Stambaugh bias, because the across asset class dimension lowers the correlation between innovations in the value spread and past return shocks.

To interpret the time variation in the value spread, let us consider the case of U.S. individual equities in the top–left panel. When the value spread is zero, value stocks are cheaper than growth stocks by their historical average amount. A positive value spread indicates that value stocks are cheaper and the cross–section of value measures is wider than normal. The same intuition applies to the other asset classes. For currencies, for instance, a large current value spread indicates that deviations from relative purchasing power parity are historically large.

The seven panels in Figure 1.A.1 present a number of episodes when the value spread is large in more than a few asset classes, such as after the collapse of the dot–com bubble and the recent financial crisis. Thus, the value spread is correlated across asset classes. This conclusion is confirmed in Panel A of Table 1.A.1, which presents the correlation matrix of value spreads. We find that the value spreads in U.S. individual equities, industries, commodities, and global equity indexes correlate strongly and positively with each other.

[Insert Table 1.A.1 about here]

These results suggest that the time variation of value spreads in different asset classes may be well captured by a small number of factors. Because the panel of value spreads is unbalanced due to limited data availability early in the sample period, we follow the procedure described in Stock and Watson (2002) to estimate the principal component factors using an iterative method based on the Expectation Maximization (EM) algorithm. Panel B of Table 1.A.1 shows that the first principal component of value spreads explains about 51% of the total variation. This first principal component is also presented in each panel of Figure 1.A.1. All value spreads load positively on the first principal component, and consistent with the correlations in Panel A, the loadings decrease from U.S individual equities to industries, equity indexes, commodities, bonds, and currencies. Consistent with these positive loadings, the correlation between a simple across–asset class average of the value spreads and the first principal component is large at 0.95.

In what follows, we refer to this first principal component as the common component of the value spreads, VS_t^{Com}. The common component explains about half of

The second principal component explains another 26% of the variation and loads heavily on currencies and bonds.

the variation in value spreads, but it is an empirical question as to what fraction of value return predictability it captures. The answer to this question is important because it determines to what extent expected returns comove across asset classes. In theory, one would expect strong comovement, but empirical evidence of this effect is scarce in the literature. For instance, following up on the evidence in Cochrane and Piazzesi (2005), who show that a single common factor extracted from forward rates describes the majority of the variance of expected bond returns, Cochrane (2011a, p. 1054) asks: "[W]hat similar patterns hold across broad asset classes?"

Figure 1 also provides evidence of asset class–specific variation in the value spreads. For instance, the value spread in global government bonds often moves in the opposite direction to the remaining asset classes. To analyze the fraction of value return predictability that is asset class–specific, we define the asset class–specific component of the value spread, VS_t^{Spec}, as the residual from a regression of the value spread in an asset class on the common component.

1.3 In–Sample Value Return Predictability

In this section, we ask whether returns to value strategies are predictable in the time series. To this end, we first analyze time series predictive regressions for each asset class. There is ample evidence for U.S. equities in the literature; however, our evidence for the other asset classes is new. Next, we analyze pooled predictive regressions to assess the joint strength of value return predictability. This pooled evidence represents a key contribution of our paper.

Panel A of Table 1.A.2 presents the unconditional performance of the High–minus–Low and rank–weighted value strategies. All returns are scaled to have an annual standard deviation of 15% to accommodate comparison. Consider first the evidence for individual equities, for which we have two signals: BM_{ExFin} and BM_{IndAdj}. Consistent with the literature, we find that the industry–adjusted value strategy performs well, with annualized Sharpe ratios (monthly Sharpe ratio $\times \sqrt{12}$) of 0.24 and 0.38 for the High–minus–Low and rank–weighted strategy, respectively. These numbers are relative to 0.14 and 0.17 for the strategy excluding financials. Overall, these Sharpe ratios are slightly lower than what is typically reported for value in the literature, because we focus on the set of relatively large and liquid stocks that cumulatively account for 90% of the total market capitalization (and value returns have generally been poor in recent years).

[Insert Table 1.A.2 about here]

For industries, commodities, currencies, global government bonds, and global stock indexes, we see that most value strategies generate a positive Sharpe ratio, but there is considerable variation in magnitude. Annualized Sharpe ratios range between 0.20 and 0.30 for the value strategies in commodities, currencies, and global stock indexes. Consistent with the literature, a value strategy using industries does not perform well unconditionally and produces a Sharpe ratio of 0.03. The Sharpe ratio of the value strategy using global government bonds is similarly small at –0.03.

1.3.1 Time Series Predictive Regressions

Panel B of Table 1.A.2 shows the results from time series predictive regressions of value returns on the value spread at forecasting horizons of $h = 1, 12$ and 24 months for all seven asset classes. We present regression coefficients, t–statistics (using Newey and West (1987a, t^{nw}) and Hodrick (1992, t^{hd}) standard errors with h lags), and R^2s.

We see that the coefficient on the value spread is positive for all asset classes and for all horizons. The evidence is strong at the annual horizon, where the coefficient estimate is significant and positive at the 10%–level in all asset classes using both Newey–West and Hodrick standard errors. Similarly strong evidence is found at the two–year horizon, with global equity indexes significant at the 10%–level and all other asset classes significant at the 5%–level. At the one–month horizon, the coefficient estimate is significant in half (one–third) of the asset classes for the High–minus–Low (rank–weighted) portfolios. So, statistically, there is considerable evidence that the value spread predicts value returns. The information in the value

Our focus on large stocks is similar to Asness et al., 2013a, and the correlation between their book–to–market strategy and our strategy excluding financial firms is 0.99.

This strategy uses traded bond futures returns and ˘5–year return as the value signal. Alternative strategies, using synthetic bond futures returns or the 5–year change in yield as value signal, perform slightly better unconditionally (see Table 1.C.2 of the Internet Appendix). 2013a also find that value strategies using global government bonds vary considerably across specifications.

spread takes longer than a month to fully materialize, however. Indeed, both the coefficient estimates and R^2s are, in most cases, increasing with the horizon.

The economic magnitudes of the coefficients on the value spread are also large. To see this, consider first the evidence for individual equities. For the High–minus–Low value strategy excluding financials, the coefficient estimates translate to an increase in monthly, annual, and bi–annual future return equal to 0.48%, 6.97%, and 16.96% (with Hodrick t–statistics of 2.20, 3.24, and 4.24, respectively) for a standard deviation increase in the value spread. The R^2 in these same regressions are 1.03%, 13.94%, and 30.86% respectively. Thus, time variation in the value spread explains almost one–third of the variation in the two–year returns of this strategy. For the industry–adjusted book–to–market strategy, the coefficient estimates and R^2 are even larger. The correlation between the value return series that excludes financials and the industry–adjusted value return series is about 0.69. This result suggests that cleaning valuation ratios from across–industry variation creates a different time series of value returns that is more predictable. For the rank–weighted portfolios, we also see economically large coefficients and R^2s, although the evidence is a bit weaker than for the decile portfolio strategy.

Recall that by standardizing the value spread, the ratio of the estimated coefficient to the intercept, b_h/a_h, measures the implied standard deviation of expected returns relative to the unconditional value premium. At all horizons, this ratio is over 2 (2.8 on average) for the High–minus–Low portfolios, and over 1.1 (1.6 on average) for the rank–weighted portfolios. We conclude that the value premium in U.S. individual equities strongly increases (decreases) as the cross–section of valuation ratios expands (compresses). To benchmark the strength of this in–sample evidence, consider that Cochrane, 2011a reports a ratio slightly below one when predicting the aggregate stock market return with the dividend yield.

Although the fact that value returns in U.S. individual equities are predictable using the value spread is not new (e.g., Asness et al., 2000a; Cohen et al., 2003a), our

In Table 1.C.3 of the Internet Appendix, we show that the value spread remains a significant predictor when we (i) extend the sample back to 1962, (ii) calculate the book–to–market ratio with annually updated market cap (as in Fama and French, 1992a), and (iii) sort stocks on the negative of the past five–year return (e.g., DeBondt and Thaler, 1985, who use a similar measure to identify undervalued firms).

evidence contributes to the literature along the following dimensions. First, we focus on a relatively small set of large and liquid stocks and extend the sample period post–2000, thus including two major recessions and the recent period of low value returns. Second, we show in the next section that the value spread predicts value returns out–of–sample. This finding is important, because even after a long history of research on the predictive relation between market returns and the dividend yield, it is unclear whether the information in the dividend yield can be used profitably in an out–of–sample setting. This lack of out–of–sample evidence has raised concerns that the in–sample predictability is spurious (Lettau and Van Nieuwerburgh, 2007; Goyal and Welch, 2008a). Third, the variation in expected value returns we document is economically large and will likely pose a challenge for standard asset pricing models to match. To see this by example, we simulate from the investment–based asset pricing model of Zhang, 2005a, which contains a time–varying value premium. Table 1.C.4 of the Internet Appendix presents the distribution of unconditional and conditional value premia obtained from 1000 simulations of the model. We see that the median ratio b_h/a_h in a regression of annual High–minus–Low value returns on the lagged value spread is 0.74. This ratio is small relative to our estimates of 3.47 (in case we exclude financials) and 2.12 (in case we use industry–adjusted book–to–market), which both fall in the far right tail of the simulated distribution.

In the remainder of Table 1.A.2, we see that the value spread predicts value returns similarly in the other asset classes, although the evidence is slightly weaker statistically. This is partly due to a lack of power in asset classes with shorter sample periods (e.g., global government bonds and global stock indexes). Let us focus on the High–minus–Low strategies for interpretation. At the annual horizon, the coefficient estimate on the value spread ranges from 4.34% ($t^{hd} = 2.09$) for industries to 7.00% ($t^{hd} = 2.36$) for global government bonds. At the two–year horizon, the coefficient estimates range from 6.86% ($t^{hd} = 1.94$) for commodities to 15.60% ($t^{hd} = 2.80$) for global government bonds. The value spread captures a considerable fraction of the variation in two–year value returns at R^2s of 16.86% for industries, 8.91% for commodities, 27.32% for currencies, 33.92% global government bonds, and 5.21%

for global stock indexes. Similar to what we find for U.S. individual equities, the ratio of the coefficient on the value spread relative to the intercept is quite large in all asset classes. This ratio is about one for currencies and commodities. The ratio is considerably larger than one in the remaining asset classes, which is partly due to the fact that the unconditional value effects are small in some cases. For instance, the unconditional average value return is only 4 bps per month for industries, which is consistent with the literature. We show that the industry value premium is large conditionally and varies over time with the value spread, just like it does in all the other classes we study. In fact, comparing the unconditional evidence in Panel A to the conditional evidence in Panel B, we conclude that the conditional variation in value premia is actually more similar across asset classes than is the unconditional value premium.

1.3.2 Pooled Predictive Regressions

We next employ pooled tests for the following value strategies: U.S. individual equities (book–to–market excluding financials and industry–adjusted book–to–market), industries, commodities, currencies, global government bonds, and global stock indexes. These pooled tests provide insight on the joint time variation in expected value premia implied by time variation in the value spread. Panel A of Table 1.A.3 presents the results for the following regression:

$$R^{x}_{c,t+1:t+h} = a_h + b_h \,\mathrm{VS}^{x}_{c,t} + e^{x}_{c,t+1:t+h}, \tag{1.4}$$

where c denotes an asset class and $x \in \{H\text{˘}L, Rank\}$. We add in these pooled tests a longer four–year horizon, $h = 48$ months, because pooling increases statistical power.

[Insert Table 1.A.3 about here]

In Panel A of Table 1.A.3, we find that the joint evidence for value return predictability is strong for both types of portfolios. For instance, for the High–minus–Low

Table 1.C.2 of the Internet Appendix shows similar evidence for global government bonds when we use the alternative value measure (5–year change in yield instead of ˘5–year return) and slightly weaker when using synthetic bond futures returns.

Because the value spread is mean zero in all asset classes, the coefficient estimate b_h is identical when we include asset class fixed effects in Equation (1.4).

portfolio, the coefficient on the value spread is significant, with a *t*–statistic above three at the monthly horizon and a *t*–statistic above five for horizons that are over a quarter. The coefficient estimates are economically large too. Looking at the ratio of the estimated coefficient to the intercept, we see that the standard deviation of expected returns implied by the value spread is about 81% to 147% (23% to 77%) larger than the unconditional value premium in the pool of High–minus–Low (rank–weighted) value strategies. For instance, the coefficient estimate is 6.41% at the annual horizon, which is relative to an unconditional average value premium of 2.64% (the intercept). Consistent with these coefficient estimates, the R^2 increases with the horizon, and exceeds 20% at the 24– and 48–month horizons. The idea that the value spread contains information for value returns at long horizons is further supported by the evidence in Figure 1.A.2. In this figure, we predict future value returns over consecutive semi–annual periods after portfolio formation. We find that the coefficient on the value spread is decreasing as time passes, but remains positive and marginally significant up to about four–and–a–half years after portfolio formation.

In Panel B of Table 1.A.3, we present an alternative way of looking at the joint strength of the value return predictability. We regress in the time series the across–asset class average value return on the across–asset class average value spread. We again see coefficient estimates on the value premium that are statistically significant and economically large. The R^2s at the 24– and 48–month horizons are even larger at over 35%, since averaging smooths out some noise in the individual value strategies. These results testify to the joint strength of the value premium predictability, but they also suggest that there is common variation in value premia across asset classes.

We run the same tests, but exclude the value strategies for individual equities. We see in Panels C and D in Table 1.A.3 that value returns in the alternative asset classes are jointly strongly predictable by the value spread, with a ratio of coefficient–to–intercept that is well above one in both the pooled and average–on–average specification. Finally, in Table 1.C.5 of the Internet Appendix, we show that the value spread predicts value returns in the pool of asset classes in both subsamples (split around June 1994). This result suggests that value return predictability is not solely driven by the highly popularized value episodes around the dot–com bubble in the late 1990's and around the recent global financial crisis.

Next, we examine whether our results are explained by time–varying exposure to a market benchmark, as in a conditional Capital Asset Pricing Model (CAPM). The literature shows that an unconditional CAPM does not explain the value premium. However, Campbell and Vuolteenaho, 2004 find that the value spread is a significant time series predictor of equity market returns. Hence, if the market beta of value strategies varies over time with the value spread, this could explain the time series variation in value premia. To see whether this is the case, we run the pooled predictive regression of value returns on the value spread, but control for market exposure in each asset class. We consider a model with constant betas, as well as a model that allows the market beta in each asset class c to vary over time with the value spread: $\beta_{MKT,c,t} = \beta_{0,c} + \beta_{1,c} VS_{c,t}$. The results are reported in Table 1.A.4. In Panel A, we use the CRSP value–weighted stock market portfolio — the most common proxy for the CAPM market portfolio in the literature — as the benchmark in all asset classes. In Panel B, we use an equal–weighted portfolio of the securities in each alternative asset class as the benchmark.

[Insert Table 1.A.4 about here]

In the first test with constant betas, we find that the estimated coefficient on the value spread is similarly large in economic magnitude and significance to what we report in Panel A of Table 1.A.3. This finding is intuitive: Since the unconditional market betas of the value strategies are small, time variation in expected market returns alone cannot explain the large amount of time variation in value premia we find. The model with time–varying betas shows that the interaction between time variation in market betas and time variation in the market risk premium cannot explain our findings either. There is no consistent pattern across asset classes in the coefficients $\beta_{1,c}$, which are mostly insignificant. Hence, the predictability of value returns due to the value spread is again largely unaffected. We conclude that our results are not driven by the predictive relation between market returns and the value spread and thus a conditional CAPM. In line with this conclusion, Table 1.C.6

of the Internet Appendix shows that the value spread is not a robust predictor of market returns in the pool of asset classes we study.

1.4 Out–of–Sample Value Timing and Rotation

In this section, we present a number of out–of–sample strategies that take advantage of the information in the value spread in real time.

1.4.1 Value Timing in Individual Equities

We construct a linear timing strategy for value in individual equities by constructing a value spread that is standardized in month t using only historical information:

$$VS_{t,His} = \frac{\sum_{s=0}^{11} VS_{t-s}/12 - \sum_{s=12}^{t-1} VS_{t-s}/(t-12)}{\sigma(VS_{1:t-12})}. \tag{1.5}$$

Thus, $VS_{t,His}$ indicates whether the average value spread over the last twelve months is historically large. We take an annual average to accommodate that return predictability using the value spread strengthens with the horizon. To ensure that our dynamic strategies are not extreme, we truncate the standardized signal at ± 2.

Table 1.A.5 presents summary performance statistics for three strategies: a unit weight strategy that captures the unconditional value premium, a linear timing strategy where $VS_{t,His}$ dollars are invested in both the long and short position of the value strategy, and a combined strategy where $1 + VS_{t,His}$ dollars are invested in the long and short position. We consider 2×2 variations of these strategies: using either (i) the book–to–market signal excluding financials or the industry–adjusted book–to–market ratio, and (ii) the High–minus–Low decile portfolio or the rank–weighted portfolio. To make the results comparable across strategies, we standardize each return series to have an ex ante annualized standard deviation of 15%. In particular, we follow 2012 and estimate ex ante variance using an exponential weighting scheme: $\sigma^2_{R_{t+1}} = \sum_{i=0}^{\infty}(1-\delta)\delta^i(R_{t-i}-\widehat{R}_{t+1})^2$, where δ is chosen so that the center of mass of the weights is two years and $\widehat{R}_{t+1}$ is the exponentially–weighted average return computed similarly. We then rescale the return on the position as follows: $R_{t+1,15\%} = \frac{R_{t+1}}{\sigma_{R_{t+1}}} \times \frac{15\%}{\sqrt{12}}$.

[Insert Table 1.A.5 about here]

We next compare the performance of the linear timing strategy to the unit weight strategy. For the High–minus–Low decile book–to–market strategy that excludes financials, we find an average return for the linear timing strategy of 60 bps ($t = 2.77$) per month, which is 62 bps higher than the average return of the unit weight strategy. For the alternative value strategies, this difference ranges from 25 bps (industry–adjusted book–to–market, rank–weighted value strategy) to 49 bps (excluding financials, rank–weighted value strategy). Because this increase in average returns is not accompanied by a proportional increase in standard deviation, the Sharpe ratio of the linear timing strategies is relatively large as well, ranging from 0.34 to 0.41 (annualized). For comparison, over the alternative value strategies, the largest Sharpe ratio for the unit weight strategy is 0.24 (industry–adjusted book–to–market, rank–weighted value strategy). In the combined strategy, the returns of its two components are summed, which yields attractive average returns ranging from 56 bps to 87 bps, but Sharpe ratios that are similar to the linear timing strategies. In all, these results suggest that investors can use the information in the value spread to time value in the stock market. Moreover, this timing strategy is an attractive complement to an unconditional value strategy.

These conclusions are supported further when we look at alphas relative to the market portfolio (of large stocks accounting for 90% of total market cap in CRSP), as well as the Fama and French (1993) three–factor model. We find that the CAPM alpha of the linear–timing strategies are large at about 55 bps and significant. This number is relative to a CAPM alpha for the unit weight strategy, which ranges from an insignificant 17 bps to a marginally significant 34 bps. This result suggests that conditional value strategies are attractive on top of an indexed market strategy and more so than an unconditional value strategy. The three–factor alpha of the linear timing strategies is also large and significant, at over 47 bps. This result suggests that the conditional value strategies using only the largest stocks are attractive even relative to unconditional value strategies using all stocks in the CRSP file.

In Table 1.C.7 of the Internet Appendix, we present results for the same strategies using alternative market cap cutoffs of 75% and 95%. Although there is some variation in magnitude, we find again that conditioning on the value spread improves performance relative to a unit weight value strategy and typically also relative to a market strategy and the Fama–French three–factor model. With the 95% cutoff, we use on average 740 stocks per month (relative to 495 using the 90% cutoff), which increases transaction costs. Including these relatively smaller stocks does increase the unconditional value premium, which is consistent with the literature. Interestingly, with the 75% cutoff, we use on average only 212 stocks per month, which lowers the transaction costs considerably.

1.4.2 Value Timing and Rotation in the Pool of Value Strategies

We next examine value timing and rotation in the pool of asset classes. To start, we run a pooled regression of value returns on a dummy variable that indicates whether the current value spread in an asset class is above the historical average:

$$R^x_{c,t+1:t+h,15\%} = a_h + b_h I_{VS^x_{c,t,His}>0} + e_{c,t+1:t+h}, \tag{1.6}$$

where c denotes an asset class and $x \in \{H\tilde{}L, Rank\}$, and $VS^x_{c,t,His}$ is defined as in Equation (1.5). The subscript indicates that we standardize each return series to have an ex ante annualized standard deviation of 15% to ensure comparability across asset classes.

Table 1.A.6 presents the results. For the one–month horizon, we find that the coefficient estimate of b is large and significant at 60 bps ($t = 2.91$) and 57 bps ($t = 2.53$) for the High–minus–Low and rank–weighted portfolios, respectively. Combined with the estimated intercept, these numbers imply that the average return of a value strategy that invests only in an asset class when $VS_{c,t,His} > 0$ equals 52 bps and 55 bps per month, respectively. These returns translate to annualized Sharpe ratios of about 0.4. In comparison, the Sharpe ratio of investing when $VS_{c,t,His} \leq 0$ is negative, although small and insignificant. This evidence suggests that investing in value in a typical asset class is only attractive when the value spread in that asset class is historically large. Finally, the regression results for longer horizons suggest

that strategies that rebalance at a lower frequency than every single month, are likely more attractive.

[Insert Table 1.A.6 about here]

We next examine strategies that rotate value across asset classes. As a benchmark, we consider an unconditional value strategy where $1/N_t$ is invested in each of N_t available value strategies (out of the maximum of seven) in each sample month t. Next, we consider a value rotation strategy where asset classes are overweighted (underweighted) when the value spread is high (low) relative to the other asset classes. We consider two alternative weighting schemes. The first rotation strategy takes a position in each asset class c in month t equal to:

$$w_{c,t}^{rot,1} = q_t \left(VS_{c,t,His} - \sum_{c=1}^{N_t} VS_{c,t,His}/N_t \right), \tag{1.7}$$

where the scalar q_t ensures that the total weight in the long and short position equals one. In the second strategy, with weights denoted $w_{c,t}^{rot,2}$, an equal weight is invested in each asset class with $VS_{c,t,His}$ above (below) the mean value spread across asset classes. We calculate performance measures for these two long–short rotation strategies, as well as for a combination with the unconditional strategy.

[Insert Table 1.A.7 about here]

The first block of results in Panel A in Table 1.A.7 is for the High–minus–Low portfolios. We find that the two rotation strategies outperform the unconditional strategy. For instance, the average return and annualized Sharpe ratio of the linear rotation strategy equal 68 bps ($t = 3.12$) and 0.52, respectively, which is large relative to 8 bps ($t = 0.74$) and 0.12 for the unconditional strategy. The equal–weighted rotation strategy performs similarly at an average return of 63 bps ($t = 3.30$) and has a Sharpe ratio of 0.55. In the second block of results for the rank–weighted value strategies, we find that the value rotation strategies outperform the unconditional strategies as well, albeit by a slightly smaller margin. The Sharpe ratio is about 0.45 for the two rotation strategies and 0.23 for the unconditional strategy. Similar to the case of individual equities, we find that the combined strategies (unconditional value plus value rotation) perform about as well as the rotation strategies in Sharpe ratio. We conclude that the value spread can be used by investors to rotate value across asset classes in real time and this strategy is attractive relative to a strategy that invests unconditionally in value in all asset classes. Thus, investing in value is most attractive in asset classes with value spreads that are large compared to other asset classes.

Table 1.A.4 also reports the abnormal return, or α, of the rotation strategies relative to an equal–weighted portfolio of the market strategies in each asset class (as shown in Panel B of Table 1.A.4). Note, this aggregate market benchmark is well–diversified and presents a tough benchmark for the dynamic strategy to beat. The value rotation strategies have lower α's than average returns, suggesting that there is some market exposure. However, the reduction is generally small (about 10 bps), such that the remaining abnormal return is economically large ($>$ 49 bps) and statistically significant. We conclude that rotation strategies may be an attractive addition to a portfolio that diversifies unconditionally across these markets. In contrast, the unconditional value strategy obtains an α that is about one–third in magnitude of the rotation strategy and is insignificant in all four cases.

In Panel B of Table 1.A.7, we present the fraction of months in which the long and short leg of the rotation strategies invest in each asset class. We see that the strategies diversify across different asset classes over time: no asset class is present in either leg for more than one–third of the sample. We next decompose the average return of the long–short rotation strategies across asset classes. For both rotation strategies, we find in Panel C that about 60% of the average return is derived from the alternative asset classes. Currencies is the asset class with the largest contribution. The value strategies using individual equities and equity indexes also contribute substantially. Thus, we conclude that not only U.S. individual equities, but also the alternative asset classes, contribute to the benefits of value rotation.

Table 1.C.8 in the Appendix presents the results from timing strategies for the alternative asset classes (analogous to Table 1.A.5). We find that the return from a linear timing strategy is non–negligible economically (ranging from about 20 to 30 bps per month) for industries, currencies, global government bonds, and global stock indexes. These effects are insignificant, however, partly due to the shorter sample period dictated by data availability. This result highlights an important difference

between value timing and rotation. Even if timing value in a specific asset class is difficult, the value spread in that asset class may contain valuable information for rotating value across asset classes. Indeed, the evidence in Table 1.A.7 suggests that comparing the value spread in currencies to other asset classes provides valuable information to determine when to go long (or short) the currency value strategy.

1.5 Common Value and Economic Drivers of Value Return Predictability

In this section, we investigate (i) the strength of comovement between the expected returns of value strategies in different asset classes, and (ii) whether this comovement is driven by economic fundamentals. Throughout this section, we discuss the results for the High–minus–Low value strategies. By and large, identical results for the rank–weighted strategies are reported in Tables 1.C.9, 1.C.10, and 1.C.11 of the Internet Appendix.

1.5.1 Common Versus Asset Class–Specific Value

We start by investigating how much predictability in value strategies is common across the different asset classes. In Table 1.A.8, we present the results from a pooled predictive regression on the two components of the value spread defined in subsection 1.2.4. The common component, denoted VS_t^{Com}, is the first principal component of value spreads. The asset class–specific component, denoted VS_t^{Spec}, is the residual from a regression of the value spread in each asset class on the common component.

[Insert Table 1.A.8 about here]

In isolation, the coefficient estimates on the common as well as the asset class–specific component of the value spread are statistically and economically significant at all horizons. Thus, both contain information about future value returns. The estimated coefficients are identical in a joint test, because the two components are orthogonal.

One can decompose the value spread arbitrarily in two orthogonal components that obtain a joint regression R^2 identical to what we present here. However, such arbitrary components will in general not predict value returns in isolation, especially a component that is restricted to not vary across asset classes.

More interesting is the relative contribution of each component to the total R^2 in the joint test. This R^2 ranges from 0.56% at the monthly horizon to 10.63% at the annual horizon, 18.78% at the two–year horizon, and 24.07% at the four–year horizon. At these horizons, the common component contributes 0.35%, 5.79%, 12.84%, and 18.45% of the explained variation, respectively. In other words, about 60% of the predictability of value returns in the pool of value strategies is driven by the common component at horizons from one month up to one year. At horizons of two and four years, the common component contributes even more at about 68% and 77%, respectively. Recall that the common component explains about half of the variation in value spreads. We thus find that it explains an even larger share of the predictability of value returns; for instance, more than two–thirds at long horizons. The asset class–specific components contribute relatively more at short horizons. This latter finding is consistent with the idea that limits to arbitrage prohibit the fast movement of money across asset classes.

In Tables 1.C.12 and 1.C.13 of the Internet Appendix, we show that these conclusions are not sensitive to the definition of the common component. A first alternative definition uses the first principal component from a standard principal component analysis performed on the panel of value spreads. However, in this case, the panel is balanced with an algorithm that recursively projects the value spread in an asset class with a shorter sample on the value spreads that are available over the full sample. A second alternative definition is the average value spread over the asset classes with available data in month t. The advantage of this definition is that the common component is directly observable and does not suffer from errors–in–variables bias. The disadvantage of this decomposition is that we assume equal loadings on the common component across asset classes. The correlation between our measure of the common component and this alternative measure is large, at 0.95, which suggests that this disadvantage should not affect the results much. For both alternative definitions and for all horizons, we find that the common component contributes about two–thirds of the predictability of value returns.

Overall, these results suggest that the common component of the value spread

The asset class–specific component is then simply the difference between the value spread in an asset class and common value.

contributes more than the asset class–specific component to value return predictability. A component of the value spread that is common across asset classes and determines about two–thirds of the variance of expected returns to value strategies is interesting from a theoretical perspective. Asset pricing models now must also explain that expected returns of value strategies rise and fall globally. As highlighted in (Cochrane, 2011a, p. 1060): "It is not enough to simply generate temporary price movements in individual securities."

1.5.2 Economic Drivers of the Components of Value

In this subsection, we analyze the economic sources of variation in the common and asset class–specific components of the value spread using state variables from recent asset pricing models. In particular, we run time series regressions of the following form:

$$VS_t^{Com} = k_0 + k_1' Z_t + u_t^{Com} \text{ and} \tag{1.8}$$

$$VS_{c,t}^{Spec} = k_0 + k_1' Z_t + u_{c,t}^{Spec}, \tag{1.9}$$

where Z_t is a particular set of state variables or risk proxies. We report the results from Equations (1.8) and (1.9) in Panels A and B of Table 1.A.9, respectively.

Intermediaries are the marginal investors in many asset markets. Hence, their marginal value of wealth is a plausible pricing kernel for a broad set of securities and may drive common variation in expected returns. Recent intermediary–based asset pricing models (e.g., He and Krishnamurthy, 2012; He and Krishnamurthy, 2013; Brunnermeier and Sannikov, 2014) show that the intermediary sector's net worth (or equivalently the reciprocal of leverage, defined as assets over equity) is the key determinant of its marginal value of wealth. We analyze the link between the aggregate leverage of financial intermediaries and the common component of the value spread in row 1 of Panel A of Table 1.A.9. We find a strong relation, with variation in leverage accounting for almost 50% of the overall variation in common value. This time series evidence complements the large and growing body of literature showing that the leverage of financial intermediaries has strong cross–sectional predictive power for returns in various asset classes (Adrian, Etula, and Muir, 2014; He, Kelly, and Manela, 2017).

[Insert Table 1.A.9 about here]

Next, guided by theory (Brunnermeier and Pedersen, 2009) and empirical evidence (Adrian and Shin, 2010) that implies a close link between funding liquidity and the balance sheet of the financial sector, we investigate the relation between illiquidity and common value. Following Nagel, 2016, we proxy for illiquidity with the repo/T–bill spread. In row 2 of Table 1.A.9, we find that illiquidity also explains considerable variation in common value with an R^2 of almost 40%. Jointly, leverage and illiquidity explain 64% of the variation in common value (row 3). Both variables enter significantly, with economically large coefficients. For a standard deviation increase in leverage and illiquidity, common value increases by 0.39 and 0.27 standard deviations, respectively.

Recent literature acknowledges that financial intermediary leverage is endogenous and its cycles may simply reflect movements in aggregate risk aversion (Campbell and Cochrane, 1999; Menzly, Santos, and Veronesi, 2004; Santos and Veronesi, 2016). Inspired by Campbell and Cochrane, 1999, who argue that the price–to–dividend ratio is nearly linear in the surplus consumption ratio, we next explore the link between common value and the dividend yield. In row 4 of Table 1.A.9, we find that the dividend yield explains lots of variation in common value, with an R^2 of almost 69%. This result is consistent with the idea that the value spread widens when risk aversion is high. We then investigate the extent to which leverage and liquidity are just a manifestation of time–varying risk aversion (as proxied by the dividend yield). We see in row 5 that intermediary leverage, illiquidity, and the dividend yield are all significant and jointly capture about three–quarters of the variation in the common component of the value spread. However, as the intimate link between leverage cycles, liquidity dry–up, and risk aversion would suggest (Santos and Veronesi, 2016), the magnitude and statistical significance of the individual coefficients falls upon joint inclusion of the variables.

Leverage is measured as the inverse of the squared intermediary capital ratio, as in 2017. This measure of leverage is based on market prices (market leverage) and, in the model of Santos and Veronesi, 2016, the debt–to–wealth ratio is monotonically decreasing in the surplus consumption ratio (see their Corollary 13).

Based on the evidence so far, we conclude that common value is large when, in bad times, intermediaries' balance sheets get shocked or aggregate risk aversion is high, or both. Consistent with this result, we see in row 6 of Table 1.A.9 that common value is higher by about 0.46 standard deviations during global recessions. In row 7, we find that this conclusion is also robust to controlling for additional state variables. Following 2017, who link the value spread in equities to business cycle risk, we include the Chicago Fed National Activity Index (CFNAI). We also include the 2015 real uncertainty index. As pointed out by Nagel, 2016, liquidity may be in part driven by the level of uncertainty, since a high level of risk can erode agents' trust that bank deposits are a good store of liquidity. Finally, we include the BAA–AAA corporate bond default spread, a popular proxy for cyclical variation in risk premia. In this "kitchen sink" regression, all three additional state variables are insignificant and the R^2 increases only marginally relative to the three–variable model in row 5 (79% vs. 74%). In all, the common value spread is high in bad times, which are modeled well as a combination of high leverage, illiquidity, and a large dividend yield.

This conclusion holds true also in changes. Rows 8 and 9 in Table 1.A.9 display the results obtained when using innovations from an AR(1) model in the common component and the state variables. We find that the innovations in common value are driven positively and significantly by innovations in these state variables. Innovations in leverage and liquidity together explain 41% of the variation in innovations in common value, whereas adding innovations in the dividend yield increases the R^2 to 57%. Among the three state variables, the dividend yield (liquidity) is relatively more (less) important.

For the asset class–specific components in Panel B in Table 1.A.9, we focus on the kitchen sink regression to have an upper bound on what risk can explain. Jointly, the risk proxies explain a considerable fraction of the variation in the asset class–specific value spread in some asset classes, with R^2's ranging from 9% for the equity value strategy excluding financials to 48% for industries. However, the loadings on individual risk proxies vary dramatically across asset classes, in both magnitude and significance.

We thank an anonymous referee for suggesting this analysis.

[Insert Table 1.A.10 about here]

Next we examine how much of the predictive ability of the common component of the value spread is captured by the part that is correlated with the risk proxies (the predicted value spread in the kitchen sink specification, $k_0 + k_1' Z_t$, of Equation (1.8)) and how much by the part that is orthogonal (the residual, u_t^{Com}). Focusing on the decomposition of R^2, we see in Panel A of Table 1.A.10 that both the explained and orthogonal part are significant in predicting value returns. The fraction of value return variation attributed to the explained part of common value increases in horizon and ranges from about two–thirds (at short horizons) to three–fourths (at the four–year horizon). Consistent with the association between these risk proxies and common value, we show in Table 1.C.14 of the Internet Appendix that the first principal component of the risk proxies predicts value returns significantly in isolation. However, it is common value that dominates in predicting value returns in a joint test. Panel B of Table 1.A.10 provides the results of a decomposition of the asset class–specific value return predictability. In contrast to the case of common value, we find that the part of the asset class–specific component of the value spread that is orthogonal to the risk proxies is relatively more important for predicting value returns than the explained part.

1.5.3 The Role of the Equity Value Spread

Although the value spread also predicts value returns outside U.S. individual equities, it is an interesting question as to how much of the value return predictability across asset classes is associated with variation in the value spread in U.S. individual equities. The fact that the dividend yield is the state variable with the largest correlation to common value suggests that U.S. individual equity valuations are relatively important. To answer this question, we conduct a pooled regression of value returns in all asset classes on the equity value spread. We report the results in Panels A and B of Table 1.C.15 of the Internet Appendix. We find that the equity value spread predicts returns about as well as the common component (see Table 1.A.8). When we exclude the two equity value returns from the test assets (Panels C and D), the value spread remains marginally significant (at the one– and two–year horizons) in both specifications, but the fraction of explained variation drops considerably. Thus, we find weak evidence for across–asset class value return predictability due to the equity value spread. This finding implies that our measure of common value extracts additional relevant information from the value spreads in the alternative asset classes.

1.5.4 Interpretation

The evidence in this section suggests that the majority of value return predictability in different asset classes is driven by a single common component. These results for common value call for a general framework, where investors shy away in bad times from holding different risky assets, such as individual equities, global stock indexes, industries, and commodities with low valuation ratios. Consequently, value spreads widen simultaneously when discount rates (and thus expected value returns) are high. The motivation is that common value return predictability is closely associated with proxies for the risk of financial intermediaries (such as market leverage and funding liquidity) and risk aversion (dividend yield). This common time–varying component of value premia is present in asset classes with potentially different investors and institutional factors.

Our analysis of the asset class–specific components of the value spread indicates the presence of additional risk and mispricing factors in time–varying value premia. For risk, we find that correlation between risk proxies, such as leverage and uncertainty, and asset class–specific value contributes to the predictability of value returns. The loadings of specific value on these risk proxies vary across asset classes, which points to heterogeneity in risk exposure as an important driver of asset class–specific value return predictability. For mispricing, we show that it is the component of the asset class–specific value spread that is orthogonal to our large set of risk proxies that contributes relatively more to the predictability of value returns. Limits to arbitrage may impair the ability of investors to undo mispricing specific to different asset classes.

In Table 1.C.16 of the Internet Appendix, we show that common value is relatively more important in the recent subsample post–1994, which is broadly consistent with these interpretations. Common value is strongly associated with proxies for the risk of financial intermediaries and financial intermediation has become progressively more important over time. Moreover, if limits to arbitrage partially drive the asset class–specific components of value return predictability, one would expect these components to become less important over time.

1.6 Conclusion

Value premia are strongly time–varying and comove across asset classes. We show that returns to value strategies in U.S. individual equities, industries, commodities, currencies, global government bonds, and global stock indexes are predictable in the time series using the value spread. This predictability is statistically significant and economically large. Our coefficient estimates suggest that expected value returns vary by at least as much as their unconditional level. To understand the drivers of this time variation, we decompose the value spread into a common component, defined as the first principal component of value spreads, and asset class–specific components. While the common component captures about half of the total variation in value spreads, it captures more — about two–thirds — of the total variation in expected value returns across asset classes. The dividend yield, intermediary leverage, and an illiquidity premium capture the bulk of the time variation in common value. Furthermore, common value return predictability is persistent and indicates that expected value returns are countercyclical. Thus, we argue that the main source of common variation in value premia is compensation for risk. On the contrary, both risk and mispricing contribute to the asset class–specific components of value return predictability.

These findings are new to the literature and are only detected in a joint examination of different asset classes. Our results confirm the basic intuition that risk premia comove strongly across asset classes, for which empirical evidence to date is scarce.

1.A Appendix

1.A.1 Figures

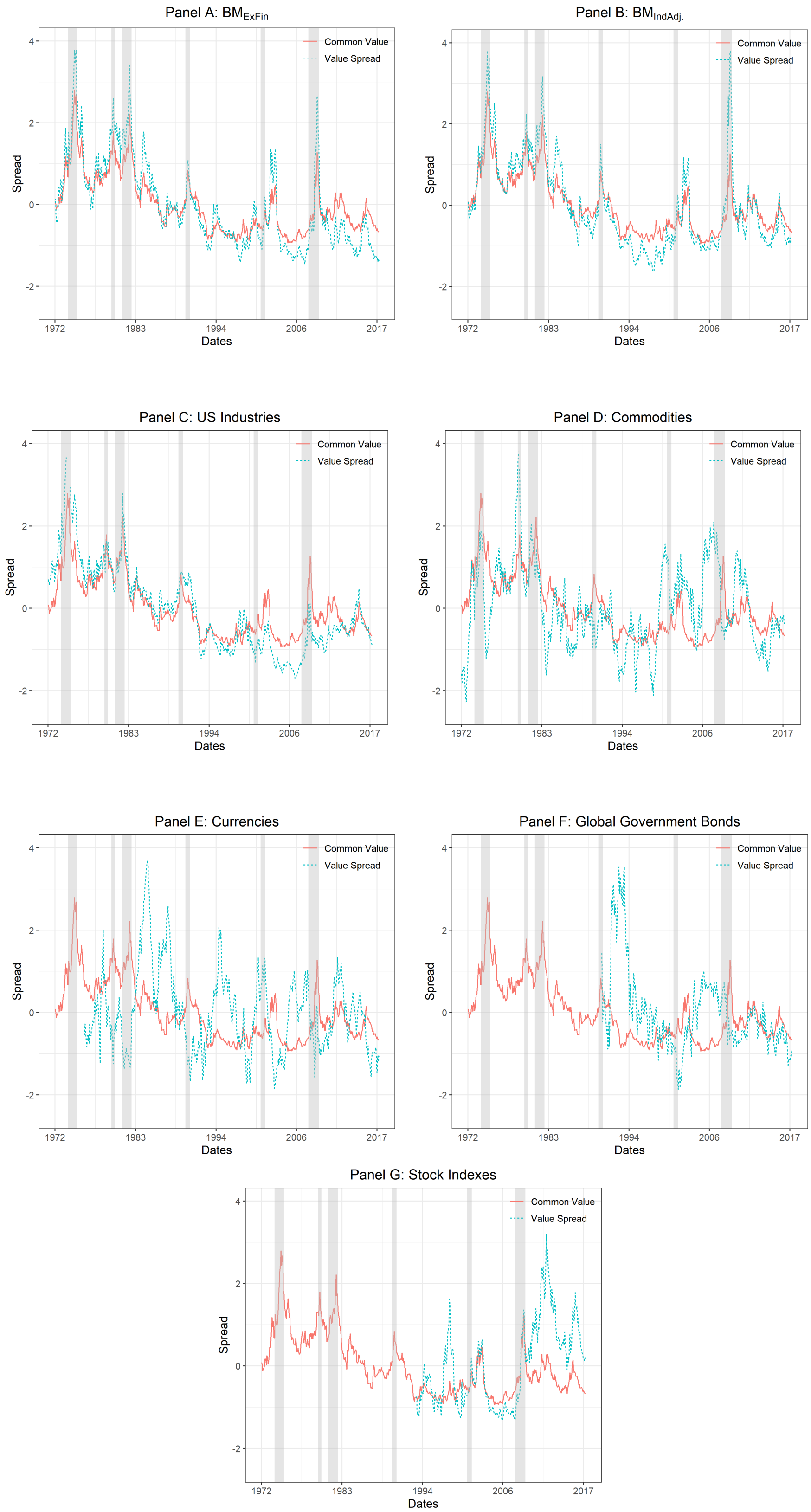

FIGURE 1.A.1: **The Value Spread in Different Asset Classes**
This figure presents the time series of standardized value spreads (in blue) for the following seven value strategies: (i) individual equities: book-to-market excluding financials, (ii) individual equities: industry adjusted book-to-market, (iii) US industries, (iv) commodities, (v) currencies, (vi) global government bonds, and (vii) global stock indexes. In each panel, we also present the time series of common value (defined as the first principal component of value spreads). The shaded areas represent NBER recessions.

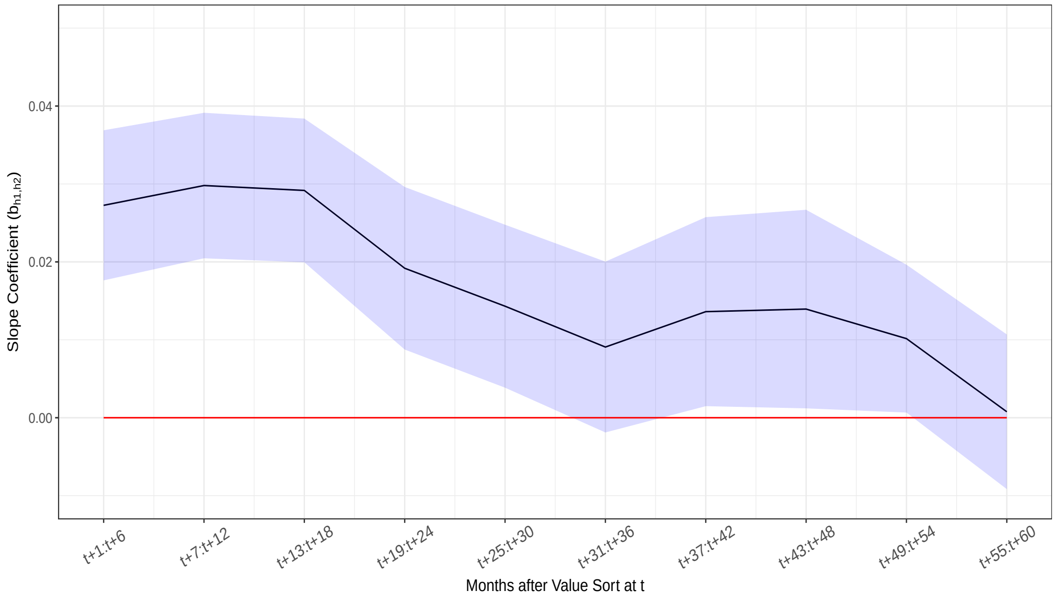

FIGURE 1.A.2: **Semi-Annual Future Value Returns on the Value Spread at Time** t
This figure presents the coefficient estimates ($\pm$ two standard errors) from pooled predictive regressions of non-overlapping semi-annual value returns on the value spread: $R_{c,t+h_1:t+h_2} = a_{h_1,h_2} + b_{h_1,h_2}\ \mathrm{VS}^x_{c,t} + e_{c,t+h_1:t+h_2}$. The semi-annual value returns range from six months ($h_1 = 1, h_2 = 6$) to five years ($h_1 = 55, h_2 = 60$) after the value spread is observed in month t. We include in the pool of value strategies the High-minus-Low value return in (i) individual equities: book-to-market excluding financials, (ii) individual equities: industry adjusted book-to-market, (iii) US industries, (iv) commodities, (v) currencies, (vi) global government bonds, and (vii) global stock indexes. The value spread is standardized.

1.A.2 Tables

TABLE 1.A.1: **Correlations and Factor Structure of Value Spreads**
Panel A of this table presents the correlation matrix of the High-minus-Low value spread in different asset classes (p-value in parentheses), with first-order autocorrelations on the diagonal. We consider two measures of value for individual equities: book-to-market excluding financial firms (BM_{ExFin}) and industry-adjusted book-to-market (BM_{IndAdj}). For seventeen industries, the value measure is the market cap-weighted book-to-market ratio. In all three cases, the data covers the period from 1972 to 2017. Market cap in the denominator of the book-to-market ratio is updated monthly and we use only the largest stocks that cumulatively account for 90 percent of the total market cap in CRSP. For commodities, the sample ranges from 1972 to 2017 and we measure value as the negative of the five-year spot return (−5-year return). For currencies, the sample ranges from 1976 to 2017 and we measure value as the inflation-adjusted negative five-year spot return (Inf. adj. return). For global government bonds, the sample ranges from 1991 to 2017 and we measure value as the negative of the five-year return of a one-month futures on a 10-year global government bond (−5-year return). For global stock indexes, the sample ranges from 1994 to 2017 and we measure value using the MSCI Book-to-Price ratio ($MSCI_{BP}$). Panel B presents the loadings of the first three principal components of the seven value spreads and the fraction of total variance explained by each component, which are extracted using the approach of Stock and Watson, 2002.

Panel A: (Auto-) Correlations

Asset Class	BM_{ExFin}	BM_{IndAdj}	US Industries	Commodities	Currencies	Bonds	Equity Indexes
BM_{ExFin}	0.97	0.95	0.86	0.34	0.04	–0.13	0.20
	(0.00)	(0.00)	(0.00)	(0.00)	(0.38)	(0.02)	(0.00)
BM_{IndAdj}		0.97	0.80	0.39	0.00	–0.17	0.40
		(0.00)	(0.00)	(0.00)	(0.96)	(0.00)	(0.00)
US Industries			0.98	0.18	–0.01	–0.08	0.59
			(0.00)	(0.00)	(0.86)	(0.13)	(0.00)
Commodities				0.95	–0.12	–0.08	–0.09
				(0.00)	(0.01)	(0.17)	(0.11)
Currencies					0.95	0.01	–0.10
					(0.00)	(0.83)	(0.10)
Bonds						0.95	–0.33
						(0.00)	(0.00)
Equity Indexes							0.97
							(0.00)

Panel B: Principal Components

Loadings	BM_{ExFin}	BM_{IndAdj}	Industries	Commodities	Currencies	Bonds	Equity Indexes	Var. Exp.
PC1	1.34	1.34	1.29	0.59	0.07	0.16	1.17	51%
PC2	–0.22	–0.09	–0.29	0.78	–1.66	–1.77	0.60	26%
PC3	0.03	0.14	–0.64	2.28	0.94	–0.01	–0.70	12%

TABLE 1.A.2: **Time Series Predictive Regressions of Value Returns on the Value Spread**
This table presents the results from predictive regressions of monthly value returns on the value spread: $R_{c,t+1:t+h} = a_h + b_h VS_{c,t} + \varepsilon_{t+1:t+h}$, in all asset classes c. Value is measured in each asset class as explained in Table 1.A.1. Value returns are calculated from two portfolio strategies: High-minus-Low ($H-L$) or rank-weighted ($Rank$). For individual equities, we sort stocks in ten value-weighted deciles. For the remaining asset classes, we sort securities in two equal-weighted portfolios, split at the median of value measures in that asset class. Panel A reports unconditional performance statistics for monthly value returns in each asset class. Panel B presents the regression results for overlapping holding period returns of $h = 1, 12, 24$ months. For the sake of comparison across asset classes, value spreads, $VS_{c,t}$, are standardized to have mean equal to zero and variance equal to one; and, value returns are scaled to have an annual standard deviation of 15%. t^{nw} and t^{hd} indicate t-statistics calculated using Newey and West, 1987a and Hodrick, 1992 standard errors, respectively.

Panel A: Unconditional Performance (Monthly Returns)

		$H-L$			$Rank$		
Asset Class	Value Measure	Avg. ret.	t	Sharpe	Avg. ret.	t	Sharpe
Ind. Equities	BM_{ExFin}	0.15	0.83	0.04	0.21	1.11	0.05
	BM_{IndAdj}	0.32	1.73	0.07	0.49	2.65	0.11
Industries	BM	0.04	0.19	0.01	0.05	0.25	0.01
Commodities	−5-year return	0.30	1.60	0.07	0.27	1.48	0.06
Currencies	Inf. adj. return	0.41	2.13	0.09	0.46	2.38	0.11
Gov't Bonds	−5-year return	–0.04	–0.16	–0.01	–0.05	–0.20	–0.01
Stock Indexes	$MSCI_{BP}$	0.22	0.87	0.05	0.32	1.27	0.08

Panel B: Predictive Regressions of Value Returns on the Value Spread

			$H-L$							$Rank$						
		h	a	b	t_a^{nw}	t_b^{nw}	t_a^{hd}	t_b^{hd}	R^2	a	b	t_a^{nw}	t_b^{nw}	t_a^{hd}	t_b^{hd}	R^2
Ind. Equities	BM_{ExFin}	1	0.15	0.48	0.80	2.35	0.84	2.20	1.03	0.21	0.35	1.06	1.53	1.11	1.52	0.47
		12	2.01	6.97	0.96	3.89	0.92	3.24	13.94	3.01	5.54	1.30	2.30	1.38	2.45	7.96
		24	4.49	16.96	1.09	4.88	1.04	4.24	30.86	7.19	15.62	1.53	2.99	1.65	3.66	22.81

Continued

			H − L							Rank						
		h	a	b	t_a^{nw}	t_b^{nw}	t_a^{hd}	t_b^{hd}	R^2	a	b	t_a^{nw}	t_b^{nw}	t_a^{hd}	t_b^{hd}	R^2
Ind. Equities	BM_{IndAdj}	1	0.32	0.65	1.65	2.76	1.73	2.75	2.09	0.49	0.54	2.52	2.02	2.65	2.04	1.40
		12	4.59	9.72	2.22	4.62	2.10	4.17	26.24	7.28	8.79	2.87	2.91	3.33	3.59	15.56
		24	9.98	22.46	2.24	4.21	2.31	5.31	39.95	16.94	22.35	3.22	3.36	3.89	4.93	31.86
US Industries	BM	1	0.04	0.20	0.19	1.07	0.19	1.00	0.03	0.05	0.18	0.24	0.87	0.25	0.83	–0.02
		12	0.29	4.34	0.15	2.58	0.13	2.09	6.44	0.41	3.67	0.18	1.78	0.19	1.65	3.63
		24	–0.10	13.17	–0.02	3.08	–0.02	3.39	16.86	0.02	11.73	0.00	2.06	0.00	2.86	11.09
Commodities	−5-year return	1	0.30	0.17	1.60	0.83	1.61	0.85	–0.02	0.27	0.11	1.47	0.52	1.49	0.54	–0.11
		12	3.10	5.08	1.55	2.62	1.42	2.28	8.65	2.72	5.35	1.27	2.56	1.24	2.47	8.78
		24	7.56	6.86	2.07	2.67	1.81	1.94	8.91	6.34	9.78	1.59	2.64	1.54	2.85	15.26
Currencies	Inf. adj. return	1	0.41	0.22	2.08	1.10	2.13	1.15	0.07	0.46	0.25	2.31	1.15	2.39	1.19	0.14
		12	6.15	6.07	2.70	2.99	2.67	3.11	11.40	6.37	6.12	2.88	3.13	2.77	3.01	11.94
		24	14.80	15.28	3.36	4.35	3.25	4.36	27.32	15.02	12.99	3.69	3.94	3.31	3.74	23.40
Gov't Bonds	−5-year return	1	–0.04	0.57	–0.16	1.70	–0.16	1.80	1.42	–0.05	0.60	–0.21	1.97	–0.20	1.96	1.59
		12	–0.96	7.00	–0.44	4.42	–0.34	2.36	18.76	–0.55	6.37	–0.28	3.99	–0.19	2.14	18.29
		24	–1.74	15.60	–0.39	5.51	–0.33	2.80	33.92	–0.75	12.32	–0.18	6.44	–0.14	2.22	26.81
Stock Indexes	$MSCI_{BP}$	1	0.22	0.44	0.90	1.68	0.88	1.57	0.67	0.32	0.36	1.25	1.28	1.29	1.22	0.33
		12	1.54	5.39	0.47	1.76	0.53	1.91	7.68	3.58	5.76	1.16	1.79	1.22	1.92	9.12
		24	1.13	7.79	0.16	1.50	0.20	1.63	5.21	6.11	8.70	1.02	1.69	1.06	1.75	9.10

TABLE 1.A.3: **Predicting Value Returns with the Value Spread: Pooled Tests**
This table reports results from joint tests that pool the value strategies across asset classes. Panel A reports regression results for the pooled predictive regression, $R_{c,t+1:t+h} = a_h + b_h VS_{c,t} + \varepsilon_{c,t+1:t+h}$. Value is measured in each asset class c as explained in Table 1.A.1. For the sake of comparison across asset classes, value spreads, $VS_{c,t}$, are standardized to have mean equal to zero and variance equal to one, whereas value returns are standardized to have a standard deviation of 15% annually. Panel B reports results of a time series regression of the cross-sectional average value return (over the seven strategies) on the cross-sectional average value spread: $\overline{R_{t+1:t+h}} = a_h + b_h \overline{VS_t} + \varepsilon_{t+1:t+h}$. Panels C and D report results for the same two specifications, but exclude the two value strategies in individual equities. We consider $h = 1, 3, 6, 12, 24, 48$ months and two portfolio weighting schemes: a High-minus-Low spreading portfolio ($H - L$) and a rank-weighted portfolio (*Rank*). The t-statistics are Hodrick, 1992 with h lags for the average-on-average time series regression and Driscoll and Kraay, 1998 with h lags for the pooled regression. The sample period is 1972 to 2017, but some alternative asset classes enter the sample only after 1972.

	$H-L$					*Rank*				
h	a	b	t_a	t_b	R^2	a	b	t_a	t_b	R^2
Panel A: Pooled Predictive Regression										
1	0.21	0.38	2.15	3.39	0.75	0.26	0.32	2.51	2.69	0.56
3	0.61	1.23	2.32	4.86	2.47	0.78	1.07	2.63	3.55	1.77
6	1.26	2.73	2.55	5.66	5.60	1.61	2.44	2.72	3.99	4.10
12	2.64	6.41	2.86	6.16	12.74	3.45	5.93	2.90	4.50	9.70
24	5.84	14.40	2.90	5.63	22.06	7.82	13.83	3.17	4.27	18.81
48	15.23	31.76	2.86	6.22	26.20	19.54	27.69	3.38	5.52	21.36
Panel B: Average Value Return on Average Value Spread										
1	0.25	0.35	2.48	2.16	2.04	0.29	0.30	2.80	1.90	1.29
3	0.73	1.14	2.43	2.39	6.47	0.87	0.95	2.78	2.06	3.88
6	1.53	2.37	2.54	2.68	14.27	1.83	2.01	2.90	2.33	7.87
12	3.30	5.30	2.74	3.51	30.06	3.96	4.66	3.16	3.02	17.00
24	7.75	12.42	3.27	4.98	48.53	9.32	11.91	3.76	4.43	35.11
48	20.57	27.88	4.61	7.08	53.25	23.81	24.19	5.04	5.45	43.74
Panel C: Excluding Value in Individual Equities (Pooled)										
1	0.20	0.28	2.13	2.80	0.43	0.22	0.26	2.29	2.58	0.37
12	2.31	5.44	2.53	5.88	9.64	2.61	5.31	2.54	5.30	8.76
24	5.15	11.74	2.81	6.09	16.05	5.68	11.23	2.81	5.10	14.98
Panel D: Excluding Value in Individual Equities (Average-on-Average)										
1	0.22	0.24	2.19	1.91	0.87	0.23	0.23	2.28	1.92	0.70
12	2.80	4.03	2.33	3.31	21.61	2.87	4.02	2.33	3.43	16.25
24	6.99	8.38	3.01	4.17	36.43	6.82	8.28	2.93	4.06	28.87

TABLE 1.A.4: **Does the CAPM Explain Time Variation in Value Returns?**
This table reports the results of pooled predictive regressions as in Table 1.A.3, but now we control for exposure to a market benchmark. We consider an unconditional specification: $R_{c,t+1} = a + bVS_{c,t} + \beta_{0,c}R_{MKT,c,t+1} + \varepsilon_{c,t+1}$ as well as a conditional alternative: $R_{c,t+1} = a + bVS_{c,t} + \beta_{0,c}R_{MKT,c,t+1} + \beta_{1,c}R_{MKT,c,t+1} \cdot VS_{c,t} + \varepsilon_{c,t+1}$, where $VS_{c,t}$ is the value spread in asset class c. $\beta_{0,c}$ captures the unconditional market exposure and $\beta_{0,c} + \beta_{1,c}VS_{c,t}$ captures the conditional market exposure of each value strategy. The market benchmark is common across asset classes in Panel A: the CRSP value-weighted stock market portfolio. The market benchmark is asset class-specific in Panel B. Value returns are calculated from two portfolio strategies: High-minus-Low ($H-L$) or rank-weighted (*Rank*). t-statistics are Driscoll and Kraay, 1998 with one lag. The full sample period is 1972 to 2017.

	a	b	$\beta_{0,\text{ExFin.}}$	$\beta_{1,\text{ExFin.}}$	$\beta_{0,\text{IndAdj.}}$	$\beta_{1,\text{IndAdj.}}$	$\beta_{0,\text{Inds.}}$	$\beta_{1,\text{Inds.}}$	$\beta_{0,\text{Com.}}$	$\beta_{1,\text{Com.}}$	$\beta_{0,\text{Cur.}}$	$\beta_{1,\text{Cur.}}$	$\beta_{0,\text{Bonds}}$	$\beta_{1,\text{Bonds}}$	$\beta_{0,\text{EqInd.}}$	$\beta_{1,\text{EqInd.}}$	R^2
Panel A: Common Market Benchmark: CRSP Market Portfolio																	
High-minus-Low ($H-L$)																	
Unconditional	0.22	0.38	–0.20		–0.03		–0.13		0.03		0.03		0.11		0.17		2.20
	(2.30)	(3.43)	(–3.33)		(–0.51)		(–2.45)		(0.62)		(0.38)		(1.94)		(2.54)		
Conditional	0.22	0.34	–0.23	0.06	–0.06	0.07	–0.11	–0.05	0.03	–0.03	0.02	0.12	0.13	0.04	0.17	0.04	2.82
	(2.27)	(3.35)	(–3.94)	(0.95)	(–1.18)	(0.96)	(–2.21)	(–1.22)	(0.71)	(–0.72)	(0.38)	(1.85)	(1.72)	(0.34)	(2.42)	(0.53)	
Rank-Weighted (*Rank*)																	
Unconditional	0.28	0.32	–0.26		–0.03		–0.14		0.05		0.00		0.08		0.22		2.72
	(2.71)	(2.74)	(–4.25)		(–0.52)		(–2.75)		(1.07)		(0.01)		(1.38)		(3.47)		
Conditional	0.27	0.30	–0.28	0.03	–0.06	0.06	–0.13	–0.03	0.06	–0.07	0.00	0.12	0.10	0.05	0.22	0.02	3.23
	(2.65)	(2.68)	(–4.86)	(0.40)	(–1.05)	(0.79)	(–2.75)	(–0.56)	(1.31)	(–1.72)	(0.03)	(1.76)	(1.38)	(0.55)	(3.46)	(0.25)	
Panel B: Asset Class-Specific Market Benchmark																	
High-minus-Low ($H-L$)																	
Unconditional	0.30	0.37	–0.20		–0.04		–0.14		–0.20		–0.24		–0.95		0.33		4.11
	(3.11)	(3.48)	(–3.38)		(–0.68)		(–2.68)		(–3.23)		(–1.83)		(–3.37)		(4.53)		
Conditional	0.29	0.34	–0.22	0.05	–0.07	0.06	–0.12	–0.07	–0.14	–0.12	–0.24	0.08	–0.96	0.45	0.32	0.09	4.94
	(3.09)	(3.47)	(–3.95)	(0.70)	(–1.26)	(0.76)	(–2.37)	(–1.55)	(–2.06)	(–2.12)	(–1.88)	(0.57)	(–3.56)	(1.85)	(4.47)	(1.25)	
Rank-Weighted (*Rank*)																	
Unconditional	0.35	0.32	–0.25		–0.02		–0.15		–0.24		–0.33		–0.65		0.39		4.96
	(3.40)	(2.79)	(–3.88)		(–0.27)		(–2.94)		(–3.76)		(–2.57)		(–2.40)		(5.95)		
Conditional	0.34	0.30	–0.24	0.00	–0.03	0.03	–0.14	–0.05	–0.19	–0.13	–0.33	0.10	–0.68	0.66	0.38	0.06	5.77
	(3.39)	(2.84)	(–4.19)	(0.00)	(–0.54)	(0.40)	(–2.88)	(–0.85)	(–2.63)	(–2.82)	(–2.65)	(0.65)	(–2.62)	(2.63)	(6.03)	(0.78)	

TABLE 1.A.5: **Value Timing in Individual Equities**
This table reports unconditional performance statistics for the monthly returns of a strategy that times value using the signal: $VS_{t,His} = \frac{\sum_{s=0}^{11} VS_{t-s}/12 - \sum_{s=12}^{t-1} VS_{t-s}/(t-12)}{\sigma(VS_{1:t-12})}$. $VS_{t,His}$ captures deviations of last year's value spread from the historical average value spread and is observable at time t. We present results for a unit weight strategy that passively captures the unconditional value premium, a linear timing strategy that invests $VS_{t,His}$ dollars in both the long and short position of the value strategy, and, finally, a combined strategy that invests $1 + VS_{t,His}$. We consider 2×2 variations of each value strategy: using either the book-to-market signal excluding financials or the industry-adjusted book-to-market ratio and either the High-minus-Low decile portfolio or the rank-weighted portfolio. To make these different value strategies comparable, we scale each value return series ex ante to have an annualized standard deviation of 15%. The sample period is 1972 to 2017.

