The X-value factor

by

Thiago de Oliveira Souza

Discussion Papers on Business and Economics
No. 2/2020

FURTHER INFORMATION
Department of Business and Economics
Faculty of Business and Social Sciences
University of Southern Denmark
Campusvej 55, DK-5230 Odense M
Denmark
E-mail: lho@sam.sdu.dk / http://www.sdu.dk/ivoe

ISSN 2596-8157

The X-value factor

Thiago de Oliveira Souza

Abstract

Value normalizes size by book equity, which is a (relatively bad) proxy for expected cash flows. X-value normalizes size by the recursive out-of-sample expectation of each firm's net income, based on its financials, with coefficients estimated by industry. Unlike value (but similarly constructed), the resulting X-value factor is unspanned by the Fama/French factors – market, size, value, investment, and profitability – individually or in different combinations (each factor and the market; all factors together; all except value). X-value spans the value and investment premiums with a Sharpe ratio of 0.57 (compared to 0.39 for value).

JEL Code: G11, G12, G14.

Keywords: Risk premiums, stock returns, Fama and French, cash flow forecasting, out of sample.

University of Southern Denmark; Danish Finance Institute. First draft: March 15, 2019. I would like to thank Tuomo Vuolteenaho.

I have nothing to disclose.

1 Introduction

There is a myriad of competing theories for the value premium. Each one offers a different explanation of the stylized fact that the ratio between a firm's book equity (BE) and its price – the book-to-market (BM) ratio – is correlated with expected returns in the cross-section of firms. In only one theory (Berk, 1995; de Oliveira Souza, 2019a), a standard present value identity explains the premium and BE is **exclusively** a proxy for expected cash flows. Under this framework, ratios that divide *any* better estimates of expected cash flows by stock prices should be more correlated with expected returns than BM is – even if the estimates completely ignore BE. This theoretical prediction is precisely what the present paper confirms empirically.

First, consider the one-period present value identity,

$$ME_{i,t} = \frac{\mathrm{E}_t\left[D_{i,t+1}\right]}{R_{i,t}}, \tag{1}$$

where $ME_{i,t}$ is the market value of equity (ME) of firm i at time t, $\mathrm{E}_t[.]$ is the expectation operator conditioned on the information at time t, $D_{i,t+1}$ is the only expected cash flow payment received from holding asset i, and $R_{i,t}$ is the (gross) equilibrium risk premium required for asset i (with zero risk-free rate without loss of generality). The negative relation between ME and equilibrium returns in Eq. (1) summarizes the explanation of Berk (1995) for the "size" premium:[1] Ranking the stocks by size, $ME_{i,t}$, is inversely related to ranking them by their (risk) premiums, $R_{i,t}$. Hence, buying small stocks and selling big stocks yields positive average returns. However, size is a bad risk proxy unconditionally (de Oliveira Souza, 2019b,a): Cross-sectional differences in expected cash flows – unrelated to risks – act as noise and randomize the size ranking (Berk, 1995). They lower the

[1] Naturally, "size" is market capitalization, not size in a strict sense (Berk, 1996).

correlation between size and risk, reducing the premium. Ideally, we would like to control for these differences.

Consider now the stock's BM, which is simply a scaled version of Eq. (1),

$$\frac{BE_{i,t}}{ME_{i,t}} = R_{i,t} \frac{BE_{i,t}}{\mathrm{E}_t\left[D_{i,t+1}\right]}, \tag{2}$$

where $BE_{i,t}$ is the BE of firm i at time t. If BE is a proxy for expected cash flows, the right-hand side of the equation is proportional to the firm's risk premium, and a cross-sectional BM ranking resembles a risk premium ranking. This solves the value premium "puzzle" according to the theory in Berk (1995) and de Oliveira Souza (2019a).

The present paper builds further on this theoretical framework: It replaces BE with a simple proxy for expected cash flows. In particular, I use predictive panel regressions, per industry as defined in Fama and French (1997), to recursively forecast one year ahead out-of-sample expected cash flows (which I assume to be related to **expected** net income) for each firm in the merged CRSP-Compustat database.[2] Next, I use this expectation (instead of BE) to normalize the size ranking in Eq. (1). This normalization creates the "expected-(cash flow)-to-market" (XM) variable,

$$\frac{X_{i,t}}{ME_{i,t}} = R_{i,t} \frac{X_{i,t}}{\mathrm{E}_t\left[D_{i,t+1}\right]} \approx c\, R_{i,t}, \tag{3}$$

where $X_{i,t} \approx c\ \mathrm{E}_t\left[D_{i,t+1}\right]$ is a proxy for the expected cash flow of firm i, multiplied by a scaling parameter common to all firms, c. The cash flow predictors are the firm's (1-year lagged) net income, total assets, number of employees, capital expenditures, R&D, advertising, and acquisitions (the latter as a robustness check due to limited data). These regressors are

[2] I focus on the simplest case and only model cash flows one period ahead under the assumption that these differences are persistent. Naturally, firms tend to generate cash flows for multiple periods. Indeed, the framework that I present is consistent with any model of expected cash flows. I present one of these possible models in Section 2 and any improvements in the cash flow model should generate even stronger results.

often used to predict cash flows in the "applied corporate finance" ("valuation") literature (Damodaran, 2008, for example).

Importantly, the XM variable is **not** a "mix" of the firm characteristics used by Fama and French (2015), for example. The variables do not even have similar interpretations. For example, pooling all firms together reveals that all regressors have positive slopes in the predictive regressions of net income, jointly or individually (Table 1, Section 