		Avg. ret.	t	Sharpe	α^{CAPM}	t_α^{CAPM}	α^{FF3}	t_α^{FF3}
Panel A: High-minus-Low ($H-L$)								
Ind. Equities ($BM_{Ex.fin.}$)	Unit Weight	–0.02	–0.08	0.00	0.17	0.89	–0.39	–2.60
	Linear Timing	0.60	2.77	0.12	0.53	2.42	0.59	2.73
	Combined	0.58	2.34	0.10	0.71	2.79	0.20	0.91
Ind. Equities ($BM_{Ind.adj.}$)	Unit Weight	0.20	1.03	0.04	0.24	1.20	–0.18	–1.09
	Linear Timing	0.60	2.66	0.11	0.55	2.39	0.48	2.12
	Combined	0.80	3.04	0.13	0.79	2.93	0.30	1.27
Panel B: Rank-Weighted ($Rank$)								
Ind. Equities ($BM_{Ex.fin.}$)	Unit Weight	0.04	0.19	0.01	0.28	1.47	–0.32	–2.35
	Linear Timing	0.52	2.35	0.10	0.53	2.33	0.50	2.23
	Combined	0.56	2.03	0.09	0.81	2.94	0.18	0.74
Ind. Equities ($BM_{Ind.adj.}$)	Unit Weight	0.31	1.61	0.07	0.34	1.70	–0.17	–1.10
	Linear Timing	0.56	2.29	0.10	0.58	2.31	0.47	1.94
	Combined	0.87	2.90	0.12	0.92	2.97	0.31	1.14

TABLE 1.A.6: **Out-of-Sample Pooled Predictive Regression**

This table reports results for pooled predictive regressions of returns of the seven value strategies on a dummy variable indicating whether the current value spread in an asset class is historically high or low. We run $R_{c,t+1:t+h,15\%} = a + bI_{VS_{c,t,His}>0} + e_{c,t+1:t+h}$, where $I_{VS_{c,t,His}>0}$ is an indicator function that equals one when the historically standardized value spread in asset class c (see Equation (1.5)) is positive, and zero otherwise. We consider returns of both High-minus-Low and rank-weighted value strategies. To make the value strategies comparable across asset classes, we scale each return series to have an ex ante annualized standard deviation of 15%. We perform this standardization in each month using only backward-looking information as detailed in Table 1.A.5. t-statistics in the pooled regressions are calculated using Driscoll and Kraay, 1998 standard errors with h lags. Panel B reports unconditional performance statistics for a value strategy that invests only in asset class c when $Value_{c,t,His} > 0$, which average return is equal to the sum of the estimated coefficients in the pooled regression at the monthly horizon ($a + b$ for $h = 1$). Conversely, the average return of a strategy that only invests in the value strategy of asset class c when $Value_{c,t,His} \leq 0$ is equal to the estimated intercept (a). The full sample period is 1972 to 2017, but some alternative asset classes enter the sample only after 1972.

Panel A: Pooled Regression on High Value Spread Dummy

	$H-L$					$Rank$				
h	a	b	t_a	t_b	R^2	a	b	t_a	t_b	R^2
1	–0.09	0.60	–0.70	2.91	0.41	–0.02	0.57	–0.12	2.53	0.37
3	–0.28	1.83	–0.83	3.36	1.15	–0.01	1.61	–0.03	2.55	0.87
6	–0.68	4.01	–1.08	3.77	2.56	0.03	3.23	0.03	2.48	1.57
12	–1.36	8.27	–1.03	3.64	4.76	–0.07	7.27	–0.04	2.67	3.34
24	–4.49	21.12	–1.42	3.56	9.82	–0.99	16.98	–0.26	2.64	6.03
48	–16.53	63.35	–1.84	3.99	17.26	–6.28	46.94	–0.64	3.43	10.07

Panel B: Implied Performance

	$H-L$				$Rank$			
	Avg. ret.	St. dev.	t	Sharpe	Avg. ret.	St. dev.	t	Sharpe
Invest when $VS_{c,t,His} > 0$	0.52	4.52	3.70	0.11	0.55	4.61	3.90	0.12
Invest when $VS_{c,t,His} \leq 0$	–0.09	4.43	–0.86	–0.02	–0.02	4.35	–0.16	–0.00

TABLE 1.A.7: **Rotating Value Strategies Across Asset Classes**

This table reports unconditional performance statistics for monthly returns of strategies that rotate value across asset classes. These strategies overweight (underweight) those asset classes where the value spread is high (low) relative to the other asset classes. As a benchmark, we consider an unconditional value strategy that invests, in each sample month t, $1/N_t$ in each of the N_t available value strategies (out of the maximum of seven). The first rotation strategy takes a position in each asset class c in month t equal to $w_{c,t}^{rot,1} = q_t(VS_{c,t,His} - Mean(VS_{c,t,His}))$, where the scalar q_t ensures that the total weight in the long and short position equal one. The second strategy, with weights denoted $w_{c,t}^{rot,2}$, invests an equal weight in each asset class with $VS_{c,t,His}$ above (below) the mean value spread across asset classes. We calculate performance measures for these two long-short rotation strategies (denoted $Rotation_{Long-Short}$) as well as for a combination with the unconditional strategy (denoted $Combined_{Long-Short}$). The reported α is relative to an unconditional market strategy, which equally-weights the market portfolio in each asset class (defined as in Table 1.A.4). The value strategy returns are scaled in each asset class to have a standard deviation of 15% using only backward-looking information. Panel B reports the fraction of the long and short leg of the two rotation strategies that is invested in each asset class (on average over time). Panel C decomposes the average return of the long-short rotation strategy across asset classes. Because we lose the first 120 months in the asset classes with the longest history as burn-in period for the value signal and we require at least four asset classes with available data, the full out-of-sample period is 1982 to 2017.

Panel A: Performance of Value Rotation Strategies

	Linear Weight ($w_{c,t}^{rot,1}$)						Equal weight ($w_{c,t}^{rot,2}$)					
	Avg. ret.	St. dev.	t	Sharpe	α	t_α	Avg. ret.	Std. dev.	t	Sharpe	α	t_α
	High-minus-Low ($H-L$)											
Unconditional	0.08	2.20	0.74	0.04	0.09	0.81	0.08	2.20	0.74	0.04	0.09	0.81
$Rotation_{Long-Short}$	0.68	4.51	3.12	0.15	0.56	2.49	0.63	3.96	3.30	0.16	0.55	2.83
Combined	0.76	4.93	3.19	0.15	0.64	2.64	0.71	4.44	3.31	0.16	0.64	2.92
	Rank-Weighted ($Rank$)											
Unconditional	0.15	2.32	1.37	0.07	0.16	1.35	0.15	2.32	1.37	0.07	0.16	1.35
$Rotation_{Long-Short}$	0.60	4.58	2.71	0.13	0.49	2.15	0.53	4.05	2.71	0.13	0.49	2.42
Combined	0.75	5.09	3.07	0.15	0.64	2.55	0.68	4.67	3.04	0.15	0.64	2.78

Panel B: % Allocation to each Asset Class in $Rotation_{Long-Short}$

	ExFin.	IndAdj.	Inds.	Com.	Cur.	Bonds	EqInd.	Sum	ExFin.	IndAdj.	Inds.	Com.	Cur.	Bonds	EqInd.	Sum
$H-L_{Long}$	0.03	0.05	0.05	0.22	0.31	0.05	0.28	1.00	0.07	0.07	0.09	0.22	0.30	0.08	0.18	1.00
$Rank_{Long}$	0.06	0.04	0.03	0.21	0.32	0.04	0.29	1.00	0.11	0.05	0.07	0.22	0.29	0.07	0.19	1.00
$H-L_{Short}$	0.20	0.20	0.23	0.12	0.14	0.09	0.03	1.00	0.20	0.20	0.21	0.12	0.11	0.11	0.04	1.00
$Rank_{Short}$	0.16	0.19	0.27	0.12	0.15	0.09	0.02	1.00	0.17	0.22	0.22	0.12	0.11	0.12	0.03	1.00

Panel C: Contribution to Avg. Ret. in $Rotation_{Long-Short}$

	ExFin.	IndAdj.	Inds.	Com.	Cur.	Bonds	EqInd.	Sum	ExFin.	IndAdj.	Inds.	Com.	Cur.	Bonds	EqInd.	Sum
$H-L$	0.11	0.16	0.01	0.00	0.27	0.02	0.11	0.68	0.14	0.13	0.03	0.03	0.23	0.00	0.08	0.63
$Rank$	0.06	0.11	0.04	0.02	0.19	0.04	0.14	0.60	0.09	0.07	0.03	0.00	0.24	–0.01	0.11	0.53

TABLE 1.A.8: **Common and Asset Class-Specific Components of the Value Spread**
This table reports results for pooled predictive regressions of High-minus-Low value returns on components of the value spread. Panel A reports the results of a pooled predictive regression on the common component of the value spread (the first principal component of the value spread in seven asset classes): $R_{c,t+1:t+h} = a_h + b_{h,Com} VS_t^{Com} + \varepsilon_{t+h}$. Panel B reports results for the asset class-specific components, which are defined as the residual in a regression of the value spread in asset class c on VS_t^{Com}: $R_{c,t+1:t+h} = a_h + b_{h,Spec} VS_{c,t}^{Spec} + \varepsilon_{t+h}$. Panel C reports the results of a pooled regression that includes the two components simultaneously. t-statistics are calculated using Driscoll and Kraay, 1998 standard errors with h lags. The sample is 1972 to 2017, but some alternative asset classes enter the sample only after 1972.

h	a	b_{Com}	b_{Spec}	t_a	$t_{b_{Com}}$	$t_{b_{Spec}}$	R^2
Panel A: Common Value							
1	0.25	0.39		2.37	2.02		0.35
3	0.73	1.28		2.63	2.71		1.18
6	1.50	2.78		2.87	3.08		2.61
12	3.17	6.43		3.25	4.16		5.79
24	7.13	16.17		3.53	5.39		12.84
48	17.43	38.58		3.85	8.83		18.45
Panel B: Specific Value							
1	0.21		0.28	2.13		2.38	0.21
3	0.61		0.96	2.26		3.40	0.73
6	1.26		2.29	2.38		4.50	1.92
12	2.64		5.67	2.41		5.49	4.84
24	5.84		10.75	2.07		5.46	5.94
48	15.23		21.20	1.80		3.29	5.62
Panel C: Common and Specific Value							
1	0.25	0.39	0.28	2.38	2.02	2.37	0.56
3	0.73	1.28	0.96	2.64	2.72	3.36	1.92
6	1.50	2.78	2.29	2.91	3.10	4.49	4.53
12	3.17	6.43	5.67	3.31	4.18	5.74	10.63
24	7.13	16.17	10.75	3.52	5.25	5.97	18.78
48	17.43	38.58	21.20	3.83	8.64	3.40	24.07

TABLE 1.A.9: **Comovement Between Risk Proxies and the Value Spread**
This table regresses components of the High-minus-Low value spread on state variables, collected in the vector Z_t, that are popular in the literature to proxy for time variation in risk premia (intermediary leverage; the illiquidity premium; the dividend yield; a global recession dummy; the default spread; real uncertainty; and, the Chicago Fed National Activity Index). Panel A reports results from time series regressions of the common component of the value spread on the risk proxies, $VS_t^{Com} = k_0 + k_1' Z_t + u_t^{Com}$. We consider both simple regressions on individual risk proxies (Specifications 1, 2, 4 and 6) and multiple regressions on sets of risk proxies (Specifications 3, 5 and 7). For specifications 3 and 5, we also run the regression in innovations, which are estimated using an AR(1)-model for both common value and the state variables. Panel B regresses the asset class-specific components of the value spread on the full set of risk proxies (as in Specification 7), $VS_{c,t}^{Spec} = k_0 + k_1' Z_t + u_t^{Spec}$. t-statistics are calculated using Newey and West, 1987a standard errors with 12 lags. The full sample period is 1972 to 2017, but some alternative asset classes enter the sample only after 1972.

	Intermediary Leverage	Illiquidity Premium	Dividend Yield	Global Recession	Default Spread	Real Uncertainty	Chicago Fed National Activity Index	R^2
				Panel A: Common Value				
1	0.52							53.02
	(6.81)							
2		0.45						40.09
		(9.56)						
3	0.39	0.27						64.39
	(6.70)	(3.83)						
4			0.59					68.70
			(10.00)					
5	0.16	0.16	0.38					73.80
	(2.94)	(2.16)	(5.37)					
6				0.46				9.88
				(2.57)				
7	0.02	0.09	0.43	0.21	0.07	0.08	0.01	78.73
	(0.30)	(1.27)	(5.91)	(2.85)	(1.07)	(1.11)	(0.26)	
3 (in AR(1)-innovations)	0.61	0.12						41.21
	(6.87)	(2.96)						
5 (in AR(1)-innovations)	0.22	0.04	0.58					57.32
	(3.23)	(1.30)	(6.58)					
				Panel B: Asset Class-Specific Value				
Ind. Equities ($BM_{Ex.Fin.}$)	–0.11	0.00	0.03	–0.04	0.08	0.08	–0.02	8.64
	(–1.27)	(0.07)	(0.44)	(–0.58)	(1.46)	(1.50)	(–0.99)	
Ind. Equities ($BM_{Ind.Adj.}$)	0.10	–0.08	–0.12	–0.14	0.14	0.04	0.02	40.73
	(1.60)	(–2.59)	(–2.15)	(–2.89)	(2.76)	(0.78)	(0.91)	
Industries	–0.30	0.09	0.29	0.03	–0.01	–0.12	–0.03	47.63
	(–4.98)	(2.64)	(4.79)	(0.35)	(–0.23)	(–2.63)	(–1.33)	
Commodities	0.15	–0.10	–0.18	–0.16	–0.42	0.52	0.15	19.29
	(0.78)	(–1.21)	(–0.99)	(–0.88)	(–2.85)	(4.19)	(2.05)	
Currencies	0.08	0.21	0.08	–0.35	0.15	–0.26	–0.15	12.35
	(0.45)	(1.21)	(0.38)	(–1.37)	(0.92)	(–1.64)	(–1.54)	
Government Bonds	–0.25	0.11	1.57	0.06	–0.09	–0.02	–0.19	35.07
	(–1.11)	(0.41)	(3.24)	(0.28)	(–0.51)	(–0.10)	(–2.16)	
Stock Indexes	0.67	0.09	–0.43	0.22	–0.23	–0.73	–0.03	44.34
	(2.42)	(0.34)	(–1.01)	(1.22)	(–1.47)	(–4.50)	(–0.35)	

TABLE 1.A.10: **Value Return Predictability Net of Risk Proxies**
In Panel A of this table, we present the results from pooled predictive regressions of High-minus-Low value returns on the explained and orthogonal components of common value; $R_{c,t+1:t+h} = a_h + b_{Com,Orth}(VS^{Com}_{c,t} - \widehat{VS^{Com}_{c,t}}) + b_{Com,Expl}\widehat{VS^{Com}_{c,t}} + \varepsilon_{c,t+1:t+h}$. The explained component of common value (denoted $\widehat{VS^{Com}_{c,t}}$) is pre-estimated by regressing the common component on the full set of risk proxies used in Table 1.A.9, and saving the predicted value. The orthogonal component of common value is the residual from this time series regression. In Panel B, we similarly decompose the asset class-specific components of the value spread into the part explained by the risk proxies and the part that is orthogonal. t-statistics in the pooled regressions are calculated using Driscoll and Kraay, 1998 standard errors with h lags. We also present the relative contribution to R^2 from the explained and orthogonal components. The full sample period is 1972 to 2017, but some alternative asset classes enter the sample only after 1972.

h	a	$b_{Com,Orth}$	$b_{Com,Expl}$	t_a	$t_{Com,Orth}$	$t_{Com,Expl}$	R^2	$R^2_{Com,Orth}$	$R^2_{Com,Expl}$
				Panel A: Common Value					
1	0.24	0.35	0.53	2.35	1.70	1.38	0.36	0.24	0.13
3	0.72	1.18	1.66	2.57	2.06	2.10	1.21	0.82	0.39
6	1.49	2.63	3.43	2.81	2.30	2.38	2.64	1.88	0.77
12	3.12	5.88	8.74	3.18	3.33	2.97	5.97	3.90	2.08
24	7.04	15.01	20.99	3.49	5.42	3.14	13.11	8.92	4.20
48	17.36	37.36	43.54	3.83	8.81	3.20	18.52	13.89	4.63
				Panel B: Asset Class-Specific Value					
1	0.21	0.11	0.34	2.13	0.50	2.51	0.24	0.01	0.23
3	0.61	0.63	1.08	2.26	1.14	3.29	0.76	0.08	0.68
6	1.26	1.33	2.63	2.38	1.24	4.52	2.05	0.17	1.88
12	2.64	4.52	6.09	2.41	2.22	4.62	4.91	0.82	4.09
24	5.84	12.40	10.14	2.07	5.35	3.67	5.99	2.14	3.85
48	15.23	34.98	15.80	1.81	5.50	1.64	6.55	4.30	2.24

1.B Variable Construction

In this section, we describe our data sources and methodology for constructing value strategies in different asset classes. We end this section with a comparison of our data to Asness et al., 2013a.

1.B.1 U.S. Individual Equities and Industries

The U.S. stock universe consists of all common equity in CRSP that trade on the NYSE, AMEX, and NASDAQ (sharecodes 10 and 11; exchange codes 1 to 3), which we match to book equity values from Compustat. Following Davis et al., 2000a, we compute book equity as shareholder's book equity, plus balance sheet deferred taxes and investment tax credit (item TXDITC) minus the book value of preferred stock. Shareholders' equity is the Compustat item SEQ if available. Otherwise, we compute shareholders' equity as common equity (item CEQ) plus the par value of preferred stock (item PSTK), or total assets (AT) minus total liabilities (LT). When TXDITC is absent, we compute deferred taxes and investment tax credit as deferred taxes (item TXDB) plus investment tax credit (item ITCB). We define the book value of preferred stock as redemption (item PSTKRV), liquidating (item PSTKL) or par value (item PSTK), depending on availability. Delisting returns realised after the last trading day of month t are considered to have accrued in month $t+1$. The sample period for individual equities starts from January 1962 and ends in December 2017. We limit the universe of U.S. equities to a liquid subset that can be traded in sizeable trading volumes at a reasonably low cost. Specifically, we rank stocks, in the cross-section, based on their end-of-month t market capitalization in descending order. We then include in our sample, stocks that cumulatively account for 90% of the total market capitalization.

We measure *value* for firm i as the ratio of book value of equity to market value of equity: $BM_{i,t} = \frac{BE_{i,t}}{ME_{i,t}}$. Book equity is updated every June using data from the previous fiscal year to ensure that the data was available to investors at the time of portfolio formation. Market values are updated monthly following Asness and Frazzini, 2013. We consider three alternative value strategies. In the first strategy, we construct our value portfolios excluding all financial firms, which we denote:

BM_{ExFin}. The motivation is that the same book-to-market ratio may signal distress for a non-financial firm, but not for a financial firm (Fama and French, 1995). The second strategy uses industry-adjusted book-to-market ratios, BM_{IndAdj}, which are calculated subtracting from each firm's book-to-market ratio the value-weighted average book-to-market ratio of the industry to which that firm belongs. We use the seventeen industry classification available on Kenneth French's webpage and construct these industry portfolios using only the restricted set of relatively large stocks. The third industry value strategy we consider sorts these seventeen industries on their value-weighted average book-to-market ratio.

1.B.2 Commodity Futures

We obtain futures price data on Crude Oil, Gasoline, Heating Oil, Natural Gas, Gas-Oil Petroleum, Coffee, Rough Rice, Orange Juice, Cocoa, Soybean Oil, Soybean Meal, Soybeans, Corn, Oats, Wheat, Cotton, Gold, Silver, Platinum, Feeder Cattle, Live Cattle, Lean Hogs from the Commodity Research Bureau and Aluminium, Nickel, Tin, Lead, Zinc, and Copper from Datastream. We calculate monthly returns as the return on the nearest-to-maturity futures contract: $R^{fut}_{i,t} = Price^{T_1}_{i,t}/Price^{T_1}_{i,t-1} - 1$, where $Price^{T_1}_{i,t}$ is the time t price of the nearest-to-maturity futures contract of commodity i. We exclude contracts that mature in month *t+1*.

For commodities, we measure value as the negative of the five year log spot return (−5-year return) as in Asness et al., 2013a. As spot prices of commodities are illiquid, we use the nearest-to-maturity futures prices to calculate the signal: −5-year return= $ln(\overline{Price^{T_1}_{i,t-60}}/Price^{T_1}_{i,t})$, where $\overline{Price^{T_1}_{i,t-60}}$ is the average price from 4.5 to 5.5 years ago to smooth out some noise. The sample period runs from January 1972 (when we have data for eleven commodities) to December 2017 (when we have data for all 28 commodities).

1.B.3 Currencies

We obtain exchange rate data (spot and one-month forward rates) from Datastream for 9 countries: Australia, Canada, Germany (replaced with the Euro from January 1999), Japan, New Zealand, Norway, Sweden, Switzerland and the U.K. We compute currency returns as:

$$R^{Cur}_{i,t+1} = (e_{i,t+1}/f_{i,t}) - 1 \tag{1.B.1}$$

where $e_{i,t}$ is the time t spot exchange rate and $f_{i,t}$ is the previous month's closing price of a one-month forward.

To measure value, we follow Asness et al., 2013a and Menkhoff et al., 2016 and use the five-year change in relative purchasing power parity, which is a natural choice to measure value in currencies. This value measure consists of two parts: (i) the negative of the five year log spot return: $-$5-year return= $ln(\overline{e_{i,t-60}}/e_{i,t})$, where $\overline{e_{i,t-60}}$ is the average spot exchange rate from 4.5 to 5.5 years ago to smooth out some noise; and, (ii) an inflation adjustment, by subtracting from $-$5-year return the five-year foreign-U.S. inflation difference. Consumer Price Indexes are from Global Financial Data, and we interpolate the quarterly Australian and New Zealand CPI estimates to get a monthly series. We end up with a sample period from February 1976 (four currencies) to December 2017 (all currencies available).

1.B.4 Global Government Bonds

The universe of government bond securities we analyze consists of Australia, Canada, Germany, Japan, Norway, New Zealand, Sweden, Switzerland, the U.K. and the U.S. The full sample starts in January 1991 and ends in December 2017. We use constant maturity, zero coupon bond yields from Jonathan Wright's webpage to calculate synthetic bond futures prices and returns and to define our value measures up to May 2009. From June 2009 to December 2017, zero coupon bond yields are from Bloomberg (see Table 1.B.1 below). We also construct traded bond index futures returns using first and second generic nearest-to-maturity futures prices from Bloomberg. These are available for six of the ten countries only (Australia, Canada, Germany, Japan, the U.K. and the U.S.). Table 1.B.1 provides Bloomberg tickers for the futures contracts we use.

Following Koijen et al., 2018, we calculate the monthly dollar excess return on a fully-collateralized, currency-hedged position in the foreign currency-denominated futures contracts (see their Appendix A). To be precise, for each bond index futures i the monthly return of the first-nearby futures strategy (that rolls at the end of the month prior to expiration) equals:

$$R^{fut}_{i,t+1} = \frac{Price^{T_n}_{i,t+1} - Price^{T_n}_{i,t}}{Price^{T_n}_{i,t}} + \frac{e_{i,t+1} - e_{i,t}}{e_{i,t}} \frac{Price^{T_n}_{i,t+1} - Price^{T_n}_{i,t}}{Price^{T_n}_{i,t}} \tag{1.B.2}$$

where $Price^{T_n}_{i,t}$ is the foreign currency price of the second-nearby generic futures contract, ($Price^{T_2}_{i,t}$), in roll-over months (which are the same for all bond indexes: March, June, September, and December) and the first-nearby generic futures contract, ($Price^{T_1}_{i,t}$), in all other months. $e_{i,t}$ is the time t exchange rate (in USD per unit of foreign currency i). Each month, we calculate the price of a synthetic one-month futures on the ten year zero coupon bond (with spot price $S^{120}_{i,t} = exp(-y^{120}_{i,t} \times 120)$) from the no-arbitrage relation:

$$Price^{1,syn}_{i,t} = S^{120}_{i,t} \times exp(y^{1}_{i,t}). \tag{1.B.3}$$

At expiration, the price of the one-month futures contract equals the spot price of a bond that matures in nine years and eleven months: $Price^{0,syn}_{i,t+1} = S^{119}_{i,t+1} = exp(-y^{119}_{i,t+1} \times 119)$, where $y^{119}_{i,t+1}$ is found by linear interpolation. As for the traded bond returns, we calculate synthetic futures returns from these prices assuming that the investor is fully-collateralized and hedges out the currency risk (denoted $R^{Syn.fut.}_{i,t+1}$).

For these global government bonds, we define two measures of value. Given that traded bond futures data is relatively scarce, we define the value measures using yield data. The first value measure is the negative of the five-year log return of the one-month future on the ten-year zero coupon bond,

$$-\text{ 5-year return} = -ln(\prod_{j=1}^{60} 1 + R^{Syn.fut.}_{i,t-j+1}). \tag{1.B.4}$$

The second value measure we consider is the five-year change in the ten-year yield (5-year Δy).

1.B.5 Global Stock Indexes

The global stock index futures data cover thirteen markets: Australia (S&P ASX 200), Canada (S&P TSE 60), France (CAC), Germany (DAX), Hong Kong (Hang Seng), Italy (FTSE MIB), Japan (Nikkei), the Netherlands (EOE AEX), Sweden (OMX), Spain (IBEX), Switzerland (SMI), the U.K. (FTSE 100) and the U.S. (S&P 500). We collect spot and (first and second generic nearest-to-maturity) futures prices from Bloomberg (see Table 1.B.1 below). Following Koijen et al., 2018, we calculate the monthly dollar excess return on a fully-collateralized, currency-hedged position in the foreign currency-denominated stock index futures contracts using Equation 1.B.2. As in Asness et al., 2013a, we use the inverse of the MSCI country-level price-to-book ratio as our measure of value for stock indexes (available from Datastream (ticker: MSBP) and denoted $MSCI_{BP}$). Requiring both past five-year returns and book-to-price to be available, we end up with a sample period from January 1994 (four markets) to December 2017 (all markets available).

1.B.6 Data Validation

Table 1.B.2 serves to validate our data and presents results for the pooled and average-on-average predictive regression for the rank-weighted value returns used in Asness et al. (2013a, AMP). The right-hand side value spreads are rank-weighted and calculated using our own data (as used in Table 1.A.3). In short, we see that our value spreads predict these alternative returns in a statistically and economically significant way. The estimated intercepts are similar to what we find, indicating that we match well the unconditional value premia of AMP over the sample period over which we observe value spreads in the different asset classes. The conditional variation in value premia due to the value spread is also similar to what we find. For instance, in the pooled specification at the annual horizon ($h = 12$), the ratio of the coefficient estimate to the intercept is 1.54 with an R^2 of 11%, which is relative to a ratio of 1.72 and an R^2 of 10% in Table 1.A.3.

There is some variation in magnitude of the coefficient estimates across the two samples, however. This variation is driven by three things. First, the set of value

We thank the authors for sharing these returns on their website.

strategies used in Table 1.B.2 excludes the individual equity strategy based on industry-adjusted book-to-market as well as the industry value strategy, which strategies AMP do not analyze. Second, AMP have returns for more assets in some markets. We do not construct returns for five additional equity indexes from a synthetic MSCI index swap instrument. Third, AMP have a longer history in some markets. For currencies, for instance, they have data for all ten countries available from 1980, whereas we have complete data only from 1990 onwards. The fact that our conclusions extend even in the presence of this disconnect is a testament to the strength of the information in the value spread.

TABLE 1.B.1: **Bloomberg Index Tickers**
The table reports the tickers for the first and second generic futures price series for global stock indexes and global government bonds from Bloomberg. To retrieve the first or second generic futures, replace "**x**" in the futures ticker with 1 and 2. For example, SP1 Index and SP2 Index are the first and second generic futures contracts for the S&P 500 and XM1 Comdty and XM2 Comdty are the first and second generic futures contracts for the Australian 10-year bond.

Country	Spot Ticker	Futures Ticker	Bond Ticker	Futures Ticker
	Stock Index Tickers		Zero Coupon Bond Tickers	
Australia	AS51 Index	XPx Index	F12710y Index	XMx Comdty
Canada	SPTSX60 Index	PTx Index	F10110y Index	CNx Comdty
France	CAC Index	CFx Index		
Germany	DAX Index	GXx Index	F91010y Index	RXx Comdty
Hong Kong	HSI Index	HIx Index		
Italy	FTSEMIB Index	STx Index		
Japan	NKY Index	NKx Index	F10510y Index	JBx Comdty
Neatherlands	AEX Index	EOx Index		
New Zealand	-	-	F25010y Index	-
Norway	-	-	F26610y Index	-
Sweden	OMX Index	QCx Index	F25910y Index	-
Spain	IBEX Index	IBx Index		
Switzerland	SMI Index	SMx Index	F25610y Index	-
U.K.	UKX Index	Zx Index	F11010y Index	Gx Comdty
U.S.	SPX Index	SPx Index	F08210y Index	TYx Comdty

TABLE 1.B.2: **Pooled Tests for Asness et al., 2013a Value Returns on our Value Spreads**
We regress rank-weighted value returns of Asness et al., 2013a (as available from `http://www.lhpedersen.com/data`) on our rank-weighted average value spreads. We consider the pooled and average-on-average specification (as in Table 1.A.3 of the paper). The number of value strategies included in each test is five: individual equities ($BM_{Ex.fin.}$); commodities, currencies, government bonds (the negative of the five year return, -5-year return); and, stock indexes ($MSCI_{BP}$). To make the asset classes comparable, we standardize the value spread and scale the value return to a standard deviation of 15% annually. The t-statistics are Hodrick, 1992 with h lags for the average-on-average time series regression and Driscoll and Kraay, 1998 with h lags for the pooled regression. The full sample period is 1972 to 2017, but some alternative asset classes enter the sample only after 1972.

	Pooled					Average-on-Average				
h	a	b	t_a	t_b	R^2	a	b	t_a	t_b	R^2
1	0.31	0.36	2.22	2.15	0.68	0.40	0.35	2.42	1.64	0.67
12	4.14	6.37	2.63	3.25	11.05	5.70	8.34	2.88	4.02	19.45
24	8.91	12.04	2.55	2.70	15.13	12.46	16.53	3.18	5.01	28.22

1.B.7 Three-Pass Regression Filter and the Value Spread

Kelly and Pruitt, 2015 propose a three-pass regression filter (3PRF) that exploits the wealth of information in a cross-section of predictor variables with a relatively short time series. Given a forecast target, the 3PRF constructs a single forecasting factor that is a linear combination of the predictor variables that are driving the forecast target itself. Importantly, the 3PRF estimator requires specifying only the number of relevant factors, regardless of the total number of common factors driving the cross-section of predictors. Practically, they use a cross-section of valuation ratios to construct a single forecasting factor for the market risk premium. We adopt the 3PRF to forecast the returns of a value-minus-growth strategy using a cross-section of portfolio-level book-to-market ratios.

In the first step of the 3PRF, we estimate time series regressions of the book-to-market ratio in month t of each decile portfolio on the forecast target, the High-minus-Low book-to-market decile return in month $t+1$. Figure 1.B.1 plots the coefficients. We observe that the coefficients are monotonically decreasing from High to Low for both value measures (book-to-market excluding financials and industry-adjusted book-to-market). This finding suggests that the High-minus-Low value spread is likely to be close to the single, optimal 3PRF forecasting factor. We confirm this intuition in Figure 1.B.2, which plots the time series of the extracted factor versus the High-minus-Low value spread. To be precise, in the second step of the 3PRF, we estimate cross-sectional regressions in each month t of ten book-to-market ratios on the ten estimated coefficients from step one. The estimated loading in this second step represents the single, optimal 3PRF forecasting factor. The High-minus-Low value spread and the optimal 3PRF forecasting factor have a correlation exceeding 0.995, a fact suggesting that the two measures contain virtually identical information.

We conclude that using the High-minus-Low value spread to predict value-minus-growth returns is not only economically sound and simple (such that is particularly suited to real-time exercises), it is also the statistically optimal way to combine the cross-section of valuation ratios of book-to-market sorted portfolios.

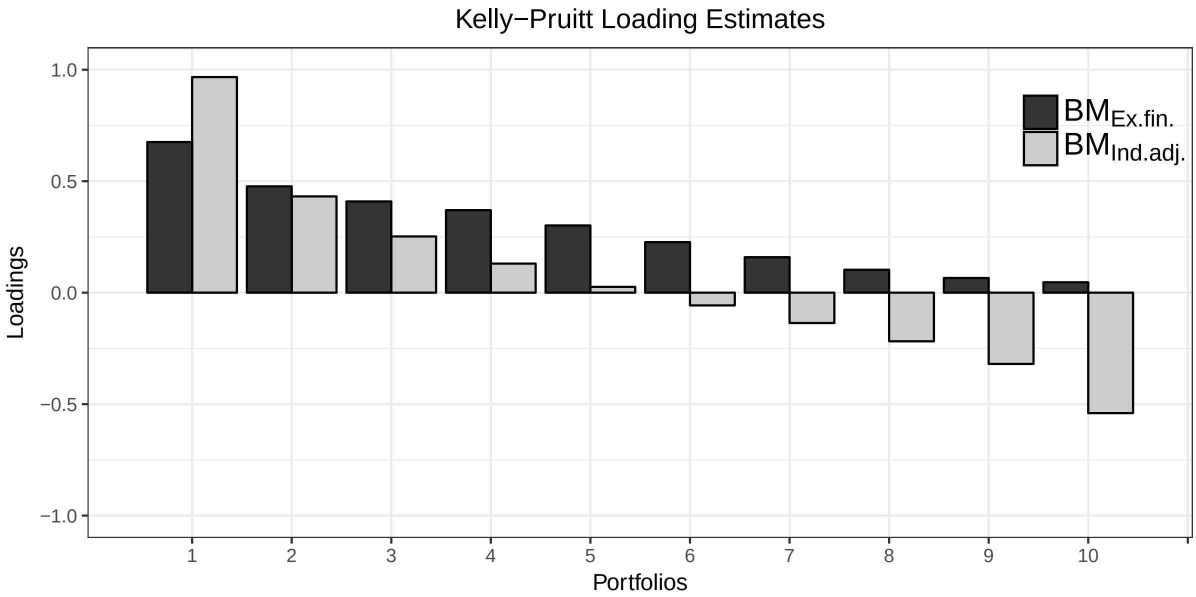

FIGURE 1.B.1: **Kelly-Pruitt Loading Estimates**
This figure shows the loadings in the first stage of the 3PRF procedure of Kelly and Pruitt, 2013. We apply their procedure to predict the High-minus-Low book-to-market decile spreading return using the valuation ratios of ten book-to-market deciles. We consider two measures of book-to-market: BM Ex. Fin. is the liquid U.S. stock sample excluding financial firms, BM Ind. Adj. uses industry-adjusted book-to-market ratios.

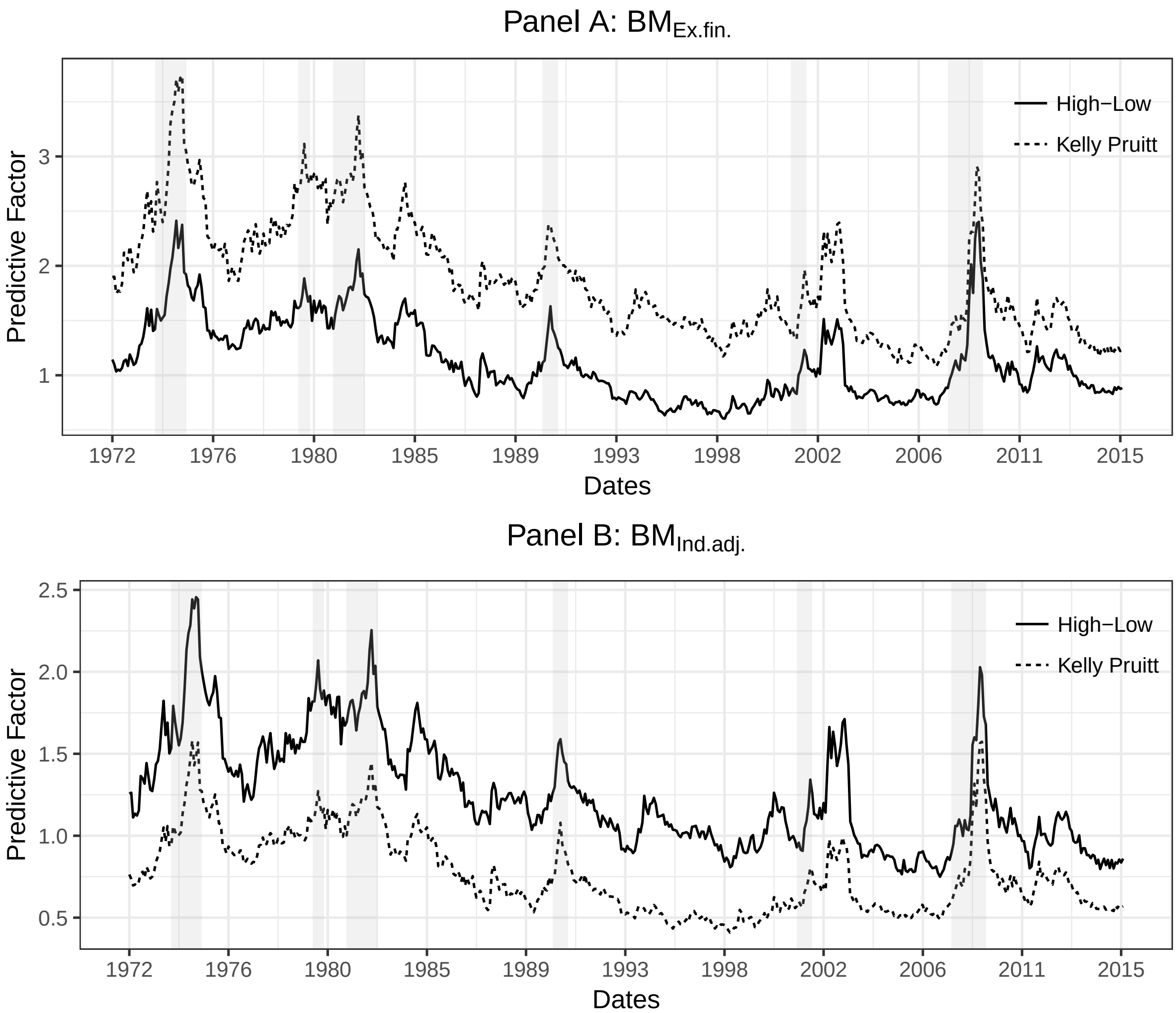

FIGURE 1.B.2: **Kelly-Pruitt and High-minus-Low Predictive Factor**
This figure compares the latent predictive factor of Kelly and Pruitt, 2013 extracted from the valuation ratios of ten book-to-market portfolios (either using all stocks except for financials (Panel A) or using industry-adjusted book-to-market ratios (Panel B)) to the High-minus-Low value spread for the sample period from 1972 to 2015. The shaded areas represent NBER recessions.

1.C Internet Appendix

1.C.1 Figures

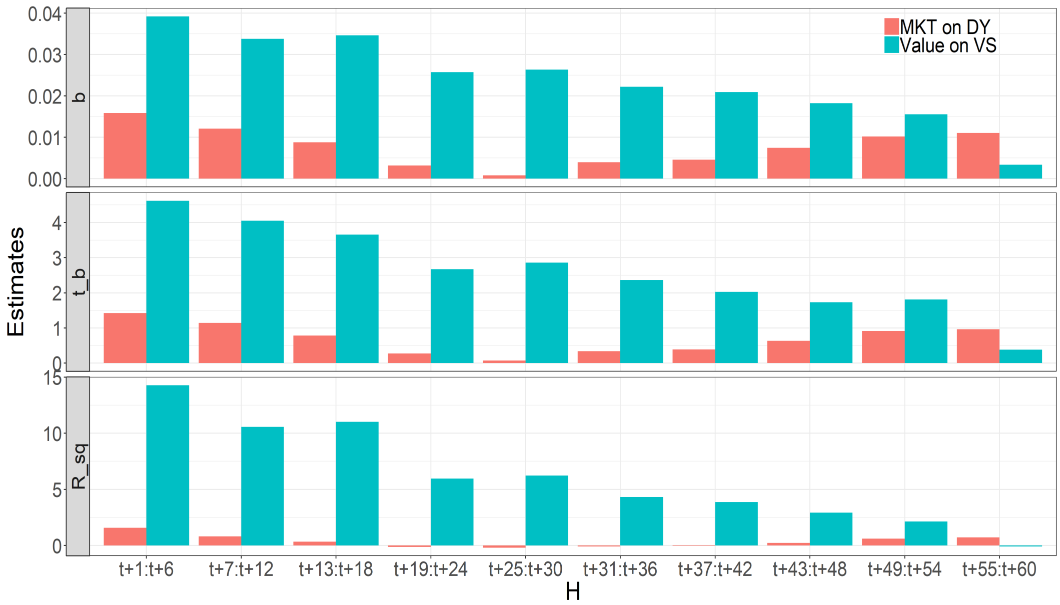

FIGURE 1.C.1: **Value Return Predictability Versus Market Return Predictability**
This figure presents the coefficient estimates (b), t-statistics (t_b), and R^2s (R_sq) from predictive regressions of non-overlapping future value returns on the lagged value spread as well as market returns on the lagged dividend yield: $R^{H-L}_{t+h_1,t+h_2} = a + bVS^{H-L}_t + \varepsilon_{VR,t+h_1:t+h_2}$ and $R^{MKT}_{t+h_1,t+h_2} = a + bDY_t + \varepsilon_{t+h_1:t+h_2}$. We consider semi-annual periods ranging from six months ($h_1 = 1, h_2 = 6$) to five years ($h_1 = 55, h_2 = 60$) in the future. The dividend yield and the value spread are standardized and t-statistics are calculated using Newey and West, 1987a standard errors.

1.C.2 Tables

TABLE 1.C.1: **Stambaugh Bias**
This table presents estimates of the Stambaugh, 1999 bias in two monthly predictive regressions, for the period from 1972 to 2017. Panel A reports results for the monthly predictive regression of market returns on the lagged dividend yield ($R^{MKT}_{t+1} = a + bDY_t + \varepsilon^{MKT}_{t+1}$, where we assume an AR(1)-process for the dividend yield: $DY_{t+1} = c + dDY_t + \varepsilon^{DY}_{t+1}$). Panel B reports results for the monthly predictive regression of the High-minus-Low value returns on the lagged High-minus-Low value spread ($R^{H-L}_{t+1} = a + bVS^{H-L}_t + \varepsilon^{H-L}_{VR,t+1}$, where we assume an AR(1)-process for the value spread: $VS^{H-L}_{t+1} = c + dVS^{H-L}_t + \varepsilon^{H-L}_{VS,t+1}$). We present coefficient estimates from these regressions as well as summary statistics for the residuals from which the *bias* can be calculated using: $-T^{-1}(1+3d) \times Cov(\varepsilon^{MKT}_{t+1}, \varepsilon^{DY}_{t+1})/(\sigma^2_{\varepsilon_{DY}})$ in Panel A (and analogously in Panel B). We see that the Stambaugh bias is small when predicting value returns with the value spread.

Panel A: Market Returns on the Dividend Yield		
	R^{MKT}_{t+1} on DY_t	DY_{t+1} on DY_t
Intercept	0.0002	0.0002
DY_t	0.1932	0.9914
$\sigma_{\varepsilon_{MKT}}$	0.0448	
$\sigma_{\varepsilon_{DY}}$	0.0015	
$Corr(\varepsilon^{MKT}_{t+1}, \varepsilon^{DY}_{t+1})$	-0.8952	
$Bias$	0.1911	

Panel B: Value Returns on the Value Spread		
	R^{H-L}_{t+1} on VS^{H-L}_t	VS^{H-L}_{t+1} on VS^{H-L}_t
Intercept	-0.0153	0.0315
VS_t	0.0157	0.9720
$\sigma_{\varepsilon_{VR}}$	0.0365	
$\sigma_{\varepsilon_{VS}}$	0.0841	
$Corr(\varepsilon^{H-L}_{VR,t+1}, \varepsilon^{H-L}_{VS,t+1})$	-0.2320	
$Bias$	0.0007	

TABLE 1.C.2: **Alternative Value Strategies using Global Government Bonds**
This table presents two robustness checks for the predictive time-series regressions of monthly value returns on the value spread in government bonds. First, we consider an alternative measure of government bond returns, derived from a synthetic one month futures on a 10-year bond. Second, we consider an alternative value measure to predict traded government bond futures returns, that is, the five-year change in 10-year bond yield (5-year Δy). Value returns are calculated from two strategies, a High-minus-Low portfolio split at the median of ranked values ($H-L$) or a rank-weighted portfolio (*Rank*). The sample starts in 1994 and ends in 2017. Panel A reports unconditional performance statistics for monthly value returns. Panel B presents the regression results for holding periods of $h = 1, 12, 24$ months. t^{nw} and t^{hd} indicate t-statistics calculated using Newey and West, 1987a and Hodrick, 1992 standard errors, respectively.

Panel A: Unconditional Performance (Monthly Returns)

		$H-L$			*Rank*		
Asset Class	Value Measure	Avg. ret.	t	Sharpe	Avg. ret.	t	Sharpe
Synthetic Gov't Bonds	−5-year return	0.15	0.63	0.04	0.29	1.19	0.07
Traded Gov't Bonds	5-year Δy	0.27	1.12	0.06	0.34	1.43	0.08

Panel B: Predictive Regressions of Value Returns on the Value Spread

			$H-L$							*Rank*						
		h	a	b	t_a^{nw}	t_b^{nw}	t_a^{hd}	t_b^{hd}	R^2	a	b	t_a^{nw}	t_b^{nw}	t_a^{hd}	t_b^{hd}	R^2
Synthetic Gov't Bonds	−5-year return	1	0.15	0.30	0.63	0.71	0.64	0.73	0.16	0.29	0.28	1.19	0.75	1.20	0.76	0.12
		12	1.35	3.22	0.88	2.10	0.49	0.84	6.45	3.04	1.91	1.78	0.90	1.08	0.49	2.03
		24	3.97	9.19	1.40	4.27	0.80	1.34	22.97	8.18	7.79	2.35	4.48	1.60	1.12	14.36
Traded Gov't Bonds	5-year Δy	1	0.27	0.46	1.15	1.22	1.13	1.33	0.80	0.34	0.69	1.45	2.10	1.43	2.17	2.20
		12	3.95	5.96	1.78	3.58	1.41	1.99	14.02	5.09	8.25	2.02	3.99	1.81	2.90	20.72
		24	10.61	11.20	2.12	3.52	2.05	2.08	20.08	12.37	17.07	2.06	4.00	2.32	3.29	30.53

TABLE 1.C.3: **Alternative Value Strategies in Individual Equities**
This table presents the results from predictive regressions of monthly High-minus-Low value returns on the value spread: $R_{t+1:t+h} = a_h + b_h VS_t + \varepsilon_{t+1:t+h}$ from 1972 to 2017. We focus on the sample excluding financial firms. In Panel A, we consider a longer sample starting in 1962. In Panel B, we update the market cap in the definition of the book-to-market ratio annually (instead of monthly). In Panel C, we consider an alternative definition of value, that is, the negative of the past five-year return.

	$H-L$							$Rank$						
h	a	b	t_a^{nw}	t_b^{nw}	t_a^{hd}	t_b^{hd}	R^2	a	b	t_a^{nw}	t_b^{nw}	t_a^{hd}	t_b^{hd}	R^2
Panel A: Sample from 1962														
1	0.13	0.48	0.75	2.56	0.78	2.39	1.09	0.22	0.33	1.27	1.56	1.34	1.56	0.45
12	1.34	6.97	0.71	4.21	0.67	3.49	14.14	2.91	5.14	1.37	2.26	1.46	2.43	6.86
24	2.6	16.38	0.7	4.85	0.66	4.42	29.51	6.64	14.38	1.54	2.79	1.67	3.61	19.31
Panel B: Annually Updated Market Cap														
1	0.17	0.47	1.04	2.79	1.04	2.75	1.04	0.27	0.39	1.56	2.23	1.60	2.26	0.67
12	1.87	5.53	0.95	3.33	0.94	2.84	8.98	3.53	4.89	1.71	2.32	1.78	2.55	6.47
24	3.53	10.49	0.79	3.11	0.89	2.82	11.59	8.32	10.86	1.83	2.79	2.09	2.94	10.7
Panel C: Negative of Five Year Return														
1	0.03	0.35	0.13	1.05	0.14	0.92	0.48	0.17	0.15	0.88	0.31	0.93	0.29	-0.06
12	-0.22	6.51	-0.10	2.00	-0.10	1.57	12.88	1.75	4.16	0.77	1.10	0.80	0.92	4.72
24	-0.29	13.20	-0.07	3.87	-0.07	2.03	23.32	4.07	9.89	0.98	2.71	0.93	1.55	12.54

TABLE 1.C.4: **Simulating from Zhang (2005)**
This table reports results from 1000 simulations of the Zhang, 2005a investment-based asset pricing model. We thank Lu Zhang for sharing the code on his website. This model endogenously generates a time-varying value spread that predicts value returns in the time series. We ask whether this model can match the variation in expected value returns observed in the data, while matching other moments of interest. Panel A reports the unconditional moments of value decile portfolios (focusing on deciles 1 (Low), 4, 7, and 10 (High) for brevity) and the High-minus-Low decile value premium. We see that our distribution is close to what is reported in Zhang (2005, Table III). Panel B reports the *Ratio* of the predictive regression coefficient to the intercept and the R^2 at the annual horizon ($h = 12$). We compare the level of these estimates from the data (as reported in Table 1.A.2) to the distribution of their counterparts estimated from simulating the model. We rank all simulations on the *Ratio* and report both the *Ratio* and R^2 at the 50, 90, 95, and 99th percentile of that distribution. The final column presents the mean across simulations as reported in Zhang, 2005a, where we have backed out the *Ratio* of 0.76 from results reported in his Table III and V.

Panel A: Unconditional Moments for Value Decile Portfolios						
	Simulated Distribution					Zhang (2005)
	1 (Low)	4	7	10 (High)	H-L	HML
Mean of avg. ret.	0.73	0.85	0.94	1.14	0.41	0.39
Mean of st. dev.	6.77	7.70	8.42	10.39	3.84	3.46

Panel B: Simulated Distribution of Annual H-L Value Premium on H-L Value Spread							
	Data		Simulated Distribution				Zhang (2005)
	BM_{ExFin}	BM_{IndAdj}	50	90	95	99	Mean
Ratio	3.47	2.12	0.74	1.39	1.64	2.50	0.76
R^2	13.94	26.24	0.20	3.32	6.89	27.26	8.84

TABLE 1.C.5: **Predicting Value Returns with the Value Spread: Pooled Tests in Subsamples**
This table is similar to Table 1.A.3 of the paper, but reports results over the first and second half of the sample.

	$H-L$					*Rank*				
h	a	b	t_a	t_b	R^2	a	b	t_a	t_b	R^2
Panel A: Pooled Predictive Regression (Jan 1972 - Jun 1994)										
1	0.35	0.25	2.24	1.42	0.38	0.35	0.22	2.31	1.20	0.29
12	3.93	5.87	2.32	3.92	11.50	3.42	5.67	1.68	2.91	9.19
24	9.99	14.21	2.80	4.10	22.28	8.50	14.45	1.97	2.84	18.62
Panel B: Pooled Predictive Regression (Jul 1994 - Dec 2017)										
1	0.18	0.45	1.27	3.05	0.71	0.26	0.40	1.61	2.31	0.55
12	1.80	6.13	1.62	3.81	8.30	3.74	6.41	2.49	3.36	7.62
24	1.67	10.98	0.93	3.09	10.52	6.45	12.07	2.42	3.23	11.66
Panel C: Average Value Return on Average Value Spread (Jan 1972 - Jun 1994)										
1	0.48	0.26	2.93	0.99	0.57	0.48	0.21	2.87	0.85	0.24
12	6.87	4.20	3.63	2.00	17.27	6.46	4.32	3.38	2.08	11.97
24	17.16	10.72	4.78	3.70	34.26	16.23	12.80	4.45	4.19	31.14
Panel D: Average Value Return on Average Value Spread (Jul 1994 - Dec 2017)										
1	0.03	0.35	0.25	2.11	2.68	0.12	0.34	0.93	1.94	2.10
12	-0.25	4.14	-0.18	2.76	27.97	1.46	4.11	0.93	2.50	18.54
24	-2.05	4.46	-0.70	1.75	17.23	2.12	4.66	0.66	1.69	11.89

TABLE 1.C.6: **Pooled Predictive Regression of Market Returns on the Value Spread**
This table presents the pooled predictive regression of Table 1.A.3 in the paper, but now we substitute the returns of the market portfolio in each asset class c on the left-hand side. We find no evidence that the (High-minus-Low) value spread predicts market returns in the pool of asset classes.

	$H-L$					*Rank*				
h	a	b	t_a	t_b	R^2	a	b	t_a	t_b	R^2
1	0.54	0.15	2.85	0.62	0.12	0.54	0.11	2.85	0.49	0.07
3	1.64	0.61	3.18	0.97	0.60	1.65	0.50	3.19	0.80	0.41
6	3.32	1.57	3.22	1.35	1.82	3.32	1.39	3.23	1.16	1.45
12	6.79	2.69	3.24	1.31	2.49	6.80	2.28	3.25	1.07	1.79
24	14.43	3.60	3.40	1.06	2.04	14.46	3.00	3.41	0.85	1.42
48	35.12	7.65	3.74	1.05	3.52	35.11	7.45	3.75	0.98	3.33

TABLE 1.C.7: **Value Timing in Individual Equities**
This table reports unconditional performance statistics for the monthly returns of an out-of-sample trading strategy similar to Table 1.A.5 in the paper, but here the strategy uses the largest stocks whose market caps cumulate to either 75% or 95% of total market cap in the CRSP file.

Market Cap Cutoff		75%							95%						
		Avg. ret.	t	Sharpe	α_{CAPM}	t_α^{CAPM}	α_{FF3}	t_α^{FF3}	Avg. ret.	t	Sharpe	α_{CAPM}	t_α^{CAPM}	α_{FF3}	t_α^{FF3}
Panel A: High-minus-Low ($H-L$)															
Ind. Equities ($BM_{Ex.fin.}$)	Unit weight	-0.13	-0.68	-0.03	0.05	0.25	-0.48	-3.09	0.08	0.42	0.02	0.26	1.36	-0.29	-1.97
	Linear Timing	0.65	2.72	0.12	0.62	2.50	0.71	2.92	0.50	2.47	0.11	0.39	1.88	0.48	2.38
	Combined	0.52	2.03	0.09	0.66	2.53	0.23	0.93	0.58	2.53	0.11	0.65	2.77	0.19	0.91
Ind. Equities ($BM_{Ind.adj.}$)	Unit Weight	0.26	1.39	0.06	0.36	1.86	-0.06	-0.36	0.23	1.17	0.05	0.25	1.27	-0.19	-1.14
	Linear Timing	0.48	2.08	0.09	0.45	1.91	0.42	1.78	0.56	2.56	0.11	0.51	2.27	0.44	1.99
	Combined	0.75	2.74	0.12	0.81	2.91	0.36	1.40	0.79	3.07	0.13	0.76	2.90	0.25	1.13
Panel B: Rank-Weighted ($Rank$)															
Ind. Equities ($BM_{Ex.fin.}$)	Unit weight	-0.03	-0.14	-0.01	0.23	1.21	-0.35	-2.44	0.14	0.74	0.03	0.38	1.97	-0.22	-1.67
	Linear Timing	0.36	1.46	0.06	0.36	1.40	0.43	1.72	0.46	2.14	0.09	0.45	2.05	0.43	2.02
	Combined	0.34	1.21	0.05	0.59	2.09	0.08	0.31	0.60	2.27	0.10	0.83	3.10	0.20	0.87
Ind. Equities ($BM_{Ind.adj.}$)	Unit Weight	0.34	1.79	0.08	0.44	2.28	-0.05	-0.34	0.45	2.33	0.10	0.45	2.27	-0.03	-0.22
	Linear Timing	0.28	1.15	0.05	0.27	1.07	0.23	0.92	0.45	1.91	0.08	0.47	1.97	0.38	1.63
	Combined	0.62	2.09	0.09	0.71	2.32	0.18	0.63	0.90	3.15	0.13	0.92	3.15	0.34	1.36

TABLE 1.C.8: **Value Timing in Alternative Asset Classes**
This table is similar to Table 1.A.5 of the paper, but uses the historically standardized value spread, $VS_{t,His}$, to time the value strategy in the alternative asset classes. We present results for a unit weight strategy that passively captures the unconditional value premium, a linear timing strategy that invests $VS_{t,His}$ dollars in both the long and short position of the value strategy, and, finally, a combined strategy that invests $1 + VS_{t,His}$. To make these different value strategies comparable, we scale each value return series ex ante to have an annualized standard deviation of 15%. For asset classes with data available from 1972, the burn-in period for calculating $VS_{t,His}$ is 120 months, whereas for the remaining asset classes the burn-in period is 60 months. The α represents the abnormal return with respect to the CRSP value-weighted market portfolio for US industries and with respect to an equal-weighted portfolio of the securities in each alternative asset class for commodities, currencies, government bonds and stock indexes.

		$H-L$				$Rank$			
		Avg. ret.	t	α	t_α	Avg. ret.	t	α	t_α
US Industries	Unit Weight	-0.05	-0.25	0.10	0.48	-0.07	-0.38	0.07	0.36
	Linear Timing	0.28	1.29	0.28	1.28	0.28	1.28	0.24	1.06
	Combined	0.23	0.92	0.38	1.49	0.21	0.82	0.31	1.19
Commodities	Unit Weight	0.35	1.69	0.36	1.74	0.45	2.18	0.47	2.27
	Linear Timing	-0.12	-0.72	-0.09	-0.52	-0.18	-1.08	-0.13	-0.81
	Combined	0.23	1.08	0.28	1.32	0.27	1.29	0.34	1.67
Currencies	Unit Weight	0.50	2.38	0.51	2.42	0.51	2.46	0.53	2.55
	Linear Timing	0.19	1.08	0.18	1.02	0.15	0.87	0.14	0.78
	Combined	0.69	2.64	0.69	2.63	0.66	2.34	0.67	2.34
Gov't Bonds	Unit Weight	-0.47	-1.76	-0.02	-0.08	-0.39	-1.47	0.01	0.02
	Linear Timing	0.23	1.28	0.15	0.80	0.22	1.09	0.15	0.75
	Combined	-0.25	-1.36	0.13	0.81	-0.18	-1.12	0.16	1.18
Stock Indexes	Unit Weight	0.24	0.88	0.07	0.27	0.38	1.35	0.18	0.73
	Linear Timing	0.36	0.98	0.26	0.69	0.34	0.91	0.21	0.56
	Combined	0.61	1.07	0.32	0.60	0.72	1.26	0.39	0.73

TABLE 1.C.9: **Common and Asset-Class-Specific Components of the Value Spread (Rank-Weighted)**
This table is identical to Table 1.A.8 of the paper, but uses rank-weighted value strategies instead of High-minus-Low value strategies. We report results for pooled predictive regressions of value returns on components of the value spread.

h	a	b_{Com}	b_{Spec}	t_a	$t_{b_{Com}}$	$t_{b_{Spec}}$	R^2
Panel A: Common Value							
1	0.30	0.33		2.69	1.82		0.30
3	0.89	1.03		2.84	2.11		0.82
6	1.84	2.19		2.92	2.15		1.66
12	3.96	5.18		3.14	2.60		3.75
24	9.07	13.81		3.49	3.27		9.72
48	21.52	31.47		4.04	5.15		14.65
Panel B: Specific Value							
1	0.26		0.26	2.49		2.09	0.17
3	0.78		0.96	2.58		3.26	0.68
6	1.61		2.49	2.62		4.44	2.02
12	3.46		6.15	2.67		5.26	4.93
24	7.82		11.78	2.63		5.03	6.43
48	19.54		21.36	2.56		4.79	5.96
Panel C: Common and Specific Value							
1	0.30	0.33	0.26	2.69	1.81	2.08	0.47
3	0.89	1.03	0.96	2.83	2.10	3.27	1.50
6	1.84	2.19	2.49	2.91	2.13	4.50	3.68
12	3.96	5.18	6.15	3.10	2.52	5.41	8.68
24	9.07	13.81	11.78	3.37	3.12	5.25	16.15
48	21.52	31.47	21.36	3.83	4.80	5.20	20.61

TABLE 1.C.10: **Comovement Between Risk-Proxies and the Value Spread (Rank-Weighted)**
This table is identical to Table 1.A.9 of the paper, but uses rank-weighted value strategies instead of High-minus-Low value strategies. We regress components of the value spread on popular proxies of risk premia.

	Intermediary Leverage	Illiquidity Premium	Dividend Yield	Global Recession	Default Spread	Real Uncertainty	Chicago Fed National Activity Index	R^2
				Panel A: Common Value				
1	0.54							51.27
	(5.98)							
2		0.49						42.19
		(11.36)						
3	0.40	0.31						64.36
	(5.48)	(4.69)						
4			0.66					77.15
			(12.30)					
5	0.08	0.15	0.51					80.10
	(1.35)	(2.40)	(7.43)					
6				0.41				7.10
				(2.13)				
7	- 0.04	0.11	0.55	0.20	0.09	0.06	- 0.02	83.53
	(-0.46)	(1.75)	(7.75)	(2.88)	(1.54)	(0.85)	(-0.41)	
3 (Innovations)	0.62	0.15						43.71
	(6.10)	(3.19)						
5 (Innovations)	0.20	0.06	0.62					62.34
	(2.83)	(1.91)	(7.26)					
				Panel B: Asset-Class-Specific Value				
Ind. Equities ($BM_{Ex.fin.}$)	-0.07	0.00	0.00	-0.02	0.05	0.07	0.01	12.05
	(- 0.98)	(- 0.12)	(0.04)	(- 0.35)	(1.45)	(1.67)	(0.61)	
Ind. Equities ($BM_{Ind.Adj.}$)	0.17	-0.08	-0.17	-0.14	0.06	0.13	0.03	60.79
	(3.34)	(-3.91)	(-3.65)	(-3.78)	(1.49)	(3.33)	(1.49)	
Industries (BM)	-0.22	0.04	0.16	0.04	-0.02	-0.05	-0.01	31.81
	(-4.82)	(1.50)	(3.24)	(0.62)	(-0.63)	(-1.19)	(-0.65)	
Commodities	0.22	-0.12	-0.21	-0.11	-0.38	0.53	0.12	19.07
	(0.99)	(-1.33)	(-0.99)	(-0.60)	(-2.41)	(4.25)	(1.56)	
Currencies	0.04	0.24	0.04	-0.37	0.10	-0.19	-0.18	11.72
	(0.21)	(1.44)	(0.18)	(-1.44)	(0.63)	(-1.28)	(-1.87)	
Government Bonds	-0.26	0.15	1.52	-0.03	-0.31	0.03	-0.21	38.21
	(-1.26)	(0.56)	(3.13)	(-0.11)	(-1.89)	(0.14)	(-2.29)	
Stock Indexes	0.53	-0.07	-0.59	0.27	-0.18	-0.67	0.02	38.73
	(2.20)	(-0.25)	(-1.29)	(1.55)	(-1.30)	(-4.28)	(0.26)	

TABLE 1.C.11: **Common Versus Specific Value Return Predictability Net of Risk Proxies (Rank-Weighted)**
This table is identical to Table 1.A.10 of the paper, but we now use rank-weighted value returns and value spreads. We regress returns on the explained (by the risk proxies) and orthogonal (to the risk proxies) components of the common component and the asset-class-specific components of the value spread.

Specification	h	a	$b_{Com,Orth}$	$b_{Com,Expl}$	t_a	$t_{Com,Orth}$	$t_{Com,Expl}$	R^2	$R^2_{Com,Orth}$	$R^2_{Com,Expl}$
Panel A: Common Value										
	1	0.29	0.25	0.83	2.63	1.26	1.90	0.41	0.14	0.26
	3	0.86	0.76	2.61	2.75	1.32	2.65	1.15	0.38	0.77
	6	1.79	1.65	5.36	2.85	1.38	2.74	2.26	0.81	1.45
	12	3.83	3.84	12.95	3.15	1.90	3.05	5.20	1.76	3.44
	24	8.80	10.86	30.98	3.70	3.14	3.51	12.30	5.12	7.18
	48	21.19	26.29	60.94	4.22	5.06	4.42	16.91	8.69	8.22
Panel B: Asset-Class-Specific Value										
	1	0.26	-0.04	0.36	2.49	-0.17	2.59	0.25	0.00	0.25
	3	0.78	0.30	1.18	2.59	0.60	3.49	0.78	0.02	0.77
	6	1.61	0.70	3.11	2.64	0.73	4.94	2.38	0.04	2.34
	12	3.45	3.94	6.95	2.68	2.06	5.09	5.16	0.54	4.62
	24	7.82	13.52	11.13	2.63	4.35	3.80	6.49	2.31	4.18
	48	19.54	34.38	16.23	2.55	7.31	2.49	6.83	4.36	2.47

TABLE 1.C.12: **Common and Asset-Class-Specific Components of the Value Spread (Alternative Principal Component)**
This table is identical to Table 1.A.8 of the paper, but uses an alternative principal component procedure to generated the common component of the value spreads. This alternative procedure extracts the first principal component from a panel of value spreads that is balanced with an algorithm that recursively projects the value spread in an asset class with a shorter sample on the value spreads that are available over the full sample.

h	a	b_{Com}	b_{Spec}	t_a	$t_{b_{Com}}$	$t_{b_{Spec}}$	R^2
Panel A: Common Value							
1	0.25	0.55		2.38	2.04		0.38
3	0.73	1.82		2.63	2.67		1.26
6	1.49	4.00		2.89	3.21		2.79
12	3.12	9.00		3.23	4.38		5.85
24	6.92	21.81		3.36	4.93		12.02
48	16.77	52.19		3.66	6.77		17.18
Panel B: Specific Value							
1	0.21		0.27	2.13	0.00	2.29	0.20
3	0.61		0.93	2.25	0.00	3.21	0.72
6	1.26		2.18	2.37	0.00	4.27	1.84
12	2.64		5.49	2.41	0.00	5.18	4.78
24	5.84		11.06	2.09	0.00	5.45	6.60
48	15.23		21.56	1.81	0.00	3.52	6.08
Panel C: Common and Specific Value							
1	0.25	0.55	0.27	2.38	2.04	2.29	0.58
3	0.73	1.82	0.93	2.64	2.65	3.22	1.97
6	1.49	4.00	2.18	2.93	3.19	4.35	4.63
12	3.12	9.00	5.49	3.31	4.47	5.54	10.63
24	6.92	21.81	11.06	3.41	5.09	5.77	18.62
48	16.77	52.19	21.56	3.63	7.40	3.66	23.26

TABLE 1.C.13: **Common and Asset-Class-Specific Components of the Value Spread (Simple Average)**
This table is identical to Table 1.A.8 of the paper, but the common component is now defined as the average value spread over the asset classes with available data in month t. The asset-class-specific component is defined as the difference between the value spread in an asset class and the simple average across asset classes.

h	a	b_{Com}	b_{Spec}	t_a	$t_{b_{Com}}$	$t_{b_{Spec}}$	R^2
Panel A: Common Value							
1	0.21	0.53		2.15	2.49		0.59
3	0.61	1.69		2.33	3.63		1.86
6	1.26	3.53		2.56	4.15		3.73
12	2.64	7.84		2.88	5.47		7.53
24	5.84	17.79		2.96	5.64		13.25
48	15.23	41.20		3.11	9.62		17.27
Panel B: Specific Value							
1	0.21		0.28	2.12		2.50	0.24
3	0.61		0.93	2.24		3.46	0.85
6	1.26		2.20	2.34		4.26	2.19
12	2.64		5.47	2.35		4.54	5.62
24	5.84		12.20	2.05		4.53	9.60
48	15.23		25.68	1.77		4.54	10.42
Panel C: Common and Specific Value							
1	0.21	0.53	0.28	2.15	2.49	2.50	0.84
3	0.61	1.69	0.93	2.33	3.63	3.46	2.70
6	1.26	3.53	2.20	2.56	4.15	4.26	5.92
12	2.64	7.84	5.47	2.88	5.47	4.54	13.15
24	5.84	17.79	12.20	2.96	5.64	4.53	22.85
48	15.23	41.20	25.68	3.11	9.62	4.54	27.69

TABLE 1.C.14: **Principal Component of Value Spreads Versus Principal Component of Risk Proxies**
This table runs a horse race between a principal component of seven risk proxies ((1) a global recession dummy; (2) the dividend yield; (3) the default spread; (4) the illiquidity premium; (5) real uncertainty; (6) intermediary leverage; and, (7) sentiment) and the common component of the value spreads, defined as the first principal component. Panel A presents the loadings of the first principal component of the risk proxies as well as the correlation with the first principal component of the value spreads. Panel B presents results from pooled predictive regressions of value returns (from seven value strategies) on the first principal component of the risk proxies: $R_{c,t+1:t+h} = a_h + b_h Macro_t^{PC1} + \varepsilon_{c,t+1:t+h}$. Panel C controls for common value. t-statistics are calculated using Driscoll and Kraay, 1998 standard errors with h lags. The full sample period is 1972 to 2017, but the alternative asset classes enter the sample only after 1972.

Panel A: Principal Component of Risk Proxies

	Intermediary Leverage	Illiquidity Premium	Dividend Yield	Default Spread	Real Uncertainty	Chicago Fed National Activity Index	Cum R^2	Corr($Macro^{PC1}$, VS^{Com})
Loadings	0.47	0.36	0.42	0.43	0.46	0.27	0.58	0.80

Panel B: Value Returns on Principal Component of Risk Proxies

h	a	$b_{Macro^{PC1}}$	t_a	$t_{b_{Macro^{PC1}}}$	R^2
1	0.23	0.11	2.25	1.45	0.20
3	0.68	0.37	2.44	1.80	0.72
6	1.40	0.84	2.65	2.20	1.76
12	2.91	1.85	2.83	3.16	3.55
24	6.46	4.54	2.68	3.59	7.48
48	16.06	10.63	2.55	3.86	10.33

Panel C: Principal Component of Risk Proxies versus Common Value

h	a	b_{Com}	$b_{Macro^{PC1}}$	t_a	$t_{b_{Com}}$	$t_{b_{Macro^{PC1}}}$	R^2
1	0.25	0.46	-0.03	2.39	1.56	-0.28	0.36
3	0.73	1.44	-0.07	2.65	2.27	-0.23	1.19
6	1.50	2.83	-0.02	2.90	2.46	-0.04	2.61
12	3.18	7.19	-0.34	3.27	3.27	-0.42	5.83
24	7.19	18.88	-1.21	3.57	3.83	-0.93	13.01
48	17.60	46.14	-3.37	3.89	5.18	-1.02	18.78

TABLE 1.C.15: **Predicting Value Returns with the Equity Value Spread**
This table is identical to Table 1.A.3 of the paper, but we replace the value spread in each asset class with the average of the value spread for the two individual equity strategies ($BM_{Ex.fin.}$ and $BM_{Ind.Adj.}$). Panels A and B report results from pooled and time-series regressions using all seven value strategies, whereas in Panels C and D we exclude the value returns for the two individual equity strategies.

	$H-L$					$Rank$				
h	a	b	t_a	t_b	R^2	a	b	t_a	t_b	R^2
Panel A: Pooled Predictive Regression										
1	0.21	0.25	2.14	1.91	0.33	0.26	0.24	2.51	1.80	0.31
12	2.64	3.99	2.74	4.04	4.95	3.45	3.43	2.89	2.60	3.25
24	5.84	9.64	2.74	3.86	9.88	7.82	8.82	3.17	2.85	7.65
Panel B: Average Value Return on Average Value Spread										
1	0.25	0.28	2.48	1.93	1.17	0.29	0.27	2.80	1.82	1.00
12	3.30	4.56	2.74	3.52	22.16	3.96	4.09	3.16	3.02	13.05
24	7.75	11.14	3.27	5.03	38.96	9.32	10.75	3.76	4.44	28.55
Panel C: Excluding Value Returns in Individual Equities (Pooled)										
1	0.20	0.09	2.13	0.76	0.05	0.22	0.14	2.29	1.24	0.10
12	2.31	1.88	2.44	2.17	1.16	2.61	1.56	2.49	1.56	0.75
24	5.15	4.61	2.65	2.81	2.48	5.68	3.65	2.76	1.89	1.58
Panel D: Excluding Value Returns in Individual Equities (Average-on-Average)										
1	0.22	0.10	2.19	0.84	0.00	0.23	0.13	2.28	1.19	0.12
12	2.80	2.32	2.33	2.00	6.99	2.87	1.90	2.33	1.65	3.48
24	6.99	5.88	3.01	2.91	17.79	6.82	4.70	2.93	2.27	9.17

TABLE 1.C.16: **Common and Asset-Class-Specific Components of the Value Spread in Subsamples**
This table is identical to Table 1.A.8 of the paper, but reports results for two subsamples split in June 1994.

h	a	b_{Com}	b_{Spec}	t_a	$t_{b_{Com}}$	$t_{b_{Spec}}$	R^2
Panel A: High-minus-Low ($H-L$)							
First Half							
1	0.41	0.11	0.26	2.41	0.35	1.33	0.22
12	4.50	4.10	6.45	2.66	1.65	3.79	8.53
24	10.01	13.68	11.46	2.54	2.64	4.28	14.58
Second Half							
1	0.42	0.85	0.28	1.68	1.97	1.89	0.67
12	4.48	10.31	4.74	3.82	5.49	4.09	7.90
24	4.60	13.83	9.38	2.48	3.85	4.04	8.27
Panel B: Rank-Weighted ($Rank$)							
First Half							
1	0.43	0.07	0.24	2.25	0.20	1.21	0.18
12	3.94	4.09	6.35	1.82	1.13	3.32	7.69
24	7.09	15.70	11.13	1.38	1.98	3.24	14.04
Second Half							
1	0.73	1.12	0.31	2.03	2.03	1.89	0.87
12	8.03	12.07	6.28	3.72	4.05	4.50	8.66
24	10.98	15.88	12.65	2.60	3.38	4.48	10.93

Chapter 2

New and Old Sorts: Implications for Asset Pricing

This chapter was written in co-authorship with Martijn Boons and Andrea Tamoni. We are grateful to Fernando Anjos, Kent Daniel (discussant) Esther Eiling, Rik Frehen, Jens Kværner, Juhani Linnainmaa, Lars Lochstoer, Tyler Muir, Dino Palazzo (discussant), Florian Peters, Simon Rottke, and seminar participants at the 2021 American Finance Association, 2020 Northern Finance Association, Rutgers Business School, Tilburg University, University of Amsterdam, Bauer College of Business – University of Houston, and Vrije Unversiteit Amsterdam for helpful comments.

2.1 Introduction

2.2 Introduction

In this paper, we study characteristic-based return predictability over horizons up to five years. Although existing literature on cross-sectional predictability almost exclusively focuses on the short-term returns to characteristic-based investing, studying longer horizon returns is interesting for various reasons. First, the horizon of most investors is considerably longer than a single month. Similarly, capital budgeting decisions of firms usually rely on discounting long-term cash flows. Furthermore, characteristics that predict returns more persistently are relatively more important for the real economy (Van Binsbergen and Opp, 2019). Finally, longer horizon returns provide a new set of useful moments to evaluate asset pricing models. These models ultimately serve to determine the discount rate used in investment decisions throughout the economy.

To study the performance of characteristic-sorted portfolios across horizons, we follow Freyberger et al., 2020a and analyze a large set of 56 characteristics that previous literature finds to predict stock returns in the cross-section. For each characteristic X, we construct value-weighted decile portfolios and track the buy-and-hold return of the high-minus-low strategy from one month up to five years after portfolio formation. This approach provides us with a three-dimensional panel of returns denoted $R_{X,(t-s),t+1}$, where $(t-s)$ refers to the sorting date and $s = 0, ..., 60$, over the sample period from 1972 to 2019. Throughout we refer to $R_{X,(t-s),t+1}$ for $s > 0$ as the return to an older sort, and to $R_{X,(t),t+1}$ as the return to the newest sort. Our paper consists of three parts that each answers important asset pricing questions using these old and new sorts.

First, What return should we expect from old sorts? and Do old sorts improve investment opportunities? In answering these questions, we take into account that persistence varies considerably *across* characteristics. This approach differentiates our work from Keloharju et al., 2019, who highlight that returns of old sorts are relatively small for the *average* characteristic (meaning that on average returns decay to zero quickly after portfolio formation). We instead show that returns of old sorts are abnormally large for some characteristics, and abnormally small for others, and this fact leads to large improvements in Sharpe ratio for investors when combining old and new sorts. Second, Do benchmark asset pricing models capture the performance differentials between old and new sorts? If not, what components of the average returns to old and new sorts are difficult to explain? Challenging models with new moments is interesting, because multi-factor models, like Fama and French, 2015, have added factors to the CAPM to improve explanatory power for cross-sections of returns at short horizons after portfolio formation. As argued in Harvey and Liu, 2019, some of the improved fit for these returns is likely due to overfitting, which may harm the performance of these models in tests using longer-term returns. Given that we construct contemporaneous returns to new and old sorts, we are able to study the across-horizon performance of benchmark asset pricing models using standard tests based on the intercepts from monthly time-series regressions. Third, Which stocks drive the performance differential between old and new sorts? and Which stocks present the largest challenge to asset pricing models? To answer these questions, we perform a novel decomposition of the newest sort into the components coming from old and new stocks, respectively. These stocks together make up the extreme high or low characteristic-sorted portfolio today, but it is only the old stocks that were in (or close to) that same characteristic-sorted portfolio in the past.

Let us describe in detail the main results for each part. Building on the popular characteristic-based model of expected returns, we derive a simple relation between the expected returns of old and new sorts. If the compensation for a characteristic is independent of how long ago the portfolio was formed, then expected high-minus-low returns decay after sorting at the same speed as the characteristic spread between the high and low portfolio. Instead, in the data, the persistence of characteristic-based return predictability often does not match the persistence of the characteristic. Moreover, we find that old sorts provide a significantly negative alpha relative to new sorts for 22 characteristics (indicating that returns decay too fast), whereas this alpha is positive and significant for another 10 characteristics (indicating that returns decay too slow). For instance, three years after portfolio

We perform a simulation study under the null to show that these results are unlikely due to chance. A variety of robustness checks confirm that our conclusions extend in subsamples and when estimating conditional alphas.

formation, a high-minus-low book-to-market strategy provides an alpha of 40 basis points per month (t-stat=2.66) relative to the newest book-to-market strategy. This alpha translates to a large improvement in Sharpe ratio: from 0.28 for the newest book-to-market sort to 0.52 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 over 100% are commonplace. We conclude that the returns to old sorts are abnormal relative to new sorts for a large subset of characteristics, and that combining old and new sorts improves the performance of characteristic-based investment strategies.

Turning to our second set of research questions, we find that benchmark asset pricing models do not fully capture the alphas between old and new sorts. To understand this failure, note that correlations between old and new sorts are generally high and decrease slowly as time passes after portfolio formation. For the median characteristic, the correlation between the return of the newest sort and the return of older sorts that were constructed one and five years ago, respectively, equals 0.85 and 0.64. While there is variation in this correlation across characteristics, we find that this variation is unrelated to the alphas between old and new sorts. When old and new sorts are highly correlated, but there is a large alpha separating these two returns, models that do a good job explaining returns of older sorts will have a hard time explaining returns of newer sorts, and vice versa.

There is an important trade-off in the number of factors, however. Small models, like the CAPM, do relatively well in pricing the older sorts (consistent with Kothari et al., 1995 and Cohen et al., 2009, among others), but are firmly rejected using the returns of newer sorts. In contrast, big multi-factor models do relatively well pricing the newer sorts, but are firmly rejected using the returns of older sorts. Among the latter set are the models of Frazzini and Pedersen (2014b), Fama and French, 2015, Hou et al. (2015a), Stambaugh and Yuan (2016), Daniel et al. (2017), and Daniel et al. (2019). 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 of characteristic-sorted portfolio returns described in this paper.

We also show that leading theoretical explanations of characteristics-based return predictability, such as Gomes et al., 2003 and Zhang, 2005b, cannot explain the performance differentials between new and old sorts.

The tension between new and old sorts is perhaps clearest when we apply principal component analysis (PCA) to the returns at each horizon. We treat the principal components extracted from the returns of the newest sorts, $R_{X,(t),t+1}$, as statistical factors in an asset pricing model (as in Kozak et al., 2019, Haddad et al., 2018, and Lettau and Pelger, 2020b). In Gibbons et al. (1989, GRS) tests, we find that these statistical factors do not price the principal components extracted from older sorts. Interestingly, the rejection in the GRS test is driven almost completely by the first principal component of returns. Consistent with this finding, our evidence suggests that adding a single factor extracted from old sorts to three statistical factors extracted from the newest sorts goes a long way in capturing the performance differentials between old and new sorts. This model also does a relatively good job at explaining variation of average returns in cross-sections of new and old sorts. In fact, it is the only model for which the cross-sectional R^2 is positive when factor risk premia are forced to match their sample average return.

Our analysis further uncovers a new dimension to the low beta anomaly, as the performance differentials between old and new sorts are strongly related to market beta. The low beta anomaly refers to the result – first documented in Black et al., 1972 – that a sort of stocks on market beta yields a large spread in beta, but not in average returns. Similarly, we sort characteristic-sorted portfolios on their market beta. We find that it is among low (high) market beta characteristics that average returns decay too fast (slow) and old sorts underperform (outperform). Existing asset pricing models do not capture this variation across market beta groups in the relative performance of old and new sorts.

Turning to our last research questions, the fact that returns immediately after portfolio formation are different from returns longer after portfolio formation, suggests that "new" stocks recently entering the extreme decile portfolio have a different contribution to returns than "old" stocks that entered the portfolio a long time ago. In line with this intuition, we find that the old-minus-new *stock* return differential lines up well with the old-minus-new *sort* differential among the 56 characteristics we study. For instance, a book-to-market strategy that uses only new stocks (that have recently seen a relatively large change in book-to-market), obtains a return that is 57 bps (t-stat = 2.18) lower than a strategy that uses only old stocks. This result obtains even though these two sets of stocks generate the same spread in book-to-market today.

In fact, even old-minus-new stock strategies that are neutral with respect to recent values of a much larger variety of characteristics (including, among others, book-to-market, size, profitability, investment, momentum, and idiosyncratic volatility) provide large abnormal returns relative to both small and big benchmark models. These abnormal returns vary in sign across characteristics and are significantly negative (positive) among low (high) market beta characteristics. This finding implies that popular explanations of the cross-section of expected returns based on recent observations of firm characteristics are incomplete: These explanations are either still missing an important characteristic or failing to account for lagged characteristics. This finding also indicates that the performance of characteristic-based investment strategies can be improved by accounting for the dynamics of characteristics at the firm-level. This insight is important because most stock-picking applications explicitly reduce the information set to the most recent values of firm characteristics.

In all, our paper makes significant progress in understanding the dynamics of expected returns in the cross-section. We show that benchmark asset pricing models uniformly fail to explain the new moments we derive from relative returns at short and long horizons after formation of a characteristic-sorted portfolio. Our results are broadly relevant for academics evaluating asset pricing models, investors trading characteristics, and managers estimating discount rates for capital budgeting decisions.

Literature

The literature on characteristics-based return predictability is vast, but almost exclusively studies the relation between characteristics and short-term returns; in contrast to this paper. In recent machine learning literature, the goal is to find the (potentially higher-order) functional form of a large set of characteristics that best predicts these

See, among many others, Brandt et al., 2009, Lewellen, 2014, Light et al., 2017, and Gu et al., 2020b as well as the factor models of BARRA (`https://www.msci.com/www/research-paper/the-barra-us-equity-model-use4/014291992`) and Bloomberg (Baturin et al., 2010) that are popular in the industry.

returns (see, e.g., Kozak et al., 2019, Freyberger et al., 2020a, and Gu et al., 2020b). Similarly, empirical tests of asset pricing models typically use both factors and test assets derived from sorting stocks on recent observations of characteristics (see, e.g., Fama and French (2015, 2018), Hou, Xue, and Zhang (2015, 2018)). We derive new moments from the returns at longer horizons after portfolio formation to challenge existing models.

Cho and Polk, 2019 and Binsbergen et al., 2021 also analyze longer-horizon returns 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. Binsbergen et al. (2021) focus on the dynamics of these price wedges at the portfolio- and firm-level and their potential for real capital misallocations; Cho and Polk, 2019 focus on the interaction between value and quality as the main determinant of price wedges in the cross-section (Cohen et al., 2009 also estimate a "price wedge" for value). In contrast to these papers, we study the relative performance of new and old sorts and highlight implications for characteristic-based investing and asset pricing models.

Chernov et al., 2018 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. Although we share the objective of generating new assets to test for (conditional) model misspecification, our paper differs in important dimensions. First, Chernov et al., 2018 use multi-period compounded returns of factors that are rebalanced annually, while we use returns at different horizons after characteristic-based portfolios are formed. Our approach is more representative of the experience of a buy-and-hold investor and relates more directly to the problem of finding the appropriate discount rate for the long-term cash flows of a firm. Second, our empirical approach is based on standard asset pricing tests (like Jensen's (1968) alpha and the GRS test), such that we do not face the usual inferential problems associated with tests that use compounded returns. Third, we study a much larger set of 56 characteristics. Fourth, we document the relative contribution of new versus old stocks to the returns from characteristic-based investing.

Our new-versus-old stock decomposition is new to the literature and different from the permanent-transitory decomposition in Keloharju et al., 2019. These authors sort the full cross-section of stocks on permanent and transitory components of characteristics and find that it is the latter component that drives return predictability for the average characteristic. Instead, our decomposition separates the stocks within the extreme high and low decile portfolios in a new and old group. The old-minus-new return differential varies in sign across characteristics, which fact remains hidden when focusing on the average characteristic. Our results imply that past values of characteristics (or changes in characteristics) contain independent information for the cross-section of stock returns. We thus extend previous work that draws a similar conclusion for book-to-market and size (Cochrane, 2011b; Gerakos and Linnainmaa, 2018b) and idiosyncratic volatility (Rachwalski and Wen, 2016).

Daniel et al., 2019 (see, also, Daniel and Titman, 1997; Herskovic et al., 2019) argue that factors can be traded more profitably by combining a factor, like the high-minus-low book-to-market portfolio (HML), with an offsetting position in a hedge portfolio that has a zero loading on the characteristic (book-to-market), but a maximum loading on the factor. 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, among which are the optimally hedged factors of Daniel et al., 2019. In fact, we reject the popular assumption (see, e.g., Cochrane, 2011b, p. 1062) that firms' loadings on the SDF are a function of current values of characteristics (size, book-to-market, profitability and investment). The reason is that our old-minus-new stock strategy is approximately neutral with respect to these characteristics, but has a non-zero average excess return. Thus, we establish a new fact, which is that the expected return of a portfolio – in a given period and holding its current loadings on characteristics fixed – depends on the time this portfolio was formed. Finally, we uncover a new across-characteristic dimension to the low beta anomaly, which contributes to recent literature that studies this anomaly across asset classes (see, e.g., Asness et al., 2012 and Frazzini and Pedersen, 2014b).

2.3 Data

We provide a detailed description of the 56 characteristics we study in Table 3.C.1. For all US common stocks traded on the NYSE, AMEX or NASDAQ from July 1964 to December 2017, 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., 2020b, we delay monthly variables by one month and annual variables by six months. We construct for each characteristic value-weighted decile portfolios split at NYSE breakpoints to reduce the influence of microcap stocks on our results. We track the buy-and-hold returns to these decile portfolios. When a stock goes missing, we reallocate the investment in this stock 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:

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

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 $s = 0, ..., 60$; and, the third subscript refers to the return realization date, $t+1$. The novelty of our method is in varying the sorting date, so that one observes *contemporaneous* returns to portfolios sorted on the same characteristic at different lags. For brevity and because some characteristics are updated only once per year, we focus on $s = 0, 12, 24, ..., 60$. Following Jegadeesh and Titman, 1993, we combine

Our choice of characteristics is based on the 62 characteristics studied in Freyberger et al., 2020a. We exclude a few characteristics that have missing observations in the beginning of the sample as well as two characteristics that measure conditional market beta, because we study in detail the link between market beta and our results in Section 2.5.3.

Thus, to predict returns for month $t+1$, the characteristics use monthly variables as they were reported at the end of month t and annual variables as they were reported at the end of month $t-6$. Using the most up-to-date characteristic values helps to differentiate new and old sorts.

For a characteristic X that predicts returns with a negative sign, like size, we sort on $-1 \times X$. Signing the characteristic-sorted portfolio returns in this way makes our results more comparable to previous work (e.g., Freyberger et al., 2020a and Haddad et al., 2018), but leaves our conclusions unchanged. If some return R_X expands the mean-variance frontier, $-R_X$ will do so as well; and, if an asset pricing model does not price R_X, it will not price $-R_X$ either.

We focus on returns up to five years after portfolio formation, because it is only for a handful of characteristics that returns beyond the five-year horizon provide an abnormal return relative to both the newest sort ($s = 0$) and the older sorts with $s \leq 60$. For instance, ten years after portfolio formation, $R_{X,(t-120),t+1}$ provides a significant abnormal return for 12 (out of 56) characteristics relative six sorts for each horizon $s > 0$ to reduce noise. Throughout, we analyze returns to characteristic-sorted portfolios from July 1974 (dictated by data availability and a burn-in period for some of our estimates) to December 2017.

To set the stage, we present in Table 2.A.1 summary statistics for portfolios sorted on book-to-market, size, investment, and profitability (which characteristics feature in the Fama and French, 2015 model). We report the average number of firms in the High plus Low portfolio as a reality check of the method. Five years after portfolio formation, the portfolios still contain about 55% of the original number of stocks, which suggests that the portfolios remain sufficiently diversified. Next, we confirm previous literature in that each of the four characteristic-sorted portfolios obtains a positive average return one month after portfolio formation, ranging from 31 bps for size to 53 bps for book-to-market.

The persistence of return predictability varies considerably across these popular characteristics, however. The book-to-market and size effects are large and (marginally) significant at all horizons up to five years after portfolio formation. In fact, both effects are largest one year after portfolio formation (at 61 bps and 50 bps, respectively), after which they slowly decrease (to 38 and 37 bps, respectively, after five years). In contrast, the profitability and investment effects are small and insignificant from two years after portfolio formation onward. Characteristic-persistence, that is persistence in the cross-sectional ranking of stocks, also varies considerably across these characteristics. For instance, the time-series average of the cross-sectional (rank-order) correlation between book-to-market at t and $t - 12$ equals 0.76, whereas this correlation is only 0.25 for investment. The higher the persistence of a characteristic, the more likely it is that a stock in the high portfolio at $t - 12$ is also in the high portfolio at t, which mechanically generates correlation between the returns of old and new sorts. Since our goal is to study the variation across characteristics in the relative performance of old and new sorts (and to understand the implications for asset pricing models), we control for across-characteristic variation in persistence in all of our tests. In contrast, Keloharju et al., 2019 focus

to $R_{X,(t),t+1}$ and for 2 characteristics relative to $R_{X,(t-60),t+1}$. There is no characteristic for which both these abnormal returns are significant.

For instance, the monthly return five years after portfolio formation is defined as: $\sum_{\tau=-2}^{3} R_{X,(t-60+\tau),t+1}/6$.

on the average characteristic, and directly compare the unconditional returns of old and new sort.

2.4 What's new in old sorts?

In this section, we explore the distinctive return and risk properties of old sorts relative to the more commonly studied new sorts. We first derive and test a simple relation between the expected return of old and new sorts, based on a characteristic-based model of expected returns. We then ask whether combining old and new sorts improves Sharpe ratios.

2.4.1 What is the expected return of an old sort?

Let $R_{X,(t),t+1}$ denote the returns to the newest sort; let $R_{X,(t-s),t+1}$ denote the returns to older sorts; and, let $X_{H-L,(t-s),t}$ denote the characteristic spread – the difference between the average characteristic in the long and short portfolio – at time t for the sort performed at $t-s, s \geq 0$. The characteristic-based model of expected returns of Fama and French, 2015, Daniel et al., 2019, and many others in the literature states that expected returns are linearly related to characteristics:

$$E_t(R_{i,t+1}) = X_t' \gamma_t. \tag{2.4.1}$$

Given a generic characteristic X, and assuming that the long-short portfolio for this characteristic is neutral with respect to other characteristics at all horizons after sorting, we have that the expected return of new and old sorts satisfy:

$$E_t(R_{X,(t),t+1}) = X_{H-L,(t),t} \gamma_{X,t} \text{ , and} \tag{2.4.2}$$

$$E_t(R_{X,(t-s),t+1}) = X_{H-L,(t-s),t} \gamma_{X,t} = \frac{X_{H-L,(t-s),t}}{X_{H-L,(t),t}} E_t(R_{X,(t),t+1}). \tag{2.4.3}$$

In words, the expected return of the old sort is a fraction $\left(\frac{X_{H-L,(t-s),t}}{X_{H-L,(t),t}}\right)$ of the expected return of the new sort, which implies that expected returns decay at the same speed as the characteristic-spread. Note that Eq. (2.4.3) is a statement about two different portfolios at the same point in time. The compensation these portfolios receive for loading on the characteristic may vary over time (through $\gamma_{X,t}$), but it is independent of when the portfolio was formed (i.e., $E(R_{X,(t),t+1})/X_{H-L,(t),t} = E(R_{X,(t-s),t+1})/X_{H-L,(t-s),t} = \gamma_{X,t}$). Such independence is implicitly assumed in most previous literature.

We acknowledge that the assumption – that portfolios remain characteristic-neutral at all horizons after sorting – may be a strong one. However, even if this assumption is violated, any difference in the returns between old and new sorts should be eliminated by controlling for the fundamental factors F_{t+1} that span the tangency portfolio. Given that the true set of fundamental factors is unknown, we control for a large set of benchmark models when we analyze the asset pricing implications from old and new sorts in Section 2.5. While in theory the relation between new and old sorts in Eq. (2.4.3) can also be rejected due to non-linearities between characteristics and expected returns, we show in Section 2.6 that non-linearities are unlikely to explain our results.

We construct strategies that invest in each month t 1\$ in the old sort and $-\frac{X_{H-L,(t-s),t}}{X_{H-L,(t),t}}$\$ in the new sort. Eq. (2.4.3) implies that the unconditional expected return of such strategies is zero:

$$E(E_t(R_{X,(t-s),t+1} - \frac{X_{H-L,(t-s),t}}{X_{H-L,(t),t}} R_{X,(t),t+1})) = 0. \tag{2.4.4}$$

To measure $X_{H-L,(t-s),t}$ ($X_{H-L,(t),t}$), we calculate in each month t the median of the characteristic X over all stocks allocated to the high and low portfolio in month $t-s$ (t) and take the difference.

To see the big picture, we present in Panels A and B of Figure 2.A.1 the average return and t-statistic of a strategy that averages over the returns from one to five years after portfolio formation. We see that there is considerable variation across characteristics, with average returns ranging from -50 to 29bps per month. For about one third of the characteristics (18 in total), the average return is negative with

A unit book-to-market spread may capture a different compensation today than it did three years ago, because characteristic premia vary over time. Our relation only states that if a portfolio formed three years ago presents a 0.5 book-to-market spread today (and zero loading on other characteristics), this portfolio should capture half the expected return of a portfolio formed today with a unit book-to-market spread (and zero loadings on other characteristics).

a t-statistic below -1.65 indicating that the decay in average returns is too fast relative to the decay in the characteristic spread. Among these negative returns, we find a number of characteristics related to (idiosyncratic) return volatility, momentum, and profitability. For 8 characteristics, the average return is positive with a t-statistic above 1.65 indicating that the decay in average returns is too slow relative to the decay in the characteristic spread. Among these positive returns, we find a number of characteristics related to illiquidity, share issuance, and value. For the remaining 30 characteristics, we cannot reject at the 10%-level the hypothesis that the expected return of the old sort is a fraction $\frac{X_{H-L,(t-s),t}}{X_{H-L,(t),t}}$ of the expected return of the new sort. Consistent with Keloharju et al., 2019, the fatter left tail of the distribution implies that returns decay too fast for the average characteristic. By controlling for persistence, and in contrast to Keloharju et al., 2019, we find that for some characteristics the return of old sorts is, if anything, too high.

In Panels C and D, we present the average return and t-statistic for each individual horizon $s = 12, 24, 36, 48, 60$. We see that there is some within-characteristic variation across horizons, but average returns are roughly increasing across characteristics from left to right at all horizons. Indeed, rank-correlations between average returns at different horizons are typically over 0.8. Thus, whether returns decay faster or slower than the characteristic spread depends on the characteristic, not the particular horizon.

We conclude that for a considerable fraction of characteristics there is a mismatch between the decay of expected returns and the characteristic-spread, rejecting the hypothesis laid out in Eq. (2.4.3). For these characteristics one may expect combinations of old and new sorts to present attractive investment strategies. However, the particular combination of the old and new sort studied in this section is not optimal in a mean-variance sense. Thus, these results likely understate how attractive combinations of old and new sorts are for investors and, therefore, how big of a challenge these combinations are for asset pricing models. In the following, we study both

We discuss a bootstrap experiment below that shows our results are unlikely due to chance.

The importance of controlling for persistence is clear from Figure 2.B.1, which plots the difference in average return of the old and new characteristic-sorted portfolios. Without controlling for persistence, virtually all characteristics generate a negative old-minus-new return difference and this difference is particularly extreme for characteristics that are known to capture relatively large short-term returns (like idiosyncratic volatility and momentum).

these issues.

2.4.2 Sharpe ratios of optimal combinations of old and new sorts

We ask whether combining old and new sorts improves Sharpe ratios relative to investing only in the newest sort. A positive answer to this question is interesting from two perspectives. First, if lagged characteristics contain independent information about expected returns, the popular characteristic-based model describing time-$(t+1)$ expected returns as a function of time-t characteristics (Eq. (2.4.1)) is misspecified. Second, improvements in Sharpe ratio are informative about the optimal way to construct a characteristic-based portfolio. Indeed, there is a variety of reasons (e.g., clientele and menu effects or specialization) for why investors such as mutual funds may target only one (or a few) characteristics, like value, and it is still an open question how best to construct such a value characteristic. We entertain the possibility that it is optimal to take into account current and lagged observations of the characteristics.

To see whether old sorts improve the mean-variance trade-off, we first estimate unconditional alphas using the full-sample regression:

$$R_{X,(t-s),t+1} = \alpha_s^u + \beta_s^u R_{X,(t),t+1} + \varepsilon_{X,(t-s),t+1}. \tag{2.4.5}$$

A significant α_s^u implies that the maximum Sharpe ratio from investing in the optimal portfolio of the old and new sort is significantly larger than the Sharpe ratio of the new sort. Also, α_s^u is the average return to a strategy that invests in $R_{X,(t-s),t+1}$ and hedges unconditionally the exposure to $R_{X,(t),t+1}$ using the hedge ratio β_s^u. One may be concerned that this hedge ratio is unknown ex ante, such that we are overstating the benefit of combining old and new sorts for investors. To address this concern, we also calculate a conditional alpha, α_s^c, as the average return to a strategy that invests in $R_{X,(t-s),t+1}$ and hedges in each month t with a position equal to the beta estimated

The optimal portfolio of the old and new sorts invests $\alpha_s^u/\sigma^2_{\epsilon_{(t-s)}}$ in $R_{X,(t-s),t+1}$ and $E(R_{X,(t),t+1})/\sigma^2_{(t)} - (\alpha_s^u/\sigma^2_{\varepsilon_{(t-s)}})\beta_s^u$ in $R_{X,(t),t+1}$.

over a 36-month historical rolling window:

$$\alpha_s^c = E(R_{X,(t-s),t+1} - \beta_{s,t}^c R_{X,(t),t+1}), \text{ with } \beta_{s,t}^c \text{ from:} \tag{2.4.6}$$

$$R_{X,(\tau-s),\tau+1} = \alpha_s + \beta_{s,t}^c R_{X,(\tau),\tau+1} + \varepsilon_{X,(t-s),\tau+1}, \ \tau = t-59 : t. \tag{2.4.7}$$

This conditional alpha represents an out-of-sample test of the relative performance of old and new sorts. In the same spirit as the test in Section 2.4.1, these alphas control for the correlation between old and new sorts, which is important when comparing the performance of old and new sorts across characteristics.

We present these alphas in Panel A of Table 2.A.2 for book-to-market, size, investment, and profitability. We find that a large number of alphas are economically and statistically significant. Let us focus on the conditional α_s^c from Eq. (2.4.6). This alpha is positive and (marginally) significant at all horizons up to five years out for book-to-market (at about 40 bps) as well as size (at about 20 bps). In contrast, we see mostly negative alphas for profitability, which are significant at about -27 bps from three to five years after portfolio formation, as well as for investment, which are (marginally) significant at about -27 bps four and five years after portfolio formation.

To appreciate how attractive optimal combinations of new and old sorts are, we present improvements in Sharpe ratio in Panel B. For book-to-market, the Sharpe ratio from investing in the newest sort, $R_{BM,(t),t+1}$, equals 0.28. The Sharpe ratio doubles to 0.56 ($= 0.28 + 0.28$) when an investment in the older sort, $R_{BM,(t-12),t+1}$, is added. Similarly, 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 double the Sharpe ratio of an investment in $R_{Size,(t),t+1}$: 0.52 versus 0.23. 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 0.12 for all but one of the sorting dates in the conditional specification. For profitability and investment, the largest increases in Sharpe ratio are observed when the return four- and five-years after portfolio formation is

The optimal portfolio invests 1.11 and -0.11 in $R_{BM,(t-12),t+1}$ and $R_{BM,(t),t+1}$, respectively, in the unconditional specification (1.21 and -0.21 on average in the conditional specification). Although optimal weights are more extreme for a few of the 56 characteristics studied below, we find in those cases that the improvement in Sharpe ratio (relative to investing only in $R_{X,(t),t+1}$) is only slightly smaller when we restrict the weights to be in the interval $[-2, +2]$.

combined with the return one-month after portfolio formation, at 0.15 and 0.11, respectively.

Among the 56 characteristics we study, large alphas and improvements in Sharpe ratio are pervasive. We present these estimates in Figures 2.A.2 and 2.A.3 respectively. To facilitate interpretation, we sort the characteristics from low to high on the conditional alphas. To see the big picture, we focus on a strategy that averages over the returns from one to five years after portfolio formation, denoted $R_{X,(t-60:t-12),t+1}$.

We see in Figure 2.A.2 that the older sorts provide large and significant alphas relative to the newest sort for over half of the characteristics we study (both in the unconditional and conditional specification). In the conditional specification, for instance, the alpha is negative and significant at the 10%-level for 22 characteristics; and positive and significant for 10 characteristics.

We see in Figure 2.A.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 suggests that for characteristics where the alpha is large in absolute magnitude, the increase in Sharpe ratio also tends to be large. For 33 (19) out of 56 characteristics, the absolute increase in Sharpe ratio is over 0.10 (0.20). For 19 (16) of these characteristics, this number translates to a relative increase in Sharpe ratio of over 100%. These findings are interesting in light of recent work by Daniel et al., 2019, who argue that factors can be traded more profitably by combining a factor, like the newest high-minus-low book-to-market portfolio (HML), with an offsetting position in a hedge portfolio. This hedge portfolio is constructed to have a maximum exposure to the factor (HML), but a zero loading on the characteristic (book-to-market). We instead argue that combining the newest portfolio with an older portfolio improves investment opportunities for a large subset of characteristics.

These results are robust. In 2.A.1 we present a simulation study that shows that a large number of significant alphas is unlikely to be generated under the null of zero alphas, even when the simulation incorporates multiple hypothesis testing and respects the correlation structure (and other moments) of the data. Also, we show

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.

in Figure 2.B.2 that alphas are largely unaffected when we control for exposure to the market. We study a broader set of benchmark factors in the next section. Figure 2.B.3 shows similar alphas when we split our sample in two halves. Figure 2.B.4 and 2.B.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 2.B.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 $t-s$.

The alphas estimated in this section are strongly correlated ($corr = 0.93$) with the average returns of the old-versus-new strategies studied in the previous subsection. The reason is clear from Figure 2.A.4, which plots the exposure of old sorts to the newest sort, β_s^u, and the average characteristic spread remaining s months after portfolio formation, $X_{H-L,(t-s)}/X_{H-L,(t)}$. These two measures of persistence are highly correlated across characteristics. At the three-year horizon, for instance, their correlation equals 0.86. This finding suggests that more persistent characteristics, such as size, generate larger correlation between old and new sorts. Less persistent characteristics, such as momentum and short-term reversal, generate substantially smaller correlation. Variation across characteristics in persistence is not what drives variation in alphas, however: Figure 2.A.4 shows no discernible pattern in persistence even though alphas are monotonically increasing from left to right. As discussed before, what drives alphas is the fact that the persistence of average returns and characteristics often do not match up.

This fact does not present a challenge to asset pricing models if old and new sorts are exposed differently to fundamental factors. However, as we will show below, exposure to the factors in benchmark asset pricing models cannot explain the alphas we document. To see the intuition for this result, we present the exposures of older sorts ($R_{X,(t-s),t+1}$, for $s = 12, ..., 60$) to the newest sort $R_{X,(t),t+1}$ in Figure 2.A.5. We see that these betas are quite large and persistent for most characteristics. The

In the setting of Section 2.4.1, the fact that β_s^u is close to $X_{H-L,(t-s)}/X_{H-L,(t)}$ implies that the fundamental factors explain a large share of the return to new sorts. See 2.A.2 for additional discussion.

median beta (across characteristics) equals 0.77 one year after portfolio formation (i.e., for $R_{X,(t-12),t+1}$) and falls only slowly to 0.55 five years after portfolio formation (i.e., for $R_{X,(t-60),t+1}$). Models that do a good job explaining the returns to a characteristic-sorted portfolio at one horizon (e.g., the newest sort) will have a hard time explaining returns at another horizon (e.g., the older sorts), when there is a large alpha separating these two highly correlated returns. This tension is at the root of the asset pricing implications we derive from old and new sorts.

2.5 Implications for asset pricing models

To study asset pricing implications, we first reduce the dimensionality of the data and extract, at each horizon, principal components from the characteristic-sorted portfolios. The motivation is that the SDF can be suitably approximated using only a few dominant principal component factors, when characteristic-sorted portfolios do not each represent an independent source of priced risk (Kozak et al., 2019; Haddad et al., 2018). Following previous literature, we treat the principal components extracted from the newest sorts – with returns one month after portfolio formation – as statistical factors in an asset pricing model. We ask if these statistical factors as well as benchmark factor models from the literature price the principal components extracted from old sorts as well as the strategies that combine old and new sorts from Section 2.4. The benchmark models are the single-factor CAPM (Sharpe, 1964, Lintner, 1965, Mossin, 1966); the three-factor model of Fama and French (1993b, FF3M); the five-factor model of Fama and French (2015, FF5M); and finally, a six-factor model including the factors in the FF5M and momentum (FF5M+MOM). We focus on these models to show the important trade off between small and big models. Results for the models of Hou et al. (2015a, HXZ), Frazzini and Pedersen (2014b, BAB), Daniel et al. (2019, DMRS), Stambaugh and Yuan (2016, SY), and Daniel et al. (2017, DHS) are consistent and summarized in Table 2.B.2 of the Internet Appendix. To provide additional economic intuition, we relate the old-versus-new return differentials to market beta at the end of this section.

Correlations between old and new sorts are slightly larger, because $R_{X,(t),t+1}$ is typically more volatile than $R_{X,(t-s),t+1}$ (the median correlation equals 0.85 for $R_{X,(t-12),t+1}$ and 0.64 for $R_{X,(t-60),t+1}$).

2.5.1 Do benchmark factor models price old and new sorts?

We extract three principal components from $R_{X,(t-s),t+1}$ at horizons $s = 0, 12, ..., 60$. At each horizon, the three principal components explain about 60% of the total variation in returns. We perform GRS tests to determine whether (i) the principal components extracted from $R_{X,(t),t+1}$ (denoted $3PC_{(t),t+1}$) or (ii) benchmark asset pricing models, price the principal components extracted from $R_{X,(t-s),t+1}$.

Table 2.A.3 presents the results. In Panel A, we see that the GRS test rejects with a p-value of 0.0017 that the statistical factor model $3PC_{(t),t+1}$ prices the three principal components extracted from $R_{X,(t-12),t+1}$. At longer horizons, the rejection is even stronger with p-values < 0.0001. The fact that three statistical factors that explain most of the variation in $R_{X,(t),t+1}$ do not price the dominant components of $R_{X,(t-s),t+1}$, confirms that the old-versus-new return differentials we document in Section 2.4 are statistically large.

For the benchmark models, several results stand out. First, small models do relatively poorly at pricing the statistical factors extracted from $R_{X,(t),t+1}$: the GRS test rejects at a p-value < 0.0001 for the CAPM and FF3M. In contrast, relatively large models do better at pricing these returns, as the GRS test does not reject at a p-value of 0.22 for the FF5M and 0.38 for the FF5M+MOM. Second, both small and large benchmark models fail to price the principal components extracted from returns at most horizons $s > 0$ after portfolio formation. However, if anything, the smaller models perform relatively well. For instance, the GRS test does not reject at the 5%-level for the CAPM and FF3M at the five-year horizon ($s = 60$), in contrast to the FF5M(+MOM).

These GRS tests confirm that existing asset pricing models fail to jointly price both the newest and older sorts. To understand which component is driving this result, we present in Table 2.A.4 the alpha for individual principal components at each horizon after portfolio formation. We estimate these alphas by running the following regressions:

$$\lambda'_{(t),z} R_{X,(t-s),t+1} = \alpha_s + \beta'_s F_{t+1} + \epsilon_{(t-s),t+1}, \tag{2.5.1}$$

Our conclusions are unchanged when we extract Risk Premium PCA factors using the approach of Lettau and Pelger, 2020b, see Table 2.B.3 of the Internet Appendix.

where F_{t+1} is a candidate factor model. Note that we apply the same loadings to returns at all horizons s after sorting, that is, the loadings $\lambda_{(t),z}$ of the z-th principal component extracted from the newest sorts with returns $R_{X,(t),t+1}$. In this way, we abstract from variation in the loadings of the z-th principal component across horizons, such that our results derive only from the relative pricing of $R_{X,(t-s),t+1}$ versus $R_{X,(t),t+1}$. Having said that, this choice does not affect our main conclusions, because the characteristic-sorted portfolio returns at different horizons are highly correlated (recall our previous discussion of Figure 2.A.5).

In Panel A, we see that the first principal component of returns provides a large and significant alpha with respect to the statistical factor model $3PC_{(t),t+1}$ at longer horizons after portfolio formation. The estimated alpha is 68 bps (t-stat = 3.69) for $s = 12$ and is larger than 100 bps for $s > 12$ (t-stat ≥ 5). For the second and third principal components, returns at longer horizons after portfolio formation are explained well by the $3PC_{(t),t+1}$ model. For the benchmark models in Panels B to E, several results stand out. Similar to the model with three statistical factors, the larger benchmark models do not price the return to the first principal component at longer horizons after portfolio formation. The alphas for FF5M and FF5M+MOM are over 100 bps at horizons $s > 12$ (t-stat > 3). These larger models do price the return of the first principal component at shorter horizons $s = 0, 12$. In contrast, the smaller benchmark models price the returns to the first principal component at longer horizons after portfolio formation ($s > 12$). These models fail completely at shorter horizons, however. For instance, the alpha at $s = 0$ is around -2.2% per month in the CAPM and FF3M. Finally, in contrast to the statistical factor model $3PC_{(t),t+1}$, the benchmark models also perform poorly pricing the second principal component of returns. The performance of small and big models again differs: while the alpha of the second principal component is negative in the CAPM (around -1% and marginally significant at all horizons $s > 0$), it is positive in the larger factor models (around 70 bps and significant at most horizons s).

In all, the rejections in the GRS test are for the largest part driven by the first principal component of characteristic-sorted portfolios. To understand why, we plot in Figure 2.A.6 the loadings of the first principal component on each characteristic (in the same order as the alphas of old versus new sorts presented in Figure 2.A.2). We see that the loadings are almost monotonically increasing from left to right, translating to a strong correlation of 0.78 between these loadings and the alphas. The strength of this relation is surprising, because the principal component loadings are determined only by the (co-) variances of the newest sorts. In light of this strong relation, it is unsurprising that a strategy that uses the principal component loadings as portfolio weights, obtains a return, say, three years after portfolio formation that is hard to explain from the point of view of returns immediately after portfolio formation.

Further intuition is provided by the summary statistics presented in Panel B of Table 2.A.3. The correlation between $\lambda'_{(t),1} R_{X,(t),t+1}$ and $\lambda'_{(t),1} R_{X,(t-36),t+1}$ is high at 0.91, whereas the average returns of these strategies is remarkably different at -60 bps versus 95 bps, respectively. Given the high correlation between these two returns, differential exposure to benchmark factors is unlikely to explain the difference in average returns.

This suggestion is confirmed in Panel F of Table 2.A.4. Here, we ask whether the benchmark models price a strategy that is long the first principal component of older sorts and short the first principal component of the newest sorts: $\lambda'_{(t),1}(R_{X,(t-36),t+1} - R_{X,(t),t+1})$. For all models and at all horizons s, this strategy provides a large and significant alpha. The larger models perform relatively well in this exercise, because the alphas decrease, for instance for $s = 36$, from 2.05% (t-stat=5.97) in the CAPM to 1.17% (t-stat = 3.52) in the FF5M+MOM. In Panel B of Table 2.B.2 of the Internet Appendix, we find a similarly large alpha for this strategy relative to three of the five alternative models: HXZ (1.23%, $t = 3.55$), BAB (1.40%, $t = 3.56$), and DMRS (1.64%, $t = 5.16$). For the remaining models, SY and DHS, the alpha of this strategy is only marginally significant at about 70 bps. Although these two models do relatively well pricing this particular combination of the older and the newest sorts, both these models fail to price the first principal component of older sorts in isolation (see Panel A of Table 2.B.2). In conclusion, none of the benchmark models we study is able to price the returns to both new and old sorts.

These loadings are robust. When we split our sample in two halves around March 1996, the loadings of the first principal component in the first half of the sample are correlated at 0.81 with the loadings of the first principal component in the second half of the sample.

This conclusion holds equally for theoretical models designed to explain characteristics-based return predictability. To see this by example, we simulate in 2.A.3 from the models of Gomes et al., 2003 and Zhang, 2005b. We find that return differentials and alphas between old and new book-to-market sorts are small in these models compared to the data.

2.5.2 What does a factor model need to price old and new sorts?

Asset pricing models should jointly price newer and older sorts. We have seen that models with factors that are based on new sorts alone fail this challenge. What about models with factors based on new and old sorts? To answer this question, we consider a four-factor model that adds to the three principal components extracted from the newest sorts, a single factor capturing the return differential between older and newer sorts. In line with the above setup, we define this factor $\lambda'_{(t),1}R_{X,(t-36),t+1}$ and denote this model $4PC_{(t,t-36),t+1}$.

In Table 2.A.5 we present alphas for the principal components at each horizon with respect to this four-factor model. We see that the alphas between old and new sorts for the first principal component are quite small, especially compared to alphas of $> 1\%$ in the three-factor model $3PC_{(t),t+1}$. Having said that, the alpha is significant in the four-factor model at about 36 bps at the longest horizons ($s = 48, 60$). Similarly, the GRS test (which considers jointly the alphas of the first three principal components) does not reject at the 5%-level at shorter horizons, but rejects marginally at the longest horizons. This result marks a big improvement over the three-factor model $3PC_{(t),t+1}$, for which the GRS test firmly rejects at all horizons with p-values well below 0.0001 (see Table 2.A.3).

The improved fit of the four-factor model is not specific to the first three principal components. In Table 2.A.5 we also show that the four-factor model is not rejected at the 5%-level at any horizon when we use as test assets the first five principal components of characteristic-sorted portfolio returns. Moreover, we ask in Table 2.A.6 whether the models capture (i) the average return of the old-versus-new strategies studied in Section 2.4.1 and (ii) the alpha of old with respect to new sorts studied in Section 2.4.2. To this end, we regress the returns of the old-versus-new strategies on the four-factor model as well as the benchmark factor models. We report the mean absolute alpha (MAA) and the number of alphas that are significant (at the 5%-level) at each horizon s. We see that our proposed model, $4PC_{(t,t-36),t+1}$, performs relatively well. For instance, at the three-year horizon ($s = 36$), we find only 7 alphas that are significant when we regress the conditionally hedged return of the old sort (see Eq. (2.4.6)) on the four factors. For comparison, this count equals 14 for the CAPM and the FF5M+MOM. Also, the mean absolute alpha is smallest for the four factor model at 10 bps versus 15 bps for the CAPM and 14 bps for the FF5M+MOM. Results are similar for the remaining strategies.

The outperformance of $4PC_{(t,t-36),t+1}$ is even starker in cross-sectional regressions. Figure 2.A.7 presents cross-sectional R^2s when the test assets are 112 characteristic-sorted portfolios (56 portfolios from (t) and $(t-s)$) and factor risk premia are forced to match their sample average returns. At all horizons s, the R^2 is negative in all four benchmark models, marginally above zero for the three-factor model $3PC_{(t),t+1}$, but large and positive at about 0.40 for the four-factor model $4PC_{(t,t-36),t+1}$. Thus, benchmark factors based on the returns of new sorts provide a poor fit in the joint cross-section of old and new sorts. In contrast, adding a single factor based on old sorts (to a model based on new sorts) goes a long way to capture old-versus-new return differentials. This finding is consistent with the fact that these pricing errors are driven by the first principal component of returns. Let us now turn to the identity of this first principal component.

2.5.3 Market beta and old-versus-new sorts

We have seen that the first principal component drives old-versus-new return differentials (see Table 2.A.4). Interestingly, there is a strong relation ($corr = 0.96$) across characteristics between the market exposure of $R_{X,(t),t+1}$ and the loading of the first principal component on $R_{X,(t),t+1}$. Thus, market beta holds promise to contain timely information about old-versus-new return differentials. To investigate this link, we

For the regression-based combination of old and new sorts, we perform a two-stage test: in the first stage we compute β_s from a regression of $R_{X,(t-s),t+1}$ on $R_{X,(t),t+1}$ and in the second stage we regress $R_{X,(t-s),t+1} - \beta_s R_{X,(t),t+1}$ on the benchmark factors. The results for the unconditional test use GMM standard errors that correct for this errors-in-variables, which correction has little impact on our conclusions.

Similar to us, Haddad et al., 2018 extract the first principal component of long-short characteristic-sorted portfolios. In contrast to us, these authors use market-neutral portfolios in the PCA. We do not clean the portfolios from market exposure to show that the market beta of a characteristic-sorted portfolio contains valuable information about the dynamics of post-formation returns.

sort the 56 long-short characteristic portfolios in three groups split at the terciles of ranked market betas in each month t. These market betas are estimated for the newest sort over a 60 month historical rolling window (i.e., using $R_{X,(t-60),t-59}$ to $R_{X,(t-1),t}$). Then, within each market beta group, we take an equal-weighted average over the characteristic-sorted portfolio returns at each horizon. To conserve space, we focus on horizons $s = 0$ and $s = 36$; these horizons are representative of the general patterns in the data.

In Panel A of Table 2.A.7, we see that average returns among high beta characteristics decrease relatively slowly after portfolio formation, from an average return of 30 bps for $s = 0$ to 25 bps for $s = 36$. In contrast, among low beta characteristics, average returns decrease considerably faster and fall from a positive 23 bps for $s = 0$ to a negative -13 bps for $s = 36$. These two facts together imply that the difference in average return between high and low beta characteristics is increasing from 7 bps ($s = 0$) to 38 bps ($s = 36$) as three years pass after portfolio formation. Thus, a first take-away from the sort on market beta is that the persistence of return predictability is larger for high market beta characteristics. As mentioned before, average returns are not the most interesting statistic in the context of old and new sorts that are differentially correlated across characteristics.

We therefore turn to the returns of the strategies that combine old and new sorts as in Sections 2.4.1 and 2.4.2. To this end, we take in each market beta group an equal-weighted average of the returns $R_{X,(t-s),t+1} - X_{H-L,(t-s)}/X_{H-L,(t)}R_{X,(t),t+1}$ (see Eq. 2.4.4) and the alphas defined as $R_{X,(t-s),t+1} - \beta_s R_{X,(t),t+1}$ (see Eqs. 2.4.5 and 2.4.6). We find that the average return is increasing from low to high market beta for all three strategies. For instance, among low beta characteristics, the return of a portfolio sorted three years ago provides a significantly negative conditional alpha relative to the newest sort of -23 bps (t-stat=-4.39). In contrast, among high beta characteristics, the same portfolio provides a positive conditional alpha of 10 bps (t-stat=2.77). These effects add up to a large alpha for the high-minus-low market beta portfolio of 33 bps (t-stat=4.42). Thus, we conclude that the relative performance of old versus new sorts is strongly related to market beta. The fact that old sorts perform relatively poorly among low market beta characteristics implies that average returns decay too fast after portfolio formation for such characteristics.

In Panel B, we ask if popular asset pricing models capture these patterns. First, we see that the smaller models, CAPM and FF3M, fail to capture the difference in returns between high and low market beta characteristics one month after portfolio formation. The CAPM and FF3M alpha for the high-minus-low market beta portfolio are significant at -40 and -42 bps, respectively. This alpha is mostly driven by a large positive alpha of about 50 bps in the low market beta group. Thus, among low beta characteristics, the average return one month after sorting is not only anomalously high relative to returns longer after portfolio formation, but also relative to the market. These small models perform much better pricing returns longer after portfolio formation. For $s = 36$, returns fall in line with the CAPM in the low market beta group, such that the alpha of the high-minus-low market beta portfolio is small and insignificant. The performance of bigger models is quite the opposite. Both the FF5M and FF5M+MOM capture the difference in returns between high and low market beta characteristics immediately after portfolio formation. However, three years after portfolio formation, the high-minus-low market beta portfolio obtains a large and significant alpha of about 43 bps (t-stat ≈ 3.5).

We thus see the same trade-off between big and small models as in Table 2.A.4. The fact that none of the benchmark models prices returns of both older and newer sorts has a familiar origin: returns of the new and old sorts are highly correlated, but their average return is different. Consistent with the evidence in Panel F of Table 2.A.4, we next confirm that all models fail to capture the return of the strategies that combine old and new sorts. The alpha of these strategies is significantly smaller among low market beta characteristics by about 30 bps (with t-stat typically well above 3) in all four benchmark models. In fact, this result holds true equally for the five alternative models considered in Panel C of Table 2.B.2 of the Internet Appendix.

Thus, we uncover a new dimension to the low beta anomaly by grouping characteristic-sorted portfolios on their market beta. We find that the relative performance differential between old and new sorts is strongly related to market beta: average returns

We already saw in the previous subsection that an unconditional CAPM cannot explain our results. Because we have now sorted the characteristics on their conditional market betas, the current evidence suggests that a conditional CAPM cannot explain our results either.

The correlation between $R_{H-L,(t),t+1}$ and $R_{H-L,(t-36),t+1}$ is high at 0.80, but their average returns are 7 bps and 38 bps, respectively (see Panel A of Table 2.A.7).

of the newest sorts are too high relative to older sorts among low market beta characteristics. This performance differential is not captured by market beta and therefore translates to a large alpha relative to the CAPM, and also relative to larger factor models.

In Panel C, we see that the four-factor model proposed in the previous subsection, $4PC_{(t,t-36),t+1}$, does go a long way to capture the relative performance differential between old and new sorts among low beta characteristics. As a result, this model eliminates almost completely the difference between high and low beta characteristics in the returns of the strategies that combine old and new sorts. Although the performance of $4PC_{(t,t-36),t+1}$ generally compares favorably to the other models studied in Table 2.A.7, the fit of the model is not perfect. For instance, high market beta characteristics provide an alpha relative to this four-factor model both one month and three years after portfolio formation (of 26 and 15 bps, respectively). Pricing both new and old sorts, and thus combinations between them, proves hard and we leave a more thorough investigation of this issue for future work. Indeed, a model that prices returns at all horizons after sorting would get price levels right (Cho and Polk, 2019; Binsbergen et al., 2021).

2.6 The contribution of new and old stocks

Our evidence so far suggests that there is a range of characteristics – that is, those with low market betas – for which the return immediately after portfolio formation is too high relative to the return longer after portfolio formation. Similarly, there are characteristics for which this return is too low. To understand which stocks drive these results, we split the stocks that are used to calculate the return of the newest *sort*, $R_{X,(t),t+1}$, into relatively new and old *stocks*. To this end, we perform a dependent double sort into ten X_t deciles and within the high and low decile into two portfolios split at the (within-portfolio) median of X_{t-36}. This decomposition ensures that the new and old portfolios (roughly) contain the same number of stocks for all characteristics. In particular, $R^{Old}_{X,(t),t+1}$ is the return of a strategy that is long (short) a value-weighted portfolio of the stocks for which, among all stocks in the highest (lowest) decile portfolio at time t, the characteristic X_{t-36} is above (below) the median value of that characteristic. The return for new stocks, $R^{New}_{X,(t),t+1}$, uses all remaining stocks in the high and low portfolio at time t. Intuitively, the new stocks are those that have seen a relatively large change in the characteristic over the last three years.

Our decomposition is new to the literature and different from the transitory-permanent decomposition of Keloharju et al., 2019. These authors decompose a firm characteristic in its historical average (the permanent component) and a residual (the transitory component). Whereas we focus on the intermediate three-year horizon, Keloharju et al., 2019 define the permanent component using a ten-year average. Moreover, we sort the stocks within the high and low decile portfolio in a new and old group, whereas Keloharju et al., 2019 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 return difference of 57 bps (t-stat = 2.18). This finding implies 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.

We first perform a reality-check of our decomposition. We have already shown that the returns of older sorts are highly correlated to the returns of the newest sort for most characteristics. One would expect this correlation to be driven by old stocks. Figure 2.A.8 plots the relative contribution to the R^2 in a joint regression of $R_{X,(t-36),t+1}$ on $R^{Old}_{X,(t),t+1}$ and $R^{New}_{X,(t),t+1}$ and confirms this intuition. For the vast majority of characteristics, the R^2 is driven by the old stocks that have extreme characteristic values in the past as well as today. Across characteristics, the median

We also allocate stocks that are recently introduced in CRSP to the new portfolio. We have considered an alternative new-versus-old decomposition that defines as old stocks only those stocks that are in the high (or low) portfolio today as well as 36 months ago. This decomposition generates large differences in the number of stocks in the old versus the new portfolio depending on the persistence of the characteristic considered.

In contrast, Gerakos and Linnainmaa, 2018b find that changes in book-to-market predict returns with a positive sign in the full cross-section of stocks, which result 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, which trend 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., 2013b).

contribution to R^2 is about three times larger for the old stocks than for the new stocks.

We now analyze the difference in average returns and alphas between the characteristic-sorted portfolios that use old or new stocks. To increase power, we test these differences across the three market beta groups of Section 2.5.3. Our decomposition in new versus old stocks is coarse, and averaging over different characteristics within each group will smooth out this noise.

Table 2.A.8 presents descriptive statistics across groups and confirms that the new and old stock portfolios are similar in important dimensions. First, the new and old stock portfolio represent a similar high-minus-low characteristic spread, which suggests that an old-minus-new strategy is roughly characteristic-neutral. For instance, on average among low market beta characteristics, the old-minus-new portfolio provides a characteristic spread that is only 3% (1.02-0.99) of the characteristic spread in the not-decomposed long-minus-short decile portfolio. Second, the difference in total market cap allocated to the new and old portfolio is small, which suggests that the new portfolio is not overpopulated by small (and therefore illiquid) stocks. For instance, among low market beta characteristics, on average 9% of total CRSP market cap is allocated to the new stocks in the high plus low portfolio, which is relative to 11% for old stocks. Third, the old and new stock portfolios are balanced in the spread between the high and low portfolio in size, book-to-market, investment and profitability. For each of these characteristics, the difference between the old and new portfolio is small. To see this, we present the characteristic spread that is obtained in a single sort of stocks on these characteristics. These spreads are larger than the old-minus-new characteristic spreads by a factor of ten (or more). Thus, any model that defines expected excess returns as a linear function of these

We focus on the grouping using market betas, because it does not require any forward-looking information. We show in Table 2.B.4 of the Internet Appendix that our conclusions are unchanged when we aggregate the characteristics using the loadings of the first principal component of the returns to new sorts, as in Section 2.5.1.

Figure 2.B.7 presents the return-differences between old and new stocks for all 56 characteristics. We see that these return differences are roughly increasing from the characteristics on the left (like idiosyncratic volatility) to the characteristics on the right (like book-to-market), consistent with the idea that the difference in returns between old and new *stocks* in the extreme portfolios contributes to the alpha between old and new *sorts* (in fact, their correlation is 0.73).

The same conclusion holds for a range of other characteristics (see Table 2.B.5 of the Internet Appendix).

four characteristics (e.g., Fama and French, 2015; Daniel et al., 2019) predicts that return differences between our old and new stock portfolios are small.

We present average returns and alphas with respect to existing factor models in Table 2.A.9. We present results for the newest characteristic-sorted portfolios ($R_{X,(t),t+1}$) as a benchmark, but our main interest is in the (difference between) the contribution of new and old stocks ($R^{New}_{X,(t),t+1}$ and $R^{Old}_{X,(t),t+1}$). To start, we see that old stocks underperform new stocks among low market beta characteristics by a significant -18 bps. This finding confirms that the negative alphas of old with respect to new sorts among low market beta characteristics (see Section 2.5.3) are driven by returns that are relatively too high immediately after portfolio formation. Among high beta characteristics, the old stocks outperform the new stocks by a significant 19 bps. The diff-in-diff between old and new stocks among high versus low beta characteristics is large and significant at 37 bps ($t = 3.17$).

Neither small (e.g., the CAPM) nor big (e.g., the FF5M+MOM) models fully capture the relative performance differential between new and old stocks. Indeed, in all models, we find a large and significant spread in the alpha of the old-minus-new strategy between high and low market beta characteristics. This alpha equals 27 bps ($t = 2.39$) in the CAPM and 44 bps ($t = 3.72$) in the FF5M+MOM. Consistent with previous evidence, we see the trade-off between small and big models. Small models, like the CAPM, capture the high-minus-low market beta spread among old stocks, but do not capture the high-minus-low market beta spread among new stocks. In contrast, big models, like the FF5M+MOM, capture the high-minus-low market beta spread among new stocks, but do not capture the high-minus-low market beta spread among old stocks. We conclude that existing asset pricing models do not jointly price new and old *stocks*.

This result provides an interesting link with our previous conclusion that existing models do not jointly price new and old *sorts*. Intuitively, old stocks capture the long-term returns to characteristic-based investing, whereas new stocks capture the short-term returns. This evidence marks an important contribution to Keloharju et al., 2019. These authors find that returns are mostly driven by the transitory (not

In the five alternative models considered in Panel D of Table 2.B.2 of the Internet Appendix, alphas are similarly large and significant and range from 35 bps ($t = 3.39$) in the DMRS model to 46 bps ($t = 3.72$) in the DHS model.

the permanent) component of the average characteristic. We instead show that the relative performance differential between new stocks, which return is likely closer to their transitory component, and old stocks, which return is likely closer to their permanent component, varies substantially and in sign across characteristics.

More generally, our evidence indicates that return spreads from stocks that have been in the extreme portfolios for a while 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. In other words, there are subsets of stocks for which the same characteristic spread is compensated with a different risk premium. In the context of Section 2.4.1, this finding indicates that the return differentials we document are unlikely due to a non-linear relation between characteristics and expected returns. Moreover, this finding contributes to 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 fixed. One may be inclined to conclude that this variation is due to differences in the loading of new and old stocks on *other* characteristics. With that prior it is surprising that the difference between new and old stocks is not captured by any of the factor models we consider, because the factors in these models are derived from sorts of stocks on these *other* characteristics. Consider, for instance, the five-factor models FF5M and DMRS. As seen in Table 2.A.8, there is not much difference between old and new stocks in the characteristics that define the factors in these models. Consequently, these models fail to capture the old-minus-new return differences. Thus, our results imply that popular explanations of the cross-section of expected returns based on recent observations of characteristics are incomplete: These explanations are either still missing an important characteristic or failing to account for lagged characteristics.

The fact that past values of, or changes over time in, characteristics contain additional independent information about expected returns among the large set of characteristics we study is interesting in and of itself. The return differentials we document imply that investors wanting to trade these characteristics should carefully consider the distinction between new and old stocks for their portfolio. Our new and old stock portfolios are tradable and require a position in much fewer stocks than the original strategies. Old stock portfolios will require relatively little rebalancing, thus lowering transaction costs even further.

2.7 Conclusion

In contrast to most existing literature that focuses on characteristic-based return predictability over short horizons, we study horizons up to five years. We uncover large return differentials across horizons, which indicate that (i) the returns of characteristic-sorted portfolios do not in general decay at the same speed as characteristics and (ii) there are large alphas in old relative to new sorts. These alphas translate to large improvements in Sharpe ratio when old and new sorts are combined in a single portfolio. Moreover, these alphas are not explained by benchmark asset pricing models, although we document an interesting trade-off between small and big models. We argue that longer horizon returns to characteristic-based investing provide a useful new set of moments to evaluate asset pricing models.

We further find that the old-versus-new return differentials are mostly driven by the first principal component of returns, which explains why a model that adds a single factor based on old sorts to a model with factors based on new sorts performs relatively well in tests with our new moments. Having said that, we leave for future work the challenging task of finding the most parsimonious factor representation that explains returns at all horizons and gets prices right (see Cochrane, 2011b, p. 1063). Finally, we show that the return differential between old and new sorts is connected to the difference in returns between "old" and "new" stocks. Our old-versus-new stock strategies are characteristic-neutral, but obtain large and significant alphas relative to benchmark asset pricing models. We thus conclude that explanations of the cross-section based on recent observations of firm characteristics are incomplete.

Our empirical evidence has implications for practitioners trading characteristics as well as empiricists testing asset pricing models. Future empirical research can use the decompositions developed in this paper – in new versus old sorts and stocks –

Relatedly, combinations of new and old sorts (such as those proposed in Section 2.4.2) are likely less expensive to trade than an investment in $R_{X,(t),t+1}$ alone. The reason is that positions in selected stocks for $R_{X,(t),t+1}$ and $R_{X,(t-s),t+1}$ are likely to cancel out (see DeMiguel et al., 2019).

(i) to shed light on the dynamics of cross-sectional return predictability and (ii) to construct test assets that go beyond short-term returns in search of a (statistical) factor model. An interesting avenue for future theoretical work is to understand the economic drivers of the post-formation dynamics of characteristic-sorted portfolio returns and, more specifically, why the expected return compensation for a portfolio's loading on a characteristic seems to depend on the time this portfolio was formed.

2.A Appendix

2.A.1 Simulation

We run a number of Monte Carlo simulations to address the multiple hypothesis testing concern that returns across the 56 characteristic-sorted portfolios are correlated, which may affect the size of our tests. In the simulations, we impose the null of zero alpha between the old and new sort, but respect the correlation structure (as well as other moments) in the data. Analogous to Eq. (2.4.5) and focusing on $s = 36$, we analyze the following regression:

$$R_{X,(t-36),t+1} = \alpha + \beta R_{X,(t),t+1} + \epsilon_{X,(t-36),t+1}. \tag{2.A.1}$$

In each of 10000 simulations, we first create for each characteristic $X = 1, ..., 56$ an artificial time-series of returns to the new sorts collected in the $T \times 56$ matrix: $R^{sim}_{(t),t+1} = [R^{sim}_{1,(t),t+1}, ..., R^{sim}_{56,(t),t+1}]$. These returns are drawn from a multivariate normal distribution, $R^{sim}_{(t),t+1} \sim N(\mu_{(t)}, \Sigma_{(t)})$, where $\mu_{(t)}$ and $\Sigma_{(t)}$ are the vector of means and the variance-covariance matrix of $R_{(t),t+1}$ in the data. Next, we create artificial returns to the old sorts, imposing zero alpha: $R^{sim}_{X,(t-36),t+1} = \beta^{sim}_{X} R^{sim}_{X,(t),t+1} + u^{sim}_{t+1}$. The exposure of the old sort to the new sort, β^{sim}_{X}, is drawn from a normal distribution with mean (standard deviation) equal to the average (standard deviation) of β in the data taken over the 56 characteristics. The residuals are drawn from a multivariate normal distribution $u^{sim}_{t+1} \sim N(0_{56}, \Sigma_u)$, where Σ_u is the variance-covariance matrix of the residuals from Eq. (2.A.1) in the data. As a benchmark, we also present results for a simulation that assumes the variance-covariance matrices, $\Sigma_{(t)}$ and Σ_u, are diagonal (rather than full), which thus ignores the correlation between characteristics. Finally, we consider 10000 bootstrap simulations that account for the fact that return(-innovations) are not normal in the data. Each bootstrap resamples (with replacement) from the original time index $t = 1, ..., T$ both the returns of the newest sort and residuals from Eq. (2.A.1) in the data. We combine these bootstrapped time-series with the simulated β^{sim}_{X}'s to create returns to the older sort.

For each set of artificial data, we estimate Eq. (2.A.1) and present below the 50, 90, 95 and 99 percentiles of the simulated distribution of the number of significant (at the 10% level) α's out of 56. In the data, we find that 23 out of 56 characteristics have an alpha (between the old sort at $s = 36$ and the newest sort at $s = 0$) that is significant at the 10% level. This number is unlikely to be generated under the null, because in 99% of the simulations that respect the correlation structure in the data (using either normal returns or bootstrapped returns) the number of significant α's is 16.

TABLE 2.A.1: **Simulated distribution of number of significant alphas**
We simulate returns to the old sort at $s = 36$ under the null of zero alpha with respect to the newest sort at $s = 0$. We report the 50, 90, 95, and 99 percentiles of the distribution (in 10000 simulations) of the number of significant alphas of the old sort with respect to the newest sort (out of a total of 56 and estimated as in Eq.(2.A.1)).

Percentiles	50	90	95	99	Data
# Significant					23
Diagonal	5	9	9	11	
Full	5	10	12	16	
Bootstrap	5	10	12	16	

2.A.2 Alphas between new and old sorts in a characteristic-based model

The characteristic-based model of Eq. (2.4.1) can be understood as stating that a unit loading on a characteristic X at time t translates to a vector of exposures $\beta_{X,t}$ to priced fundamental factors F_{t+1} with expected returns $\mu_{F,t}$, such that $\gamma_{X,t} = \beta'_{X,t}\mu_{F,t}$. Next, we use this insight to understand the connection between the approaches of Section 2.4.1 and 2.4.2, respectively. For ease of exposition, let us consider an unconditional setting.

Starting from Eqs. (2.4.2) and (2.4.3) and assuming that realized returns on both the new and old sorts contain factor risk as well as uncorrelated residual risk, we have:

$$R_{X,(t),t+1} = X_{H-L,(t)}(\beta'_X F_{t+1}) + e_{(t),t+1} \text{ and} \tag{2.A.2}$$

$$R_{X,(t-s),t+1} = X_{H-L,(t-s)}(\beta'_X F_{t+1}) + e_{(t-s),t+1}. \tag{2.A.3}$$

In a regression of the old on the new sort: $R_{X,(t-s),t+1} = \alpha_s + \beta_s R_{X,(t),t+1} + \epsilon_{(t-s),t+1}$, we then have that:

$$\beta_s = \frac{X_{H-L,(t-s)}}{X_{H-L,(t)}} \times (1 - \frac{Var(e_{(t),t+1})}{Var(R_{X,(t),t+1})}) \tag{2.A.4}$$

$$\alpha_s = X_{H-L,(t-s)} \beta'_X \mu_F \frac{Var(e_{(t),t+1})}{Var(R_{X,(t),t+1})}. \tag{2.A.5}$$

Thus, when the fundamental factors explain all variation in the returns of the new sorts, $\beta_s = \frac{X_{H-L,(t-s)}}{X_{H-L,(t)}}$ (and $\alpha_s = 0$). If the fundamental factors do not explain all variation, β_s will be a downward biased estimate of $\frac{X_{H-L,(t-s)}}{X_{H-L,(t)}}$. Note that this result implies that there is a small improvement in Sharpe ratio from combining new and old sorts (i.e., $\alpha_s > 0$) even when returns decay at the same speed as the characteristic, due to diversification of residual risk. What we are interested in is the variation in this improvement across characteristics, however.

2.A.3 Theoretical explanations of characteristic-based return predictability

Many production-based asset pricing theories seek to explain cross-sectional return patterns associated with characteristics. In the model of Gomes et al., 2003, current size and book-to-market predict returns in the cross-section, because these characteristics are correlated to the firm's true conditional market beta. This model shares an important feature with alternative theories of characteristic-based return predictability (such as Berk et al., 1999 and Zhang, 2005b): the focus is on relating one-period ahead expected returns to current characteristics. To see whether our results present a challenge for such models, we analyze where our empirical estimates lie in the simulated distribution generated under the null of the models of Gomes et al., 2003 and Zhang, 2005b.

In Panel A of Table 2.A.2, we report the average return of old (focusing on the three-year horizon, $s = 36$) and new portfolios sorted on simulated book-to-market ratios. Both models generate a high-minus-low book-to-market spread one month

A potential concern is that these models are at a disadvantage because their economies are stationary, precluding entry and exit of firms. Recall, however, that the alphas between old and new sorts are similar when we define both returns using only those firms that are in the sample at t and $t-36$ (see Figure 2.B.6).

after portfolio formation. This spread is relatively small in the model of Gomes et al., 2003, consistent with the original study. More important in the context of our paper, we see that return predictability in both models fades relatively fast as time passes after portfolio formation. As a result, we find in Panels B and C that the models generate returns for the strategies that combine old and new sorts that are too small relative to what we find in Sections 2.4.1 and 2.4.2, respectively. For instance, our estimate of the unconditional alpha between old and new book-to-market sorts is 36 bps and lies above the 99th percentile of both simulated distributions. Similarly, in Panel D we see that neither model generates as large a difference between old and new book-to-market stocks (defined in Section 2.6) as we observe in the data.

In all, we reject the models, because neither model generates the return differentials between old and new book-to-market *sorts* and *stocks* that we observe in the data.

TABLE 2.A.2: **Simulating from Gomes, Kogan and Zhang (2003) and Zhang (2005)**
This table reports results from 1500 simulations of the models in Gomes et al., 2003 and Zhang, 2005b. We thank the authors for sharing the code on their websites. These models endogenously generate a positive spread in returns between high and low book-to-market stocks. We ask whether these models can match the relative performance of old and new *sorts* and *stocks* that we observe for book-to-market in the data, while matching other moments of interest. Therefore, we run these simulations using the same parameters as those used in the original studies. Panel A reports the average returns of the newest, $R_{BM,(t),t+1}$, and older, $R_{BM,(t-36),t+1}$, high-minus-low book-to-market portfolios. Panel B reports the average return to a strategy that invests in each month t 1\$ in the old sort and $-\frac{X_{H-L,(t-36),t}}{X_{H-L,(t),t}}$\$ in the new sort. Panel C reports the intercept from a regression of the old sort on the new sort. Panel D reports the difference in average returns between old, $R^{Old}_{BM,(t),t+1}$, and new, $R^{New}_{BM,(t),t+1}$, stocks that together make up the newest sort. In each panel, we report the percentiles of the simulated distribution as well as our estimate in the data.

		1	5	10	50	90	95	99	Data
				Panel A: New and old sorts					
Z05	$R_{BM,(t),t+1}$	0.149	0.372	0.459	0.603	0.725	0.770	0.884	0.530
	$R_{BM,(t-36),t+1}$	0.048	0.075	0.089	0.137	0.186	0.199	0.226	0.491
GKZ03	$R_{BM,(t),t+1}$	-0.002	0.043	0.056	0.112	0.158	0.169	0.182	0.530
	$R_{BM,(t-36),t+1}$	-0.066	-0.041	-0.036	0.008	0.051	0.057	0.070	0.491
				Panel B: Old-versus-new sort ($R_{X,(t-36),t+1} - \frac{X_{H-L,(t-s),t}}{X_{H-L,(t),t}} \times R_{X,(t),t+1}$)					
Z05		0.013	0.045	0.061	0.113	0.160	0.171	0.195	0.200
GKZ03		-0.143	-0.097	-0.085	-0.042	0.001	0.009	0.027	0.200
				Panel C: Old-versus-new sort ($R_{X,(t-36),t+1} - \beta^u \times R_{X,(t),t+1}$)					
Z05		-0.077	-0.055	-0.043	-0.005	0.046	0.063	0.099	0.361
GKZ03		-0.096	-0.057	-0.051	-0.009	0.035	0.039	0.051	0.361
				Panel D: Old-minus-new stocks ($R^{Old}_{BM,(t),t+1}$ - $R^{New}_{BM,(t),t+1}$)					
Z05		-0.459	-0.080	0.053	0.201	0.288	0.309	0.356	0.575
GKZ03		-0.216	-0.179	-0.128	-0.029	0.078	0.107	0.144	0.575

2.A.4 Figures

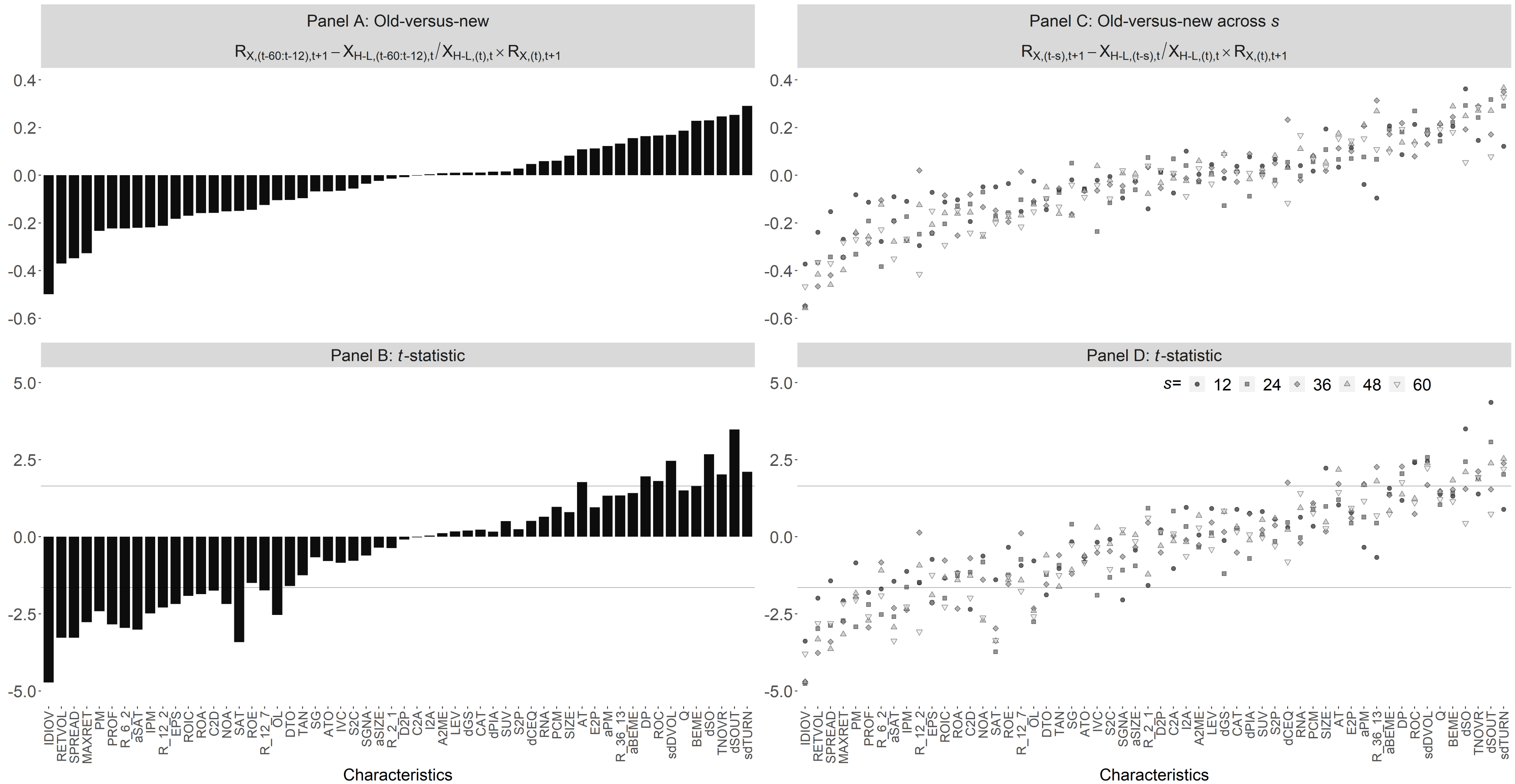

FIGURE 2.A.1: **Average returns of old versus new sorts**

This figure presents for all 56 characteristics defined in Table 3.C.1, the average return to strategies that invest in each month t 1\$ in the old sort and $-\frac{X_{H-L,(t-s),t}}{X_{H-L,(t),t}}$\$ in the new sort (see Eq. (2.4.3)). The expected return of these strategies is zero if returns of characteristic-sorted portfolios decay at the same speed as the characteristic spread. Panel A reports the average return for a single combination of the five old sorts: $R_{X,(t-60:t-12),t+1} = 1/5(R_{X,(t-12),t+1} + R_{X,(t-24),t+1} + ... + R_{X,(t-60),t+1})$. Panel C reports the average return for each individual horizon $s = 12, 24, 36, 48, 60$. Panel B and D report the associated White et al., 1980 heteroskedasticity-consistent t-statistics. We calculate the characteristic spread as the median characteristic in the high portfolio, $X_{H,(t-s),t}$, minus the median characteristic in the low portfolio, $X_{L,(t-s),t}$, at time t and for $s \geq 0$. To facilitate interpretation, we sort the characteristics from low to high average return in Panel A. The sample period runs from July 1974 to December 2017.

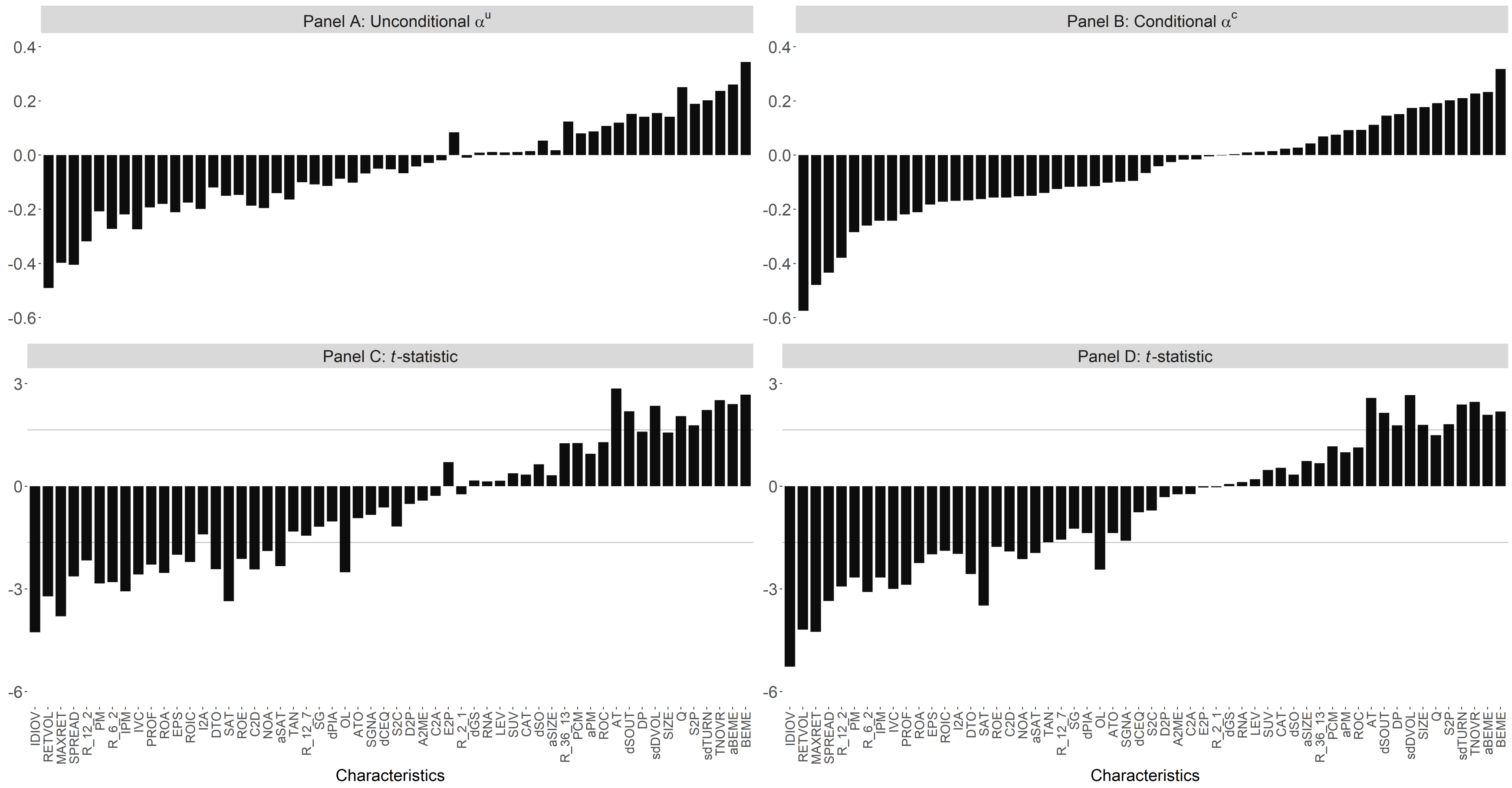

FIGURE 2.A.2: **Alphas between old and new sorts**
This figure presents the unconditional (α^u, Panel A) and conditional (α^c, Panel C) alpha (with associated White et al., 1980 heteroskedasticity-consistent t-statistics in Panels B and D) of the old sorts with respect to the newest sort for 56 characteristics. We report this alpha for a single combination of five old sorts: $R_{X,(t-60:t-12),t+1} = 1/5(R_{X,(t-12),t+1} + R_{X,(t-24),t+1} + ... + R_{X,(t-60),t+1})$, such that it represents the abnormal return from one to five years after portfolio formation. The unconditional alpha is estimated using a single time-series regression of $R_{X,(t-60:t-12),t+1}$ on $R_{X,(t),t+1}$ (see Eq. (2.4.5)). The conditional alpha is calculated as the average return to a strategy that invests in $R_{X,(t-60:t-12),t+1}$ but hedges in each month t the conditional exposure to $R_{X,(t),t+1}$ (see Eq. (2.4.6)). To facilitate interpretation, we sort the characteristics from low to high α^c. The sample period runs from July 1974 to December 2017.

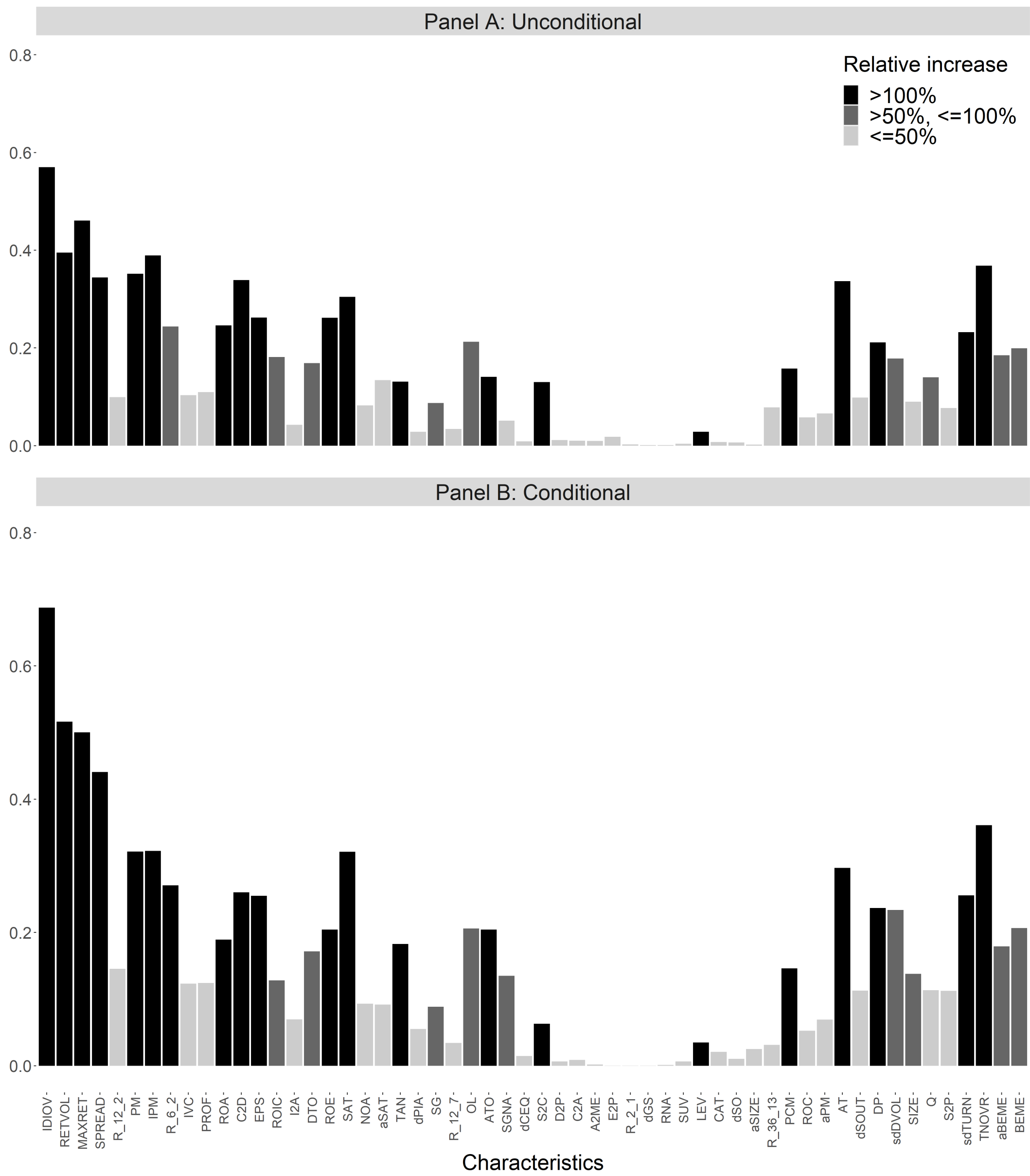

FIGURE 2.A.3: **Increases in Sharpe ratio**
This figure presents the maximum improvement in Sharpe ratio from combining the newest sort ($R_{X,t,t+1}$) with a single combination of five old sorts: ($R_{X,(t-60:t-12),t+1}$). The improvement is defined as: Max. Sharpe($R_{X,(t-60:t-12),t+1}$, $R_{X,(t),t+1}$) - Sharpe($R_{X,(t),t+1}$). We color code these improvements to highlight the relative increase in Sharpe ratio (defined as: improvement/Sharpe($R_{X,(t),t+1}$)-1). To calculate the maximum Sharpe ratio, we add to the return of the newest sort either the return of the old sort (Panel A) or the conditionally hedged return of the old sort (Panel B, see Eq. (2.4.6)).

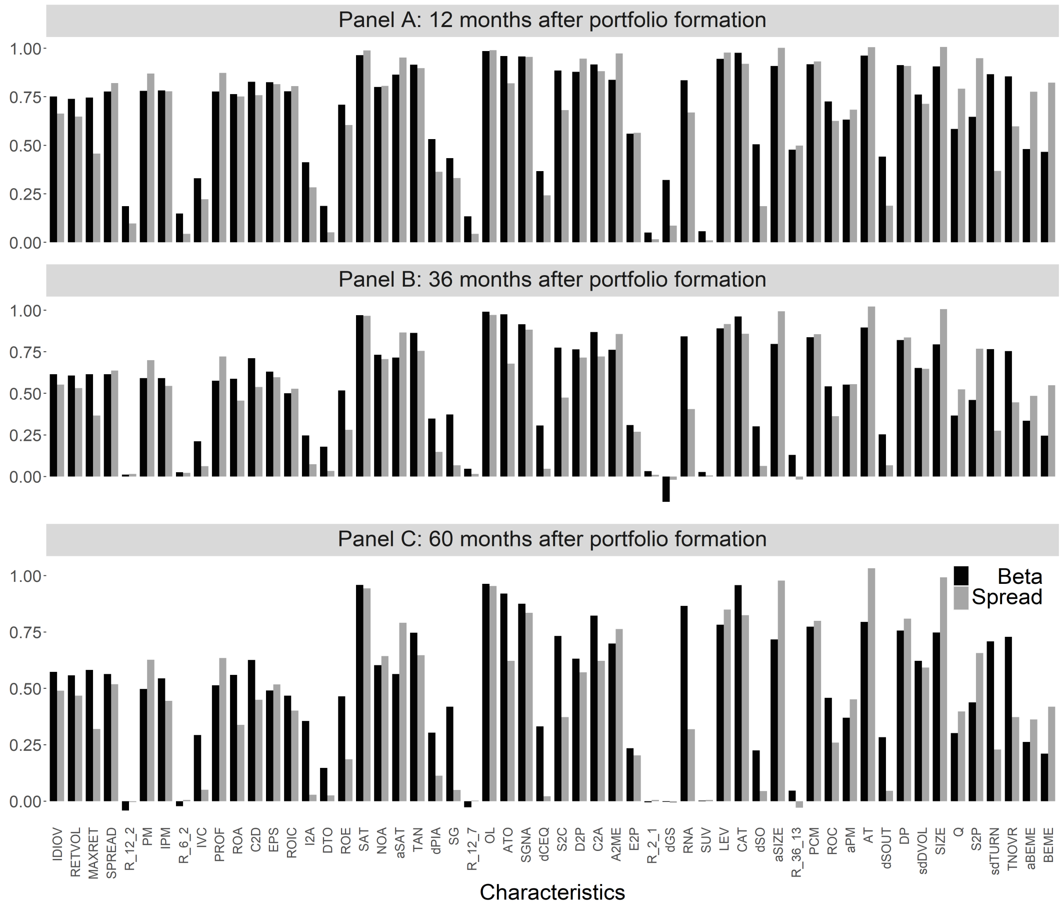

FIGURE 2.A.4: **Persistence and the beta of old to new sorts**
This figure presents the persistence of the 56 characteristics we study (in the same order as the conditional alphas from Figure 2.A.2) as well as the beta in a regression of the old on the new sort, β_s^u, for $s = 12, 36, 60$. Persistence is measured as the high-minus-low characteristic spread that remains s months after portfolio formation as a fraction of the same spread at portfolio formation. The sample period runs from July 1974 to December 2017.

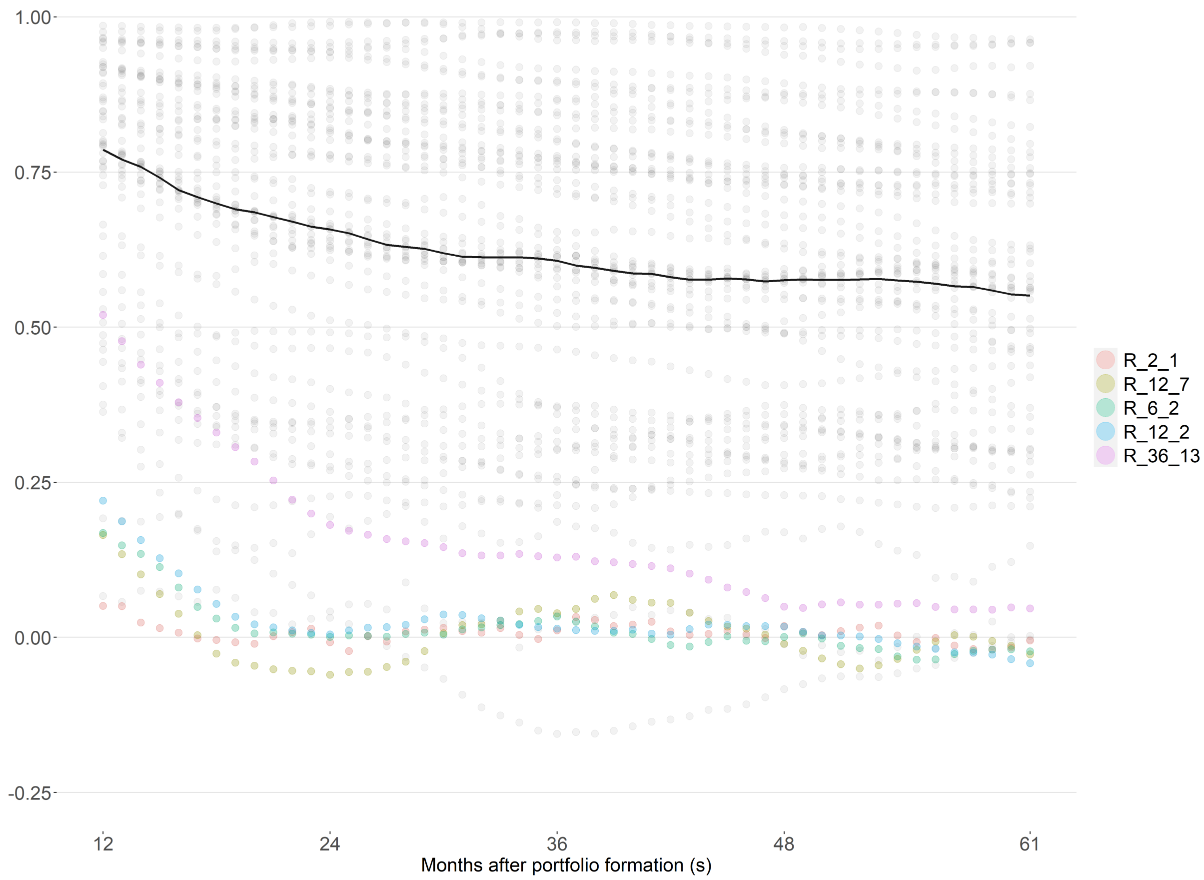

FIGURE 2.A.5: **Betas of old with respect to new sorts**
This figure presents for all 56 characteristics the beta in a regression of older sorts, with returns $R_{X,(t-s),t+1}$ for $s = 12, ..., 60$, on the newest sort, with return $R_{X,(t),t+1}$. The solid line highlights the median across characteristics. We highlight the relatively low betas for past-return-based characteristics (as defined in Table 3.C.1). The sample period runs from July 1974 to December 2017.

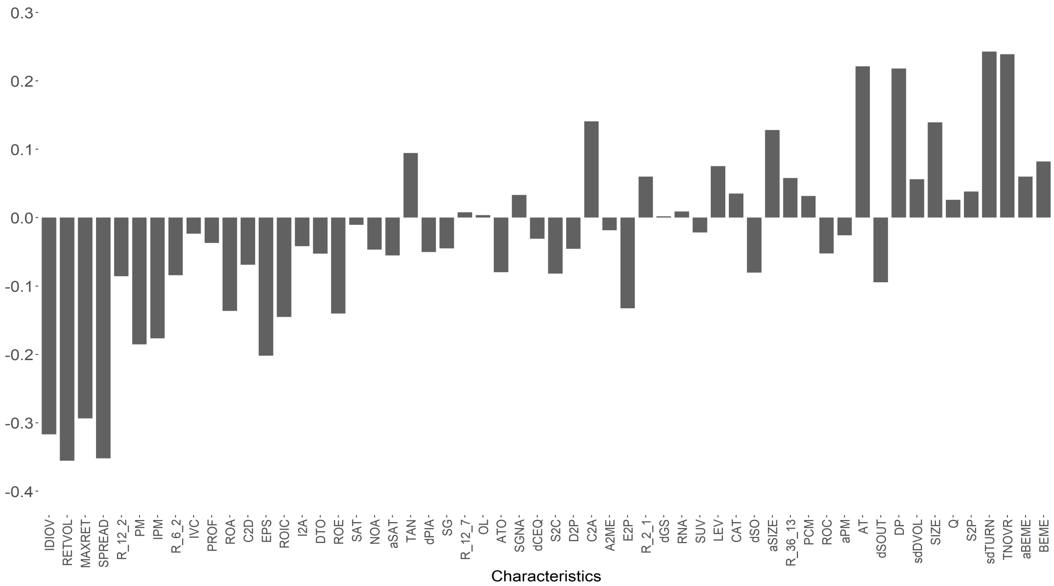

FIGURE 2.A.6: **Principal component loadings**
This figure plots the loadings of the first principal component extracted from returns of the newest sorts, $R_{X,(t),t+1}$, from July 1974 to December 2017 and in the same order as the conditional alphas from Figure 2.A.2.

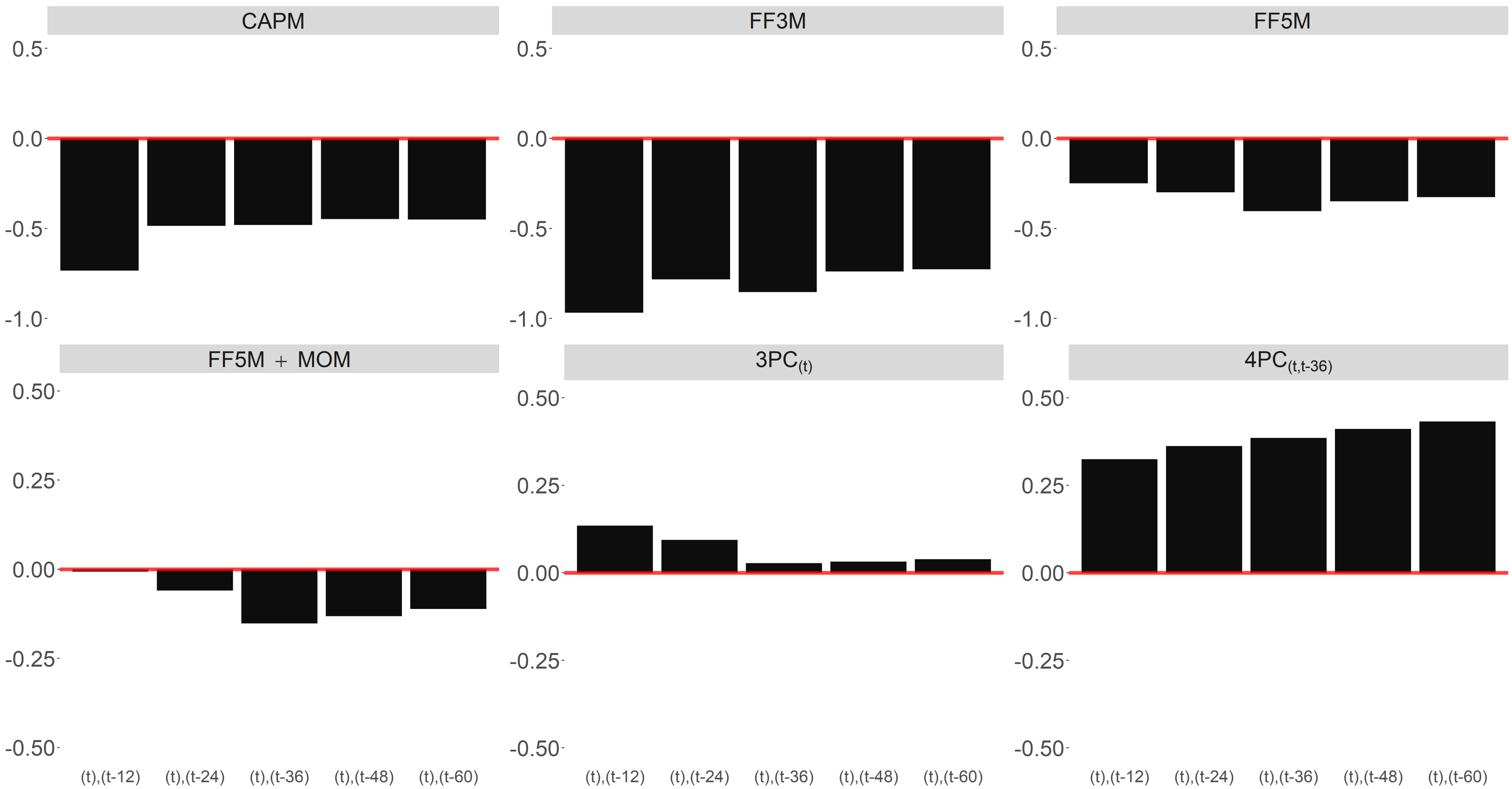

FIGURE 2.A.7: **Cross-sectional** R^2
This figure presents for each factor model (defined as in Table 2.A.5) the cross-sectional R^2 when the test assets are 112 characteristic-sorted portfolios (56 returns to new sorts, $R_{X,(t),t+1}$, and 56 returns to old sorts, $R_{X,(t-s),t+1}$, with sample average return denoted μ) and the factor risk premia are set equal to the sample average factor returns (denoted μ_F). Thus, $R^2 = 1 - var(\mu - \beta_F' \mu_F)/var(\mu)$. The sample period runs from July 1974 to December 2017.

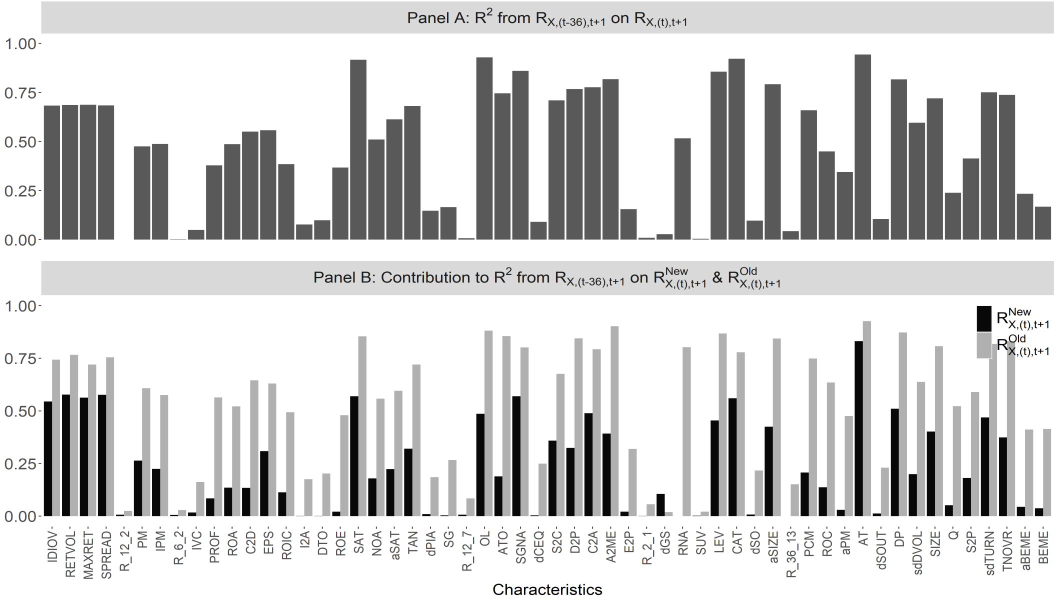

FIGURE 2.A.8: **Contribution to R^2 from old and new stocks**
In Panel A of this figure we report the R^2 from a regression of an old sort, $R_{X,(t-36),t+1}$, on the newest sort, $R_{X,(t),t+1}$. Panel B plots the relative contribution to the R^2 in a regression of the old sort, $R_{X,(t-36),t+1}$, on the new and old stock components, $R^{New}_{X,(t),t+1}$ and $R^{Old}_{X,(t),t+1}$, which together make up the return of the newest sort. The sample period runs from July 1974 to December 2017.

2.A.5 Tables

TABLE 2.A.1: **Summary statistics of old and new sorts for four popular characteristics**
This table reports summary statistics for old and new sorts on book-to-market, size, profitability, and investment. We track the returns of long-short decile portfolios (value-weighted and split at NYSE breakpoints) for each characteristic from one to sixty months after portfolio formation. For each long-short portfolio, we report the average number of firms in the high plus low portfolio, the average high-minus-low return and its associated White et al., 1980 heteroskedasticity consistent t-statistic. The sample period runs from July 1974 to December 2017.

		Book-to-market			Size			Profitability			Investment		
		Firms	Ret.	t-stat	Firms	Ret.	t-stat	Firms	Ret.	t-stat	Firms	Ret.	t-stat
New sort	$R_{X,(t),t+1}$	1155	0.53	1.86	2121	0.31	1.55	1085	0.43	3.28	1343	0.51	3.48
Old sorts	$R_{X,(t-12),t+1}$	1029	0.61	3.12	1882	0.50	2.48	960	0.26	2.14	1193	0.22	1.81
	$R_{X,(t-24),t+1}$	917	0.51	2.87	1661	0.41	2.05	848	0.14	1.15	1054	0.09	0.71
	$R_{X,(t-36),t+1}$	821	0.49	2.88	1476	0.32	1.69	756	0.01	0.12	937	0.02	0.18
	$R_{X,(t-48),t+1}$	738	0.51	3.04	1316	0.37	2.04	677	0.01	0.10	837	0.00	0.00
	$R_{X,(t-60),t+1}$	670	0.38	2.21	1187	0.37	1.99	615	0.00	-0.01	756	-0.08	-0.57

TABLE 2.A.2: **Alphas of old versus new sorts and improvements in Sharpe ratio**
Panel A of this table reports the alpha of old with respect to new sorts on book-to-market, size, profitability, and investment. 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: $R_{X,(t-s),t+1} = \alpha_s^u + \beta_s^u R_{X,(t),t+1} + \epsilon_{X,(t-s),t+1}$ (see Eq. (2.4.5)). The conditional alpha, α^c, is calculated as the average return of a strategy that invests in $R_{X,(t-s),t+1}$ but hedges in each month t the conditional exposure to $R_{X,(t),t+1}$. Following Eq. (2.4.6), we estimate this exposure over a 60 month historical rolling window. White et al., 1980 heteroskedasticity consistent t-statistics are reported in parentheses. Panel B reports the Sharpe ratio of returns immediately after portfolio formation, 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 1974 to December 2017.

	Book-to-market			Size			Profitability			Investment		
	Panel A: Alphas between old and new sorts											
	α^u	β^u	α^c	α^u	β^u	α^c	α^u	β^u	α^c	α^u	β^u	α^c
$R_{X,(t-12),t+1}$	0.36	0.47	0.42	0.22	0.91	0.26	-0.08	0.78	-0.10	0.01	0.41	-0.02
	(2.68)	(7.97)	(3.14)	(2.67)	(18.53)	(3.08)	(-1.28)	(15.85)	(-1.74)	(0.06)	(11.91)	(-0.23)
$R_{X,(t-24),t+1}$	0.33	0.32	0.39	0.15	0.85	0.20	-0.13	0.62	-0.15	-0.05	0.28	-0.10
	(2.30)	(6.54)	(2.72)	(1.45)	(16.98)	(1.96)	(-1.56)	(9.89)	(-1.91)	(-0.45)	(6.50)	(-0.80)
$R_{X,(t-36),t+1}$	0.36	0.25	0.40	0.07	0.79	0.12	-0.23	0.58	-0.27	-0.10	0.25	-0.17
	(2.38)	(4.86)	(2.66)	(0.74)	(18.42)	(1.23)	(-2.39)	(9.72)	(-3.04)	(-0.80)	(6.49)	(-1.38)
$R_{X,(t-48),t+1}$	0.39	0.23	0.40	0.13	0.76	0.19	-0.22	0.55	-0.27	-0.19	0.37	-0.23
	(2.57)	(4.90)	(2.68)	(1.40)	(17.75)	(1.98)	(-2.30)	(10.21)	(-3.06)	(-1.46)	(7.01)	(-1.85)
$R_{X,(t-60),t+1}$	0.26	0.22	0.26	0.14	0.75	0.19	-0.23	0.52	-0.28	-0.27	0.36	-0.31
	(1.68)	(4.92)	(1.68)	(1.29)	(14.65)	(1.69)	(-2.19)	(7.38)	(-2.91)	(-1.92)	(5.57)	(-2.39)
	Panel B: Improvements in Sharpe ratio											
		u	c		u	c		u	c		u	c
	Sharpe($R_{X,(t),t+1}$)											
$R_{X,(t),t+1}$		0.28			0.23			0.50			0.53	
	Max. Sharpe($R_{X,(t-s),t+1}$, $R_{X,(t),t+1}$) - Sharpe($R_{X,(t),t+1}$)											
$R_{X,(t-12),t+1}$		0.19	0.28		0.21	0.29		0.04	0.05		0.00	0.00
$R_{X,(t-24),t+1}$		0.16	0.25		0.08	0.15		0.05	0.05		0.00	0.01
$R_{X,(t-36),t+1}$		0.17	0.24		0.02	0.08		0.12	0.14		0.01	0.03
$R_{X,(t-48),t+1}$		0.19	0.24		0.08	0.15		0.11	0.15		0.05	0.07
$R_{X,(t-60),t+1}$		0.09	0.12		0.07	0.12		0.10	0.13		0.08	0.11

TABLE 2.A.3: **Do factor models price principal components at longer horizons after portfolio formation?**

In Panel A of this table, we present F-statistics and p-values from GRS tests, where the test assets are the first three principal components extracted from characteristic-sorted portfolio returns at different horizons s after portfolio formation ($\lambda'_{(t-s),z=1,2,3}R_{X,(t-s),t+1}$). We ask whether these returns are priced by one of five models. The first is a statistical factor model that uses the first three principal components extracted from $R_{X,(t),t+1}$ as factors, denoted $3PC_{(t),t+1}$. Next, we also consider the single-factor CAPM (Sharpe, 1964, Lintner, 1965, Mossin, 1966); the three-factor model of Fama and French (1993b, FF3M); the five-factor model of Fama and French (2015, FF5M); and a six-factor model that augments FF5M with momentum (FF5M+MOM). These returns are taken directly from Kenneth French's website. Panel B presents the summary statistics for the first principal component extracted from $R_{X,(t-s),t+1}$. The sample runs from July 1974 to December 2017.

Panel A: GRS tests										
$\lambda'_{(t-s),z=1,2,3}R_{X,(t-s),t+1}$	$3PC_{(t),t+1}$		CAPM		FF3M		FF5M		FF5M+MOM	
s	F-stat	p-val	F-stat	p-val	F-stat	p-val	F-stat	p-val	F-stat	p-val
0			7.90	0.0000	8.43	0.0000	1.48	0.2203	1.02	0.3827
12	5.13	0.0017	5.66	0.0008	5.39	0.0012	3.50	0.0154	2.99	0.0306
24	9.65	0.0000	2.57	0.0539	2.96	0.0321	5.24	0.0014	5.46	0.0011
36	9.70	0.0000	2.24	0.0823	3.05	0.0281	6.10	0.0004	5.85	0.0006
48	10.50	0.0000	2.26	0.0806	2.84	0.0372	6.94	0.0001	6.95	0.0001
60	9.45	0.0000	1.44	0.2296	2.22	0.0844	6.42	0.0003	6.53	0.0002

Panel B: Summary statistics for the first principal component							
$\lambda'_{(t-s),1}R_{X,(t-s),t+1}$	Avg. Ret.	Correlations					
s		0	12	24	36	48	60
0	-0.60	1.00					
12	0.37	0.96	1.00				
24	0.87	0.93	0.98	1.00			
36	0.95	0.91	0.96	0.98	1.00		
48	1.02	0.90	0.95	0.97	0.98	1.00	
60	1.01	0.90	0.94	0.95	0.96	0.98	1.00

TABLE 2.A.4: **Alphas of individual principal components at longer horizons after portfolio formation**

This table presents the intercept (α) and associated t-statistic (based on White et al., 1980 heteroskedasticity consistent standard errors) from regressing the returns at longer horizons after portfolio formation for each of three principal component strategies on five candidate factor models (Panels A to E; see Table 2.A.3 for model definitions). These returns are defined as the linear combination of (i) the loadings $\lambda_{(t),z}$ of the z-th principal component extracted from $R_{X,(t),t+1}$ and (ii) characteristic-sorted portfolio returns at horizons $s = 12, 24, ..., 60$ after portfolio formation (see Section 2.5.1 for more detail). Panel F presents the alpha from a regression of returns to an old-minus-new strategy (i.e., a simple long-short combination of $R_{X,(t-s),t+1}$ and $R_{X,(t),t+1}$, weighted by $\lambda_{(t),1}$) on the same factor models. The sample period runs from July 1974 to December 2017.

$\lambda'_{(t),z}R_{X,(t-s),t+1}$	PC1 ($z=1$)		PC2 ($z=2$)		PC3 ($z=3$)	
s	α	t-stat	α	t-stat	α	t-stat
Panel A: $3PC_{(t),t+1}$						
12	0.68	3.69	0.04	0.22	-0.39	-1.35
24	1.13	5.06	0.17	0.66	-0.28	-0.95
36	1.27	4.98	0.23	0.73	-0.01	-0.02
48	1.37	5.32	0.07	0.22	-0.21	-0.80
60	1.37	5.11	0.09	0.27	-0.41	-1.51
Panel B: CAPM						
0	-2.32	-3.26	-0.33	-0.46	1.43	3.34
12	-1.08	-1.67	-1.25	-2.05	-0.03	-0.10
24	-0.44	-0.71	-1.08	-1.88	-0.04	-0.12
36	-0.27	-0.45	-0.87	-1.58	0.35	1.13
48	-0.17	-0.29	-1.02	-1.91	0.20	0.58
60	-0.14	-0.26	-0.96	-1.74	-0.04	-0.13
Panel C: FF3M						
0	-2.18	-4.46	1.69	3.41	0.82	2.28
12	-0.86	-2.03	0.68	2.37	-0.33	-1.08
24	-0.12	-0.30	0.60	1.76	-0.48	-1.69
36	0.09	0.25	0.65	1.79	-0.15	-0.53
48	0.20	0.54	0.41	1.10	-0.36	-1.21
60	0.16	0.45	0.41	1.05	-0.59	-2.03
Panel D: FF5M						
0	-0.33	-0.85	1.21	2.05	0.50	1.29
12	0.61	1.83	0.81	2.90	-0.34	-1.05
24	1.13	3.32	0.80	2.31	-0.53	-1.68
36	1.32	3.92	0.80	2.03	-0.20	-0.62
48	1.36	4.32	0.58	1.42	-0.52	-1.71
60	1.28	4.03	0.69	1.63	-0.80	-2.47
Panel E: FF5M+MOM						
0	0.10	0.26	0.03	0.08	-0.38	-1.75
12	0.64	1.89	0.70	2.41	-0.59	-1.84
24	1.10	3.17	0.88	2.54	-0.56	-1.78
36	1.27	3.68	0.88	2.24	-0.25	-0.77
48	1.35	4.20	0.66	1.62	-0.62	-1.99
60	1.30	3.94	0.80	1.90	-0.78	-2.26

Continued

Panel F: Old-minus-new sorts								
$\lambda'_{(t),1}(R_{X,(t-s),t+1} - R_{X,(t),t+1})$	CAPM		FF3M		FF5M		FF5M+MOM	
	α	t-stat	α	t-stat	α	t-stat	α	t-stat
12	1.24	5.02	1.32	5.67	0.94	3.63	0.54	2.27
24	1.88	6.26	2.06	7.22	1.46	4.64	1.00	3.52
36	2.05	5.97	2.27	6.87	1.65	4.70	1.17	3.52
48	2.15	6.08	2.37	7.09	1.69	4.81	1.25	3.60
60	2.18	6.01	2.34	6.86	1.61	4.69	1.20	3.43

TABLE 2.A.5: **Does a statistical four-factor model capture the returns of old sorts?**
This table presents results from asset pricing tests for the principal components extracted from new and old characteristic-sorted portfolios. We report the intercept (α, with accompanying t-statistic) from simple regressions of each of the first five principal components on a four factor model, $4PC_{(t,t-36),t+1}$, that augments the first three principal components extracted from $R_{X,(t),t+1}$ with the first principal component extracted from $R_{X,(t-36),t+1}$. We also report the GRS test statistic and associated p-values from pricing the first three or first five principal components extracted from the returns at each horizon. The sample period runs from July 1974 to December 2017.

$\lambda'_{(t),z}R_{X,(t-s),t+1}$	PC1 ($z=1$)		PC2 ($z=2$)		PC3 ($z=3$)		PC4 ($z=4$)		PC5 ($z=5$)		GRS_3		GRS_5	
s	α	t-stat	α	t-stat	α	t-stat	α	t-stat	α	t-stat	F-stat	p-val	F-stat	p-val
12	0.06	0.47	0.01	0.04	-0.59	-1.99	0.31	0.94	0.24	0.97	1.33	0.26	1.41	0.22
24	0.22	1.67	0.18	0.70	-0.54	-1.82	0.10	0.33	0.19	0.82	2.01	0.11	1.41	0.22
36	0.00	-0.08	0.46	1.39	-0.26	-0.95	0.09	0.28	0.44	1.90				
48	0.28	2.19	0.29	0.85	-0.42	-1.49	0.18	0.57	0.26	1.08	2.48	0.06	1.82	0.11
60	0.44	2.37	0.24	0.65	-0.40	-1.46	0.01	0.03	0.01	0.05	3.37	0.02	2.02	0.07

TABLE 2.A.6: **Do factor models capture the alpha between old and new sorts?**
This table reports the mean absolute alpha (MAA) and number of test statistics significant at the 5%-level (#, out of 56 characteristics) from simple regressions of the returns to old-versus-new characteristic-sorted portfolio strategies on a number of factor models. Panel A reports results for the benchmark factor models; CAPM, FF3M, FF5M and FF5M+MOM and Panel B reports results for the statistical factor models; $3PC_{(t),t+1}$ and $4PC_{(t,t-36),t+1}$. Building on Sections 2.4.1 and 2.4.2, we consider three old-versus-new strategies that each invest 1$ in the old sort and short the newest sort for an amount equal to $\frac{X_{H-L,(t-s),t}}{X_{H-L,(t),t}}$ (the characteristic spread that remains s months after portfolio formation), β^u (the unconditional beta from a regression of the old on the new sort), and β^c_t (the conditional beta from a regression of the old on the new sort estimated over a 60-month rolling window), respectively. The sample period runs from July 1974 to December 2017.

	$\frac{X_{H-L,(t-s),t}}{X_{H-L,(t),t}}$		β^u		β^c_t		$\frac{X_{H-L,(t-s),t}}{X_{H-L,(t),t}}$		β^u		β^c_t	
s	MAA	#	MAA	#	MAA	#	MAA	#	MAA	#	MAA	#
	Panel A: Benchmark factor models											
	CAPM						FF3M					
12	0.12	9	0.13	12	0.14	15	0.10	6	0.11	9	0.12	10
24	0.14	12	0.16	16	0.16	14	0.13	13	0.14	13	0.15	14
36	0.14	16	0.15	17	0.15	14	0.14	12	0.15	14	0.15	14
48	0.15	14	0.17	20	0.18	17	0.14	15	0.16	14	0.17	18
60	0.14	12	0.16	17	0.16	18	0.13	8	0.15	15	0.15	15
	FF5M						FF5M+MOM					
12	0.10	9	0.11	10	0.11	8	0.12	13	0.11	12	0.11	10
24	0.14	13	0.14	12	0.13	13	0.15	17	0.14	15	0.14	14
36	0.15	14	0.15	13	0.14	14	0.16	12	0.15	14	0.14	14
48	0.16	15	0.17	16	0.17	16	0.17	17	0.16	15	0.17	14
60	0.17	15	0.17	12	0.16	14	0.18	16	0.18	15	0.17	18
	Panel B: Statistical models											
	$3PC_{(t),t+1}$						$4PC_{(t,t-36),t+1}$					
12	0.10	10	0.10	8	0.11	11	0.08	6	0.08	5	0.10	7
24	0.14	17	0.14	15	0.14	14	0.10	8	0.09	8	0.10	8
36	0.14	15	0.14	15	0.14	16	0.10	7	0.09	6	0.10	7
48	0.15	19	0.16	18	0.17	19	0.10	11	0.09	10	0.10	10
60	0.15	19	0.16	19	0.16	20	0.10	6	0.09	8	0.10	10

TABLE 2.A.7: **Alphas between old and new sorts across market beta groups**
This table presents average returns and alphas of old-versus-new strategies across three market beta groups. We sort the 56 characteristics using the market beta of the newest characteristic-sorted portfolio estimated over a 60 month rolling window. Characteristic-sorted portfolio returns are equal-weighted within each market beta group. We report results for the returns to five strategies, where $R_{X,(t),t+1}$ is the return to the new sort, $R_{X,(t-36),t+1}$ is the return to the old sort (three years after portfolio formation), and the remaining strategies combine old and new sorts: each investing 1$ in the old sort and shorting the new sort for an amount equal to $\frac{X_{H-L,(t-s),t}}{X_{H-L,(t),t}}$ (the characteristic spread that remains s months after portfolio formation), β^u (the unconditional beta from a regression of the old on the new sort), and β^c_t (the conditional beta from a regression of the old on the new sort estimated over a 60-month rolling window), respectively. Panel A reports average returns to these strategies, whereas Panels B and C report the intercept from regressing each strategy's returns on benchmark and statistical factor models, respectively. t-statistics use White et al., 1980 heteroskedasticity consistent standard errors. The sample period runs from July 1974 to December 2017.

	Low	Mid	High	H-L	Low	Mid	High	H-L
	Panel A: Average returns							
	Avg. ret.				t-stat			
$R_{X,(t),t+1}$	0.23	0.37	0.30	0.07	1.60	5.89	2.86	0.29
$R_{X,(t-36),t+1}$	-0.13	0.13	0.25	0.38	-1.15	3.25	2.69	1.91
$R_{X,(t-36),t+1} - \frac{X_{H-L,(t-s),t}}{X_{H-L,(t),t}} \times R_{X,(t),t+1}$	-0.19	0.02	0.10	0.29	-3.11	0.47	2.63	3.33
$R_{X,(t-36),t+1} - \beta^u \times R_{X,(t),t+1}$	-0.22	-0.01	0.10	0.32	-3.90	-0.17	2.89	4.05
$R_{X,(t-36),t+1} - \beta^c_t \times R_{X,(t),t+1}$	-0.23	-0.01	0.10	0.33	-4.39	-0.17	2.77	4.42
	Panel B: Alphas in benchmark factor models							
	α				t-stat			
	CAPM							
$R_{X,(t),t+1}$	0.53	0.43	0.13	-0.40	4.64	7.06	1.39	-2.07
$R_{X,(t-36),t+1}$	0.08	0.16	0.10	0.02	0.84	3.75	1.22	0.13
$R_{X,(t-36),t+1} - \frac{X_{H-L,(t-s),t}}{X_{H-L,(t),t}} \times R_{X,(t),t+1}$	-0.11	0.02	0.07	0.18	-1.99	0.60	1.73	2.17
$R_{X,(t-36),t+1} - \beta^u \times R_{X,(t),t+1}$	-0.18	-0.01	0.08	0.26	-3.23	-0.15	2.29	3.32
$R_{X,(t-36),t+1} - \beta^c_t \times R_{X,(t),t+1}$	-0.20	-0.01	0.07	0.27	-3.83	-0.17	2.12	3.73
	FF3M							
$R_{X,(t),t+1}$	0.50	0.32	0.08	-0.42	5.46	5.75	1.13	-2.75
$R_{X,(t-36),t+1}$	0.03	0.11	0.09	0.07	0.35	2.56	1.50	0.53
$R_{X,(t-36),t+1} - \frac{X_{H-L,(t-s),t}}{X_{H-L,(t),t}} \times R_{X,(t),t+1}$	-0.18	0.01	0.08	0.26	-3.62	0.27	2.28	3.66
$R_{X,(t-36),t+1} - \beta^u \times R_{X,(t),t+1}$	-0.23	-0.01	0.08	0.31	-4.80	-0.31	2.30	4.50
$R_{X,(t-36),t+1} - \beta^c_t \times R_{X,(t),t+1}$	-0.24	-0.01	0.06	0.30	-4.79	-0.22	1.92	4.38

Continued

	Low	Mid	High	High-Low	Low	Mid	High	High-Low
	α				t-stat			
	FF5M							
$R_{X,(t),t+1}$	0.20	0.20	0.23	0.03	2.41	3.59	3.20	0.19
$R_{X,(t-36),t+1}$	-0.20	0.05	0.25	0.44	-2.82	1.06	4.03	3.68
$R_{X,(t-36),t+1} - \frac{X_{H-L,(t-s),t}}{X_{H-L,(t),t}} \times R_{X,(t),t+1}$	-0.26	-0.01	0.12	0.37	-4.73	-0.24	3.04	4.92
$R_{X,(t-36),t+1} - \beta^u \times R_{X,(t),t+1}$	-0.28	-0.01	0.10	0.37	-4.90	-0.36	2.60	4.80
$R_{X,(t-36),t+1} - \beta^c_t \times R_{X,(t),t+1}$	-0.25	-0.01	0.07	0.32	-4.34	-0.27	2.00	4.20
	FF5M+MOM							
$R_{X,(t),t+1}$	0.13	0.22	0.23	0.10	1.32	3.55	2.89	0.61
$R_{X,(t-36),t+1}$	-0.19	0.04	0.24	0.43	-2.71	0.96	3.85	3.53
$R_{X,(t-36),t+1} - \frac{X_{H-L,(t-s),t}}{X_{H-L,(t),t}} \times R_{X,(t),t+1}$	-0.24	-0.03	0.10	0.34	-4.31	-0.74	2.47	4.25
$R_{X,(t-36),t+1} - \beta^u \times R_{X,(t),t+1}$	-0.24	-0.04	0.08	0.32	-4.19	-0.89	1.96	3.87
$R_{X,(t-36),t+1} - \beta^c_t \times R_{X,(t),t+1}$	-0.22	-0.04	0.06	0.28	-3.68	-0.91	1.48	3.38
	Panel C: Alphas in statistical factor models							
	$3PC_{(t),t+1}$							
$R_{X,(t),t+1}$	0.11	0.31	0.32	0.21	1.75	6.07	5.29	1.96
$R_{X,(t-36),t+1}$	-0.19	0.11	0.28	0.48	-3.45	2.72	5.20	4.80
$R_{X,(t-36),t+1} - \frac{X_{H-L,(t-s),t}}{X_{H-L,(t),t}} \times R_{X,(t),t+1}$	-0.22	0.00	0.12	0.34	-4.74	0.02	3.05	4.70
$R_{X,(t-36),t+1} - \beta^u \times R_{X,(t),t+1}$	-0.23	-0.02	0.10	0.33	-4.65	-0.50	2.70	4.42
$R_{X,(t-36),t+1} - \beta^c_t \times R_{X,(t),t+1}$	-0.24	-0.02	0.09	0.33	-4.60	-0.53	2.53	4.40
	$4PC_{(t,t-36),t+1}$							
$R_{X,(t),t+1}$	0.06	0.25	0.26	0.21	0.97	4.96	4.12	1.84
$R_{X,(t-36),t+1}$	-0.03	0.12	0.15	0.18	-0.59	3.01	3.02	2.12
$R_{X,(t-36),t+1} - \frac{X_{H-L,(t-s),t}}{X_{H-L,(t),t}} \times R_{X,(t),t+1}$	-0.06	0.04	0.03	0.09	-1.79	1.09	0.97	1.87
$R_{X,(t-36),t+1} - \beta^u \times R_{X,(t),t+1}$	-0.05	0.01	0.02	0.07	-1.51	0.45	0.59	1.37
$R_{X,(t-36),t+1} - \beta^c_t \times R_{X,(t),t+1}$	-0.06	0.02	0.02	0.08	-1.64	0.62	0.61	1.51

TABLE 2.A.8: **Descriptives for new and old stock portfolios across market beta groups**
This table presents descriptive statistics for the new and old stock portfolios, which together make up the newest characteristic-sorted portfolio. Each statistic is presented for three groups of characteristics sorted on market beta (equal-weighting the characteristics within each group). We present for both the old and new stock portfolio: (i) the high-minus-low characteristic spread as a fraction of the characteristic spread in the not-decomposed newest characteristic-sorted portfolio; (ii) the total market cap (as a fraction of total CRSP market cap) in the high and low portfolio; (iii) the high-minus-low difference in market cap (as a fraction of total CRSP market cap); and (iv-vi) the difference in median book-to-market, profitability, and investment between the high and low portfolio. To put these differences in perspective, the table also reports the characteristic spread that is obtained in a single sort of stocks on these characteristics.

	Market beta			
	Low	Mid	High	
	Characteristic spread			
New	0.99	1.04	0.90	
Old	1.02	0.98	1.04	
Single sort on characteristic				1
	% CRSP market cap (High + Low)			
New	0.09	0.09	0.10	
Old	0.11	0.11	0.15	
Single sort on Size				0.61
	% CRSP market cap (High - Low)			
New	0.00	0.00	0.01	
Old	0.01	0.02	0.05	
Single sort on Size				0.58
	Book-to-market (BM, High - Low)			
New	-0.13	-0.09	0.01	
Old	-0.14	-0.14	-0.05	
Single sort on BM				1.98
	Profitability ($PROF$, High - Low)			
New	0.18	0.25	0.20	
Old	0.20	0.26	0.22	
Single sort on $PROF$				2.13
	Investment ($I2A$, High - Low)			
New	14.72	16.13	3.72	
Old	10.70	11.81	2.65	
Single sort on $I2A$				85.89

TABLE 2.A.9: **The relative performance of new versus old stocks**

This table presents results for the returns of new versus old stocks across three market beta groups. The returns of new ($R^{New}_{X,t,t+1}$) and old ($R^{Old}_{X,t,t+1}$) stocks together make up the return to the newest sort ($R_{X,t,t+1}$) and are derived from a dependent double sort. The old stock component is the return to a strategy that goes long the subset of stocks for which, among all stocks in the High portfolio at time t, the characteristic X is above the median value of that characteristic 36 months ago. Conversely, this strategy goes short the subset of stocks for which, among all stocks in the Low portfolio at time t, the characteristic X is below the median value of that characteristic 36 months ago. The new stock component uses all remaining stocks in the High and Low portfolio at time t. All returns are equal-weighted within market beta groups. We report the intercept from regressing the market beta-sorted portfolios on benchmark asset pricing models as defined in Table 2.A.3. t-statistics use White et al., 1980 heteroskedasticity consistent standard errors. The sample period runs from July 1974 to December 2017.

	Market beta				Market beta			
	Low	Mid	High	H-L	Low	Mid	High	H-L
	Avg. ret.				t-stat			
$R_{(t)t+1}$	0.23	0.37	0.30	0.07	1.60	5.89	2.86	0.29
$R^{New}_{X,t,t+1}$	0.31	0.35	0.19	-0.12	2.43	5.05	1.82	-0.53
$R^{Old}_{X,t,t+1}$	0.13	0.40	0.38	0.25	0.77	5.32	3.63	0.94
$R^{Old}_{X,t,t+1}$ - $R^{New}_{X,t,t+1}$	-0.18	0.06	0.19	0.37	-2.14	0.94	3.11	3.17
	α				t-stat			
	CAPM							
$R_{(t)t+1}$	0.53	0.43	0.13	-0.40	4.64	7.06	1.39	-2.07
$R^{New}_{X,t,t+1}$	0.56	0.40	0.03	-0.53	5.43	5.95	0.33	-2.92
$R^{Old}_{X,t,t+1}$	0.47	0.48	0.21	-0.26	3.30	6.61	2.28	-1.17
$R^{Old}_{X,t,t+1}$ - $R^{New}_{X,t,t+1}$	-0.09	0.09	0.18	0.27	-1.11	1.41	3.07	2.39
	FF3M							
$R_{(t)t+1}$	0.50	0.32	0.08	-0.42	5.46	5.75	1.13	-2.75
$R^{New}_{X,t,t+1}$	0.55	0.28	-0.03	-0.58	6.20	4.39	-0.39	-3.82
$R^{Old}_{X,t,t+1}$	0.41	0.35	0.18	-0.22	3.65	5.54	2.42	-1.31
$R^{Old}_{X,t,t+1}$ - $R^{New}_{X,t,t+1}$	-0.15	0.08	0.21	0.36	-1.96	1.21	3.59	3.37
	FF5M							
$R_{(t)t+1}$	0.20	0.20	0.23	0.03	2.41	3.59	3.20	0.19
$R^{New}_{X,t,t+1}$	0.29	0.18	0.10	-0.19	3.41	2.64	1.30	-1.25
$R^{Old}_{X,t,t+1}$	0.04	0.20	0.32	0.29	0.38	3.36	4.28	1.86
$R^{Old}_{X,t,t+1}$ - $R^{New}_{X,t,t+1}$	-0.26	0.02	0.22	0.47	-3.43	0.29	3.44	4.31
	FF5M+MOM							
$R_{(t)t+1}$	0.13	0.22	0.23	0.10	1.32	3.55	2.89	0.61
$R^{New}_{X,t,t+1}$	0.22	0.23	0.13	-0.10	2.19	3.09	1.36	-0.53
$R^{Old}_{X,t,t+1}$	-0.04	0.19	0.30	0.34	-0.35	2.96	3.81	2.00
$R^{Old}_{X,t,t+1}$ - $R^{New}_{X,t,t+1}$	-0.26	-0.04	0.18	0.44	-3.40	-0.56	2.60	3.72

2.B Internet Appendix

2.B.1 Figures

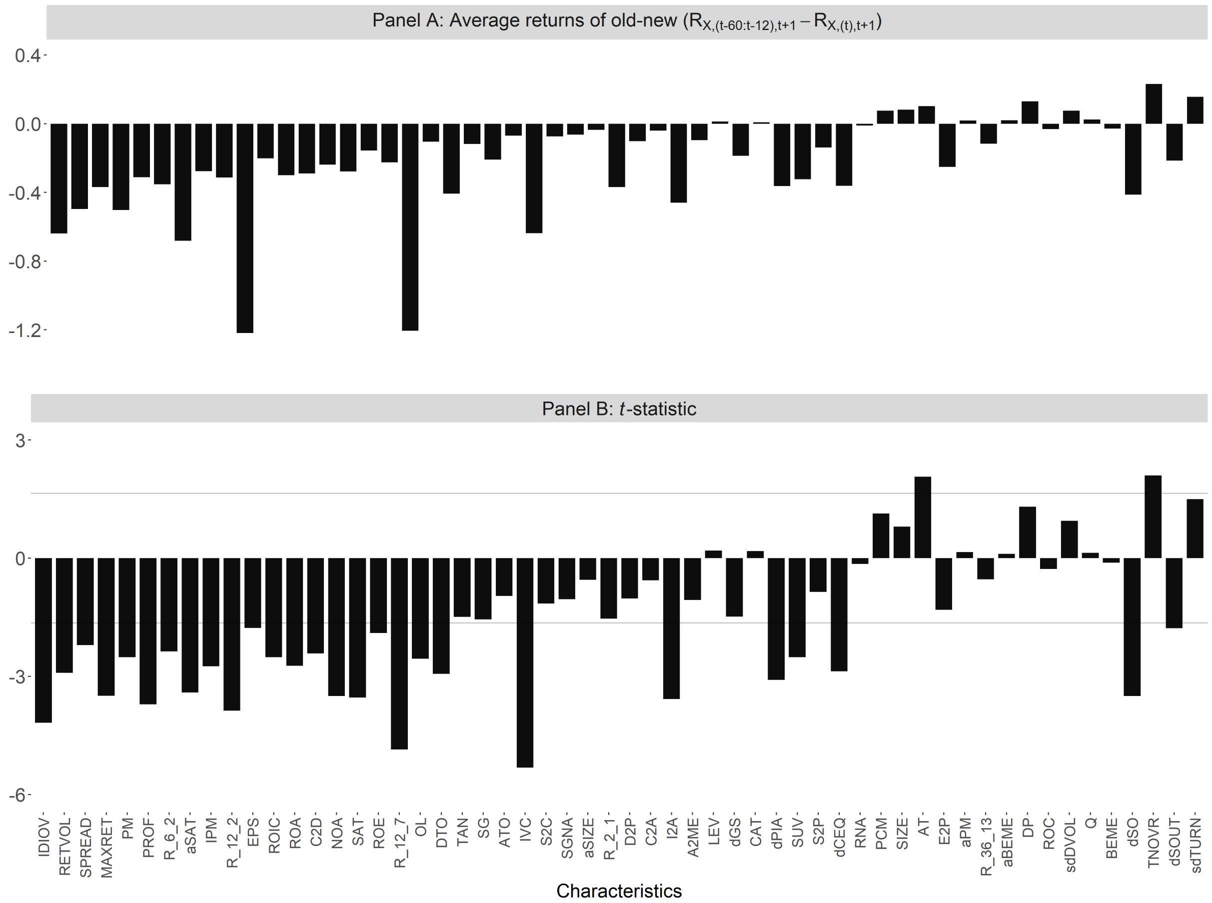

FIGURE 2.B.1: **Average returns of old-minus-new sorts**
This figure presents the average return (and t-statistic) of old-minus-new sorts, where the old sort is a single combination of five older sorts: $R_{X,(t-60:t-12),t+1} = 1/5(R_{X,(t-12),t+1} + R_{X,(t-24),t+1} + ... + R_{X,(t-60),t+1})$. To facilitate interpretation, the characteristics are sorted in the same order as the average returns from Figure 2.A.1.

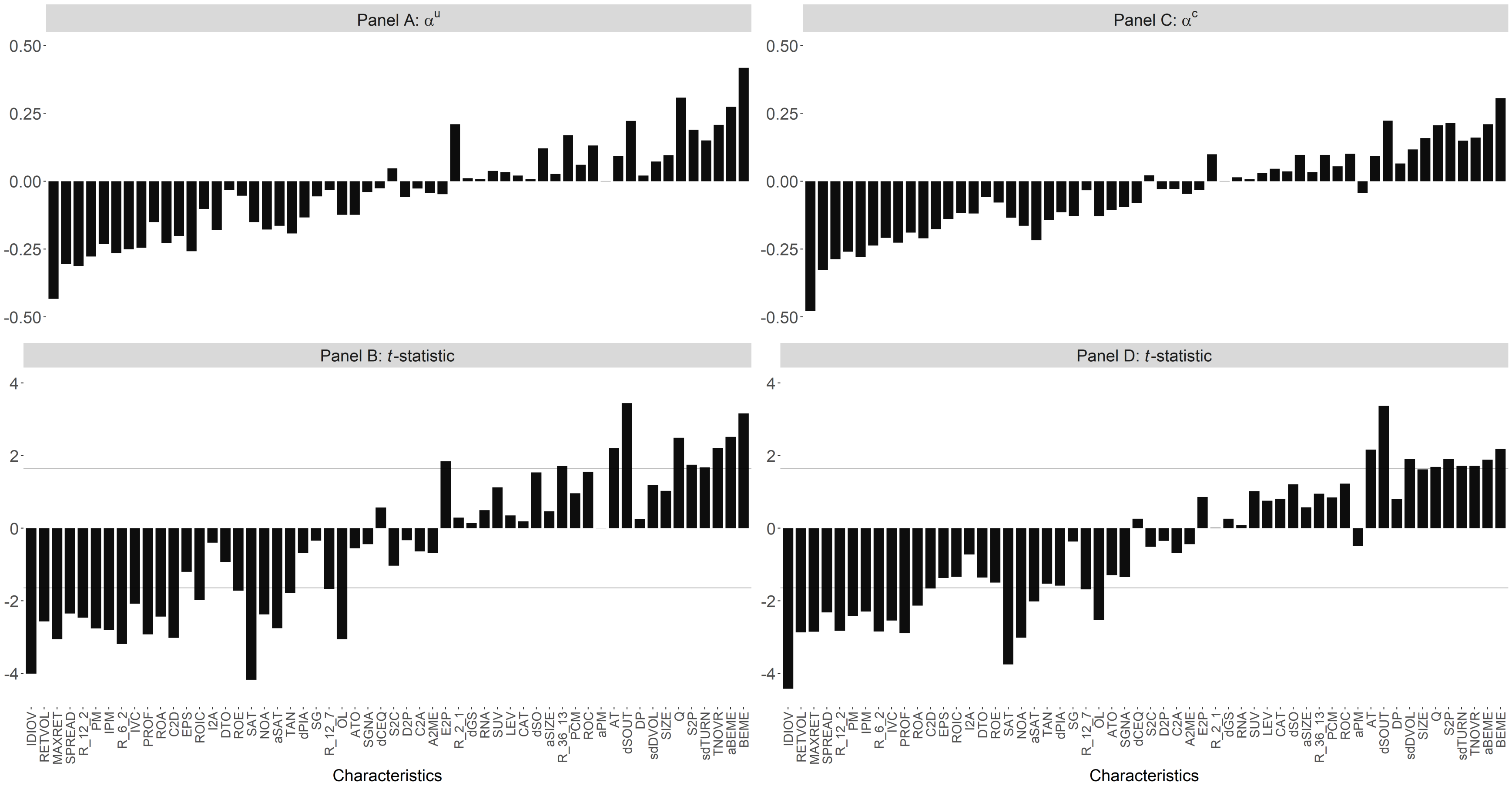

FIGURE 2.B.2: **Alphas between old and new sorts controlling for market exposure**
This figure presents the unconditional and conditional alpha of the older sorts (with return $R_{X,(t-60:t-12),t+1}$) relative to the newest sort (with return $R_{X,(t),t+1}$) when we control for exposure to the market as in the CAPM. To facilitate interpretation, the characteristics are sorted in the same order as the conditional alphas, α^c, from Figure 2.A.2.

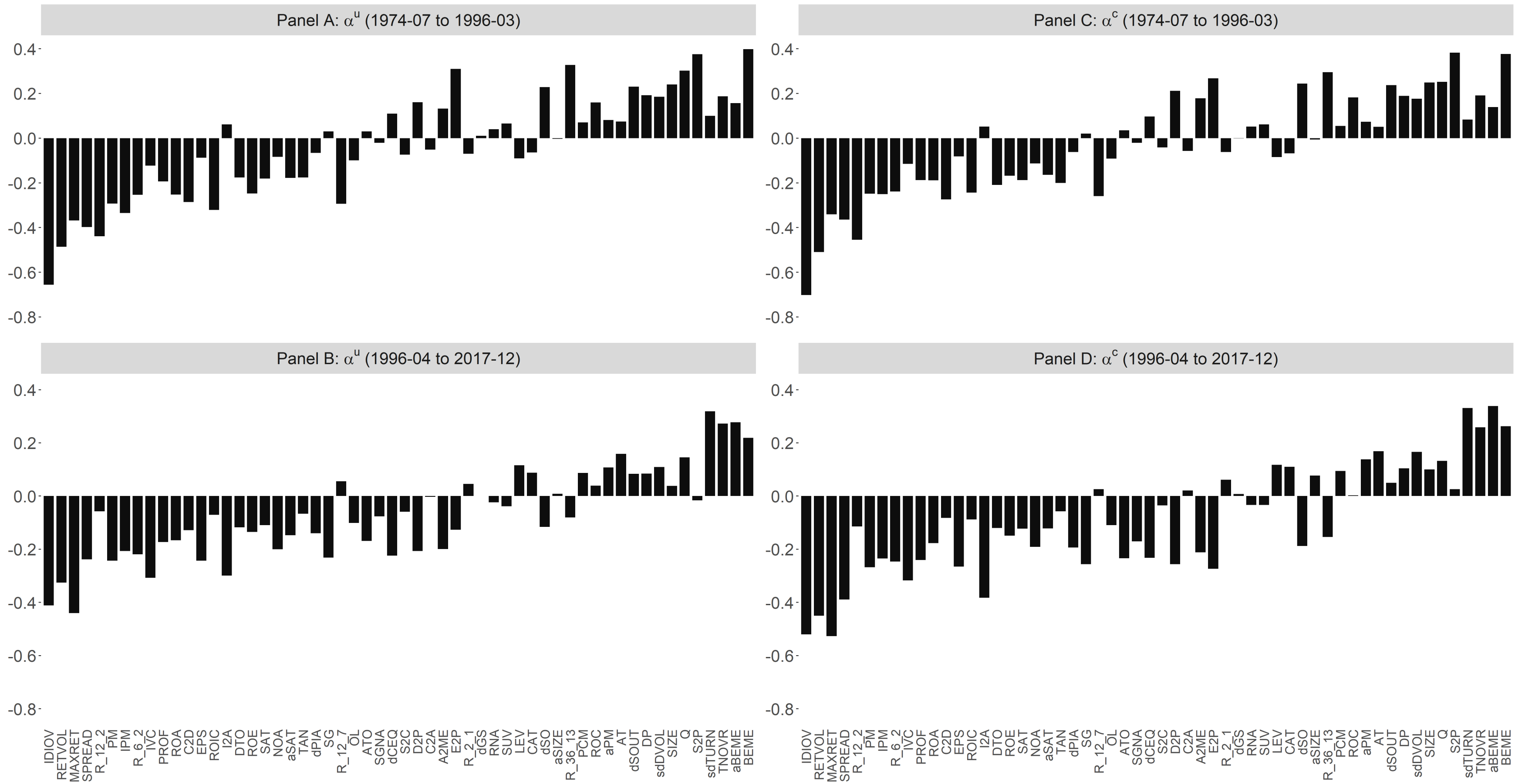

FIGURE 2.B.3: **Alphas between old and new sorts in subsamples**
This figure presents the unconditional and conditional alpha of the older sorts (with return $R_{X,(t-60:t-12),t+1}$) relative to the newest sort (with return $R_{X,(t),t+1}$) over two subsamples split around March 1996. To facilitate interpretation, the characteristics are sorted in the same order as the conditional alphas, α^c, from Figure 2.A.2.

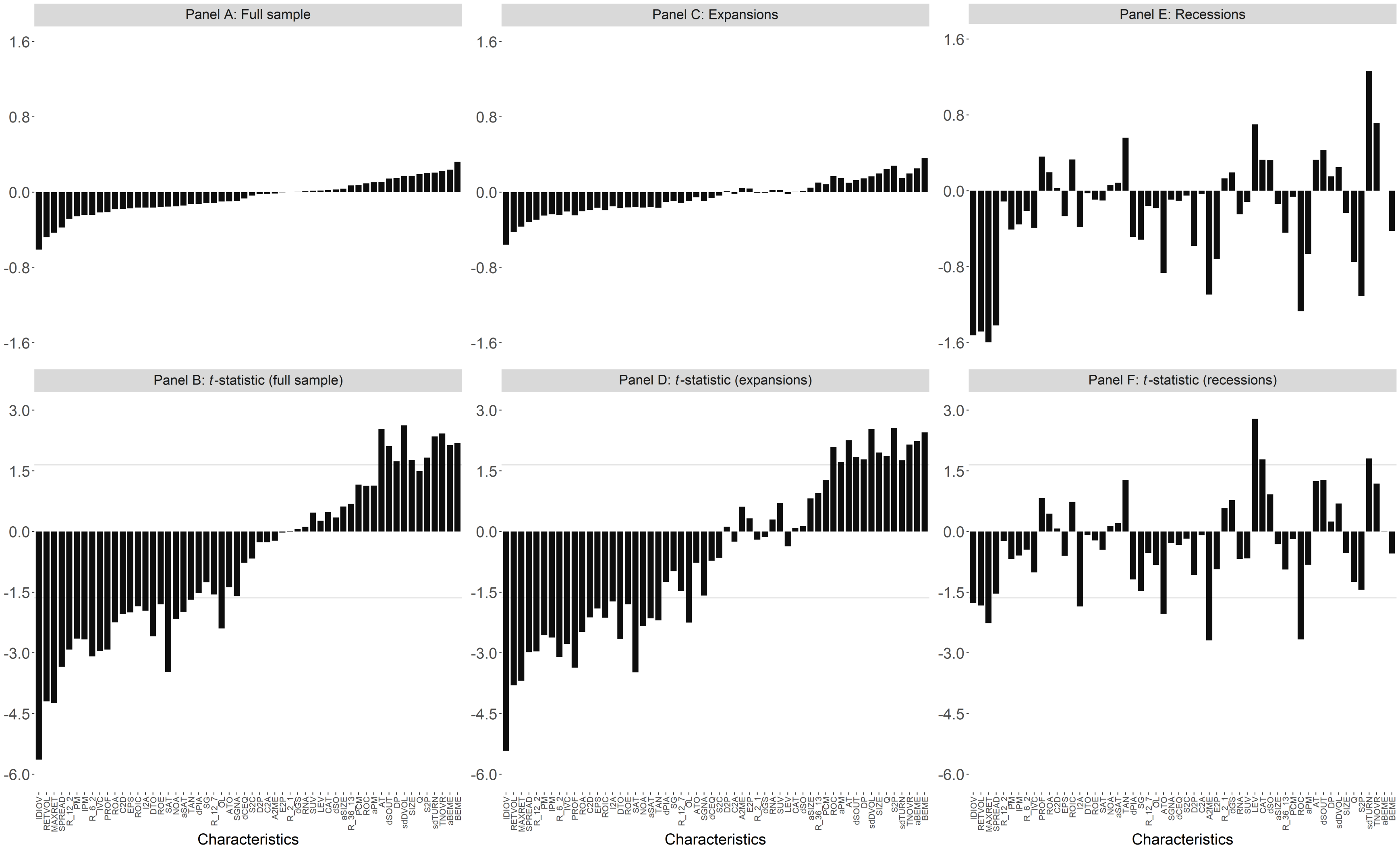

FIGURE 2.B.4: **Alphas between old and new sorts in expansions versus recessions**
This figure presents the conditional alpha of the older sorts (with return $R_{X,(t-60:t-12),t+1}$) relative to the newest sort (with return $R_{X,(t),t+1}$) in NBER expansions and recessions. To facilitate interpretation, the characteristics are sorted in the same order as the conditional alphas, α^c, from Figure 2.A.2.

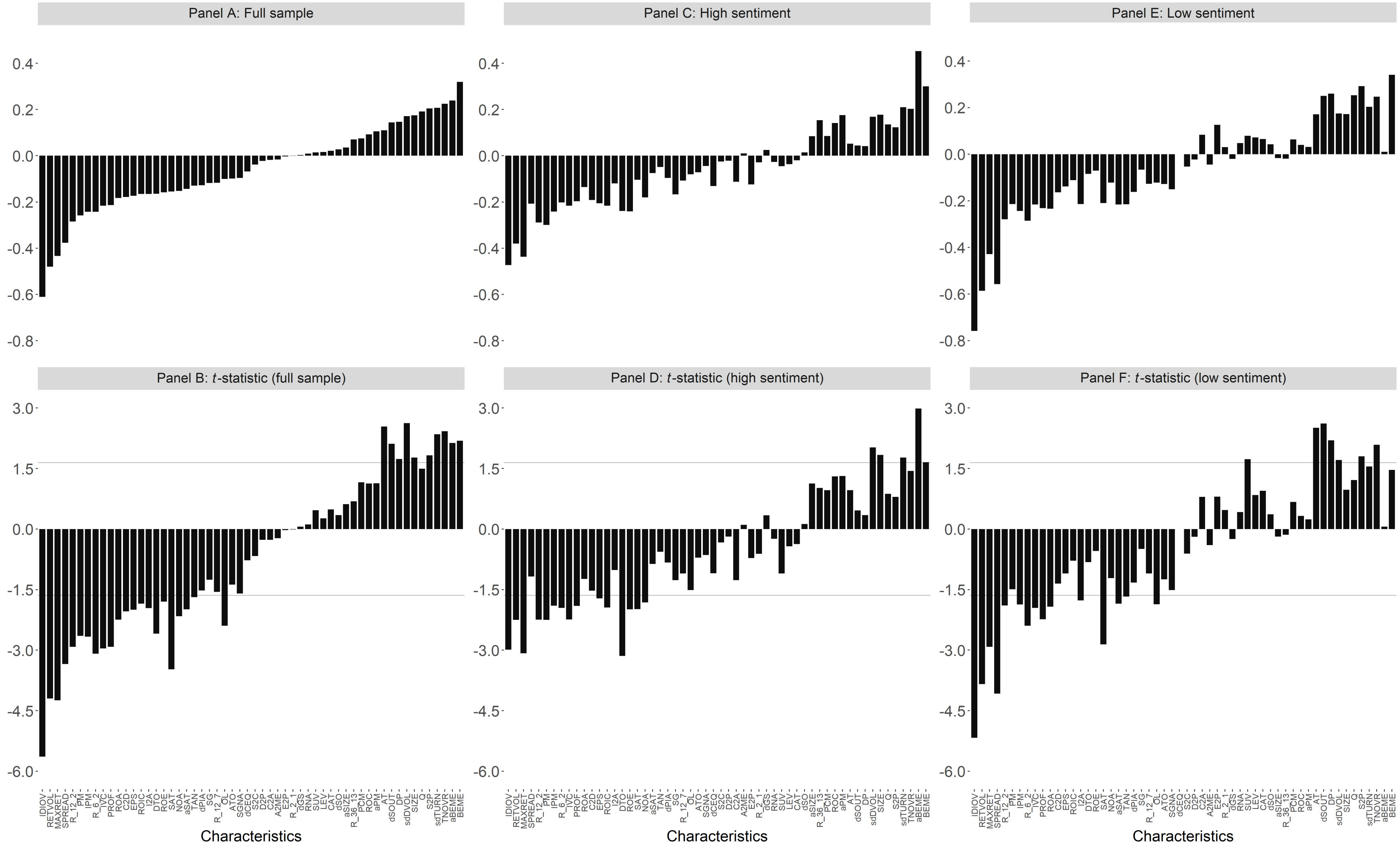

FIGURE 2.B.5: **Alphas between old and new sorts in high versus low sentiment regimes**
This figure presents the conditional alpha of the older sorts (with return $R_{X,(t-60:t-12),t+1}$) relative to the newest sort (with return $R_{X,(t),t+1}$) in high and low sentiment regimes. We follow Stambaugh et al., 2012 and define a high-sentiment month as one in which the sentiment index of Baker and Wurgler, 2006 is above its historical mean. To facilitate interpretation, the characteristics are sorted in the same order as the conditional alphas, α^c, from Figure 2.A.2.

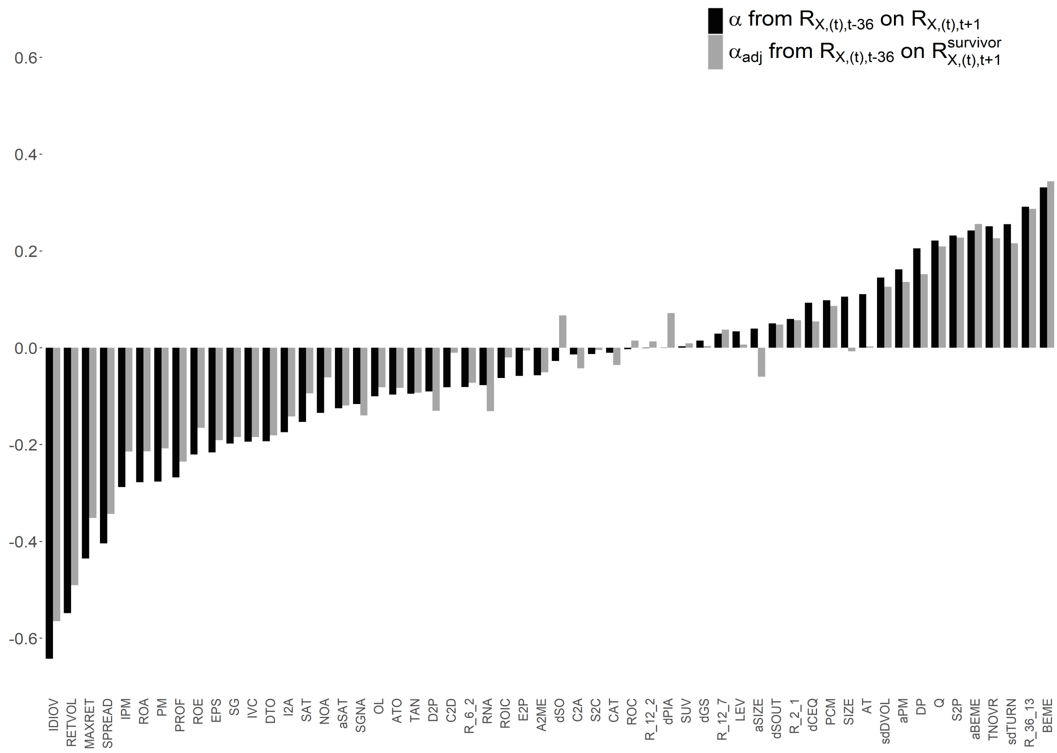

FIGURE 2.B.6: **Survivorship bias-adjusted alphas**
This figure presents the unconditional alpha of the old sort with return $R_{X,(t-36),t+1}$ relative to the newest sort with return $R_{X,(t),t+1}$ (estimated using the regression in Eq. (2.4.5)). It also reports results for a survivorship bias-adjusted version of the newest sort, for which case we exclude from the high and low portfolio at time t all stocks that were not in the CRSP file at $t-36$. In this way, we condition on firm survival on both sides of the regression.

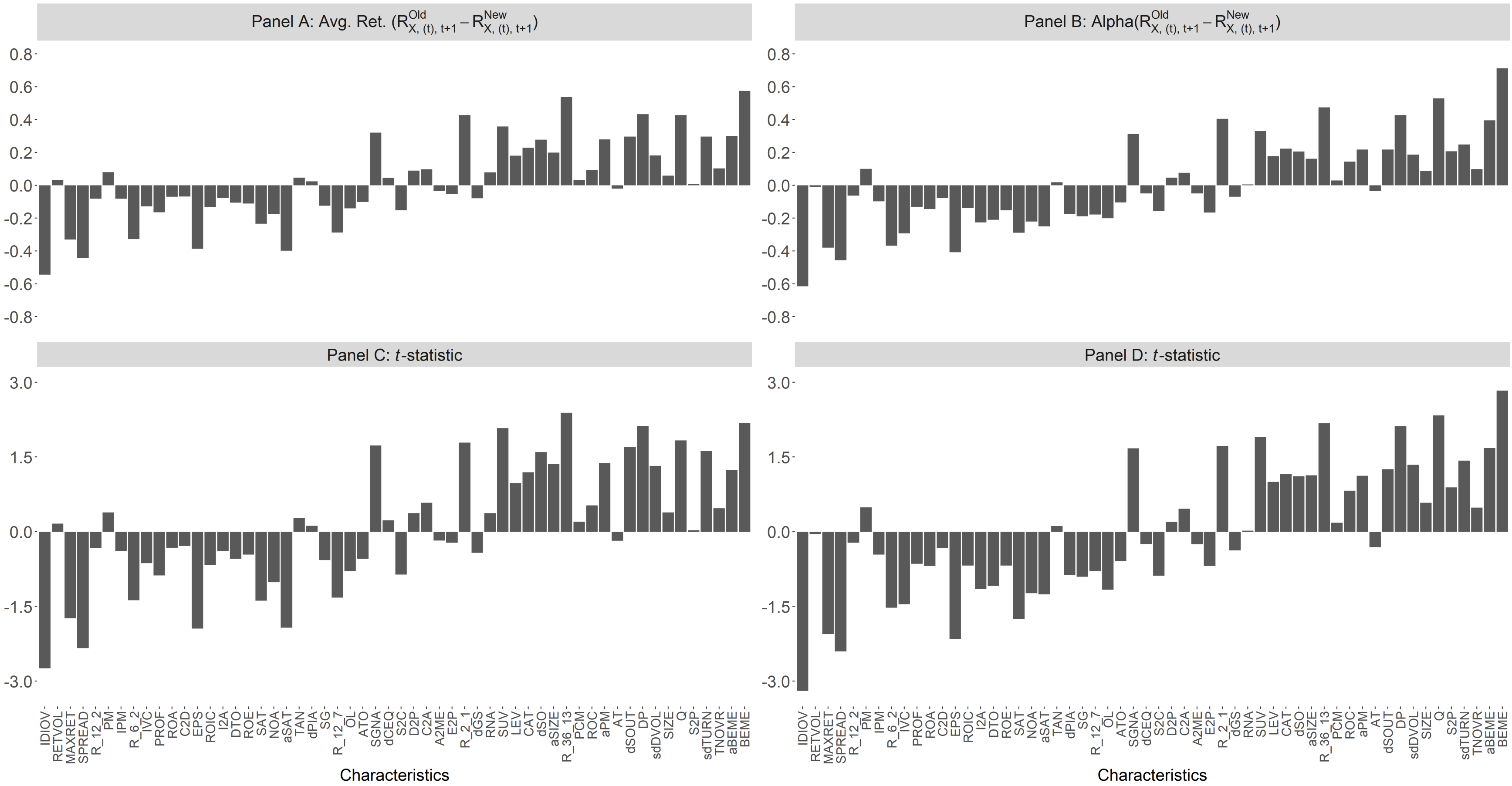

FIGURE 2.B.7: **The average return of old versus new stocks in characteristic-sorted portfolios**
In Panel A of this figure, we present the difference in average return between old ($R^{Old}_{X,(t),t+1}$) and new ($R^{New}_{X,(t),t+1}$) stocks, which together make up the newest long-short characteristic-sorted portfolio ($R_{X,(t),t+1}$). Panel B presents the intercept from a regression of old-minus-new returns on the newest sort. Panels B and D present White et al., 1980 t-statistics. The sample period runs from July 1974 to December 2017.

2.B.2 Tables

TABLE 2.B.1: **Characteristics**
This table lists the characteristics used in this paper. For each characteristic, we present the associated acronym, the original source and the definition of the characteristic.

Acronym	Author(s)	Definition
A2ME	Bhandari, 1988	Total assets (at) over market capitalization (prc x shrout)
AT	Gandhi and Lustig, 2015	Total assets (at)
ATO	Soliman, 2008	Net sales (sales) over lagged net operating assets. Net operating assets is the difference between operating assets and operating liabilities. Operating Assets is total assets (at) minus cash and short-term investments (che) minus investments and other advances (ivao). Operating Liabilities is total assets (at) minus debt in current liabilities (dlc) minus long-debt debt (dltt) minus minority interest (mib) minus preferred stock (pstk) minus common equity (ceq).
BEME (BM)	Davis et al., 2000b	Book equity to market equity. Book equity is shareholders' equity (seq), (if missing, common equity (ceq) plus preferred stock (pstk), if missing, total assets (at) minus total liabilities (lt)), plus deferred taxes and investment tax credit (txditc) minus preferred stock (pstrkrv), (if missing, liquidation value, (pstkl), if missing par value (pstk)). Market value of equity is shares outstanding (shrout) times price (prc).
aBEME	Asness et al., 2000c	$BEME$ minus average industry $BEME$. Industry level is defined as the Fama-French 48 industries.
C2A	Palazzo, 2012	Cash and short-term investments (che) to total assets (at).
C2D	Ou and Penman, 1989	Cashflow to debt. Cashflow is the sum of income and extraordinary items (ib) and depreciation and amortization (dp). And debt is to total liabilities (lt).
CAT	Haugen and Baker, 1996	Sales (sale) to lagged total assets (at).
D2P	Litzenberger and Ramaswamy, 1979	Debt to price. Debt is long-term debt (dltt) plus debt in current liabilities (dlc). Market capitalization is the product of shares outstanding (shrout) and price (prc).
dCEQ	Richardson et al., 2005	Annual % change in book value of equity (ceq).
dGS	Abarbanell and Bushee, 1997	% change in gross margin minus % change in sales (sale). Gross margin is the difference in sales (sale) and cost of goods sold (cogs).
dPIA	Lyandres et al., 2008	Change in property, plants and equipment (ppegt) and inventory (invt) over lagged total assets (at).
dSO	Fama and French, 2008	Log change in the product of shares outstanding (csho) and the adjustment factor (ajex).
dSOUT	Pontiff and Woodgate, 2008	Annual % change in shares outstanding (shrout).
DP	Litzenberger and Ramaswamy, 1979	Sum of monthly dividend over the last 12 months to last month's price (prc).

Continued

Acronym	Author(s)	Definition
DTO	Garfinkel, 2009	Daily volume (vol) to shares outstanding (shrout) minus the daily market turnover and detrended by the 180 trading day median. To address the double counting of volume for NASDAQ securities, we follow Anderson and Dyl, 2005 and scale down the volume of NASDAQ securities by 50% before and by 38% after 1997.
E2P	Basu, 1983	Income before extraordinary items (ib) to market capitalization (prc x shrout).
EPS	Basu, 1977	Income before extraordinary items (ib) to shares outstanding (shrout).
I2A (INV)	Cooper et al., 2008	Annual % change in total assets (at).
IDIOV	Ang et al., 2006	Standard deviation of the residuals from a regression of excess returns on the Fama and French, 1993b three-factor model.
IPM		Pre-tax income (pi) over sales (sale).
IVC	Thomas and Zhang, 2002	Annual change in inventories (invt) in the last two fiscal years over the average total assets (at) over the last two fiscal years.
LEV	Lewellen, 2015	long-term debt (dltt) plus current liabilities (dlc) over the sum of long term debt (dltt), debt in current liabilities (dlc) and stockholders equity (seq).
MAXRET	Bali et al., 2011	Maximum daily return in the previous month.
NOA	Hirshleifer et al., 2004	Operating assets minus operating liabilities to lagged total assets (at). Operating assets is total assets (at) minus cash and short term investments (che) minus investment and other advances (ivao). Operating liabilities is total assets (at) minus debt in current liabilities (dlc) minus long-term debt (dltt) minus minority interest (mib) minus preferred stock (pstk) minus common equity (ceq).
OL	Novy-Marx, 2011	Sum of cost of goods sold (cogs) and selling, general and administrative expense (xsga) over total assets (at).
PCM	Gorodnichenko and Weber, 2016	Net sales (sale) minus cost of goods sold (cogs) all scaled by net sales (sale).
PM	Soliman, 2008	Operating Income after depreciation (oiadp) to sales (sale).
aPM	Soliman, 2008	PM minus average industry PM. Industry level is defined as the Fama-French 48 industries.
PROF	Ball et al., 2015	Gross profitability (gp) over book equity as defined in $BEME$.
Q		Total assets (at) plus market value of equity (shrout x prc) minus common equity (ceq) minus deferred taxes (txdb) all scaled by total assets (at).
R_12_2	Fama and French, 1996b	Cumulative return from 12 months to 2 months ago.
R_12_7	Novy-Marx, 2012	Cumulative return from 12 months to 7 months ago.
R_2_1	Jegadeesh, 1990	Lagged one month return.
R_36_13	De Bondt and Thaler, 1985	Cumulative return from 36 months to 13 months ago.
R_6_2	Jegadeesh and Titman, 1993	Cumulative return from 6 months to 2 months ago.
RETVOL	Ang et al., 2006	Standard deviation of residuals from a regression of excess returns on a constant using one month of daily data. We require there to be at least 15 non-missing observations.

Continued

Acronym	Author(s)	Definition
RNA	Soliman, 2008	Operating income after depreciation (oiadp) scaled by lagged net operating assets. Net operating assets is operating assets minus operating liabilities. Operating assets is total assets (at) minus cash and short term investments (che) minus investment and other advances (ivao). Operating liabilities is total assets (at) minus debt in current liabilities (dlc) minus long-term debt (dltt) minus minority interest (mib) minus preferred stock (pstk) minus common equity (ceq).
ROA	Balakrishnan et al., 2010	Income before extraordinary items (ib) to lagged total assets (at).
ROC	Chandrashekar and Rao, 2009	Market value of equity (shrout x prc) plus long-term debt (dltt) minus total assets (at) all over cash and short-term investments (che).
ROE	Haugen and Baker, 1996	Income before extraordinary items (ib) to lagged book-value of equity.
ROIC	Brown and Rowe, 2007	Earnings before interest and taxes (ebit) less non-operating income (nopi) to the sum of common equity (ceq), total liabilities (lt), and cash and short-term investments (che).
S2C	Ou and Penman, 1989	Net sales (sale) to cash and short-term investments (che).
S2P	Lewellen, 2015	Net sales (sale) to market capitalization (shrout x prc).
SAT	Soliman, 2008	Sales (sale) to total assets (at).
aSAT	Soliman, 2008	SAT minus average industry SAT. Industry level is defined as the Fama-French 48 industries.
sdDVOL	Chordia et al., 2001	Standard deviation of residuals from a regression of daily volume (vol) on a constant. Use one month of daily data requiring at-least 15 non-missing observations.
sdTURN	Chordia et al., 2001	Standard deviation of residuals from a regression of daily turnover on a constant. Turnover is volume (vol) divided by shares outstanding (shrout). Use one month of daily data requiring at-least 15 non-missing observations.
SG	Lakonishok et al., 1994	% growth rate in sales (sale).
SGNA		Selling, general and administrative expenses (XSGA) to net sales (sale).
SIZE	Fama and French, 1992b	Price (prc) times shares outstanding (shrout) .
aSIZE	Asness et al., 2000c	$SIZE$ minus average industry $SIZE$. Industry level is defined as the Fama-French 48 industries.
SPREAD	Chung and Zhang, 2014	Average daily bid-ask spread in the previous month.
SUV	Garfinkel, 2009	Difference between actual volume and predicted volume. Predicted volume is from a regression of previous month's daily volume on a constant and the absolute values of positive and negative previous month's returns. Unexplained volume is standardized by the standard deviation of the residuals from the regression.
TAN	Hahn and Lee, 2009	Tangibility is defined as (0.715 x total receivables (rect) + 0.547 x inventories (invt) + 0.535 x property, plant and equipment (ppent) + cash and short-term investments (che)) / total assets (at).
TNOVR	Datar et al., 1998	Volume (vol) over shares outstanding (shrout).

TABLE 2.B.2: **Overview of results for alternative asset pricing models**
This table presents an overview of alphas from old versus new *sorts* and *stocks* relative to the models of Hou et al. (2015a, HXZ), Frazzini and Pedersen (2014b, BAB), Daniel et al. (2019, DMRS), Stambaugh and Yuan (2016, SY), and Daniel et al. (2017, DHS). In Panel A, we report the alpha of the first principal component of returns at all horizons $s = 0, 12, ..., 60$ after sorting (defined as $\lambda'_{(t),1} R_{X,(t-s),t+1}$, see Table 2.A.4). In Panel B, we report the alpha of a strategy that is long the first principal component of older sorts and short the first principal component of the newest sort (analogous to Panel F of Table 2.A.4). In Panel C, we report the alpha of an old-minus-new sort strategy that is long (short) an equal-weighted portfolio of the conditionally hedged returns (defined as in Eq. (2.4.6)) of high (low) market beta characteristics (analogous to Table 2.A.7). In Panel D, we report the alpha of a strategy that is long (short) an equal-weighted portfolio of the old-minus-new stock return differences among high (low) market beta characteristics (analogous to Table 2.A.9). t-statistic are based on White et al., 1980 heteroskedasticity consistent standard errors. The sample period runs from July 1974 to the end of the sample over which the factors are available. We thank the authors for sharing the factor data.

	HXZ		BAB		DMRS		SY		DHS	
	α	t	α	t	α	t	α	t	α	t
Panel A: First principal component of old and new sorts										
$\lambda'_{(t),1} R_{X,(t-s),t+1}$										
0	0.21	0.45	-0.32	-0.43	-1.14	-2.89	0.35	0.73	1.31	2.16
12	0.82	1.76	0.59	0.79	-0.14	-0.44	0.51	1.14	1.75	3.27
24	1.23	2.60	1.01	1.37	0.40	1.29	0.83	1.89	1.99	3.92
36	1.44	3.27	1.08	1.55	0.50	1.60	1.06	2.49	2.01	3.96
48	1.56	3.79	1.19	1.77	0.58	1.97	1.10	2.78	1.94	4.03
60	1.40	3.46	1.15	1.83	0.58	1.91	1.07	2.70	1.80	3.69
Panel B: Old-minus-new sorts										
$\lambda'_{(t),1}(R_{X,(t-s),t+1} - R_{X,(t),t+1})$										
12	0.61	2.36	0.91	3.16	1.00	4.27	0.16	0.68	0.45	1.50
24	1.02	3.36	1.34	3.87	1.54	5.49	0.48	1.69	0.68	1.90
36	1.23	3.55	1.40	3.56	1.64	5.16	0.71	2.00	0.70	1.70
48	1.34	3.78	1.51	3.87	1.72	5.30	0.75	2.03	0.64	1.55
60	1.19	3.39	1.48	3.86	1.72	5.11	0.72	1.87	0.50	1.23
Panel C: Old-versus-new sort strategies (High-minus-Low market beta)										
$R_{X,(t),t+1}$	0.08	0.45	0.10	0.49	-0.06	-0.43	0.22	1.19	0.47	2.34
$R_{X,(t-36),t+1}$	0.45	3.30	0.45	2.46	0.26	2.24	0.43	3.18	0.63	4.05
$R_{X,(t-36),t+1} - \frac{X_{H-L,(t-s),t}}{X_{H-L,(t),t}} \times R_{X,(t),t+1}$	0.33	4.01	0.34	3.76	0.30	4.25	0.28	3.08	0.31	3.53
$R_{X,(t-36),t+1} - \beta^u \times R_{X,(t),t+1}$	0.35	4.26	0.33	3.52	0.26	3.82	0.31	3.47	0.31	3.73
$R_{X,(t-36),t+1} - \beta^c_t \times R_{X,(t),t+1}$	0.28	3.43	0.31	3.56	0.31	4.51	0.25	2.74	0.30	3.56
Panel D: Old-minus-new stocks (High-minus-Low market beta)										
$R^{New}_{X,t,t+1}$	-0.12	-0.63	-0.07	-0.33	-0.24	-1.66	0.03	0.17	0.25	1.17
$R^{Old}_{X,t,t+1}$	0.33	1.81	0.31	1.37	0.11	0.69	0.42	2.25	0.71	3.40
$R^{Old}_{X,t,t+1} - R^{New}_{X,t,t+1}$	0.45	3.74	0.38	2.98	0.35	3.39	0.39	3.03	0.46	3.72

TABLE 2.B.3: **Alphas of alternative principal components**
This table is identical to Table 2.A.4 in the main text, except that the principal component loadings are extracted using the method of Lettau and Pelger, 2020a; Lettau and Pelger, 2020b.

$\lambda'_{(t),z}R_{X,(t-s),t+1}$	PC1 ($z=1$)		PC2 ($z=2$)		PC3 ($z=3$)	
s	α	t-stat	α	t-stat	α	t-stat
			Panel A: $3PC_{(t),t+1}$			
12	0.33	1.74	-0.89	-3.66	0.02	0.07
24	0.70	3.04	-1.00	-3.32	-0.12	-0.45
36	0.85	3.25	-0.88	-2.44	-0.25	-0.98
48	1.02	3.82	-0.72	-1.94	0.04	0.16
60	1.07	3.82	-0.75	-1.97	0.28	1.20
			Panel B: CAPM			
0	-2.51	-3.53	0.45	0.63	-2.46	-6.50
12	-1.17	-1.81	1.28	2.10	-0.55	-1.78
24	-0.50	-0.81	1.10	1.93	-0.37	-1.39
36	-0.35	-0.58	0.92	1.66	-0.71	-2.68
48	-0.23	-0.40	1.06	1.96	-0.47	-1.72
60	-0.18	-0.33	0.97	1.76	-0.11	-0.45
			Panel C: FF3M			
0	-2.29	-4.70	-1.59	-3.24	-2.03	-6.59
12	-0.89	-2.11	-0.66	-2.30	-0.45	-1.62
24	-0.14	-0.34	-0.57	-1.69	-0.18	-0.75
36	0.07	0.17	-0.60	-1.66	-0.48	-1.88
48	0.18	0.49	-0.38	-1.02	-0.19	-0.80
60	0.17	0.46	-0.41	-1.02	0.13	0.60
			Panel D: FF5M			
0	-0.43	-1.12	-1.09	-1.88	-1.63	-4.90
12	0.58	1.76	-0.76	-2.75	-0.38	-1.34
24	1.13	3.30	-0.75	-2.17	-0.11	-0.44
36	1.30	3.87	-0.73	-1.86	-0.40	-1.40
48	1.36	4.35	-0.54	-1.32	0.04	0.15
60	1.31	4.12	-0.67	-1.58	0.36	1.56
			Panel E: FF5M+MOM			
0	0.01	0.02	0.06	0.17	-0.86	-4.21
12	0.63	1.85	-0.65	-2.30	-0.14	-0.48
24	1.10	3.16	-0.83	-2.41	-0.09	-0.35
36	1.26	3.66	-0.81	-2.08	-0.37	-1.25
48	1.37	4.26	-0.62	-1.52	0.13	0.52
60	1.33	4.03	-0.78	-1.84	0.34	1.41

Continued

Panel F: Old-minus-new sorts								
$\lambda'_{(t),1}(R_{X,(t-s),t+1} - R_{X,(t),t+1})$	CAPM		FF3M		FF5M		FF5M+MOM	
	α	t-stat	α	t-stat	α	t-stat	α	t-stat
12	1.34	5.39	1.40	6.01	1.01	3.92	0.62	2.58
24	2.01	6.66	2.16	7.54	1.56	4.91	1.09	3.80
36	2.16	6.30	2.36	7.13	1.73	4.89	1.25	3.72
48	2.28	6.47	2.47	7.42	1.79	5.10	1.36	3.87
60	2.33	6.46	2.46	7.26	1.73	5.05	1.32	3.75

TABLE 2.B.4: **The relative performance of new versus old stocks (alternative weighting)** This table is similar to Table 2.A.9, but weights the characteristics-sorted portfolios (as well as its new and old stock components) using the loadings of the first principal component (see Section 2.5.1).

	Avg. Ret.	t-stat
$R_{(t)t+1}$	0.38	0.69
$R^{New}_{X,t,t+1}$	0.67	1.33
$R^{Old}_{X,t,t+1}$	0.03	0.05
$R^{Old}_{X,t,t+1}$ - $R^{New}_{X,t,t+1}$	-0.64	-2.61
	α	t-stat
	CAPM	
$R_{(t)t+1}$	1.48	3.26
$R^{New}_{X,t,t+1}$	1.64	3.95
$R^{Old}_{X,t,t+1}$	1.20	2.29
$R^{Old}_{X,t,t+1}$ - $R^{New}_{X,t,t+1}$	-0.44	-1.81
	FF3M	
$R_{(t)t+1}$	1.39	4.46
$R^{New}_{X,t,t+1}$	1.60	5.15
$R^{Old}_{X,t,t+1}$	1.03	3.04
$R^{Old}_{X,t,t+1}$ - $R^{New}_{X,t,t+1}$	-0.57	-2.70
	FF5M	
$R_{(t)t+1}$	0.21	0.85
$R^{New}_{X,t,t+1}$	0.49	1.93
$R^{Old}_{X,t,t+1}$	-0.16	-0.61
$R^{Old}_{X,t,t+1}$ - $R^{New}_{X,t,t+1}$	-0.65	-3.00
	FF5M+MOM	
$R_{(t)t+1}$	-0.06	-0.26
$R^{New}_{X,t,t+1}$	0.22	0.86
$R^{Old}_{X,t,t+1}$	-0.43	-1.59
$R^{Old}_{X,t,t+1}$ - $R^{New}_{X,t,t+1}$	-0.65	-2.88

TABLE 2.B.5: **Descriptives for new and old stock portfolios across market beta groups (alternative characteristics)**
This table is similar to Table 2.A.8 but reports results for momentum (R_12_2), short-term reversal (R_2_1), idiosyncratic volatility ($IDIOV(\%)$), and turnover (DTO).

	Market beta			
	Low	Mid	High	
	Momentum (R_12_2)			
New	0.04	0.04	0.08	
Old	0.04	0.05	0.07	
Single sort on R_12_2				1.27
	Short-term reversal (R_2_1)			
New	0.01	0.00	0.02	
Old	0.01	0.01	0.02	
Single sort on R_2_1				0.35
	Idiosyncratic volatility ($IDIOV(\%)$)			
New	0.37	-0.15	-0.13	
Old	0.20	-0.35	-0.20	
Single sort on $IDIOV$				3.74
	Turnover (DTO)			
New	0.10	0.35	-0.03	
Old	0.10	0.33	-0.05	
Single sort on DTO				6.26

Chapter 3

Machine learning and return predictability across firms, time and portfolios

I am especially grateful to Martijn Boons, Andrea Tamoni and Melissa Prado for their guidance and comments. I thank Giorgio Ottonello, Fernando Anjos, Irem Demirci, Miguel Ferreira, Andre Castro Silva and seminar participants at Nova SBE, VU Amsterdam, Virginia Tech, HEC Paris, Stockholm School of Economics, Frankfurt School of Finance and Management, and Indiana University for helpful comments. This work was funded by FCT Portugal. The author has no conflicts of interest to disclose and all remaining errors are mine.

Introduction

In this paper, I study how incorporating economic priors in the specification of a machine learning model improves the resulting model's predictive accuracy with respect to predicting equity returns. I do this by comparing the forecasting accuracy of a neural network model that imposes several restrictions reflecting recent findings in the financial economics literature to an alternative neural network model that is much simpler in its structure. The simple model I benchmark against is the best performing neural network model, as specified in Gu et al., 2020b. I study the predictability of returns to individual US equities, 56 long-short characteristic sorted-portfolios, and the value-weighted market portfolio over multiple horizons.

This paper shows that the stylized facts from the financial economics literature have an integral role to play in guiding the application of machine learning to finance. This conclusion stems from the number and nature of improvements one observes when comparing my proposed model (the economically restricted model) to the benchmark model. First, in predicting individual equity returns, I find that forecasts from the benchmark model explain about 0.58% (out-of-sample R^2) of the variation in next month's returns, whereas forecasts from the economically restricted model can explain about 0.99%—close to a two-fold increase.

Second, I show that investors who employ return forecasts from the economically restricted neural network model enjoy large and robust economic gains. For instance, using individual equity return predictions' for the following month, a long-short portfolio that buys (sells) the 10% highest (lowest) expected return stocks has an annualized average return of 15.82%. This estimate corresponds to a Sharpe ratio of 0.78, a certainty equivalent of 11.73, and a Fama and French, 2018 6 factor-alpha of 11.05%, all annualized. This strategy only trades the 500 largest market-capitalized firms each month, thus generating the gains from the most liquid stocks in the cross-section. The annualized average return of the strategy falls to about 1% when one uses forecasts from the benchmark model.

This result is particularly interesting because it shows that the economically restricted model's improved predictive accuracy is not concentrated among small stocks.

All results pertain to the out-of-sample period; January 1995 to December 2018.

Given the small annualized return the strategy produces when one conditions on forecasts from the benchmark model, it is evident that the benchmark model extracts a non-trivial fraction of its predictive accuracy from small and difficult to arbitrage stocks (see Avramov et al., 2020).

Third, I find that to produce forecasts that robustly generalize beyond individual equities, restrictions implied by findings from the financial economics literature are crucial. When predicting returns to the value-weighted market portfolio, forecasts from the economically restricted model predict time-series variation in monthly returns as far out as three years in the future. Additionally, when predicting returns to 56 long-short characteristic sorted portfolios, the aggregate forecasts predict time-series variation in next month's returns for 53 of the 56 portfolios. On the other hand, forecasts from the benchmark model fail to robustly predict time-series variation in returns to both the aggregate market and long-short characteristic sorted portfolios.

A natural question one would ask at this point is, "Along which dimension does the economically restricted model help improve stock return forecasts?" This question has so far received little attention in this emerging literature. I shed light on this by decomposing stock return forecasts into two components; a fraction explaining variations in a level factor (the equally-weighted market return) and a fraction explaining variations in returns over and above the cross-sectional mean (relative stock returns).

I find that the improvement in forecasting accuracy primarily comes from predicting better the relative stock return component. For this component, forecasts from the economically restricted model explain about 0.55% of the variations in next month's returns, while the benchmark model only explains 0.16%-a more than three-fold improvement. The models are comparable in their ability to explain variations in the level component. Forecasts from the economically restricted model explain about 0.43 % of the variations in the level factor, while forecasts from the benchmark model explain 0.41 %.

Most papers in the literature study cross-sectional return predictability over the next month or at most over the following year (see Kozak, 2019, Gu et al., 2020b, and

I decompose returns ($r_{i,t}$) into a level factor; captured by the cross-sectional average return, $(N_t^{-1})\sum_{i\in t} r_{i,t}$ and a slope factor; captured by the cross-sectional dispersion around the mean, $r_{i,t}^{RR} = r_{i,t} - (N_t^{-1})\sum_{i\in t} r_{i,t}$.

Freyberger et al., 2020b). Papers that study returns further into the future, conditional on what we know today, tend to study exclusively the time-series properties of returns with annual holding periods. Therefore, this paper is among the first to study monthly stock return predictability as far as ten years into the future and documents new evidence on the cross-sectional and time-series properties of conditional expected stock returns across horizons. The two main results are that 1) stock return predictability decreases over the horizon and 2) the nature of stock predictability in the short run is very different from the long-run.

First, stock returns are much more predictable in the short-run than in the long-run. Forecasts from the economically restricted neural network model can explain about 0.99% (out-of-sample R^2) of the variations in next month's stock returns. However, this estimate falls to about 0.28% when predicting returns five years into the future and about 0.13% when predicting returns ten years into the future.

Second, the nature of short-run stock return predictability is very different from long-run stock return predictability. Accounting for the inherent factor structure underpinning stock returns, I find that a large fraction of the observed stock return predictability across horizons comes from predicting variations in the equally-weighted market return or level component in the pool of stocks. When predicting next month's return, about 43% of the variation explained (0.43% out of 0.99%) comes from explaining variations in the level component. For forecasts that pertain to months that are at least one year in the future, over 95% of the variation explained comes from explaining variations in this component.

To summarize, I find that the forecasts' ability to explain cross-sectional variation in returns is only present in the short-run. While in the long-run, stock return predictability entirely comes from predicting variations in the level component in the pool of stocks.

The empirical asset pricing literature has shown that firm characteristics are correlated with subsequent stock returns, but evidence on how well these characteristics or combinations thereof proxy for conditional expected returns is scarce (see for example, Basu, 1977, Jegadeesh and Titman, 1993, and Sloan, 1996). Measuring relative stock returns as the stock return in excess of the cross-sectional mean, I find that a one percent relative return forecast on average predicts a 0.97 percentage point increase in next month's relative stock return. Similar to the conclusions drawn from the out-of-sample R^2 analysis, I find that this estimate decreases as the horizon increases. In predicting the monthly relative stock return realized one-year in the future, this estimate falls to 0.79 and further down to 0.24 when predicting monthly relative returns two years in the future.

I find similarly robust estimates for the value-weighted market portfolio and long-short characteristic sorted portfolio returns. On average, a one percent demeaned market forecast predicts a 1.80 percentage point increase in market return. The estimates are statistically significant for monthly forecasts for up to three years in the future. A one percent demeaned long-short portfolio forecast, on average, predicts about a 2.00 percentage point increase in long-short portfolio returns. The estimates are statistically significant for forecasts up to a year into the future.

The findings in this paper are both important and interesting for the following reasons. First, conditional expected returns for multiple future dates are analogous to a term structure of implied discount rates (cost of capital) conditional on an information set observed today. These discount rates are of particular importance to firms when evaluating investment opportunities with cash-flows maturing over multiple future dates. The proposed model in this paper is one way of using project-specific characteristics observed today as a basis for coming up with consistent (implied) discount rates to help evaluate such investment opportunities.

Second, the return predictability literature is only beginning to tackle the question of whether or not long-short characteristic sorted portfolio returns are predictable over time (see 2018 and 2020). Reporting the Sharpe ratio of such portfolios only tells us that the cross-sectional variation in a characteristic generates an unconditional spread in returns but not whether the time-series variation in the returns to such a portfolio is predictable. I find that the economically restricted neural network forecast can predict time-series variation in next month's return for over 90% of the long-short portfolios I study. This result is important because the returns to factor portfolios can be low or negative for prolonged periods (see Israel et al., 2020). Having access to conditional expected return estimates for these portfolios should aid investors in making their portfolio timing decisions. More generally, improving short-and long-run expected return estimates is essential because these estimates serve as fundamental inputs in tactical and strategic portfolio decisions, respectively.

Finally, Martin and Nagel, 2019 consider a world where agents have to condition on thousands of potentially relevant variables to forecast returns. If agents (investors) are uncertain about how exactly cash-flows relate to these predictors, then a factor zoo will naturally emerge. A world not too dissimilar from our own. Bryzgalova et al., 2019 show that complex non-linearities exist between firm characteristics and stock returns. Taking these two facts together, agents will need learning algorithms that can efficiently handle large dimensional predictors while simultaneously learning the non-linearities that exist therein. Gu et al., 2020b show that neural networks are the best learning algorithm for this problem. This paper shows that incorporating economic restrictions in the neural network design robustly enhances their predictive ability.

Literature

This work is related to the emerging literature in economics and finance using machine learning methods to answer economic questions that are fundamentally predictive. 2016 show that deep neural networks are strong predictors of mortgage repayment, delinquency, and foreclosures. Butaru et al., 2016 use regression trees to predict the probability of consumer credit card delinquencies and defaults. 2020b use the adaptive group LASSO to study which subset of 62 characteristics provides incremental information about the cross-section of expected returns. The spline methodology the authors use cannot easily accommodate higher-order interactions between covariates (characteristics). However, deep neural networks, the learning algorithm used in this paper, easily approximates higher-order non-linear interactions between covariates (see Goodfellow et al., 2016). 2019 estimate the stochastic discount factor using neural networks and find that a model that bakes in economic restrictions outperforms all other benchmarks in an out-of-sample setting. Like these authors, I show that designing neural network models using financial economic priors does generate robust forecasts, although the proposed models different. The economically restricted neural network model I propose is similar to the autoencoder model of Gu et al., 2020a. Gu et al., 2020a primarily study the asset pricing implications of their model for next month returns, I study return predictability across time and portfolios.

This work primarily extends the literature on stock return predictability. I show that a neural network architecture design that imposes restrictions reflecting findings in the financial economics literature improves stock return forecasts out-of-sample. Lewellen, 2015 studies expected returns across stocks as a linear function of firm-level characteristics and finds that the forecasts generated by the linear model explain some variation in returns. The proposed framework in this paper allows for high-dimensional non-linear interactions between characteristics and also imposes a Lasso penalty to remove non-essential return predictors in the information set I condition on. 2020b show that allowing for non-linear interactions between characteristics help improve the forecasting accuracy of ML models. Specifically, the authors show that firm-level characteristics can be combined with macroeconomic variables using different machine learning methods to predict returns better. I show that the information set we condition on is not only informative of return realizations for the next month but extends much further out into the future. This finding is important because Van Binsbergen and Opp, 2019 argue only characteristics that predict persistently generate substantial economic distortions. Finally, I show that relative stock return predictability is short-lived. Specifically, machine learning forecasts for return realizations beyond one year into the future are no better than a zero forecast in discriminating between high and low expected return firms; a result that suggests that longer-run discount rates converge across firms (see 2019).

The results in this paper also contribute to the literature that studies aggregate market return predictability. Cochrane, 2008 studies market return predictability and provides evidence that the dividend-yield predicts time-series variation in the equity risk premium. Goyal and Welch, 2008b study market return predictability in the time-series using macroeconomic variables and show that the historical average market return is a challenging benchmark to beat. I show that a neural network model that adheres to economic theory robustly out-performs the historical equity return in predicting time-series variation in monthly market returns as far as three years into the future. Engelberg et al., 2019 aggregate 140 individual firm-characteristics, including the dividend-yield, and ask how many of these aggregates can predict market returns. The authors find that the aggregated cross-sectional variables that appear to be statistically significant in predicting market returns when examined in isolation are no longer significant in a multiple testing framework. I find that we can distill the predictive information in individual firm-characteristics into a single measure of expected stock return using machine learning methods. Aggregating this single variable into a market forecast predicts time-series variation in market returns as far as three years (statistically significant at the 5% level) into the future.

My results also contribute to the stream of literature that studies time-series predictability of returns to characteristic sorted portfolios. Cohen et al., 2003b predict returns to the value portfolio. Cooper et al., 2004 and Daniel and Moskowitz, 2016 both study time-series predictability of the returns to the momentum portfolio. Similar to Haddad et al., 2020, my framework allows me to study a much larger cross-section of long-short portfolios while entertaining a large dimensional conditioning information set. Specifically, I contribute to the literature by showing that long-short portfolio forecasts formed from stock return forecasts generated by a neural network model can predict time-series variation in 53 of 56 long-short portfolios (32 of 56 are statistically significant at the 5% level). I also show that imposing economic restrictions on the corresponding machine learning model is essential in producing the forecasts that generalize to the cross-section of long-short characteristic sorted portfolios.

3.1 Empirical Framework and Data

In this section, I detail the assumptions underlying the empirical exercise in this paper.

3.1.1 Factor Model

I assume that stock returns are conditionally priced by a linear combination of J factors, $F_{t+1} = [f_{1,t+1}, f_{2,t+1}, ..., f_{J,t+1}]$.

Assumption 1. *A conditional factor model holds such that:*

$$r_{i,t+1} = \beta'_{i,t} F_{t+1} + \varepsilon_{i,t+1} \tag{3.1.1}$$

where $r_{i,t+1}$ is the stock return of firm i at time $t+1$, $\beta_{i,t}$ is a $J \times 1$ dimensional vector of conditional factor loadings and $\varepsilon_{i,t+1}$ is an independent identically distributed normal random process, $\mathcal{N}(0, \sigma_{i,\varepsilon})$.

My interest in this paper is to learn a set of expected return functions, $\mathbb{E}_{t-h+1}[r_{i,t+1}]$, where $h \in H = \{1, 2, 3, 13, 37, 61, 91, 121\}$, conditional on some information set, I_{t-h+1}. Supposing this is month t, I predict returns for the following month, $t+1$, by conditioning on the information set I_t and generate return forecasts with the function, $\mathbb{E}_t[r_{i,t+1}]$. To predict returns one year from next month, *1+13*, I condition on the information set observed today, I_t, and generate return forecasts with the function, $\mathbb{E}_t[r_{i,t+13}] = \mathbb{E}_{t-12}[r_{i,t+1}]$.

3.1.2 Economically restricted model

Guided by economic theory, I introduce the following assumptions to pin down the structural nature of the expectation functions.

Assumption 2. *Expected stock returns are linear in conditional betas and conditional price of risks:*

$$\mathbb{E}_{t-h+1}[\beta'_{i,t}]\mathbb{E}_{t-h+1}[F_{t+1}] \approx b^*_h(\cdot)' * f^*_h(\cdot) \tag{3.1.2}$$

where $b^*_h(\cdot)$ is a function that approximates the time $t+1$ expected conditional risk exposures of firm i and $f^*_h(\cdot)$ is a function that approximates the time $t+1$ expected conditional price of risk, all conditional on the information set, I_{t-h+1}. The crucial assumption here is that expected returns is the sum of the product of conditional risk loadings (betas) and the corresponding conditional price of risk. This restriction is standard in the literature and follows from assuming that the SDF is linear or approximately linear in a set of unknown parameters.

I can impose this linearity assumption only because I model the conditional price of risk and conditional beta exposures separately. This separation also allows me to treat conditioning information more in line with findings in the literature. Specifically, I treat characteristic realizations as being informative of risk loadings as in Cosemans et al., 2016, Chordia et al., 2017 and Kelly et al., 2019 and treat the conditional price of risk as arising from linear combinations of trade-able portfolios formed from sorts on characteristics similar to factor definitions in Fama and French, 1996b, Hou et al., 2015b and Stambaugh and Yuan, 2017.

The conditional price of risk function

The conditional price of risk function, $f_h^*(\cdot)$, is initialized with a $(P+2)$-dimensional column vector of portfolio average returns, $\bar{r}_{p,t-h+1}$, when predicting returns for time $t+1$. This vector comprises an expanding window average return of long-short portfolios formed from sorts on the P firm-level characteristics. I concatenate this vector with the expanding window average return of the equally-weighted market and the risk-free assets. I compute all expanding window averages using portfolio returns starting from January 1965 up to time $t-h+1$. I define the conditional price of risk function as:

$$\mathbb{E}_{t-h+1}[F_{t+1}]' = \bar{r}_{p,t-h+1}W_{0,h} + b_{0,h} \tag{3.1.3}$$

where $W_{0,h} \in \mathbb{R}^{58\times 3}$ and $b_{0,h} \in \mathbb{R}^{1\times 3}$ are unknown parameters to be estimated. This parameterization allows for the pricing function to be dense in the space of portfolio and security returns (58 average returns) and simultaneously remain sparse in pricing factor (3 latent factors).

From Kozak et al., 2020, we know that a handful of latent factors are enough to explain a significant fraction of the variations observed in realized returns. Guided by this finding, I set the number of pricing factors to 3. It is worth mentioning that the small number of factors I impose does not restrict the resulting approximator to the same space as a three principal component (PC) model. This is because factor loadings in Equation (3.1.2) are time-varying as opposed to the statistic loadings in

For each forecasting horizon in H, we estimate a different expectation function denoted by the subscript h.

Picking J between 3 and 10 does not qualitatively change the results but increases the time it takes the models to converge a PC model. Similar to Kelly et al., 2019 and Gu et al., 2020a, I find that restricting the model to one or two latent factors is too restrictive.

I do not allow for non-linear interactions between portfolio returns in determining the factor returns because I require the factor returns to be spanned by the returns of the underlying 58 portfolios. I construct each long-short characteristic sorted portfolio by fixing portfolio weights as the rank-normalized characteristic realizations at some time t. I then go long one dollar and short another dollar. All the long-short portfolios I consider are therefore spanned by the stocks in the cross-section.

The expected conditional beta function

The expected conditional beta exposure function, $b_h^*(\cdot)$, is initialized with a P-dimensional vector of rank-normalized firm characteristics, $p_{i,t}$, when predicting returns for time $t+h$. I assume that characteristic realizations at time t are informative of their time $t+h$ realizations.. I approximate the beta exposures as:

$$Y_{1,h} = \psi(p_{i,t-h+1}W_{0,h} + b_{0,h}) \tag{3.1.4}$$

$$Y_{2,h} = \psi(Y_{1,h}W_{1,h} + b_{1,h}) \tag{3.1.5}$$

$$\mathbb{E}_{t-h+1}[\beta_{i,t}]' = Y_{2,h}W_{2,h} + b_{2,h} \tag{3.1.6}$$

where $W_{0,h} \in \mathbb{R}^{56\times1024}$, $W_{1,h} \in \mathbb{R}^{1024\times1024}$, $W_{2,h} \in \mathbb{R}^{1024\times3}$, $b_{0,h} \in \mathbb{R}^{1\times1024}$, $b_{1,h} \in \mathbb{R}^{1\times1024}$ and $b_{2,h} \in \mathbb{R}^{1\times3}$ are unknown parameters to be estimated. ψ is the relu non-linearity; $\psi(\cdot) = max(y,0)$. This parameterization of the beta exposure function allows me to project the 56 firm-characteristics into a higher dimensional (1024-dimensional) feature space where new features are easier to learn and project the resulting feature set back to the 3-dimensional latent pricing factor space (see Recanatesi et al., 2019). By allowing the nodes in the first layer of the model to be greater than the size of the input vector, I also maintain the universal approximation property of the deep neural network model (see Johnson, 2019).

I rank-normalize all firm characteristics in a cross-section at time t to the interval [-1,1]

Given that some characteristics are highly persistent, this is not a controversial claim (see Baba-Yara et al., 2020) Replacing the time t realizations with rolling window means does not change the results.

Even though I initialize all conditional beta exposure functions with the same characteristic vector, the resulting $J \times 1$ vector of conditional betas can differ across horizons. To see this, consider the relation between the momentum characteristic and expected returns. Momentum is positively related to realized returns for time period $t+1$ but negatively related to realized returns for time period $t+13$ (the reversal characteristic). Therefore, the learned relationship between the same characteristic and realized returns at different horizons by the neural network model will be different.

In the asset pricing literature, betas (risk loadings) are mostly specified as unconditional scaling functions that load on factor portfolio returns. Although this parameterization restricts the resulting model, it is still preferred to the conditional alternative because it is easier to estimate. Given that I estimate most of the unknown parameters of the model using stochastic gradient descent, I do not pay a steep estimation cost by preferring a conditional beta model to an unconditional model.

Additionally, by allowing for beta to be time-varying, the resulting predictive model is much more general in that beta responds to evolving firm characteristics. Consider a growth firm in the initial part of our sample transitioning to a value firm by the end of the sample. By allowing firm characteristics to inform conditional betas, the firm's risk loading (beta) on a particular factor can similarly transition from a low value to a high value across these two distinct regimes. Compare this to the unconditional beta model, which would have to be a scaler that captures the average risk loading of both the growth and value phases of the firm.

Besides the beta conditionality, I also allow for nonlinear interactions between firm characteristics via the ψ non-linearities. This specification is motivated by recent findings in the literature that shows that non-linearities between firm characteristics matter in explaining variations in firm returns. Bryzgalova et al., 2019 find that allowing for non-linearities through conditional sorting improves the resulting mean-variance frontier in the space of characteristic sorted portfolios. Gu et al., 2020a find that allowing for non-linearities results in an autoencoder asset pricing model that prices 87 out of 95 factor portfolios they consider.

3.1.3 A simple neural network model

I consider a simpler forecasting model that approximates the product of expected conditional price of risk and expected risk loadings with minimal assumptions coming from economic theory. Specifically, I estimate:

$$\mathbb{E}_{t-h+1}[r_{i,t+1}] \approx g_h^*(z_{i,t-h+1}) \tag{3.1.7}$$

where $g_h^*(\cdot)$ is some real-valued deterministic function of $P+M$ real variables, $z_{i,t-h+1}$. $z_{i,t-h+1} = [p_{i,t-h+1} : q_{t-h+1}]$, where $p_{i,t-h+1}$ is firm specific and q_{t-h+1} is the same across firms. I specify $p_{i,t-h+1}$ as a 56-vector of firm level characteristics, the same as in the expected conditional beta exposures function, and concatenate it with an M-dimensional, q_{t-h+1}, aggregate variables as in Gu et al., 2020b.

The difference between this forecasting model and the one I propose is that it does not model the conditional beta exposures and conditional price of risk functions separately. It approximates the expected return function directly while skipping all intermediary restrictions. This is the best performing machine learning model in Gu et al., 2020b and so serves as a natural benchmark for the more restricted model I propose. It is simpler in that it makes very little structural assumptions about how the different constituents of the information set interact in informing return expectations.

Following Gu et al., 2020b, I approximate Equation (3.1.7) using a three-layer feedforward neural network, which is defined as:

$$Y_{1,h} = \psi(z_{i,t-h+1}W_{0,h} + b_{0,h}) \tag{3.1.8}$$

$$Y_{2,h} = \psi(Y_{1,h}W_{1,h} + b_{1,h}) \tag{3.1.9}$$

$$Y_{3,h} = \psi(Y_{2,h}W_{2,h} + b_{2,h}) \tag{3.1.10}$$

$$\mathbb{E}_{t-h+1}[r_{i,t+1}] = Y_{3,h}W_{3,h} + b_{3,h} \tag{3.1.11}$$

where $W_{0,h} \in \mathbb{R}^{64\times 32}$, $W_{1,h} \in \mathbb{R}^{32\times 16}$, $W_{2,h} \in \mathbb{R}^{16\times 8}$, $W_{3,h} \in \mathbb{R}^{8\times 1}$, $b_{0,h} \in \mathbb{R}^{1\times 32}$, $b_{1,h} \in \mathbb{R}^{1\times 16}$, $b_{2,h} \in \mathbb{R}^{1\times 8}$ and $b_{3,h} \in \mathbb{R}$ are unknown parameters, θ, to be estimated.

Feedforward networks are the main building blocks of much more complicated neural networks. Among the five feedforward neural network models that Gu et al., 2020b study, the three-layer deep neural network out-performs along several dimensions.

ψ is a non-linear function (relu) applied element-wise after linearly transforming an input vector, either $z_{i,t-h+1}$ or $Y_{k,h}$.

Despite its flexibility, this simple forecasting model imposes some important restrictions on the estimation problem. The function, $g_h^*(\cdot)$, depends neither on i nor t but only h. By maintaining the same functional form over time and across firms for some time-period h, the model leverages information from the entire firm-month panel. This restriction significantly reduces the number of parameters I need to estimate and increases the resulting estimates' stability. This restriction is loose in that I re-estimate $g_h^*(\cdot)$ every two years, which means that each subsequent 24 month set of stock forecasts for some particular horizon h comes from a slightly different approximation of $g_h^*(\cdot)$. Finally, the specification also assumes that the same information set is I_t is relevant for making predictions for all horizons in H.

3.1.4 Loss Function

I estimate Equation (3.1.2) and Equation (3.1.7) by minimizing the mean squared error loss function with an l_1 penalty:

$$\mathcal{L}(\theta) = (N_t T)^{-1} \sum_{i=1}^{N_t} \sum_{t=1}^{T} \left(R_{t+1} - \hat{R}_{t+1}\right)^2 + \lambda_1 \left|\left|\theta\right|\right|_1 \tag{3.1.12}$$

where R_{t+1} is a vector of stock returns for time t, $\tilde{R}_{t+1}$ is a vector of predicted returns for all N_t firms in the cross section at time t, θ is the vector of model parameters. I minimize the empirical loss function over a pool of firm-month observations. I choose hyper-parameters such as λ_1 via a validation set. All hyper-parameters are detailed in in Appendix 3.D.

3.1.5 Estimation

I use the AdaBound learning algorithm from Luo et al., 2019 to estimate the unknown parameters (θ).

In addition to the l_1 penalty, I use batch normalization to help prevent internal covariate shifts across layers during training, (see Ioffe and Szegedy, 2015). I train

AdaBound leverages the rapid training process of the more popular adaptive optimizers such as Adam, Kingma and Ba, 2014, and generalizes like the classic stochastic gradient descent optimizer. Also, AdaBound has theoretical convergence guarantees which other optimizers such as ADAM lack.

the model on a batch size of randomly sampled 10000 firm-month observations per iteration. I estimate the model over 100 epochs, where an epoch represents a complete cycle through all of the training data. I stop training before the 100th epoch if the validation set does not increase after five subsequent epochs. Further details of the learning algorithm are provided in Appendix 3.D.

Sample Splitting

The dataset starts from January 1965 and ends in December 2018. I employ a rolling window estimation scheme by splitting the dataset into three parts; training, validation, and testing.

[Insert Figure 3.A.1 about here]

In predicting returns for month $t+1$ using information available up to time t, I estimate the model using 15 years of data starting from January 1975 and ending in December 1989. I choose hyper-parameters by comparing estimated model performance over a validation dataset starting from January 1990 to December 1994. I use the optimal model to make one-month ahead return predictions from January 1995 to December 1996. Figure 3.A.1 illustrates this exercise. I move the training, validation, and testing set forward by two years and repeat the process.

In predicting returns for month $t+2$ using information available up to time t, I estimate the model using 15 years of data starting from December 1974 and ending s in November 1989. I choose optimal hyper-parameters by comparing estimated model performance over a validation dataset from December 1989 to November 1994. I use the optimal model to make two-month ahead predictions starting from December 1994 to November 1996. This ensures that when comparing model performance across horizons, I am always comparing returns realized between January 1995 to December 1996, thereby aligning return realization dates across prediction periods, H. Similar to $t+1$, I move the training, validation, and test set forward by two years and repeat the process.

I always predict returns for the out-of-sample period; January 1995 to December 2018. As discussed above, I do this by shifting the conditioning information further into the past. This allows me to maintain the same training, validation and testing data size (in months) across horizons. Although this allows me to compare forecasts from different horizons for the same out-of-sample period, the subset of firms I am comparing across horizons is different. This is because firms enter and exit the CRSP file over time. Consider two different horizon forecasts for the month January 1995. The one month ahead forecast will condition on firms alive in December 1994. Whereas, the five year-ahead monthly forecast will condition on firms alive in December 1989. The trade-off I make is to align my setup more with a real-time setting, where agents form expectations for all future horizons in H, conditional on what they observe at the time.

I choose to estimate monthly forecasts because this allows us to bring the standard financial econometric tools to the problem and side step the econometric issues inherent in using compounded returns.

3.1.6 Data

I obtain monthly market data for US common stocks traded on AMEX, NASDAQ, and NYSE stock exchanges from CRSP. I match market data with annual and quarterly fundamental data from COMPUSTAT. I build a set of 56 firm-level characteristics from this panel. The characteristic definitions are from Freyberger et al., 2020b and Green et al., 2017. I obtain the one-month risk-free rate from Kenneth French's website. To avoid forward-looking bias, I follow the standard practice in the literature and delay monthly, quarterly and annual characteristics, by a month, four months, and six months respectively similar to Green et al., 2017; Gu et al., 2020b. To be included in the sample for some month t, a firm must have at least 30 non-missing characteristic observations. I rank-normalize the characteristics to the interval [-1,1] and replace missing values with zero.

The aggregate variable set, q_t, I use is from Goyal and Welch, 2008b, namely the S&P 500 dividend-to-price ratio, the S&P 12-month earnings-to-price ratio, the S&P 500 book-to-market ratio, net equity expansion, stock variance, the term spread, the default spread, and the treasury-bill rate. I condition on this set of aggregate variables to keep the simple model in-line with the specification in Gu et al., 2020b.

The details of the characteristics are provided in Table 3.C.1.

I would like to thank Amit Goyal for making this series available on his website.

Conditioning the simpler model on the same aggregate variables in as in Equation (3.1.3) leads to qualitatively poorer results.

3.2 Neural network forecasts in the cross-section of stocks

This section examines how incorporating economic theory in designing a neural network forecasting model helps improve return forecasts. I do this by comparing the forecasting accuracy of the economically restricted neural network model to that of the simple model in the cross-section of stocks across horizons. Additionally, I decompose the forecasts of both models to shed light on the cross-sectional and time-series prediction properties of the models.

The standard statistic I use to assess the predictive performance of these forecasts is the out-of-sample R Squared (R^2_{OOS}), which is defined as:

$$R^2_{OOS} = 1 - \frac{\sum_{(t)\,\in\,oss}(R_t - \tilde{R}_{t,1})^2}{\sum_{(t)\,\in\,oss}(R_t - \tilde{R}_{t,2})^2} \tag{3.2.1}$$

where R_t is the time t vector of realized stock returns, $\tilde{R}_{t,1}$ is a vector of forecasts from model 1 and $\tilde{R}_{t,2}$ is a vector of forecasts from model 2. Intuitively, the statistic compares the forecasting error of model 1 ($(R_t - \tilde{R}_{t,1})^2$), to that of model 2 ($(R_t - \tilde{R}_{t,2})^2$). If the forecasting error of model 1 is smaller than that of model 2, then R^2_{OOS} will be positive. A positive R^2_{OOS} therefore means that forecasts from model 1 improve upon forecasts from model 2.

I formally test the null hypothesis that forecasts from model 1 are no different from forecasts from model 2 in explaining variations in stock returns using the Clark-West (2007) test with Newey-West (1987) adjusted standard errors.

3.2.1 Can neural networks predict stock returns across horizons?

To answer this question, I define forecasts for model 1 as forecasts from the neural network models. I compare each models forecast to a zero prediction benchmark; $\tilde{R}_{t,2} = 0$. The results from this exercise answer the question, "How much variation in realized returns are explained by the neural network forecasts?"

[Insert Table 3.A.1 about here]

Panel A of Table 3.A.1 reports results for both the economically restricted model and the simple model. All the R^2_{OOS} estimates are positive and statistically significant across horizons. In general, both models' ability to explain variations in stock returns monotonically decrease the further into the future the forecasts pertain. Whereas the economically restricted model can explain about 0.99% of the variation in next month's return, it can only explain about 0.13% of the variation in ten-year returns. Similarly, the simple model can explain about 0.58% of the variation in next month's return, and this falls to 0.18% of the variations in return ten years in the future.

Comparing the models on the variations in returns they explain in next month's return, the economically restricted model explains close to twice the variation explained by the simple model; 0.99% against 0.58%. In explaining variations in stock returns further in the future, the simple model explains a slightly larger fraction; 0.18% against 0.13%.

3.2.2 Disentangling the composite R^2_{OOS}

The R^2_{OOS} tells us how much variation in returns the forecasts from model 1 explain when the benchmark model (model 2) is a zero prediction model. The results show that both models can predict stock returns across horizons. However, the R^2_{OOS}, as defined above, fails to tell us along which dimension of stock returns these estimates forecast well. The forecast may be predicting stock returns well because they predict the level factor in stocks. Or they could additionally be predicting time-series variation in the cross-sectional dispersion in stock returns. Given that a strong factor structure holds in the pool of stocks, it is instructive that we disentangle the R^2_{OOS} to shed light on this.

I assume a two-factor structure holds for the stock return forecasts. I fix the first factor as the equally-weighted market forecast and allow the second factor to subsume all other priced factors in the cross-section. This parameterization allows me to decompose return forecasts from some model m for a firm i at some time t into

Kozak et al., 2020 show that an asset pricing model of a similar form explains a significant fraction of the variations in returns two parts:

$$r_{m,i,t} = (N_t^{-1}) \sum_{k \in t} r_{1,k,t} + r_{m,i,t}^{RR} \tag{3.2.2}$$

where $(N_t^{-1}) \sum_{k \in t} r_{1,k,t}$ captures the cross-sectional mean forecast of model m and $r_{m,i,t}^{RR}$ captures the cross-sectional variation in forecasts across firms. The return to each firm i ($r_{m,i,t}$) is therefore made up of the cross-sectional level factor $((N_t^{-1}) \sum_{k \in t} r_{1,k,t})$ and firm specific relative return ($r_{m,i,t}^{RR}$).

I further decompose the relative forecast ($r_{m,i,t}^{RR}$) into an unconditional component ($\mu_{1,i}^{RR}$) and a conditional component ($\tilde{r}_{1,i,t}^{RR}$). Specifically, I decompose $r_{m,i,t}^{RR}$ as follows:

$$r_{m,i,t}^{RR} = \mu_{m,i}^{RR} + \tilde{r}_{m,i,t}^{RR} \tag{3.2.3}$$

where $\tilde{r}_{1,i,t}^{RR}$ is mean zero (by construction) and captures the relative (residual) time-series forecasts of model m. $\mu_{1,i}^{RR}$ is the average firm i forecast over the out-of-sample period and captures the unconditional relative (residual) stock forecast. This parameterization allows me to study the time-series predictability of relative stock returns absent the unconditional component. See section 3.E in the Appendix for more details on the decomposition.

Panel B of Table 3.A.1 reports the results for the decomposition of the R_{OOS}^2 against a zero prediction benchmark. For both models, the ability of their forecasts to explain time-series variation in relative stock return is only present for short-run months. Neither model can explain time-series variation in relative stock returns realized beyond one year in the future.

However, the amount of time-series variation in relative stock returns the models can explain is very different. Whereas the simple model explains about 0.20% of time-series variation in next month's relative stock return, the economically restricted model explains about 0.71%, a more than three-fold improvement. In predicting monthly relative stock returns one year in the future, the simple model explains about 0.03% of time-series variation in relative stock returns against 0.08% for the economically restricted model.

For both models, a large fraction of the reported composite R_{OOS}^2 comes from explaining variations in the level factor in stock returns. For the economically restricted model, about 40% of the composite R^2_{OOS} (0.43% out of 0.99%) comes from explaining variations in next month's level factor. For the simple model, this figure is 70% (0.41 out of 0.58%). For all other future forecasting periods, more than 90% of the composite R^2_{OOS} comes from the models' ability to explain variations in level factor, with little to negative (R^2_{OOS}) contributions coming from explaining variations in relative stock returns.

The results show that intermediate and long-run forecasts from the neural network models are very different from short-run forecasts. Whereas short-run predictions can discriminate between high and low expected return stocks (relative stock returns) in addition to forecasting the level factor, intermediate and longer-run forecasts only explain variations in the cross-sectional average return (level factor).

3.2.3 An alternative benchmark

Results from the decomposition of the R^2_{OOS} with respect to the zero prediction benchmark show that the dominant factor that the forecasts are predicting is the equally-weighted market return. This result suggests that an alternative benchmark that does reasonably well along this particular dimension of returns should be tougher for the neural network forecasts to beat. From Goyal and Welch, 2008b, we know that one such example is the historical average market return. I define this benchmark's $t+h$ stock return forecast as the time t average equally-weighted market return computed using data from 1926.

[Insert Table 3.A.2 about here]

The results are reported in Table 3.A.2. For short-run months, I find a more than 30% reduction in the composite R^2_{OOS} compared to the zero-prediction model. From this result, we can conclude that the historical average market return is a challenging benchmark, even in the pool of individual stocks. In the long-run, I find an increase in the composite R^2_{OOS} compared to the zero-prediction model. This result means that the zero-prediction model remains the tougher benchmark for longer run returns. This finding is explained by the fact that more than 40% of firms alive at any

Results for other alternative models are in the Table 3.B.1 of the Internet Appendix.

period t fall out of the sample by $t + 60$. Thus, the historical average market return computed as a function of firms alive at some t, will be a poor estimate of the longer-run unconditional average return.

Comparing the R^2_{OOS} estimates of the simple model to the economically restricted model across horizons and benchmarks, it is evident that economic restrictions generally improve the forecasts. In predicting next month's return, the economically restricted model has an R^2_{OOS} of about 0.64%, whereas the simple model has an R^2_{OOS} of about 0.20%. In predicting returns ten-years into the future, the economically restricted model has an R^2_{OOS} of 0.62%, and the simple model has an R^2_{OOS} of about 0.55%.

Taken together, the results in this section show that incorporating economic restrictions improves the ability of a neural network model to predict stock returns. This improvement is most evident in the ability of the forecasts to explain time-series variations in relative stock returns over the short-run.

3.3 Predicting market and long-short portfolio returns

The previous section shows that incorporating economic theory in designing a neural network architecture improves return forecasts in the cross-section of stocks. Since individual stock forecasts can easily be aggregated to forecast market returns and returns to long-short characteristic sorted portfolios, it is natural to ask if the model that incorporates economic theory generalizes better along these dimensions than the simple model. That is the central question I answer in this section.

3.3.1 Can the forecasts predict market returns?

To answer this question, I define the market forecast as the value-weighted monthly stock forecast for period $t + h$ and define the market return as the value-weighted monthly cross-sectional average stock return of firms in the CRSP file at time $t + h$.To capture the pure effect of different forecasts, I always use market-caps from time t but allow the forecasts to vary across horizons. I compute the R^2_{OOS} with respect to two benchmarks; a zero-prediction model and the historical average-market return.

[Insert Table 3.A.3 about here]

The R^2_{OOS} of the neural network against a zero prediction benchmark tells us how much variation in market returns the forecasts explain. The results in Table 3.A.3 show that both models can robustly explain market returns across all horizons, I consider. The economically restricted model can explain a larger fraction of the variation in market returns compared to the simple model, especially for short-run to intermediate horizons (up to three years). For example, in predicting next returns, the economically restricted model explains about 5.35% of the variation in market returns while the simple model explains about 2.05%.

Decomposing the R^2_{OOS} into a time-series variation and an unconditional return component shows that less than 35% of the variation explained in market returns across horizons pertains to the ability of both models to explain time-series variations in returns. For instance, in predicting market returns one year into the future, 1.51% of the 4.90% composite R^2_{OOS} comes from the ability of the economically restricted model forecasts' to explain time-series variation in market returns. The rest comes from matching the unconditional market return in the out-of-sample period.

Focusing on the more challenging historical average market return benchmark, we see that the simple model's market return forecasts offer no improvements. For all horizons and dimensions of market returns, this model fails to improve upon the historical average market forecast. The story is different for the economically restricted model. This model fails to improve upon the historical average market forecast in predicting the unconditional market return in the out-of-sample period. However, its ability to out-perform the historical average market return forecast in predicting time-series variation in market returns is large and statistically significant at the 5% level up to three years in the future.

3.3.2 Can forecasts predict long-short portfolios returns?

The positive and statistically significant R^2_{OOS} in rows 1 and 3 in panel B of Table 3.A.1 suggest that both neural network forecasts should be able to forecast returns to long-short portfolios. This is because this dimension of the decomposed R^2_{OOS} is related to predicting time-series variation in relative stock returns. And this translates into returns of long-short portfolios. However, we can not make conclusive statements from the results in Table 3.A.1 because the R^2_{OOS} are computed with respect to the entire cross-section of stocks, whereas long-short portfolios only buy and sell a fraction of stocks that are most of the time in the tails of the return distribution. Additionally, long-short portfolios are mostly value-weighted as and not equally-weighted as in Table 3.A.1.

To answer the question, I sort stocks on the five characteristics in the Fama and French (2018) factor model; book-to-market, investment, size, operating profit, and momentum. For characteristics computed from balance sheet or income statement variables, I update them at the end of June of year s using the characteristic observations from the fiscal year-end $s - 1$. For characteristics computed only from CRSP variables, I update them at the end of each month and re-balance accordingly. I form decile portfolios from the sorts and value-weight to reduce the effect of small stocks. The return (forecast) to the long-short portfolio is the value-weighted return (forecast) of portfolio ten minus the value-weighted return (forecast) to portfolio one.

Similar to analyzing market return predictability, I decompose the (R^2_{OOS}) to investigate time-series and unconditional forecasting accuracy of the long-short characteristic portfolio forecasts.

[Insert Table 3.A.4 about here]

Results for the simple neural network model are reported in Table 3.A.5. The alternative model is the zero-prediction model. Even against this much weaker benchmark, the simple model fails to robustly explain any variation in returns to long-short characteristic sorted portfolios. For almost all reported horizons and across all five long-short portfolios, the R^2_{OOS} is negative. For the few horizons and portfolios where the estimate is positive, it is seldom statistically significant.

[Insert Table 3.A.5 about here]

Table 3.A.5 reports results for the economically restricted model. For this model, the benchmark is the historical average long-short portfolio return computed using data from 1964. The model does a much better job predicting returns to long-short portfolios than the simple model, despite the more challenging benchmark. Focusing

To be included in a sort, a firm must have a neural network forecast and non-missing observations for return and characteristic being sorted on.

on forecasts for time $t + 1$, I find positive R^2_{OOS} for all five long-short portfolios as against two for the simple model. For four of these five portfolios, the R^2_{OOS} is statistically significant at the 5 % level. The decomposed R^2_{OOS} shows that the forecasting power of the economically restricted model is driven by its ability to better predict time-series variation in returns to these long-short portfolios.

To show how pervasive this finding is, I expand the universe of long-short portfolios to all 56 characteristics that I condition on in the beta function (Equation (3.1.4)) and focus on forecasts for month $t + 1$. Figure 3.A.2 reports the results.

[Insert Figure 3.A.2 about here]

In panel A, I find that 53 (32) of the 56 long-short portfolios have positive (and statistically significant) composite R^2_{OOS}. Similar to the results above, most of the composite R^2_{OOS} is driven by the forecasts' ability to predict time-series variation in returns to long-short portfolios. In panel B, I find that for 53 of the 56 long-short portfolios the neural network forecasts improve upon the the benchmarks ability to predict time-series variation in returns. For 31 portfolios, this improvement is statistically significant at the 5%.

[Insert Figure 3.A.3 about here]

Moving beyond month $t + 1$ forecasts, I report results for other horizons in Figure 3.A.3. To keep things compact, I only report the fraction of long-short portfolios with positive composite R^2_{OOS}, positive contributions coming from forecasting time-series variation in returns, and positive contributions coming from predicting the unconditional long-short portfolio return. Panel B reports the fractions that are both positive and significant. I observe that although the neural network forecasts predict a majority of long-short portfolio returns in short-run months, the fraction that is significant precipitously drops to zero when I use forecasts older than three months. From this, I conclude that the more timely the information set I condition, the more accurate the forecasts predict time-series variations in returns to long-short portfolios.

The results in this section show that machine learning guided by economic theory can lead to significant improvements in predicting returns that robustly generalize beyond the cross-section of stocks. Specifically, such a model can predict time-series variation in monthly market returns up to three years into the future.

Additionally, the model can predict time-series variation in next month returns for 53 (32 are statistically significant) out of 56 long-short portfolios.

3.4 Neural network forecasts and conditional expected returns

The previous sections show that the economically restricted model explains significant variation in stock returns. This ability generalizes to market returns and long-short portfolio returns. This section analyzes how well the economically restricted model forecasts line up with conditional expected returns across firms, portfolios, and time.

The standard tool in the literature used in this specific analysis is time-series predictive regressions (see among others Cochrane, 2008, and Lewellen, 2015). The slope coefficient from regressing demeaned forecasts on returns is informative of how well the forecasts line up with conditional expected returns. We are interested in predictions that get the conditional direction of returns right. If the slope coefficient is positive and statistically different from zero, then it fulfills this requirement. Additionally, we are interested in unbiased return forecasts, that is, models for which the slope coefficient is indistinguishable from one. For such models, a one percent forecast on average translates into a one percent return.

[Insert Table 3.A.6 about here]

Panel A of Table 3.A.6 report results from regressing demeaned relative stock returns on realized stock returns. The results generally confirm the conclusions from the decomposed out-of-sample R^2 analysis. For the short-run months, $t+1$ up to $t+13$, I can reject the null hypothesis that the forecasts fail to predict time-series variation in relative stock returns. This is because the 95% confidence interval of the slope coefficient is strictly positive. For $t+1$, the forecasts are unbiased because the 95% confidence interval of the slope coefficient includes one. Specifically, a one percent relative stock forecast on average translates into a 0.97 percentage point increase in next month's realized relative stock return. The model over predicts time-series variation in relative stock returns for all other short-run months because the confidence intervals are strictly less than one but positive.

From these results, we can conclude that the model's forecasts line up well with expected stock returns for the next month's returns but over-predict stock returns for all other months.

Panel B of Table 3.A.6 shows that the market forecasts, up to intermediate-term months, on average, do line up with expected market returns. For months $t + 1$ up to $t + 37$, the slope coefficients from regressing demeaned market forecasts on market returns are positive and statistically different from zero. The estimates are around 1.50, meaning a one percent market return forecast translates into a 1.50 percentage point increase in market return. And the 95% confidence interval of the slope coefficient includes one for these specific monthly forecasts.

Panel C of Table 3.A.6 reports results for long-short portfolios. Slope coefficients for months $t + 1$, $t + 3$, $t + 13$ are positive and statistically different from zero. This means the aggregate neural network forecasts from the economically restricted model can predict time-series variation in returns to long-short portfolios for short-run months. These forecasts for long-short portfolios do not generally line up well with conditional expected returns. A one percent forecast on average translates into about a 2 percentage point realized return to the typical long-short Fama and French, 2018 5 model characteristic sorted portfolio. We cannot reject the null hypothesis that the slope coefficients for $t + 1$, $t + 3$ and $t + 13$ are unbiased.

3.5 Optimal Portfolios

This section introduces several optimal trading strategies that highlight the practical usefulness of the neural network forecasts. We show that an investor using these forecasts in a pseudo-real-time setting over the out-of-sample period enjoys significant improvements measured by average returns, Sharpe ratios, risk-adjusted returns, and certainty equivalents.

I define the certainty equivalent with respect to an investor with a mean-variance utility function and a risk aversion parameter of 2. Specifically, I compute the certainty equivalent return of a strategy as:

$$CE = \bar{r}_h^p - \frac{\gamma}{2}\sigma_{p,h}^2 \tag{3.5.1}$$

where $\sigma_{p,h}$ is the sample standard deviation of the strategy. The certainty equivalent can be interpreted as the risk-free return that a mean-variance investor with a risk-aversion coefficient of γ would consider equivalent to employing this strategy. Alternatively, it can be viewed as a fee that an investor is willing to pay to use the information inherent in our forecast. I report the certainty equivalent annualized and in percentages.

3.5.1 Optimal timing strategies

I consider a strategy that times a risky security by levering up and down the position in the security based on whether conditional expected returns are high or low. The previous section showed that the forecasts from the economically restricted neural network model explain time-series variation in returns for most of the portfolios we consider. Therefore, we should expect these forecasts to be informative of when to lever up and down based on return expectations for the future.

For each month, t, I use the conditional expected return forecast from the economically restricted neural network model to calculate the Markowitz optimal weight to be invested in the risky asset as:

$$w_{t,h} = \frac{\tilde{r}_{t,h} - r^f_{t+1}}{\gamma\sigma(\tilde{r}_{1:t-1,h})} \tag{3.5.2}$$

where γ is the risk aversion coefficient, which I set to 2. I fix the conditional standard deviation estimate ($\sigma(\tilde{r}_{1:t-1,h})$) at an annualized value of 15 % across securities because of two main reasons; 1) to remove the impact of volatility timing from the exercise (see **muir**) and 2) because the forecasting model does not produce a conditional standard deviation estimate. At the end of each month, I compute the timing portfolio return as:

$$r^p_{t,h} = w_{t,h}r_{t+1,h} - (1 - w_{t,h})r^f_{t+1} \tag{3.5.3}$$

and iterate until the end of the out-of-sample period, December 2018.

The optimal market timing portfolio

The first trading strategy I consider tries to time the value-weighted market return by deciding how much to invest between the market and a risk-free asset using the aggregated forecast for the market. I restrict the market to the 500 largest market capitalized firms at each time t. For each month t, the strategy invests $w_{t,h}$ in the value-weighted market and $1 - w_{t,h}$ in the risk-free asset.

[Insert Table 3.A.7 about here]

Panel A of Table 3.A.7 reports the results. A buy and hold strategy that is fully invested in the market over the sample period makes a annualized average return of 10.39 % with a certainty equivalent of 8.18. The return to this strategy is fully explained by the CAPM and the Fama and French, 2018 5 factor model. A timing strategy that uses the most recent market forecasts, $t+1$, earns an annualized average return of 17.21 % with a certainty equivalent of 12.86. Timing the market with predictions that are a month old, $t+2$ to two years old, $t+25$ all out-perform the buy and hold strategy. Generally, the more timely the forecasts are, the higher their accuracy in predicting time-series variation in market returns. We can see this from the higher certainty equivalents and average returns for periods $t+1$ and lower estimates for much older forecasts such as $t+121$.

The optimal characteristic timing portfolio

The second timing strategy I consider tries to time an equally-weighted portfolio of book-to-market, size, investment, profitability, and momentum long-short portfolios. For each month t, the strategy invests $w_{t,h}$ in this equally-weighted portfolio and $1 - w_{t,h}$ in the risk-free asset.

Panel B of Table 3.A.7 reports the result for the timing strategy. Similar to previous results, I find that the spread in returns generated by the timing strategy is strongest when the forecast is much closer to the re-balancing month t. Forecasts that are older than two years to the date of re-balancing generate negative spreads. Characteristic timing, like all other strategies considered in this section, requires timely

Although the portfolio we are timing is equally-weighted, the individual long-short characteristic sorted portfolios are all value-weighted.

information. The most timely forecast $t + 1$ almost triples the buy and hold certainty equivalent, average returns, and alphas.

Timing individual characteristic sorted portfolios

The third timing strategy I consider tries to time individual value-weighted portfolios formed from sorts on the characteristics I condition on. The time t weight $w_{t,h}$ invested in each portfolio is determined as in the other timing strategies. I only report results for the $t + 1$ forecasts for the sake of space..

[Insert Figure 3.A.4 about here]

Panel A of Figure 3.A.4 reports the annualized average return of the difference in returns between the timing portfolio and a buy and hold variant. For 54 of the 56 long-short portfolios, timing leads to an improvement in average returns. For 23 of the 56, this improvement is statistically significant at the 5% level. 37 of the 56 portfolios show a greater than 5% improvement in annualized returns. Panel B reports the Sharpe ratio of the difference in returns between the timing portfolio and the buy and hold. For 46 of the 56 long-short portfolios, the increase in the annualized Sharpe ratio is at least 0.20 units. Finally, in Panel C, I report the alpha from regressing the timing portfolio returns on the Fama and French, 2018 6 factor model. For 34 of the 56 portfolios, the timing portfolio returns are not fully explained by this asset pricing model.

The results from this timing exercise show that the neural network forecasts predict well time-series variation in relative stock returns for short-run months. And this predictive ability strongly generalizes to predicting time-series variation in long-short portfolios. Remember that long-short portfolios zero out the level factor in stocks by construction. So, forecasts that predict stock returns well only because they predict the level factor cannot predict return variations in long-short portfolios.

3.5.2 Optimal rotation strategies

The alternate set of strategies that I consider rotates across securities in the cross-section of stocks and long-short characteristic sorted portfolios. Strategies of this type allow us to gauge how accurately the relative stock return forecasts' discriminate between high expected return stocks and low expected return stocks in the cross-section.

For each period t, I sort all securities in a particular cross-section on their forecasted return, buy (sell) the top (bottom) 10%. I repeat these sorts H times for each forecasting period. If the timeliness of the information set is important for accurately discriminating between high and low expected return securities in a cross-section, then the strategies that use forecasts for short-run periods (h $\in \{1, 2, 3, 13, \}$) should out-perform strategies that use much longer-run forecasts.

[Insert Table 3.A.8 about here]

Long-short stocks strategy

This is a long-short strategy that buys (sells) the value-weighted portfolio of the 10% highest (lowest) expected return stocks within the 500 largest market capitalized firms in the cross-section of stocks at some time t.

The results are presented in panel A of Table 3.A.8. Re-balancing the long-short stock portfolio each month using the $t + 1$ forecasts produces a certainty equivalent of 11.73, larger than the 8.13 for the buy and hold investor. This re-balanced portfolio also generates returns that are not explained by the CAPM or the Fama and French, 2018 5 factor model. Therefore, the long-short portfolio's improved performance does not come from loading on the fundamental factors in either asset pricing model. Generally, the certainty equivalents, average returns, and alphas fall with the horizon. The right way to think about this is that using older forecasts to re-balance the long-short portfolio comes at a cost. Remember, the portfolios are re-balanced monthly, and the horizon dimension is captured by how old the forecast are. A monthly re-balanced long-short portfolio using forecasts for period $t + 120$ means re-balancing with forecasts that are ten years old.

Characteristics rotation strategy

This cross-section is made up of the five characteristics in the Fama and French, 2018 6 factor model. The strategy buys (sells) an equally-weighted portfolio of the two long-short characteristic sorted portfolios with the highest (lowest) expected returns. Individual long-short portfolios are value-weighted.

The results are presented in panel B of Table 3.A.8. The rotation strategy using forecasts for period $t+1$ almost double the certainty equivalent, average returns, and alphas of the benchmark strategy that buys and holds all five portfolios. The gains from using the forecasts fall as I use older forecasts. Whereas forecasts for period $t+1$ generate an annualized average return of 17.06 %, forecasts for period $t+120$ generate an annualized average return of -3.22 %. Expanding this cross-section to the 56 long-short portfolios does not change the conclusions.

3.6 Predictability decay over time

Ben-Rephael et al., 2015 show that the characteristic liquidity premium has significantly declined over time. Chordia et al., 2013 study the predictive accuracy of a broader set of characteristics in two sub-periods and find significant attenuation in the predictive accuracy of the characteristics in the second sub-period. The decreasing predictive accuracy of characteristics is not much challenged in the literature. However, the hypothesis put forward to explain the phenomena are many and varied. Chordia et al., 2011 argue this result may be the effect of institutional activity, bringing about more efficient price formation. McLean and Pontiff, 2016 argue that the result may be due to popularization coming from academic research.

Given that I condition the forecasting model on an overlap of these characteristic realizations, it is essential to answer the question, "How has return predictability evolved over the sample period?" If the information set has become less informative about expected returns, we should expect to see decreasing out-of-sample R^2 over time. To answer this question, I compute a two-year rolling out-of-sample R^2 from January 1997 to December 2018. I present the results for forecasts generated for month $t+1$.

[Insert Figure 3.A.5 about here]

Panel A of Figure 3.A.5 reports the rolling out-of-sample R^2 with respect to predicting time-series variation in relative stock returns. The figure shows that a large fraction of the forecasts' ability to explain time-series variation in relative stock returns comes from the initial part of the sample. The average out-of-sample R^2 in the first half of the sample is about twice as large in the second half. From this, I can conclude that the predictive ability of the information set has waned over time. However, I cannot conclude that the forecasts have entirely lost their ability to predict time-series variation in relative stock returns even though the estimates are negative at the end of the sample. This is because a similar negative streak is present between 2007 and 2009, after which positive estimates remerged.

The results for the market show that the rolling estimates are much more volatile. This may be because the time-series variation in market returns is tougher to predict or because the sample from which I compute the rolling estimate is smaller. Similar to the case of individual stocks, the forecasts' ability to predict time-series variation in returns is much stronger in the first half of the sample than the second. At the end of the sample, the rolling estimates are negative.

3.7 Variable importance across horizons

In this section, I investigate which covariates matter the most in generating return forecasts across horizons. I use the notion of Shapley values from Lundberg and Lee, 2017. The authors show that Shapley values generalize many competing measures of model explainability with respect to neural networks.

Shapley regression values are measures of covariate importance for linear models that are robust to multicollinearity (Lundberg and Lee, 2017). It is a feature importance measure that requires re-estimation of a model on all possible covariate subsets $S \subseteq F$, where F is the set of all covariates. It assigns a value to each covariate representing the marginal effect on the model prediction of including that feature.

To compute this marginal effect for some model f, consider the model $f_{S\cup\{k\}}$ trained with all covariates present and model f_S with feature k withheld. The predictions from the two models are then compared for some observation x_S, $f_{S\cup\{k\}}(x_{S\cup\{k\}}) - f_{S\cup\{k\}}(x_S)$, to compute the marginal effect for that observation. This effect is computed for all possible subsets $S \subseteq F\backslash k$. Shapley values are then computed as the weighted average of all possible differences:

$$\phi_i = \sum_{S \subseteq F \setminus k} \frac{|S|!(|F| - |S| - 1)!}{|F|!} \left[f_{S \cup \{k\}}(x_{S \cup \{k\}}) - f_{S \cup \{k\}}(x_S) \right] \tag{3.7.1}$$

Shapley additive explanation (SHAP) values represent an easily computable approximation of Eq. (3.7.1). These values provide the unique covariate importance measure I use.

[Insert Figure 3.A.6 about here]

Figure 3.A.6 reports the overall ranking of characteristics for select horizons. I estimate SHAP values for each observation in our sample and then average across characteristics. Characteristics are ordered so that the highest-ranked is at the bottom and lowest-ranked characteristics (out of the top four) at the top. Blue represents a negative contribution to the forecast and red a positive contribution. In analyzing the results in this section, it is important to remember the previous section's results. In the cross-section of stocks, the dominant factor we predict is the level factor.

The results show that the same three variables are the most important drivers of return forecasts across horizons. The most important variable is dividend-yield (DP). A unit increase in this variable positively predicts returns in the short and intermediate run. It is not surprising that the most dominant predictive variable in our conditioning information is the one variable that the literature has shown to predict market returns, the dividend-yield (see, Goyal and Welch, 2008b,Cochrane, 2011b and Ferreira and Santa-Clara, 2011).

The second most important variable is debt-to-price. A unit increase in this variable leads to a reduction in return forecasts across horizons. Intuitively, this results suggests that as the average leverage in the cross-section of stocks increase, expected returns falls.

The third important variable is closeness of last month's price to last 52 week close (CL2HG), a trend related measure, is the third. The fourth measure is depreciation and amortization to total assets.

Although one may be tempted to interpret all these estimates and their impact on return forecasts with respect to only the cross-section of stocks, it is important to

See Lundberg and Lee, 2017 for detail

remember that there exists a factor structure inherent in this cross-section. However, it is not clear how to decompose variable importance along similar lines as in Eq. (3.2.2). One solution is to exclude a subset of characteristics and then compare the reduction in the decomposed out-of-sample R^2. However, systematically dropping and including characteristics is leads to a combinatorial explosion in the number of tests. And so without strong priors on what clusters of characteristics matter for which dimension of returns, this solution is computationally infeasible.

3.8 Conclusion

This paper primarily shows that incorporating economic theory in a neural network model architectural design significantly improves the resulting model's forecasting ability. I find that the improvements mainly come from the resulting model's ability to explain time-series variations in returns. The improved model explains almost twice as much of the variation in the following month's returns compared to a much simpler neural network model.

I show that the model that strictly adheres to economic theory produces forecasts that robustly generalize beyond the cross-section of stocks compared to the simpler model. Specifically, the aggregate market forecast robustly predicts time-series variations in market returns up to three years into the future. And, the aggregate long-short portfolio forecasts predict time-series variation in the following month's return for 53 out of the 56 long-short portfolios. The simpler neural network fails along these two dimensions.

I disentangle the nature of the stock forecasts and show that short-run stock return predictability is very different from long-run predictability. Monthly forecasts for up to one year into the future, predict cross-sectional variations in stock returns and time-series variation in relative stock returns, in addition to predicting the level factor in returns. In contrast, forecasts for periods beyond one year only predict variations in the level factor.

Studying the time-series and cross-sectional properties of conditional expected returns across multiple horizons is important for many reasons; 1) such studies naturally produce new test portfolios for examining asset-pricing models. 2) they uncover new stylized facts about expected returns that constitute a new set of moments for emerging theoretical models to match. 3) their findings are practically useful because more accurate stock return forecasts allow us to devise better trading strategies and find better costs of capital estimates for future cash-flows.

In light of this paper's findings, it would be interesting in future research to answer the question, "Why does cross-sectional return predictability decay so quickly along the horizon dimension?" One possible hypothesis is that most of the predictable cross-section variation in next month's return is due to mispricing, which is then corrected quickly. And so beyond the following month, there is very little cross-sectional mispricing information in the information set we are conditioning on. Another hypothesis is that today's firms are very different from their future selves if we compare them based on their characteristic realizations'. If this is true, then characteristic realizations today will have a lot less to say about future cross-sectional dispersion in returns the further in the future we want to predict.

3.A Appendix

3.A.1 Robustness Checks

In this section, I consider alternative benchmarks and additional checks of the robustness of the results in the paper.

3.A.2 Alternative benchmarks

To highlight the general performance of the neural network forecasts, I consider two additional benchmarks from the literature. These benchmark models forecast monthly firm returns as a function of historical individual firm returns. The first model predicts the time $t + h$ return of firm i as the average firm i return computed over a five-year rolling window; the five-year rolling window firm average return benchmark. The second model predicts the time $t + h$ return of firm i as the average firm i return from the start of the sample up to time t; the expanding window firm average return benchmark.

[Insert Table 3.B.1 about here]

Panels A and B of Table 3.B.1 report results for the alternative benchmarks described above. The R^2_{OOS} is positive for all future periods and statistically significant for both models. Prediction accuracy is higher for short-run periods and falls with the horizon. The positive R^2_{OOS} of the neural network forecasts against these benchmarks are much larger than the estimates against the zero prediction benchmark because firm average returns are very noisy estimates of expected firm returns.

Comparing the R^2_{OOS} estimates of the simple model to the economically restricted model across horizons and benchmarks, it is evident that economic restrictions generally improve the forecasts. Most of this improvement is concentrated in the short-run. In predicting next month's return, the economically restricted model has an R^2_{OOS} of about 2.18% against the five-year rolling window benchmark, whereas the simple model has an R^2_{OOS} of about 1.76%. In predicting returns ten-years into the future, the economically restricted model has an R^2_{OOS} of 0.63% against the expanding window firm average return benchmark, and the simple model has an R^2_{OOS} of about 0.64%.

3.A.3 Figures

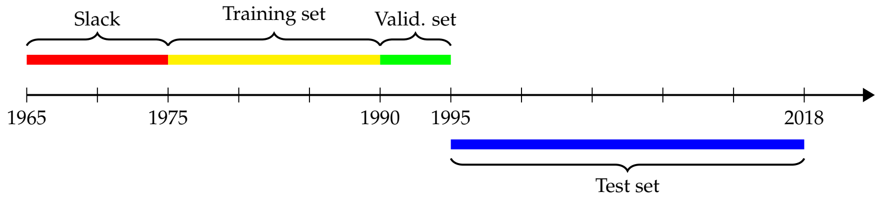

FIGURE 3.A.1: **Sample Splitting time-line**
This figure presents a time-line for sample splitting scheme in the paper.

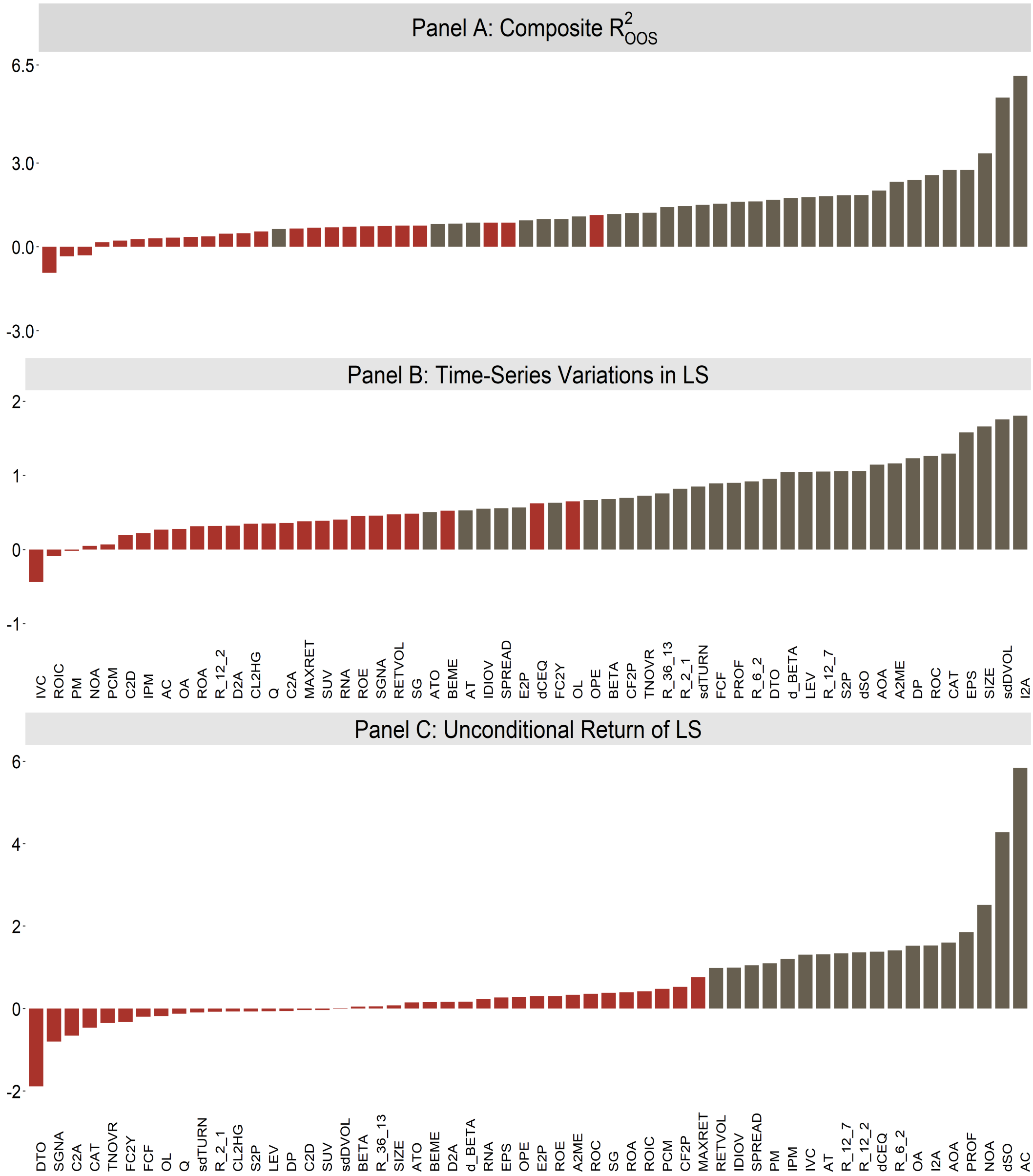

FIGURE 3.A.2: **Forecasting returns of long-short portfolios**
This figure reports the out-of-sample R^2 (R^2_{OOS}) of monthly re-balanced long-short portfolios formed from sorts on 56 characteristics listed in Table 3.C.1 using forecasts from time $t-h$, where $h \in \{1,2,3,13,25,37,49,61,121\}$. Grey bars represent statistically significant R^2_{OOS} using the the Clark-West (2007) test at the 5% level. Panel A reports the composite R^2_{OOS}, Panel B reports the R^2_{OOS} contribution that comes from the ability of the neural network forecasts to better predict time-series variation in returns to long-short portfolios, and Panel C reports the R^2_{OOS} contribution that comes from the ability of the neural network forecasts to better predict the unconditional returns to long-short portfolios. The alternative model is the historical portfolio return computed from an expanding window mean with data from 1965. The sample period is from January 1995 to December 2018.

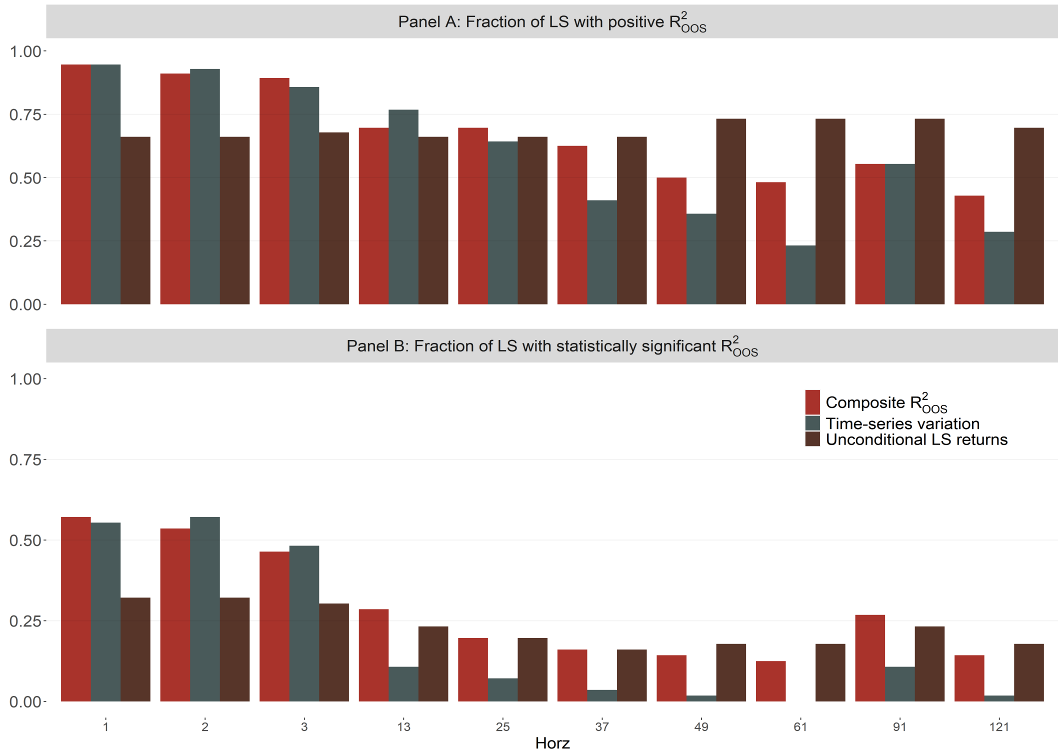

FIGURE 3.A.3: **Forecasting returns of long-short portfolios using forecasts for different horizons.**
This figure reports statistics for long-short portfolios formed from sorts on 56 characteristics listed in Table 3.C.1. We report in panel A the fraction of 56 long-short portfolios where the neural network forecast has a positive R^2_{OOS} with respect to the historical portfolio return computed from an expanding window mean with data from 1965. Panel B reports the fraction of 56 long-short portfolios where the forecasts from the neural network model are statistically better than the forecasts from the zero-prediction model at the 5% level using the the Clark-West (2007) test. For each prediction $t-h$, we report fractions that pertain to results for the composite R^2_{OOS}, the contributions coming from better predicting time-series variations in return (grey), and improvements coming from predicting the unconditional portfolio return (brown). The sample period is from January 1995 to December 2018.

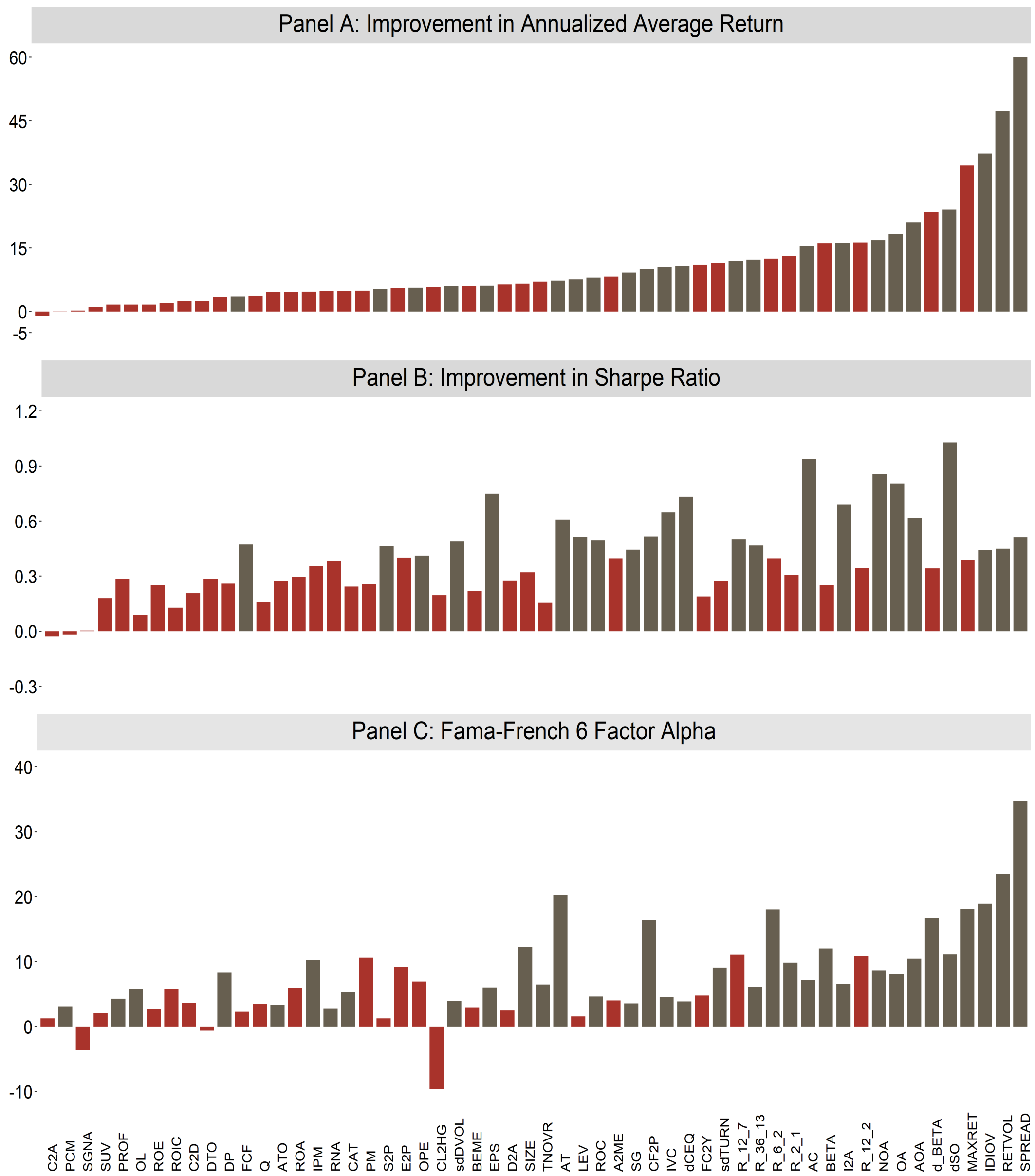

FIGURE 3.A.4: **Timing long-short characteristic sorted portfolios**
This figure reports the performance statistics of a portfolio that times each of the 56 long-short portfolios defined in Figure 3.A.2 and the risk-free asset. For Panels A and B, we report the performance statistics for the difference in returns between holding the timing portfolio and the buy and hold alternative. Panel A reports the annualized average return, Panel B reports the Sharpe ratio improvement, and panel C reports the Fama and French, 2018 6 factor model alpha (annualized) from regressing the returns of the timing portfolio on the pricing factors. Grey bars in Panels A and B represent long-short portfolios for which the improvement in returns is statistically significant at the 5% level. Grey bars in Panel C represent long-short portfolio returns with a statically significant alpha at the 5 % level with respect to the Fama-French 6 factor model. The sample period is from January 1995 to December 2018.

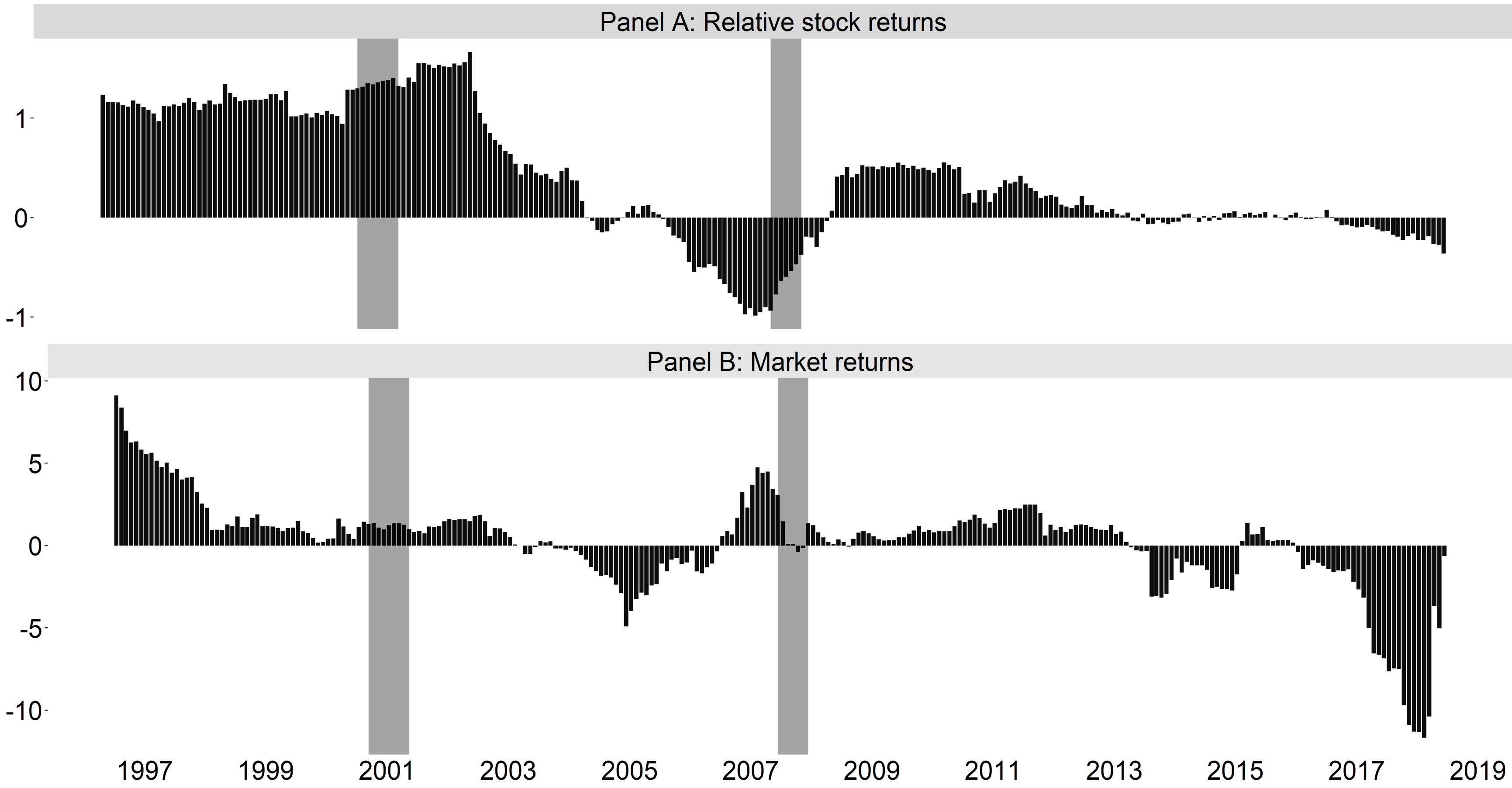

FIGURE 3.A.5: **Two-year rolling out-of-sample R^2**
This figure reports a two-year rolling out-of-sample R^2 (R^2_{OOS}) for demeaned monthly return forecasts from the economically restricted neural network model. The rolling estimates capture the changes in the neural network forecasts' ability to predict time-series variations in stock returns. In panel A, we report this estimate for the pool of stocks, and in panel B, we report this estimate for the value-weighted market. The alternative model is a zero prediction model and the period is $t+1$. The sample period is from January 1995 to December 2018.

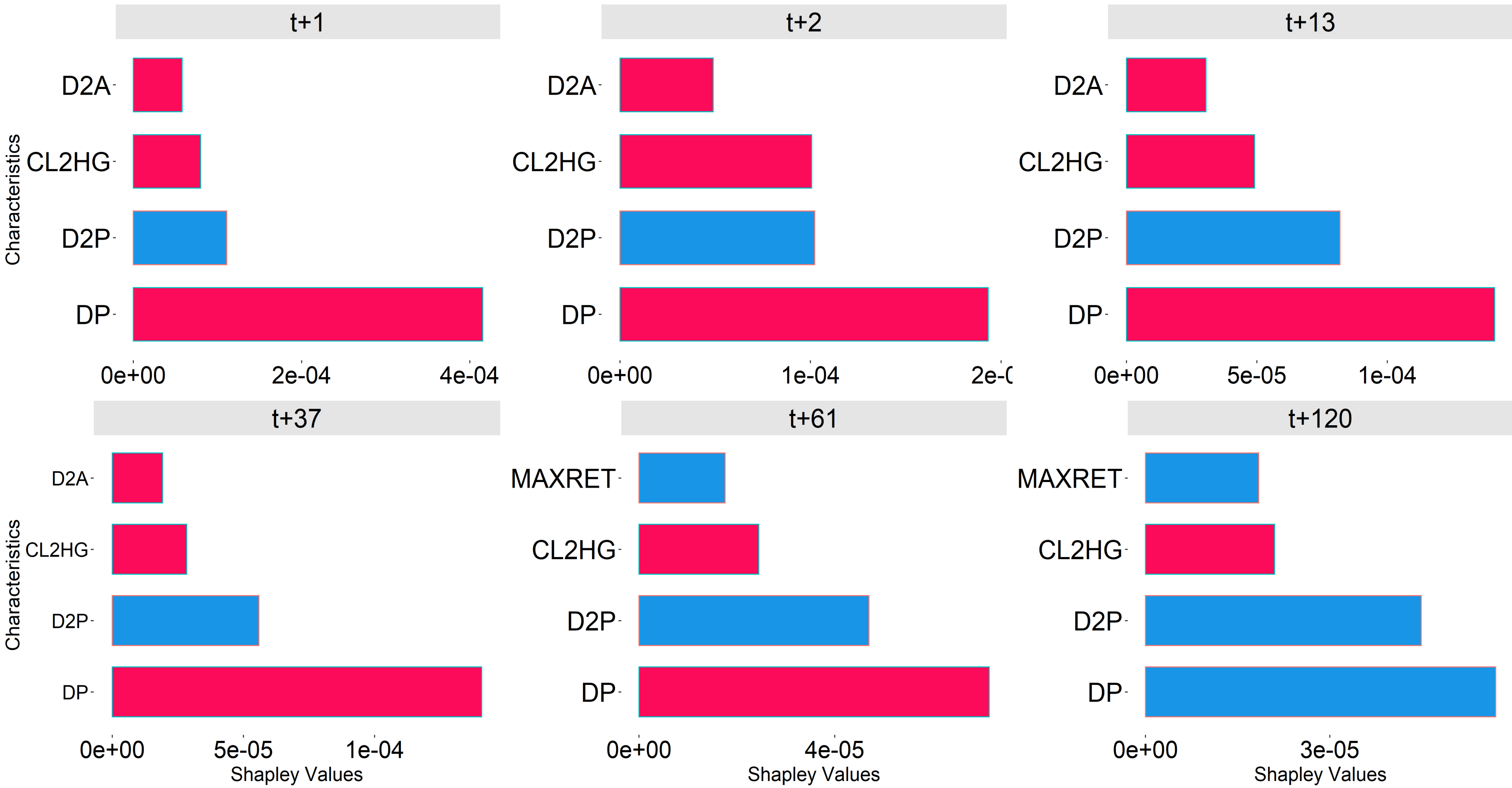

FIGURE 3.A.6: **Variable importance with Shapley values**
This figure reports the top four variables that increase return forecasts the most. A one-unit increase in a variable changes the return forecast by the corresponding Shapley value. A red (blue) bar represents a positive (negative) change in return forecasts. Shapley values are computed for each firm-month observation in the out-of-sample period and averaged over all observations. The sample period is from January 1995 to December 2018.

3.A.4 Tables

TABLE 3.A.1: **Predicting stock returns across horizons**
This table presents R^2_{OOS} estimates for firm-level return forecasts generated at time t for month $t+h$ where $h \in \{1, 2, 3, 13, 37, 61, 91, 121\}$. The R^2_{OOS} is defined with respect to a zero prediction benchmark. Panel A reports the R^2_{OOS} for a neural network forecasting model that adheres strictly to economic theory as defined in Equation (3.1.2) and a simpler neural network forecasting model with no economic restrictions as defined in Eq. (3.1.8). Panel B reports the decomposition of the R^2_{OOS} into three parts. The time-series variation dimension captures the forecasting model's ability to improve on forecasts from the benchmark model with respect to predicting time-series variation in relative stock returns. The unconditional relative stock return dimension captures the forecasting model's ability to improve upon on forecasts from the benchmark model with respect to predicting the unconditional relative stock return. Finally, the cross-sectional mean return dimension captures the forecasting model's ability to improve upon forecasts from the benchmark model with respect to predicting the level factor in stock returns; the equally-weighted market return. Bold fonts indicate horizons for which the forecasts from the neural network model and alternative model are statistically different at the 5% level or better using the Clark-West (2007) test. The sample period is from January 1995 to December 2018.

h	1	2	3	13	37	61	91	121
Panel A: Composite R^2_{OOS}								
Economically restricted model	**0.99**	**0.49**	**0.40**	**0.35**	**0.28**	**0.28**	**0.24**	**0.13**
Simple model	**0.58**	**0.41**	**0.40**	**0.37**	**0.35**	**0.34**	**0.27**	**0.18**
Panel B: Decomposed R^2_{OOS}								
	Economically restricted model							
Time-series variation	**0.71**	**0.16**	**0.09**	**0.08**	-0.06	-0.05	-0.07	-0.09
Unconditional rel. stock ret.	-0.16	-0.03	-0.02	-0.08	0.03	0.02	0.04	0.03
Cross-sectional mean return	**0.43**	**0.36**	**0.33**	**0.36**	**0.32**	**0.31**	**0.27**	**0.20**
	Simple model							
Time-series variation	**0.20**	**0.01**	**0.02**	**0.03**	-0.04	-0.04	-0.07	-0.07
Unconditional rel. stock ret.	-0.04	-0.01	-0.02	-0.05	-0.01	-0.00	0.00	0.00
Cross-sectional mean return	**0.41**	**0.41**	**0.41**	**0.40**	**0.40**	**0.39**	**0.33**	**0.24**

TABLE 3.A.2: **Historical average market return benchmark**
This table is similar to Table 3.A.1 but the benchmark model is the historical average equally-weighted market return computed from an expanding window using data from 1926. Panel A reports the R^2_{OOS} for a neural network forecasting model that adheres strictly to economic theory as defined in Equation (3.1.2), and panel B reports results for a simpler neural network forecasting model with no economic restrictions as defined in Eq. (3.1.8). Bold fonts indicate horizons for which the forecasts from the neural network model and alternative model are statistically different at the 5% level or better using the Clark-West (2007) test. The sample period is from January 1995 to December 2018.

h	1	2	3	13	37	61	91	121
Panel A: Economically restricted model								
Composite R^2_{OOS}	**0.64**	**0.14**	**0.04**	0.00	**0.05**	**0.16**	**0.37**	**0.62**
Time-series variation	**0.71**	**0.16**	**0.09**	**0.08**	-0.06	-0.05	-0.07	-0.09
Unconditional rel. stock ret.	-0.16	-0.03	-0.02	-0.08	0.03	0.02	0.04	0.03
Cross-sectional mean return	**0.08**	**0.01**	-0.02	**0.01**	**0.08**	**0.19**	**0.40**	**0.69**
Panel B: Simple model								
Composite R^2_{OOS}	**0.20**	**0.03**	**0.01**	-0.00	**0.04**	**0.17**	**0.32**	**0.55**
Time-series variation	**0.20**	**0.01**	**0.02**	**0.03**	-0.04	-0.04	-0.07	-0.07
Unconditional rel. stock ret.	-0.04	-0.01	-0.02	-0.05	-0.01	-0.00	0.00	0.00
Cross-sectional mean return	**0.03**	**0.03**	**0.02**	**0.02**	**0.09**	**0.21**	**0.39**	**0.61**

TABLE 3.A.3: **Predicting market returns across horizons**
This table presents out-of-sample (R^2_{OOS}) estimates for market forecasts for different months in the future $t+h, h \in \{1, 2, 3, 13, 37, 61, 91, 121\}$, conditional on information observed at time t. Panel A reports results for a neural network forecasting model that adheres strictly to economic theory as defined in Equation (3.1.2). Panel B reports results for a simpler neural network forecasting model with no economic restrictions as defined in Eq. (3.1.8). The market forecasts are defined as the value-weighted cross-sectional average stock forecasts and the market return as the value-weighted cross-sectional average stock return. I report results for two benchmark models, a zero-prediction model that predicts a return of zero for all horizons and the historical equity market benchmark that forecasts market return as the average market return using data from July 1926 upto time t. I decompose the R^2_{OOS} into contributions coming from the neural network forecasts ability to explain better time-series variation (Time-series variation) and the unconditional market return (Uncond. market ret.). Bold fonts indicate horizons for which the forecasts from the neural network model and alternative model are statistically different at the 5% level or better using the Clark-West (2007) test. The sample period is from January 1995 to December 2018.

h	1	2	3	13	37	61	91	121
Panel A: Economically restricted model								
	Zero prediction benchmark							
Composite R^2_{OOS}	**5.35**	**5.30**	**5.17**	**4.90**	**4.42**	**3.05**	**3.48**	**2.77**
Time-series variation	**1.92**	**1.82**	**1.69**	**1.51**	**1.10**	-0.03	0.19	-0.42
Uncond. market ret.	**3.42**	**3.49**	**3.49**	**3.39**	**3.31**	**3.08**	**3.30**	**3.19**
	Historical average market benchmark							
Composite R^2_{OOS}	**2.05**	**2.03**	**1.89**	**1.64**	1.09	-0.40	0.13	-0.63
Time-series variation	**2.06**	**1.97**	**1.84**	**1.68**	**1.21**	-0.04	0.26	-0.39
Uncond. market ret.	-0.01	0.06	0.05	-0.04	-0.12	-0.36	-0.13	-0.24
Panel B: Simple model								
	Zero prediction benchmark							
Composite R^2_{OOS}	**1.43**	**2.53**	**2.37**	**3.08**	**3.65**	**4.14**	**3.47**	**3.72**
Time-series variation	0.21	0.48	0.26	0.54	0.60	0.70	0.01	0.03
Uncond. market ret.	**1.23**	**2.06**	**2.10**	**2.54**	**3.05**	**3.44**	**3.46**	**3.69**
	Historical average market benchmark							
Composite R^2_{OOS}	-2.29	-1.13	-1.31	-0.52	0.04	0.47	-0.16	0.07
Time-series variation	0.29	0.59	0.36	0.70	0.73	0.75	0.10	0.09
Uncond. market ret.	-2.59	-1.72	-1.67	-1.22	-0.69	-0.28	-0.25	-0.02

TABLE 3.A.4: **Predicting long-short characteristic sorted portfolio returns (Simple model)**
This table reports the out-of-sample R^2 (R^2_{OOS}) estimates for monthly rebalanced long-short portfolios formed from sorts on book-to-market (BEME), investments (INV), size (SIZE), mom (Momentum), and profitability (PROF). The results are for the simple neural network forecasting model, as defined in Equation (3.1.8). The benchmark model is the zero prediction model. Bold fonts indicate horizons for which the forecasts from the simple neural network model and alternative model are statistically different at the 5% level or better using the Clark-West (2007) test. The sample period is from January 1995 to December 2018.

h	1	2	3	13
	BEME			
Composite R^2_{OOS}	-1.94	-1.53	-2.10	-0.78
Time-series variation	-2.64	-1.33	-1.48	-0.32
Unconditional LS Ret.	0.70	-0.20	-0.62	-0.46
	INV			
Composite R^2_{OOS}	-0.78	**2.91**	**3.60**	0.41
Time-series variation	0.15	0.89	0.97	0.63
Unconditional LS Ret.	-0.92	**2.02**	**2.62**	-0.22
	SIZE			
Composite R^2_{OOS}	0.44	0.61	0.48	-0.15
Time-series variation	0.45	0.88	0.89	0.11
Unconditional LS Ret.	-0.01	-0.27	-0.41	-0.26
	MOM			
Composite R^2_{OOS}	-1.47	-0.98	-2.60	-2.29
Time-series variation	-1.24	-0.53	-1.31	-0.96
Unconditional LS Ret.	-0.23	-0.44	-1.29	-1.34
	PROF			
Composite R^2_{OOS}	-0.56	0.14	0.28	1.23
Time-series variation	-0.74	-0.06	0.28	1.00
Unconditional LS Ret.	0.18	0.20	0.00	0.23

TABLE 3.A.5: **Predicting long-short portfolio returns (Economically restricted model)**
This table reports the out-of-sample R^2 (R^2_{OOS}) estimates for monthly rebalanced long-short portfolios formed from sorts on book-to-market (BEME), investments (INV), size (SIZE), mom (Momentum), and profitability (PROF). The results are for the neural network forecasting model that adheres strictly to economic theory as defined in Equation (3.1.2). The benchmark model is the historical average return of the long-short portfolio computed from an expanding window using data from 1964. Bold fonts indicate horizons for which the forecasts from the neural network model and alternative model are statistically different at the 5% level or better using the Clark-West (2007) test. The sample period is from January 1995 to December 2018.

h	1	2	3	13	37	61	91	121
				BEME				
Composite R^2_{OOS}	0.68	0.80	0.73	0.81	-0.30	-0.17	-0.34	-0.23
Time-series variation	0.52	0.62	0.51	-0.03	-0.30	-0.13	-0.20	0.01
Unconditional LS Ret.	0.15	0.18	0.22	0.84	-0.00	-0.04	-0.14	-0.24
				SIZE				
Composite R^2_{OOS}	**1.74**	**1.48**	**1.32**	**0.71**	0.14	0.01	-0.45	-0.56
Time-series variation	**1.66**	**1.43**	**1.28**	**0.78**	0.10	-0.06	-0.41	-0.44
Unconditional LS Ret.	0.08	0.05	0.04	-0.08	0.04	0.07	-0.04	-0.12
				PROF				
Composite R^2_{OOS}	**0.95**	**0.97**	0.75	0.55	-0.06	0.07	0.24	-0.12
Time-series variation	**0.67**	**0.72**	**0.51**	0.28	-0.03	-0.05	0.15	-0.14
Unconditional LS Ret.	0.28	0.24	0.25	0.27	-0.03	0.12	0.09	0.01
				INV				
Composite R^2_{OOS}	**3.34**	**3.49**	**3.23**	**1.93**	0.30	0.17	0.00	0.36
Time-series variation	**1.81**	**1.95**	**1.67**	0.56	0.07	-0.13	-0.09	-0.01
Unconditional LS Ret.	**1.53**	**1.55**	**1.56**	**1.37**	**0.23**	0.30	0.10	0.37
				MOM				
Composite R^2_{OOS}	**1.34**	**0.84**	0.43	**0.41**	-0.21	-0.10	-0.25	-0.50
Time-series variation	0.22	0.18	0.22	0.28	0.02	-0.02	-0.04	-0.14
Unconditional LS Ret.	**1.12**	**0.66**	0.20	**0.14**	-0.24	-0.08	-0.21	-0.37

TABLE 3.A.6: **Time-series regression of returns on forecasts**
This table reports the results from a time-series regression of returns on the economically restricted model's demeaned forecasts. Panel A reports results from regressing stock returns on demeaned relative stock forecasts. Panel B reports results from regressing the value-weighted market return on demeaned value-weighted market forecasts. Panels C and D report results from regressing long-short portfolio returns on demeaned long-short portfolio forecasts. β_0 and β_1 are the intercept and slope coefficients from these regressions. For each of the estimates, we report the 95% confidence interval constructed from Newey-West 1987 corrected standard errors. The sample period is from January 1995 to December 2018.

h	1	2	3	13	25	37	49	61	91	121
					Panel A: Pool of stocks					
β_0	1.08	1.08	1.08	1.02	0.93	0.83	0.78	0.72	0.59	0.47
(β_0^l, β_0^u)	(1.05,1.11)	(1.05,1.11)	(1.05,1.11)	(1.00,1.05)	(0.90,0.96)	(0.81,0.86)	(0.76,0.80)	(0.70,0.74)	(0.57,0.60)	(0.46,0.49)
β_1	0.97	0.80	0.68	0.79	0.24	-0.02	0.03	0.11	-0.12	-0.04
(β_1^l, β_1^u)	(0.94,1.00)	(0.76,0.84)	(0.64,0.72)	(0.74,0.85)	(0.19,0.30)	(-0.09,0.05)	(-0.04,0.09)	(0.04,0.17)	(-0.19,-0.05)	(-0.10,0.02)
R^2	0.71	0.17	0.11	0.09	0.01	0.00	0.00	0.00	0.00	0.00
$NOS.$	1507557	1508884	1510083	1519230	1525769	1529073	1532204	1536016	1543943	1535857
					Panel B: Value-weighted market					
β_0	0.79	0.79	0.79	0.79	0.79	0.79	0.79	0.79	0.79	0.79
(β_0^l, β_0^u)	(0.28,1.30)	(0.28,1.30)	(0.28,1.30)	(0.28,1.30)	(0.28,1.30)	(0.28,1.30)	(0.28,1.30)	(0.27,1.30)	(0.27,1.30)	(0.27,1.30)
β_1	1.78	1.58	1.34	1.69	1.45	1.63	1.66	0.42	0.44	-0.22
(β_1^l, β_1^u)	(0.44,3.12)	(0.24,2.92)	(0.09,2.60)	(0.23,3.16)	(0.11,2.79)	(0.01,3.24)	(-0.24,3.57)	(-1.61,2.45)	(-1.14,2.02)	(-2.17,1.73)
R^2	2.18	1.79	1.53	1.53	1.49	1.05	0.82	0.05	0.08	0.02
$NOS.$	282	282	282	282	282	282	282	282	282	282
					Panel C: Pool of long-short portfolios					
β_0	0.31	0.31	0.29	0.28	0.25	0.23	0.23	0.27	0.26	0.13
(β_0^l, β_0^u)	(-0.04,0.65)	(-0.03,0.65)	(-0.05,0.63)	(-0.05,0.61)	(-0.08,0.57)	(-0.09,0.56)	(-0.08,0.55)	(-0.05,0.58)	(-0.05,0.56)	(-0.16,0.42)
β_1	1.96	2.20	2.07	2.29	2.12	0.35	-0.73	-0.98	-0.99	-2.93
(β_1^l, β_1^u)	(0.88,3.04)	(1.19,3.21)	(0.98,3.16)	(0.67,3.90)	(-0.34,4.58)	(-2.87,3.56)	(-4.67,3.22)	(-4.67,2.71)	(-4.32,2.34)	(-6.34,0.48)
R^2	1.74	1.81	1.48	0.66	0.23	0.00	0.01	0.02	0.03	0.22
$NOS.$	282	282	282	282	282	282	282	282	282	282

Continued

h	1	2	3	13	25	37	49	61	91	121
Panel D: Individual long-short portfolios										
BEME										
β_1	1.39	2.73	2.45	0.57	-1.69	-8.14	-9.46	-3.97	-10.64	-0.21
(β_1^l, β_1^u)	(-1.33,4.11)	(-1.12,6.58)	(-1.69,6.60)	(-4.86,6.01)	(-10.29,6.90)	(-18.32,2.05)	(-18.59,-0.34)	(-11.92,3.98)	(-22.79,1.51)	(-13.18,12.77)
R^2	0.56	1.04	0.77	0.02	0.09	1.16	1.62	0.27	0.95	0.00
SIZE										
β_1	4.99	4.36	3.80	6.55	4.20	3.67	3.49	-0.93	-6.85	-6.17
(β_1^l, β_1^u)	(1.98,8.00)	(1.55,7.17)	(1.11,6.49)	(1.20,11.91)	(-1.67,10.07)	(-5.70,13.04)	(-7.75,14.74)	(-9.19,7.32)	(-13.13,-0.56)	(-12.56,0.22)
R^2	4.60	3.51	2.80	2.83	0.78	0.22	0.10	0.01	1.29	1.23
PROF										
β_1	4.72	5.40	4.65	3.44	2.95	0.38	2.44	-1.60	5.85	-4.49
(β_1^l, β_1^u)	(0.46,8.98)	(1.42,9.38)	(0.11,9.18)	(-1.24,8.12)	(-2.56,8.47)	(-6.85,7.60)	(-8.36,13.24)	(-10.77,7.57)	(-3.72,15.42)	(-15.44,6.45)
R^2	1.75	2.13	1.32	0.59	0.26	0.00	0.06	0.03	0.51	0.27
INV										
β_1	4.03	5.27	4.66	3.01	2.94	1.52	-2.55	-2.52	-1.26	0.31
(β_1^l, β_1^u)	(1.50,6.57)	(2.38,8.16)	(1.66,7.65)	(-0.75,6.76)	(-2.07,7.95)	(-3.95,7.00)	(-9.43,4.34)	(-8.61,3.57)	(-8.27,5.75)	(-8.96,9.58)
R^2	4.25	5.85	4.53	1.01	0.46	0.08	0.18	0.13	0.04	0.00
MOM										
β_1	1.41	2.40	1.85	4.25	4.92	3.88	3.96	-1.31	-1.58	-9.04
(β_1^l, β_1^u)	(-1.88,4.70)	(-1.74,6.54)	(-2.56,6.25)	(-2.13,10.62)	(-4.43,14.27)	(-10.24,18.01)	(-12.25,20.17)	(-18.03,15.41)	(-17.78,14.62)	(-25.93,7.85)
R^2	0.34	0.50	0.28	0.67	0.38	0.10	0.09	0.01	0.02	0.39
$NOS.$	282	282	282	282	282	282	282	282	282	282

TABLE 3.A.7: **Timing Strategies**
This table reports performance statistics for portfolios formed from sorts on neural network forecasts. I report the annualized average return (Avg. ret), annualized Sharpe ratio (Sharpe), the annualized certainty equivalent (Utility) in percentages for an agent with a mean-variance utility function, and a risk aversion parameter of 2, the CAPM and Fama-French 6 factor model alphas. The two timing strategies invests: $w_{t,h} = \frac{\tilde{r}_{t,h} - r^f_{t+1}}{\gamma\sigma^2(\tilde{r}_{1:t-1,h})}$ in a risky security and $1 - w_{t,h}$ in the risk-free asset. I fix $\sigma(\tilde{r}_{1:t-1,h})$ at $\frac{0.15}{\sqrt{12}}$ to make the results comparable across forecasts. Panel A reports results for a strategy where the expected return on the value-weighted market ($\tilde{r}_{t,h}$) is computed as a function of different forecasts ($h \in \{1, 2, 3, 13, 25, 37, 49, 61, 121\}$) from the neural network model. We restrict the cross-section to the 500 largest stocks. The buy and hold strategy buys all 500 stocks in this restricted cross-section for each month t. Panel B reports results for a strategy where the expected return on an equally-weighted portfolio of the five long-short portfolios (value-weighted) formed from the characteristics in the Fama and French, 2018 model. The buy and hold strategy buys all five portfolios in each month t whereas the timing strategies lever up and down this portfolio based on expected next month returns. The sample period is 1995 to 2018.

	Utility	Avg. ret	Sharpe	α_{capm}	α_{FF6}
		Panel A: Market			
Buy & hold	8.18	**10.39**	0.70	0.23	0.04
h					
1	12.86	**17.21**	0.83	**8.41**	4.14
2	12.47	**15.79**	0.87	**8.53**	3.49
3	12.76	**15.70**	0.92	**8.95**	3.40
13	10.54	**12.33**	0.92	**7.31**	**4.64**
25	10.60	**12.61**	0.89	**7.43**	**4.77**
37	9.19	**10.39**	0.95	**5.76**	**4.53**
49	7.75	**8.83**	0.85	**4.50**	**3.97**
61	5.31	**6.66**	0.57	3.01	2.03
91	7.04	**8.15**	0.77	**3.42**	**3.37**
121	4.77	**6.86**	0.47	2.97	4.12
		Panel B: Equally-weighted portfolio of long-short portfolios			
Buy & hold	3.02	**3.88**	0.42	**5.73**	1.70
h					
1	5.04	**6.72**	0.52	**9.05**	**5.28**
2	4.78	**6.22**	0.52	**8.05**	**4.48**
3	4.40	**5.78**	0.49	**7.57**	**3.90**
13	3.48	**3.82**	0.65	**4.79**	**2.76**
25	2.78	**2.92**	0.79	**3.49**	**2.39**
37	2.33	**2.40**	0.95	**2.78**	**1.94**
49	2.39	**2.44**	1.12	**2.79**	**1.91**
61	2.39	**2.43**	1.25	**2.68**	**2.07**
91	2.14	**2.17**	1.13	**2.44**	**1.65**
121	1.62	**1.64**	1.13	**1.78**	**1.25**

TABLE 3.A.8: **Rotation strategies**
This table reports performance statistics, as in Table 3.A.8. The strategy in panel A goes long (short) the value-weighted portfolio of the top (bottom) 10% of stocks with the highest (lowest) forecasted return for each month t. I restrict the cross-section of stocks to the 500 largest market-cap firms at each time t. The buy and hold strategy buys all 500 stocks in this restricted cross-section for each month t. In Panel B, the strategy buys (sells) an equally-weighted portfolio of the two characteristic-sorted long-short portfolios with an expected return above (below) the median forecast for each month t. The buy and hold strategy buys all long-short portfolios in each month t. I restrict the characteristic-sorted long-short portfolios to the five Fama and French, 2018 characteristics. To highlight the importance of timely conditioning information, we consider variants of the strategy where the expected month t stock return, $\mathbb{E}_{t-h}[R_t]$, comes from different horizons, $h \in \{1, 2, 3, 13, 25, 37, 49, 61, 121\}$. The sample period is 1995 to 2018.

	Utility	Avg. ret	Sharpe	α_{capm}	α_{FF6}
		Panel A: Stocks			
Buy & hold	8.18	**10.39**	0.70	0.23	0.04
Rotation (h)					
1	11.73	**15.82**	0.78	**21.19**	**11.05**
2	8.74	**13.89**	0.61	**20.01**	**8.37**
3	3.98	**9.15**	0.40	**15.27**	3.46
13	0.26	6.26	0.26	**11.22**	0.07
25	-0.76	3.69	0.18	7.75	0.33
37	-1.60	0.45	0.03	2.02	-0.04
49	-1.43	0.62	0.04	2.08	1.80
61	0.85	2.74	0.20	4.99	2.46
91	-0.57	1.50	0.10	2.89	1.63
121	-0.14	1.43	0.11	0.96	0.17
		Panel B: Long-short portfolios			
Buy & hold	3.02	**3.88**	0.42	**5.73**	1.70
Rotation (h)					
1	10.45	**17.06**	0.66	**24.51**	**9.94**
2	12.77	**19.41**	0.75	**26.36**	**12.47**
3	9.80	**16.65**	0.64	**23.88**	**10.15**
13	3.01	8.96	0.37	**14.27**	4.20
25	2.04	7.62	0.32	**11.90**	5.79
37	-5.75	-1.06	-0.05	0.25	-1.46
49	-6.33	-1.27	-0.06	0.65	-0.91
61	-6.85	-2.50	-0.12	-0.62	-3.48
91	-8.01	-3.80	-0.19	-2.48	-4.38
121	-7.00	-3.22	-0.17	-2.40	-2.96

3.B Internet Appendix

3.B.1 Figures

3.B.2 Tables

TABLE 3.B.1: **Alternative benchmarks**
This table presents out-of-sample R^2 (R^2_{OOS}) estimates for firm-level return forecasts generated by neural network models at time t for month $t+h$ where $h \in \{1, 2, 3, 13, 25, 37, 49, 61, 121\}$. Panel A reports the R^2_{OOS} for a neural network forecasting model that adheres strictly to economic theory as defined in Equation (3.1.2), and panel B reports results for a simpler neural network forecasting model with no economic restrictions as defined in Eq. (3.1.8). I present the R^2_{OOS} for the neural network model against two alternative models; a five-year rolling window firm average return model, and an expanding window firm average return model. Bold fonts indicate horizons for which the forecasts from the neural network model and alternative model are statistically different at the 5% level or better using the Clark-West (2007) test. The sample period is from January 1995 to December 2018.

h	1	2	3	13	37	61	91	121
Panel A: Economically restricted model								
Five-Year Rolling Window	**2.18**	**1.78**	**1.71**	**1.75**	**1.34**	**1.54**	**1.54**	**1.58**
Expanding Window	**1.26**	**0.81**	**0.72**	**0.74**	**0.58**	**0.59**	**0.62**	**0.63**
Panel B: Simple model								
Five-Year Rolling Window	**1.76**	**1.64**	**1.62**	**1.58**	**1.28**	**1.51**	**1.42**	**1.38**
Expanding Window	**0.91**	**0.75**	**0.72**	**0.70**	**0.59**	**0.68**	**0.65**	**0.64**

3.C Variable Construction

TABLE 3.C.1: **Characteristics**
This table lists the characteristics used in this paper. For each characteristic, we present the associated acronym, the original source and the definition of the characteristic.

Acronym	Author(s)	Definition
A2ME	Bhandari, 1988	Total assets (at) over market capitalization (prc x shrout)
AC	Sloan, 1996	Change in operating working capital per split adjusted share from fiscal year $t-2$ to $t-1$ to book equity, (BEME), per share. Operating working capital per split-adjusted share is defined as current assets (ACT) minus cash and short-term investments (che) minus current liabilites (lct) minus debt in current liabilities (dlc) minus income taxes payable (txp).
AOA	Bandyopadhyay et al., 2010	Absolute value of OA
AT	Gandhi and Lustig, 2015	Total assets (at)
ATO	Soliman, 2008	Net sales (sales) over lagged net operating assets. Net operating assets is the difference between operating assets and operating liabilities. Operating Assets is total assets (at) minus cash and short-term investments (che) minus investments and other advances (ivao). Operating Liabilities is total assets (at) minus debt in current liabilities (dlc) minus long-debt debt (dltt) minus minority interest (mib) minus preferred stock (pstk) minus common equity (ceq).
BEME (BM)	Davis et al., 2000b	Book equity to market equity. Book equity is shareholders' equity (seq), (if missing, common equity (ceq) plus preferred stock (pstk), if missing, total assets (at) minus total liabilities (lt)), plus deferred taxes and investment tax credit (txditc) minus preferred stock (pstrkrv), (if missing, liquidation value, (pstkl), if missing par value (pstk)). Market value of equity is shares outstanding (shrout) times price (prc).
BETA	Frazzini and Pedersen, 2014b	The product of the correlation between stock excess returns and market excess returns and the ratio of volatilities. Ratio of volatilities is the volatility of stock excess returns to the volatility of market excess returns. Volatility is computed from the standard deviations of daily log excess returns over a one-year horizon requiring at least 120 observations. Correlations is computed using overlapping three-day log excess returns over a five-year period requiring at least 750 non-missing observations.
$BETA_d$	Lewellen and Nagel, 2006	The sum of the regression coefficients of daily excess returns on market excess returns and the lag of market excess returns.
C2A	Palazzo, 2012	Cash and short-term investments (che) to total assets (at).
C2D	Ou and Penman, 1989	Cashflow to debt. Cashflow is the sum of income and extraordinary items (ib) and depreciation and amortization (dp). And debt is to total liabilities (lt).

Continued

Acronym	Author(s)	Definition
CAT	Haugen and Baker, 1996	Sales (sale) to lagged total assets (at).
CF2P	Desai et al., 2004	Cashflow to book value of equity is the ratio of net income (ni), depreciation and amortization (dp) less change in working capital (wcapch) and capital expenditure (capx) over the book-value of equity (BEME).
CL2HG	George and Hwang, 2004	ratio of last month closing price to the max closing price over the last 52 weeks.
D2A	Gorodnichenko and Weber, 2016	Depreciation and amortization (dp) to total assets (at).
D2P	Litzenberger and Ramaswamy, 1979	Debt to price. Debt is long-term debt (dltt) plus debt in current liabilities (dlc). Market capitalization is the product of shares outstanding (shrout) and price (prc).
dCEQ	Richardson et al., 2005	Annual % change in book value of equity (ceq).
dGS	Abarbanell and Bushee, 1997	% change in gross margin minus % change in sales (sale). Gross margin is the difference in sales (sale) and cost of goods sold (cogs).
dPIA	Lyandres et al., 2008	Change in property, plants and equipment (ppegt) and inventory (invt) over lagged total assets (at).
dSO	Fama and French, 2008	Log change in the product of shares outstanding (csho) and the adjustment factor (ajex).
dSOUT	Pontiff and Woodgate, 2008	Annual % change in shares outstanding (shrout).
DP	Litzenberger and Ramaswamy, 1979	Sum of monthly dividend over the last 12 months to last month's price (prc).
DTO	Garfinkel, 2009	Daily volume (vol) to shares outstanding (shrout) minus the daily market turnover and de-trended by the 180 trading day median. To address the double counting of volume for NASDAQ securities, we follow Anderson and Dyl, 2005 and scale down the volume of NASDAQ securities by 50% before and by 38% after 1997.
E2P	Basu, 1983	Income before extraordinary items (ib) to market capitalization (prc x shrout).
EPS	Basu, 1977	Income before extraordinary items (ib) to shares outstanding (shrout).
FC2Y	D'Acunto et al., 2018	Ratio of selling, general and administrative expenses (xsgs), research and development expenses (xrd) and advertising expenses (xad) to net sales.
FCF	Hou et al., 2011	Ratio of net income (ni), depreciation and amortization (dp) less change in working capital (wcapch) and capital expenditure (capx) over book value of equity as defined in BEME.
I2A (INV)	Cooper et al., 2008	Annual % change in total assets (at).
IDIOV	Ang et al., 2006	Standard deviation of the residuals from a regression of excess returns on the Fama and French, 1993b three-factor model.
IPM		Pre-tax income (pi) over sales (sale).
IVC	Thomas and Zhang, 2002	Annual change in inventories (invt) in the last two fiscal years over the average total assets (at) over the last two fiscal years.
LEV	Lewellen, 2015	long-term debt (dltt) plus current liabilities (dlc) over the sum of long term debt (dltt), debt in current liabilities (dlc) and stockholders equity (seq).

Continued

Acronym	Author(s)	Definition
MAXRET	Bali et al., 2011	Last months stock price (prc) over previous 52 week max price.
NOA	Hirshleifer et al., 2004	Operating assets minus operating liabilities to lagged total assets (at). Operating assets is total assets (at) minus cash and short term investments (che) minus investment and other advances (ivao). Operating liabilities is total assets (at) minus debt in current liabilities (dlc) minus long-term debt (dltt) minus minority interest (mib) minus preferred stock (pstk) minus common equity (ceq).
OA	Sloan, 1996	Changes in non-cash working capital minus depreciation (dp), all scaled by lagged total assets (at). Changes in non-cash working capital is difference in current assets (act) minus difference in cash and short-term investments (che) minus difference in current liabilities (lct) minus difference in debt in current liabilities (dlc) minus difference in taxes payable (txp).
OL	Novy-Marx, 2011	Sum of cost of goods sold (cogs) and selling, general and administrative expense (xsga) over total assets (at).
PCM	Gorodnichenko and Weber, 2016	Net sales (sale) minus cost of goods sold (cogs) all scaled by net sales (sale).
PM	Soliman, 2008	Operating Income after depreciation (oiadp) to sales (sale).
PROF	Ball et al., 2015	Gross profitability (gp) over book equity as defined in $BEME$.
Q		Total assets (at) plus market value of equity (shrout x prc) minus cash and short-term investments (ceq) minus deferred taxes (txdb) scaled by total assets (at).
$R_{12,2}$	Fama and French, 1996b	Cumulative return from 12 months to 2 months ago.
$R_{12,7}$	Novy-Marx, 2012	Cumulative return from 12 months to 7 months ago.
$R_{2,1}$	Jegadeesh, 1990	Lagged one month return.
$R_{36,13}$	De Bondt and Thaler, 1985	Cumulative return from 36 months to 13 months ago.
$R_{6,2}$	Jegadeesh and Titman, 1993	Cumulative return from 6 months to 2 months ago.
RETVOL	Ang et al., 2006	Standard deviation of residuals from a regression of excess returns on a constant using one month of daily data. We require there to be at least 15 non-missing observations.
RNA	Soliman, 2008	Operating income after depreciation (oiadp) scaled by lagged net operating assets. Net operating assets is operating assets minus operating liabilities. Operating assets is total assets (at) minus cash and short term investments (che) minus investment and other advances (ivao). Operating liabilities is total assets (at) minus debt in current liabilities (dlc) minus long-term debt (dltt) minus minority interest (mib) minus preferred stock (pstk) minus common equity (ceq).
ROA	Balakrishnan et al., 2010	Income before extraordinary items (ib) to lagged total assets (at).

Continued

Acronym	Author(s)	Definition
ROC	Chandrashekar and Rao, 2009	Market value of equity (shrout x prc) plus long-term debt (dltt) minus total assets (at) all over cash and short-term investments (che).
ROE	Haugen and Baker, 1996	Income before extraordinary items (ib) to lagged book-value of equity.
ROIC	Brown and Rowe, 2007	Earnings before interest and taxes (ebit) less non-operating income (nopi) to the sum of common equity (ceq), total liabilities (lt), and cash and short-term investments (che).
S2P	Lewellen, 2015	Net sales (sale) to market capitalization (shrout x prc).
sdDVOL	Chordia et al., 2001	Standard deviation of residuals from a regression of daily volume (vol) on a constant. Use one month of daily data requiring at-least 15 non-missing observations.
sdTURN	Chordia et al., 2001	Standard deviation of residuals from a regression of daily turnover on a constant. Turnover is volume (vol) times shares outstanding (shrout). Use one month of daily data requiring at-least 15 non-missing observations.
SG	Lakonishok et al., 1994	% growth rate in sales (sale).
SGNA		Selling, general and administrative expenses (XSGA) to net sales (sale).
SIZE	Fama and French, 1992b	Price (prc) times shares outstanding (shrout) .
SPREAD	Chung and Zhang, 2014	Average daily bid-ask spread in the previous month.
SUV	Garfinkel, 2009	Difference between actual volume and predicted volume. Predicted volume is from a regression of previous month's daily volume on a constant and the absolute values of positive and negative previous month's returns. Unexplained volume is standardized by the standard deviation of the residuals from the regression.
TNOVR	Datar et al., 1998	Volume (vol) over shares outstanding (shrout).

3.D Details of Algorithms

Denote the penalized loss function of the neural network model as $\mathcal{L}(\theta;\cdot)$. The standard method for finding the optimal parameters (θ^*) that minimizes $\mathcal{L}(\theta;\cdot)$ is stochastic gradient descent (SGD), Goodfellow et al., 2016. Minimizing this loss function with SGD is slow and inefficient because it is a first-order optimization procedure. In this study we use a recent variant of SGD called AdaBound, Luo et al., 2019 which uses second-order information and has theoretical convergence guarantees.

We initialize θ by sampling θ_0 from $\mathcal{N}(0, n_h^{-1})$ where n_h is the size of the input vector of layer h. A single training step t, consists of a randomly sampling 10000 firm-month observations from the training set and running Algorithm 1.

Algorithm 1: AdaBound Variant of Stochastic Gradient Descent

Initialization : $\theta_0 \sim \mathcal{N}(0, n_h^{-1})$. $\alpha = 10^{-1}$. $m_0 = 0$. $v_o = 0$;

while *θ_t not coverge* **do**

- $t \leftarrow t + 1$;
- $g_t = \nabla_\theta \mathcal{L}_t(\theta_{t-1};\cdot)$;
- $m_t = \beta_1 m_{t-1} + (1 - \beta_{1,t}) g_t$;
- $v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2$;
- $V_t = diag(v_t)$;
- $\hat{\eta}_t = Clip(\alpha/\sqrt{V_t}, \eta_l(t), \eta_u(t))$;
- $\eta_t = \hat{\eta}_t/\sqrt{t}$;
- $\theta_t = \prod_{diag(\eta_t^{-1})}(\theta_{t-1} - \eta_t \odot m_t)$;

end

Result: Final parameter estimate $\theta_{\tilde{t}}$;

where $Clip(.)$ is a clipping function that bounds the learning rate (α) to the interval $[\eta_l, \eta_u]$.

Algorithm 2: Batch Normalization

Input : Values of x for each activation over a single batch $B = \{x_1, x_2, x_3, .., x_N\}$;

$\mu_B \leftarrow \frac{1}{N} \sum_{i=1}^{N} x_i$;

$\sigma_B^2 \leftarrow \frac{1}{N} \sum_{i=1}^{N} (x_i - \mu_B)^2$;

$\hat{x}_i \leftarrow \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}}$;

$y_i \leftarrow \gamma x_i + \beta = BN_{\gamma,\beta(x_i)}$;

Result: $y_i = BN_{\gamma,\beta(x_i):i=1,2,3,\ldots,N}$;

TABLE 3.D.1: **Hyper-parameters**
This table reports all hyper-parameters in the paper and the range in which they were chosen via the validation set.

Name	Range
l_1 Penalty (λ_1)	$(10^{-1}, 10^{-6})$
Learning rate	$(10^{-2}, 10^{-4})$
Batch size	10000
Epochs	100
Adabound parameters	Default (amsbound=True)
Patience	5
Ensemble	100

3.E Decomposing the out-of-sample R^2

We start with the definition of the out-of-sample R^2 (R^2_{OOS});

$$R^2_{OOS} = 1 - \frac{\sum_{(t)\,\in\,oss}(R_t - R_{t,1})^2}{\sum_{(t)\,\in\,oss}(R_t - R_{t,2})^2} \tag{3.E.1}$$

We define the relative stock return forecast of some firm i at time t, as the time t return forecast for firm i minus the time t cross-sectional mean forecast.

Specifically, we re-define stock return forecast as follows:

$$r_{1,i,t} = \mu^{RR}_{1,i} + (N_t^{-1}) \sum_{i \in t} r_{1,i,t} + \tilde{r}^{RR}_{1,i,t} \tag{3.E.2}$$

$$r_{2,i,t} = \mu^{RR}_{2,i} + (N_t^{-1}) \sum_{i \in t} r_{1,i,t} + \tilde{r}^{RR}_{2,i,t} \tag{3.E.3}$$

This definition allows model one to improve upon forecast from model two along three possible dimensions. First, model one can improve the forecasts of model two with respect to predicting time-series variation in relative stock returns; $\tilde{r}^{RR}_{1,i,t}$ vs. $\tilde{r}^{RR}_{2,i,t}$. Second, the improvement maybe in predicting better the unconditional relative stock returns; $\mu^{RR}_{1,i}$ vs. $\mu^{RR}_{2,i}$. Finally, the improvement maybe in predicting the cross-sectional mean better; $(N_t^{-1}) \sum_{i \in t} r_{1,i,t}$ vs. $(N_t^{-1}) \sum_{i \in t} r_{1,i,t}$. The decomposition therefore allows us to directly pin-point along which dimension one forecaster is doing better than another.

We claim that:

$$R^2_{OOS} = R^2_{TS} + R^2_{UN} + R^2_{CS} \tag{3.E.4}$$

where R^2_{TS}; captures the ability of model 1 to improve on forecasts from model 2 with respect to predicting time series variation in relative stock returns, R^2_{UN}; captures the ability of model 1 to improve on forecasts from model 2 with respect to predicting the unconditional relative stock returns and R_{CS}; cross-sectional mean dimension (time series plus unconditional cross-sectional market return).

We begin by expanding R^2_{TS}:

$$R^2_{TS} = 1 - \frac{\sum r_{i,t} - \mu_{2,i,t,RR} - \tilde{r}_{1,i,t,RR} - \sum_t r_{2,i,t}{}^2}{\sum (r_{i,t} - r_{2,i,t})^2} \tag{3.E.5}$$

$$A_1 = r_{i,t};\ \ B_2 = \mu_{2,i,t,RR};\ \ C_1 = \tilde{r}_{1,i,t,RR};\ \ D_2 = \sum r_{2,i,t} \tag{3.E.6}$$

$$R^2_{TS} = 1 - \frac{\sum (A_1 - B_2 - C_1 - D_2)^2}{\sum (r_{i,t} - r_{2,i,t})^2} \tag{3.E.7}$$

$$R^2_{TS} = 1 - \frac{\sum (A_1^2 - 2A_1B_2 - 2A_1C_1 - 2A_1D_2 + B_2^2 + 2B_2C_1 + 2B_2D_2 + C_1^2 + 2C_1D_2 + D_2^2)}{\sum (r_{i,t} - r_{2,i,t})^2} \tag{3.E.8}$$

Expanding R^2_{UN}:

$$R^2_{UN} = 1 - \frac{\sum r_{i,t} - \mu_{1,i,t,RR} - \tilde{r}_{2,i,t,RR} - \sum_t r_{2,i,t}{}^2}{\sum (r_{i,t} - r_{2,i,t})^2} \tag{3.E.9}$$

$$A_1 = r_{i,t};\ \ B_1 = \mu_{1,i,t,RR};\ \ C_2 = \tilde{r}_{2,i,t,RR};\ \ D_2 = \sum r_{2,i,t} \tag{3.E.10}$$

$$R^2_{UN} = 1 - \frac{\sum (A_1 - B_1 - C_2 - D_2)^2}{\sum (r_{i,t} - r_{2,i,t})^2} \tag{3.E.11}$$

$$R^2_{UN} = 1 - \frac{\sum (A_1^2 - 2A_1B_1 - 2A_1C_2 - 2A_1D_2 + B_1^2 + 2B_2C_2 + 2B_1D_2 + C_2^2 + 2C_2D_2 + D_2^2)}{\sum (r_{i,t} - r_{2,i,t})^2} \tag{3.E.12}$$

Expanding R^2_{CS}:

$$R^2_{CS} = 1 - \frac{\sum (r_{i,t} - \mu_{2,i,t,RR} - \tilde{r}_{2,i,t,RR} - \sum_t r_{1,i,t})^2}{\sum (r_{i,t} - r_{2,i,t})^2} \tag{3.E.13}$$

$$A_1 = r_{i,t};\ \ B_2 = \mu_{2,i,t,RR};\ \ C_2 = \tilde{r}_{2,i,t,RR};\ \ D_1 = \sum r_{1,i,t} \tag{3.E.14}$$

$$R^2_{CS} = 1 - \frac{\sum (A_1 - B_2 - C_2 - D_1)^2}{\sum (r_{i,t} - r_{2,i,t})^2} \tag{3.E.15}$$

$$R^2_{CS} = 1 - \frac{\sum (A_1^2 - 2A_1B_2 - 2A_1C_2 - 2A_1D_1 + B_2^2 + 2B_2C_2 + 2B_2D_1 + C_2^2 + 2C_2D_1 + D_1^2)}{\sum (r_{i,t} - r_{2,i,t})^2} \tag{3.E.16}$$

From the expansions above we have:

$$
\begin{aligned}
&R^2_{TS} + R^2_{UN} + R^2_{CS} = ...\\
&3 - \frac{\sum(A_1^2 - 2A_1B_2 - 2A_1C_1 - 2A_1D_2 + B_2^2 + 2B_2C_1 + 2B_2D_2 + C_1^2 + 2C_1D_2 + D_2^2)}{\sum\left(r_{i,t} - r_{2,i,t}\right)^2}\\
&- \frac{\sum\left(A_1^2 - 2\,A_1B_1 - 2A_1C_2 - 2A_1D_2 + B_1^2 + 2\,B_1C_2 + 2B_1D_2 + C_2^2 + 2C_2D_2 + D_2^2\right)}{\sum\left(r_{i,t} - r_{2,i,t}\right)^2}\\
&- \frac{\sum\left(A_1^2 - 2\,A_1B_2 - 2A_1C_2 - 2A_1D_1 + B_2^2 + 2\,B_2C_2 + 2B_2D_1 + C_2^2 + 2C_2D_1 + D_1^2\right)}{\sum\left(r_{i,t} - r_{2,i,t}\right)^2}
\end{aligned}
$$

Cross-products with C, where subscripts are different fall out because we define the time varying relative return forecasts to have mean zero. We further assume; $\tilde{r}_{2,i,t,RR} \perp \sum_t r_{1,i,t}$ and $\tilde{r}_{1,i,t,RR} \perp \sum_t r_{2,i,t}$.

$$
\begin{aligned}
&R^2_{TS} + R^2_{UN} + R^2_{CS} = ...\\
&3 - \frac{\sum\left(A_1^2 - 2A_1C_1 + C_1^2 - 2\,A_1B_1 + B_1^2 - 2A_1D_1 + D_1^2\right)}{\sum\left(r_{i,t} - r_{2,i,t}\right)^2}\\
&- \frac{\sum\left(A_1^2 + 2C_2D_2 + D_2^2 + B_2^2 + 2\,B_2C_2 + C_2^2 - 2A_1C_2 - 2\,A_1B_2 - 2A_1D_2\right)}{\sum\left(r_{i,t} - r_{2,i,t}\right)^2}\\
&- \frac{\sum\left(A_1^2 + B_2^2 + 2B_2D_2 + D_2^2 + C_2^2 - 2A_1C_2 - 2A_1D_2 - 2\,A_1B_2\right)}{\sum\left(r_{i,t} - r_{2,i,t}\right)^2}
\end{aligned}
$$

Add $(2B_1C_1 - 2B_1C_1)$, $(2B_1D_1 - 2B_1D_1)$, and $(2C_1D_1 - 2C_1D_1)$ to the first fraction, $(2A_1D_2 - 2A_1D_2)$ to the second fraction and $(2B_2C_2 - 2B_2C_2)$ and $(2C_2D_2 - 2C_2D_2)$ to the last fraction.

$$
\begin{aligned}
&R^2_{TS} + R^2_{UN} + R^2_{CS} = ...\\
&[1.1]3 - \frac{\sum(A_1^2 - 2A_1C_1 + C_1^2 - 2A_1B_1 + B_1^2 - 2A_1D_1 + D_1^2 + (2B_1C_1 - 2B_1C_1) + (2B_1D_1 - 2B_1D_1) + (2C_1D}{\sum\left(r_{i,t} - r_{2,i,t}\right)^2}\\
&[1.1]- \frac{\sum(A_1^2 + 2C_2D_2 + D_2^2 + B_2^2 + 2B_2C_2 + C_2^2 - 2A_1C_2 - 2A_1B_2 - 2A_1D_2) + (2B_2D_2 - 2B_2D_2)}{\sum\left(r_{i,t} - r_{2,i,t}\right)^2}\\
&[1.1]- \frac{\sum(A_1^2 + B_2^2 + 2B_2D_2 + D_2^2 + C_2^2 - 2A_1C_2 - 2A_1D_2 - 2A_1B_2) + (2B_2C_2 - 2B_2C_2) + (2C_2D_2 - 2C_2D_2)}{\sum\left(r_{i,t} - r_{2,i,t}\right)^2}
\end{aligned}
$$

$$
\begin{aligned}
&R^2_{TS} + R^2_{UN} + R^2_{CS} = ... \\
&[1.1]3 - \frac{\sum(A_1^2 - 2A_1C_1 + C_1^2 - 2A_1B_1 + B_1^2 - 2A_1D_1 + D_1^2 + 2B_1C_1 + 2B_1D_1 + 2C_1D_1)}{\sum(r_{i,t} - r_{2,i,t})^2} \\
&[1.1] - \frac{\sum(A_1^2 + 2C_2D_2 + D_2^2 + B_2^2 + 2B_2C_2 + C_2^2 - 2A_1C_2 - 2A_1B_2 - 2A_1D_2 + 2B_2D_2)}{\sum(r_{i,t} - r_{2,i,t})^2} \\
&[1.1] - \frac{\sum(A_1^2 + B_2^2 + 2B_2D_2 + D_2^2 + C_2^2 - 2A_1C_2 - 2A_1D_2 - 2A_1B_2 + 2B_2C_2 + 2C_2D_2)}{\sum(r_{i,t} - r_{2,i,t})^2} \\
&[1.1] - \frac{\sum(2B_1C_1 + 2B_1D_1 + 2C_1D_1 + 2B_2D_2 + 2B_2D_2 + 2B_2C_2 + 2C_2D_2)}{\sum(r_{i,t} - r_{2,i,t})^2}
\end{aligned}
$$

All cross-terms with C fall out:

$$
\begin{aligned}
&R^2_{TS} + R^2_{UN} + R^2_{CS} = ... \\
&[1.1]3 - \frac{\sum(A_1 - B_1 - C_1 - D_1)^2}{\sum(r_{i,t} - r_{2,i,t})^2} - \frac{\sum(A_1 - B_2 - C_2 - D_2)^2}{\sum(r_{i,t} - r_{2,i,t})^2} - \frac{\sum(A_1 - B_2 - C_2 - D_2)^2}{\sum(r_{i,t} - r_{2,i,t})^2} - \frac{\sum(2B_1D_1 + 4B_2D_2)}{\sum(r_{i,t} - r_{2,i,t})^2}
\end{aligned}
\tag{3.E.17}
$$

$$
\begin{aligned}
&R^2_{TS} + R^2_{UN} + R^2_{CS} = ... \\
&[1.1]3 - \frac{\sum(A_1 - B_1 - C_1 - D_1)^2}{\sum(r_{i,t} - r_{2,i,t})^2} - \frac{\sum(A_1 - B_2 - C_2 - D_2)^2}{\sum(r_{i,t} - r_{2,i,t})^2} - \frac{\sum(A_1 - B_2 - C_2 - D_2)^2}{\sum(r_{i,t} - r_{2,i,t})^2}
\end{aligned}
\tag{3.E.18}
$$

Where $r_{2,i,t} = B_2 + C_2 + D_2$ and $r_{1,i,t} = B_1 + C_1 + D_1$.

$$
R^2_{TS} + R^2_{UN} + R^2_{CS} = 3 - \frac{\sum(A_1 - B_1 - C_1 - D_1)^2}{\sum(r_{i,t} - r_{2,i,t})^2} - 1 - 1
$$

Putting it all together we have:

$$R^2 = R_{TS}^2 + R_{UN}^2 + R_{CS}^2 = 1 - \frac{\sum (r_{i,t} - r_{1,i,t})^2}{\sum (r_{i,t} - r_{2,i,t})^2}$$

where $\tilde{r}_{t,i,1}$ and $\tilde{r}_{t,i,2}$ have mean zero. It then follows that re-writing $R_{t,1}$ as $(\mu_{i,1} - \mu_{i,1}) + \tilde{r}_{t,i,1} + \mu_{i,2}$, and comparing this to forecasts from model 2 allows us to focus on the time-series variations in the forecasts. This is because, this specific re-write forces model 1 and model 2 to be equal in their ability to explain the unconditional stock return (all equal to mu_2), and the difference in forecasting ability now comes from $\tilde{r}_{t,i,1}$ versus $\tilde{r}_{t,i,2}$. Re-writing model 1 forecasts as $\mu_{i,1} + (\tilde{r}_{t,i,1} - \tilde{r}_{t,i,1}) + \tilde{r}_{t,i,2}$ forces the two models to match in explaining the time-series variation in returns ($\tilde{r}_{t,i,2}$), and differ in their ability to explain the unconditional stock return ($\mu_{i,2}$ vs $\mu_{i,1}$) over the sample.

Bibliography

Abarbanell, Jeffrey S and Brian J Bushee (1997). Fundamental analysis, future earnings, and stock prices, *Journal of accounting research* 35.1, pp. 1–24.

Adrian, Tobias, Erkko Etula, and Tyler Muir (2014). Financial Intermediaries and the Cross-Section of Asset Returns, *The Journal of Finance* 69.6, pp. 2557–2596.

Adrian, Tobias and Hyun Song Shin (2010). Liquidity and leverage, *Journal of Financial Intermediation* 19.3, pp. 418 –437. ISSN: 1042-9573.

Anderson, Anne-Marie and Edward A Dyl (2005). Market structure and trading volume, *Journal of Financial Research* 28.1, pp. 115–131.

Ang, Andrew, Robert J Hodrick, Yuhang Xing, and Xiaoyan Zhang (2006). The cross-section of volatility and expected returns, *Journal of Finance*.

Asness, Cliff and Andrea Frazzini (2013). The Devil in HML's Details, *Journal of Portfolio Management* 39.4, pp. 49–68.

Asness, Cliff, Jacques Friedman, Robert J. Krail, and John M. Liew (2000a). Style Timing: Value vs. Growth, *Journal of Portfolio Management* 26.1, pp. 50–60.

Asness, Cliff, John Liew, Lasse H. Pedersen, and Ashwin Thapar (2018). *Deep value*. Working paper. AQR Capital Management.

Asness, Cliff S., Burt Porter, and L. Stevens Ross (2000b). *Predicting Stock Returns Using Industry-Relative Firm Characteristics*. Working paper. AQR Capital Management.

Asness, Clifford, Swati Chandra, Antti Ilmanen, and Ronen Israel (2017). Contrarian factor timing is deceptively difficult, *The Journal of Portfolio Management* 43.5, pp. 72–87.

Asness, Clifford S, Andrea Frazzini, and Lasse H Pedersen (2012). Leverage aversion and risk parity, *Financial Analysts Journal* 68.1, pp. 47–59.

Asness, Clifford S., Tobias J. Moskowitz, and Lasse Heje Pedersen (2013a). Value and momentum everywhere, *Journal of Finance* 68.3, pp. 929–985.

Asness, Clifford S, Tobias J Moskowitz, and Lasse Heje Pedersen (2013b). Value and momentum everywhere, *The Journal of Finance* 68.3, pp. 929–985.

Asness, Clifford S, R Burt Porter, and Ross L Stevens (2000c). Predicting stock returns using industry-relative firm characteristics, *Available at SSRN 213872*.

Avramov, Doron, Si Cheng, and Lior Metzker (2020). Machine learning versus economic restrictions: Evidence from stock return predictability, *Available at SSRN 3450322*.

Baba-Yara, Fahiz, Martijn Boons, and Andrea Tamoni (2018). Value return predictability across asset classes and commonalities in risk premia, *Review of Finance, (forthcoming)*.

— (2020). New and Old Sorts: Implications for Asset Pricing, *Available at SSRN 3529140*, p. 69.

Baker, Malcolm and Jeffrey Wurgler (2006). Investor sentiment and the cross-section of stock returns, *The Journal of Finance* 61.4, pp. 1645–1680.

Balakrishnan, Karthik, Eli Bartov, and Lucile Faurel (2010). Post loss/profit announcement drift, *Journal of Accounting and Economics* 50.1, pp. 20–41.

Bali, Turan G, Nusret Cakici, and Robert F Whitelaw (2011). Maxing out: Stocks as lotteries and the cross-section of expected returns, *Journal of Financial Economics* 99.2, pp. 427–446.

Ball, Ray, Joseph Gerakos, Juhani T Linnainmaa, and Valeri V Nikolaev (2015). Deflating profitability, *Journal of Financial Economics* 117.2, pp. 225–248.

Bandyopadhyay, Sati P, Alan G Huang, and Tony S Wirjanto (2010). The accrual volatility anomaly, *Unpublished Manuscript, University of Waterloo*.

Basu, Sanjoy (1977). Investment performance of common stocks in relation to their price-earnings ratios: A test of the efficient market hypothesis, *The journal of Finance* 32.3, pp. 663–682.

— (1983). The relationship between earnings' yield, market value and return for NYSE common stocks: Further evidence, *Journal of financial economics* 12.1, pp. 129–156.

Baturin, Nick, S Gollamudi, S Persad, and H Zhang (2010). US equity fundamental factor model, *Bloomberg Portfolio and Risk Analytics*.

Ben-Rephael, Azi, Ohad Kadan, and Avi Wohl (2015). The diminishing liquidity premium, *Journal of Financial and Quantitative Analysis*, pp. 197–229.

Berk, Jonathan B, Richard C Green, and Vasant Naik (1999). Optimal investment, growth options, and security returns, *The Journal of Finance* 54.5, pp. 1553–1607.

Bhandari, Laxmi Chand (1988). Debt/equity ratio and expected common stock returns: Empirical evidence, *The journal of finance* 43.2, pp. 507–528.

Binsbergen, Jules van, Martijn Boons, Christian Opp, and Andrea Tamoni (2021). Dynamic Asset (Mis)Pricing: Build-up vs. Resolution Anomalies, *Available at SSRN*.

Black, Fischer, Michael C Jensen, and Myron Scholes (1972). The capital asset pricing model: Some empirical tests, *Studies in the theory of capital markets* 81.3, pp. 79–121.

Boudoukh, Jacob, Matthew Richardson, and Robert F Whitelaw (2006). The myth of long-horizon predictability, *Review of Financial Studies* 21.4, pp. 1577–1605.

Brandt, Michael W, Pedro Santa-Clara, and Rossen Valkanov (2009). Parametric portfolio policies: Exploiting characteristics in the cross-section of equity returns, *The Review of Financial Studies* 22.9, pp. 3411–3447.

Brown, David P and Bradford Rowe (2007). The productivity premium in equity returns, *Available at SSRN 993467*.

Brunnermeier, Markus and Lasse Pedersen (2009). Market Liquidity and Funding Liquidity, *Review of Financial Studies* 22.6, pp. 2201–2238.

Brunnermeier, Markus K. and Yuliy Sannikov (2014). A Macroeconomic Model with a Financial Sector, *American Economic Review* 104.2, pp. 379–421.

Bryzgalova, Svetlana, Markus Pelger, and Jason Zhu (2019). Forest through the trees: Building cross-sections of stock returns, *Available at SSRN 3493458*.

Butaru, Florentin, Qingqing Chen, Brian Clark, Sanmay Das, Andrew W Lo, and Akhtar Siddique (2016). Risk and risk management in the credit card industry, *Journal of Banking & Finance* 72, pp. 218–239.

Campbell, John Y. and John Cochrane (1999). Force of habit: A consumption-based explanation of aggregate stock market behavior, *Journal of Political Economy* 107, pp. 205–251.

Campbell, John Y and John H Cochrane (2000). Explaining the poor performance of consumption-based asset pricing models, *Journal of Finance* 55.6, pp. 2863–2878.

Campbell, John Y and Tuomo Vuolteenaho (2004). Bad beta, good beta, *American Economic Review* 94.5, pp. 1249–1275.

Chandrashekar, Satyajit and Ramesh KS Rao (2009). The productivity of corporate cash holdings and the cross-section of expected stock returns, *McCombs Research Paper Series No. FIN-03-09*.

Chen, Luyang, Markus Pelger, and Jason Zhu (2019). Deep learning in asset pricing, *Available at SSRN 3350138*.

Chernov, Mikhail, Lars A Lochstoer, and Stig RH Lundeby (2018). *Conditional dynamics and the multi-horizon risk-return trade-off*. Tech. rep. National Bureau of Economic Research.

Cho, Thummim and Christopher Polk (2019). Asset Pricing with Price Levels, *Available at SSRN 3499681*.

Chordia, Tarun, Amit Goyal, and Jay A Shanken (2017). Cross-sectional asset pricing with individual stocks: betas versus characteristics, *Available at SSRN 2549578*.

Chordia, Tarun, Richard Roll, and Avanidhar Subrahmanyam (2011). Recent trends in trading activity and market quality, *Journal of Financial Economics* 101.2, pp. 243–263.

Chordia, Tarun, Avanidhar Subrahmanyam, and V Ravi Anshuman (2001). Trading activity and expected stock returns, *Journal of Financial Economics* 59.1, pp. 3–32.

Chordia, Tarun, Avanidhar Subrahmanyam, and Qing Tong (2013). Trends in the cross-section of expected stock returns, *SSRN eLibrary*.

Chung, Kee H and Hao Zhang (2014). A simple approximation of intraday spreads using daily data, *Journal of Financial Markets* 17, pp. 94–120.

Clark, Todd E and Kenneth D West (2007). Approximately normal tests for equal predictive accuracy in nested models, *Journal of econometrics* 138.1, pp. 291–311.

Cochrane, John H (2008). The dog that did not bark: A defense of return predictability, *The Review of Financial Studies* 21.4, pp. 1533–1575.

Cochrane, John H. (2011a). Presidential Address: Discount Rates, *Journal of Finance* 66.4, pp. 1047–1108.

Cochrane, John H (2011b). Presidential address: Discount rates, *The Journal of finance* 66.4, pp. 1047–1108.

Cochrane, John H. and Monika Piazzesi (2005). Bond Risk Premia, *American Economic Review* 95.1, pp. 138–160.

Cochrane, John Howland (2001). *Asset Pricing* Princeton University Press. XVII, 530.

Cohen, Randolph B. and Christopher Polk (1998). *The impact of industry factors in stock returns*. Working paper. Kellogg School of Management.

Cohen, Randolph B, Christopher Polk, and Tuomo Vuolteenaho (2003b). The value spread, *The Journal of Finance* 58.2, pp. 609–641.

Cohen, Randolph B., Christopher Polk, and Tuomo Vuolteenaho (2003a). The Value Spread, *Journal of Finance* 78.2, pp. 609–642.

Cohen, Randolph B, Christopher Polk, and Tuomo Vuolteenaho (2009). The price is (almost) right, *The Journal of Finance* 64.6, pp. 2739–2782.

Cooper, Michael J, Huseyin Gulen, and Michael J Schill (2008). Asset growth and the cross-section of stock returns, *The Journal of Finance* 63.4, pp. 1609–1651.

Cooper, Michael J, Roberto C Gutierrez Jr, and Allaudeen Hameed (2004). Market states and momentum, *The journal of Finance* 59.3, pp. 1345–1365.

Cosemans, Mathijs, Rik Frehen, Peter C Schotman, and Rob Bauer (2016). Estimating security betas using prior information based on firm fundamentals, *The Review of Financial Studies* 29.4, pp. 1072–1112.

Daniel, Kent, David Hirshleifer, and Avanidhar Subrahmanyam (1998). Investor Psychology and Security Market Under- and Overreactions, *Journal of Finance* 53.6, pp. 1839–1885.

Daniel, Kent, David Hirshleifer, and Lin Sun (2017). *Short-and long-horizon behavioral factors*. Tech. rep. National Bureau of Economic Research.

Daniel, Kent and Tobias J Moskowitz (2016). Momentum crashes, *Journal of Financial Economics* 122.2, pp. 221–247.

Daniel, Kent and Sheridan Titman (1997). Evidence on the characteristics of cross sectional variation in stock returns, *the Journal of Finance* 52.1, pp. 1–33.

Daniel, Kent D, Lira Mota, Simon Rottke, and Tano Santos (2019). The Cross-Section of Risk and Return, *Review of Financial Studies* 18-4.

Datar, Vinay T, Narayan Y Naik, and Robert Radcliffe (1998). Liquidity and stock returns: An alternative test, *Journal of Financial Markets* 1.2, pp. 203–219.

Davis, James L, Eugene F Fama, and Kenneth R French (2000a). Characteristics, covariances, and average returns: 1929 to 1997, *Journal of Finance* 55.1, pp. 389–406.

— (2000b). Characteristics, covariances, and average returns: 1929 to 1997, *The Journal of Finance* 55.1, pp. 389–406.

De Bondt, Werner FM and Richard Thaler (1985). Does the stock market overreact?, *The Journal of finance* 40.3, pp. 793–805.

DeBondt, Werner F. M. and Richard Thaler (1985). Does the Stock Market Overreact?, *Journal of Finance* 40.3, pp. 793–805.

DeMiguel, Victor, Alberto Martin-Utrera, Francisco J Nogales, and Raman Uppal (2019). A transaction-cost perspective on the multitude of firm characteristics, *Available at SSRN 2912819*.

Desai, Hemang, Shivaram Rajgopal, and Mohan Venkatachalam (2004). Value-glamour and accruals mispricing: One anomaly or two?, *The Accounting Review* 79.2, pp. 355–385.

Driscoll, John C and Aart C Kraay (1998). Consistent covariance matrix estimation with spatially dependent panel data, *Review of economics and statistics* 80.4, pp. 549–560.

D'Acunto, Francesco, Ryan Liu, Carolin Pflueger, and Michael Weber (2018). Flexible prices and leverage, *Journal of Financial Economics* 129.1, pp. 46–68.

Engelberg, Joseph, R David McLean, Jeffrey Pontiff, and Matthew Ringgenberg (2019). "Are cross-sectional predictors good market-level predictors?" *American Finance Association Annual Meeting Paper*.

Fama, Eugene F. and Kenneth R. French (1992a). The Cross-Section of Expected Stock Returns, *Journal of Finance* 47.2, pp. 427–465.

Fama, Eugene F and Kenneth R French (1992b). The cross-section of expected stock returns, *the Journal of Finance* 47.2, pp. 427–465.

Fama, Eugene F. and Kenneth R. French (1993a). Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33.1, pp. 3–56. URL: http://EconPapers.repec.org/RePEc:eee:jfinec:v:33:y:1993:i:1:p:3-56.

Fama, Eugene F and Kenneth R French (1993b). Common risk factors in the returns on stocks and bonds, *Journal of financial economics* 33.1, pp. 3–56.

Fama, Eugene F and Kenneth R French (1995). Size and book-to-market factors in earnings and returns, *Journal of Finance* 50.1, pp. 131–155.

— (1996a). Multifactor explanations of asset pricing anomalies, *Journal of Finance* 51.1, pp. 55–84.

— (1996b). Multifactor explanations of asset pricing anomalies, *The journal of finance* 51.1, pp. 55–84.

Fama, Eugene F. and Kenneth R. French (1998). Value versus Growth: The International Evidence, *Journal of Finance* 53.6, pp. 1975–1999. ISSN: 1540-6261.

Fama, Eugene F and Kenneth R French (2008). Dissecting anomalies, *The Journal of Finance* 63.4, pp. 1653–1678.

— (2015). A five-factor asset pricing model, *Journal of financial economics* 116.1, pp. 1–22.

— (2018). Choosing factors, *Journal of Financial Economics* 128.2, pp. 234–252.

Ferreira, Miguel A and Pedro Santa-Clara (2011). Forecasting stock market returns: The sum of the parts is more than the whole, *Journal of Financial Economics* 100.3, pp. 514–537.

Frazzini, Andrea and Lasse Pedersen (2014a). Betting against beta, *Journal of Financial Economics* 111.1, pp. 1–25.

Frazzini, Andrea and Lasse Heje Pedersen (2014b). Betting against beta, *Journal of Financial Economics* 111.1, pp. 1–25.

Freyberger, Joachim, Andreas Neuhierl, and Michael Weber (2020a). Dissecting characteristics nonparametrically, *The Review of Financial Studies* 33.5, pp. 2326–2377.

— (2020b). *Dissecting characteristics nonparametrically*. Tech. rep. 5, pp. 2326–2377.

Froot, K. A. and T. Ramadorai (2005). Currency Returns, Intrinsic Value, and Institutional-Investor Flows, *The Journal of Finance* 60.3, pp. 1535–1566.

Gandhi, Priyank and Hanno Lustig (2015). Size anomalies in US bank stock returns, *The Journal of Finance* 70.2, pp. 733–768.

Garfinkel, Jon A (2009). Measuring investors' opinion divergence, *Journal of Accounting Research* 47.5, pp. 1317–1348.

George, Thomas J and Chuan-Yang Hwang (2004). The 52-week high and momentum investing, *The Journal of Finance* 59.5, pp. 2145–2176.

Gerakos, Joseph and Juhani T Linnainmaa (2018a). Decomposing value, *The Review of Financial Studies* 31.5, pp. 1825–1854.

— (2018b). Decomposing value, *The Review of Financial Studies* 31.5, pp. 1825–1854.

Gibbons, Michael R, Stephen A Ross, and Jay Shanken (1989). A test of the efficiency of a given portfolio, *Econometrica: Journal of the Econometric Society*, pp. 1121–1152.

Golubov, Andrey and Theodosia Konstantinidi (2016). *Where is the risk in value? Evidence from a market-to-book decomposition*. Working paper. Rotman School of Management.

Gomes, Joao, Leonid Kogan, and Lu Zhang (2003). Equilibrium cross section of returns, *Journal of Political Economy* 111.4, pp. 693–732.

Goodfellow, Ian, Yoshua Bengio, Aaron Courville, and Yoshua Bengio (2016). *Deep learning* vol. 1. MIT press Cambridge.

Gorodnichenko, Yuriy and Michael Weber (2016). Are sticky prices costly? Evidence from the stock market, *American Economic Review* 106.1, pp. 165–99.

Goyal, Amit and Ivo Welch (2008a). A comprehensive look at the empirical performance of equity premium prediction, *Review of Financial Studies* 21.4, pp. 1455–1508.

— (2008b). A comprehensive look at the empirical performance of equity premium prediction, *Review of Financial Studies* 21.4, pp. 1455–1508.

Green, Jeremiah, John RM Hand, and X Frank Zhang (2017). The characteristics that provide independent information about average us monthly stock returns, *The Review of Financial Studies* 30.12, pp. 4389–4436.

Gu, Shihao, Bryan Kelly, and Dacheng Xiu (2020a). Autoencoder asset pricing models, *Journal of Econometrics*.

— (2020b). Empirical asset pricing via machine learning, *The Review of Financial Studies* 33.5, pp. 2223–2273.

Haddad, Valentin, Serhiy Kozak, and Shrihari Santosh (2017). *Predicting relative returns*. Working paper. National Bureau of Economic Research.

— (2018). The economics of factor timing, *Available at SSRN 2945667*.

— (2020). *Factor timing*. Tech. rep. 5, pp. 1980–2018.

Hahn, Jaehoon and Hangyong Lee (2009). Financial constraints, debt capacity, and the cross-section of stock returns, *The Journal of Finance* 64.2, pp. 891–921.

Harvey, Campbell R and Yan Liu (2019). Lucky factors, *Available at SSRN 2528780*.

Haugen, Robert A and Nardin L Baker (1996). Commonality in the determinants of expected stock returns, *Journal of financial economics* 41.3, pp. 401–439.

He, Zhiguo, Bryan Kelly, and Asaf Manela (2017). Intermediary asset pricing: New evidence from many asset classes, *Journal of Financial Economics* 126.1, pp. 1–35.

He, Zhiguo and Arvind Krishnamurthy (2012). A Model of Capital and Crises, *The Review of Economic Studies* 79.2, pp. 735–777.

— (2013). Intermediary Asset Pricing, *American Economic Review* 103.2, pp. 732–70.

Herskovic, Bernard, Alan Moreira, and Tyler Muir (2019). Hedging risk factors, *Available at SSRN 3148693*.

Hirshleifer, David, Kewei Hou, Siew Hong Teoh, and Yinglei Zhang (2004). Do investors overvalue firms with bloated balance sheets?, *Journal of Accounting and Economics* 38, pp. 297–331.

Hodrick, Robert J (1992). Dividend yields and expected stock returns: Alternative procedures for inference and measurement, *Review of Financial Studies* 5.3, pp. 357–386.

Hong, Harrison and Jeremy C. Stein (1999). A Unified theory of underreaction, momentum trading, and overreaction in asset markets, *Journal of Finance* 54.6, pp. 2143–2184.

Hou, Kewei, G Andrew Karolyi, and Bong-Chan Kho (2011). What factors drive global stock returns?, *The Review of Financial Studies* 24.8, pp. 2527–2574.

Hou, Kewei, Haitao Mo, Chen Xue, and Lu Zhang (2018). Which factors?, *Review of Finance* 23.1, pp. 1–35.

Hou, Kewei, Chen Xue, and Lu Zhang (2015a). Digesting anomalies: An investment approach, *The Review of Financial Studies* 28.3, pp. 650–705.

— (2015b). Digesting anomalies: An investment approach, *The Review of Financial Studies* 28.3, pp. 650–705.

Ioffe, Sergey and Christian Szegedy (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift, *arXiv preprint arXiv:1502.03167*.

Israel, Ronen, Kristoffer Laursen, and Scott A Richardson (2020). Is (systematic) value investing dead?, *Available at SSRN*.

Israel, Ronen and Tobias J. Moskowitz (2013). The role of shorting, firm size, and time on market anomalies, *Journal of Financial Economics* 108.2, pp. 275–301.

Jegadeesh, Narasimhan (1990). Evidence of predictable behavior of security returns, *The Journal of finance* 45.3, pp. 881–898.

Jegadeesh, Narasimhan and Sheridan Titman (1993). Returns to buying winners and selling losers: Implications for stock market efficiency, *The Journal of finance* 48.1, pp. 65–91.

Johnson, J (2019). "Deep, skinny neural networks are not universal approximators". *7th International Conference on Learning Representations, ICLR 2019*.

Jurado, Kyle, Sydney C. Ludvigson, and Serena Ng (2015). Measuring uncertainty, *American Economic Review* 105.3, pp. 1177–1216.

Kelly, Bryan and Seth Pruitt (2013). Market expectations in the cross-section of present values, *The Journal of Finance* 68.5, pp. 1721–1756.

— (2015). The three-pass regression filter: A new approach to forecasting using many predictors, *Journal of Econometrics* 186.2, pp. 294–316. DOI: 10.1016/j.jeconom.2015.02. URL: https://ideas.repec.org/a/eee/econom/v186y2015i2p294-316.html.

Kelly, Bryan T, Seth Pruitt, and Yinan Su (2019). Characteristics are covariances: A unified model of risk and return, *Journal of Financial Economics*.

Keloharju, Matti, Juhani T Linnainmaa, and Peter M Nyberg (2019). Long-Term Discount Rates Do Not Vary Across Firms, *Available at SSRN 3125502*.

Kingma, Diederik P and Jimmy Ba (2014). Adam: A method for stochastic optimization, *arXiv preprint arXiv:1412.6980*.

Koijen, Ralph S.J., Hanno Lustig, and Stijn Van Nieuwerburgh (2017). The cross-section and time series of stock and bond returns, *Journal of Monetary Economics* 88, pp. 50 –69.

Koijen, Ralph S.J., Tobias J. Moskowitz, Lasse Heje Pedersen, and Evert B. Vrugt (2018). Carry, *Journal of Financial Economics* 127.2, pp. 197–225.

Kothari, Sagar P, Jay Shanken, and Richard G Sloan (1995). Another look at the cross-section of expected stock returns, *The journal of finance* 50.1, pp. 185–224.

Kozak, Serhiy (2019). Kernel Trick for the Cross-Section, *Available at SSRN 3307895*.

Kozak, Serhiy, Stefan Nagel, and Shrihari Santosh (2019). Shrinking the cross-section, *Journal of Financial Economics*.

— (2020). Shrinking the cross-section, *Journal of Financial Economics* 135.2, pp. 271–292.

Lakonishok, Josef, Andrei Shleifer, and Robert W Vishny (1994). Contrarian investment, extrapolation, and risk, *The journal of finance* 49.5, pp. 1541–1578.

Lettau, Martin, Matteo Maggiori, and Michael Weber (2014). Conditional risk premia in currency markets and other asset classes, *Journal of Financial Economics* 114.2, pp. 197–225.

Lettau, Martin and Markus Pelger (2020a). Estimating latent asset-pricing factors, *Journal of Econometrics* 218.1, pp. 1–31.

— (2020b). Factors that fit the time series and cross-section of stock returns, *The Review of Financial Studies* 33.5, pp. 2274–2325.

Lettau, Martin and Stijn Van Nieuwerburgh (2007). Reconciling the Return Predictability Evidence, *Review of Financial Studies* 21.4, pp. 1607–1652.

Lewellen, Jonathan (1999). The time-series relations among expected return, risk, and book-to-market, *Journal of Financial Economics* 54.1, pp. 5–43.

— (2004). Predicting returns with financial ratios, *Journal of Financial Economics* 74.2, pp. 209–235.

— (2014). The cross section of expected stock returns, *Forthcoming in Critical Finance Review*.

— (2015). The cross section of expected stock returns, *Critical Finance Review*.

Lewellen, Jonathan and Stefan Nagel (2006). The conditional CAPM does not explain asset-pricing anomalies, *Journal of financial economics* 82.2, pp. 289–314.

Li, Yan, David Ng, and Bhaskaran Swaminathan (2014). *Predicting time-varying value premium using the implied cost of capital*. Working paper. SSRN.

Liew, Jimmy and Maria Vassalou (2000). Can book-to-market, size and momentum be risk factors that predict economic growth?, *Journal of Financial Economics* 57.2, pp. 221–245.

Light, Nathaniel, Denys Maslov, and Oleg Rytchkov (2017). Aggregation of information about the cross section of stock returns: A latent variable approach, *The Review of Financial Studies* 30.4, pp. 1339–1381.

Lintner, John (1965). The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets, *The Review of Economics and Statistics*, pp. 13–37.

Litzenberger, Robert H and Krishna Ramaswamy (1979). The effect of personal taxes and dividends on capital asset prices: Theory and empirical evidence, *Journal of financial economics* 7.2, pp. 163–195.

Lundberg, Scott M and Su-In Lee (2017). "A unified approach to interpreting model predictions". *Advances in neural information processing systems*, pp. 4765–4774.

Luo, Liangchen, Yuanhao Xiong, Yan Liu, and Xu Sun (2019). Adaptive gradient methods with dynamic bound of learning rate, *arXiv preprint arXiv:1902.09843*.

Lyandres, Evgeny, Le Sun, and Lu Zhang (2008). The new issues puzzle: Testing the investment-based explanation, *The Review of Financial Studies* 21.6, pp. 2825–2855.

Martin, Ian and Stefan Nagel (2019). *Market Efficiency in the Age of Big Data*. Tech. rep. National Bureau of Economic Research.

McLean, R David and Jeffrey Pontiff (2016). Does academic research destroy stock return predictability?, *The Journal of Finance* 71.1, pp. 5–32.

Menkhoff, Lukas, Lucio Sarno, Maik Schmeling, and Andreas Schrimpf (2016). Currency value, *Review of Financial Studies* 30.2, pp. 416–441.

Menzly, Lior, Tano Santos, and Pietro Veronesi (2004). Understanding Predictability, *Journal of Political Economy* 112.1, pp. 1–47.

Moreira, Alan and Tyler Muir (2017). Volatility-Managed Portfolios, *Journal of Finance* 72.4, pp. 1611–1644.

Moskowitz, Tobias J., Yao Hua Ooi, and Lasse Heje Pedersen (2012). Time series momentum, *Journal of Financial Economics* 104.2, pp. 228–250.

Mossin, Jan (1966). Equilibrium in a capital asset market, *Econometrica: Journal of the econometric society*, pp. 768–783.

Nagel, Stefan (2016). The liquidity premium of near-money assets, *Quarterly Journal of Economics* 131.4, pp. 1927–1971.

Nagel, Stefan and Kenneth J Singleton (2011). Estimation and evaluation of conditional asset pricing models, *Journal of Finance* 66.3, pp. 873–909.

Neuhierl, Andreas and Michael Weber (2017). *Monetary momentum*. Working paper. University of Chicago Booth.

Newey, Whitney K. and Kenneth D. West (1987a). A simple positive semi-definite heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55.3, pp. 703–708.

Newey, Whitney K and Kenneth D West (1987b). Hypothesis testing with efficient method of moments estimation, *International Economic Review*, pp. 777–787.

Novy-Marx, Robert (2011). Operating leverage, *Review of Finance* 15.1, pp. 103–134.

— (2012). Is momentum really momentum?, *Journal of Financial Economics* 103.3, pp. 429–453.

Ou, Jane A and Stephen H Penman (1989). Financial statement analysis and the prediction of stock returns, *Journal of accounting and economics* 11.4, pp. 295–329.

Palazzo, Berardino (2012). Cash holdings, risk, and expected returns, *Journal of Financial Economics* 104.1, pp. 162–185.

Pontiff, Jeffrey and Artemiza Woodgate (2008). Share issuance and cross-sectional returns, *The Journal of Finance* 63.2, pp. 921–945.

Rachwalski, Mark and Quan Wen (2016). Idiosyncratic risk innovations and the idiosyncratic risk-return relation, *The Review of Asset Pricing Studies* 6.2, pp. 303–328.

Recanatesi, Stefano, Matthew Farrell, Madhu Advani, Timothy Moore, Guillaume Lajoie, and Eric Shea-Brown (2019). Dimensionality compression and expansion in Deep Neural Networks, *arXiv preprint arXiv:1906.00443*.

Richardson, Scott A, Richard G Sloan, Mark T Soliman, and Irem Tuna (2005). Accrual reliability, earnings persistence and stock prices, *Journal of accounting and economics* 39.3, pp. 437–485.

Rytchkov, Oleg (2010). Expected returns on value, growth, and HML, *Journal of Empirical Finance* 17.4, pp. 552–565.

Santos, Tano and Pietro Veronesi (2016). *Habits and leverage*. Working paper. National Bureau of Economic Research, Inc. URL: https://ideas.repec.org/p/nbr/nberwo/22905.html.

Sharpe, William F (1964). Capital asset prices: A theory of market equilibrium under conditions of risk, *The journal of finance* 19.3, pp. 425–442.

Sirignano, Justin, Apaar Sadhwani, and Kay Giesecke (2016). Deep learning for mortgage risk, *arXiv preprint arXiv:1607.02470*.

Sloan, Richard G (1996). Do stock prices fully reflect information in accruals and cash flows about future earnings?, *Accounting review*, pp. 289–315.

Soliman, Mark T (2008). The use of DuPont analysis by market participants, *The Accounting Review* 83.3, pp. 823–853.

Stambaugh, Robert F. (1999). Predictive regressions, *Journal of Financial Economics* 54.3, pp. 375–421.

Stambaugh, Robert F, Jianfeng Yu, and Yu Yuan (2012). The short of it: Investor sentiment and anomalies, *Journal of Financial Economics* 104.2, pp. 288–302.

Stambaugh, Robert F. and Yu Yuan (2016). Mispricing Factors, *The Review of Financial Studies* 30.4, pp. 1270–1315.

Stambaugh, Robert F and Yu Yuan (2017). Mispricing factors, *The Review of Financial Studies* 30.4, pp. 1270–1315.

Stock, J. and M. Watson (2002). Macroeconomic forecasting using diffusion indexes, *Journal of Business & Economic Statistics* 20.2, pp. 147–162.

Thomas, Jacob K and Huai Zhang (2002). Inventory changes and future returns, *Review of Accounting Studies* 7.2-3, pp. 163–187.

Valkanov, Rossen (2003). Long-horizon regressions: Theoretical results and applications, *Journal of Financial Economics* 68.2, pp. 201–232.

Van Binsbergen, Jules H and Christian C Opp (2019). Real anomalies, *The Journal of Finance* 74.4, pp. 1659–1706.

Vayanos, Dimitri and Paul Woolley (2013). An institutional theory of momentum and reversal, *Review of Financial Studies* 26.5, pp. 1087–1145.

Vuolteenaho, Tuomo (2002). What Drives Firm-Level Stock Returns?, *Journal of Finance* 57.1, pp. 233–264. DOI: 10.1111/1540-6261.00421.

White, Halbert et al. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity, *econometrica* 48.4, pp. 817–838.

Zhang, Lu (2005a). The value premium, *Journal of Finance* 60.1, pp. 67–103.

— (2005b). The value premium, *The Journal of Finance* 60.1, pp. 67–103.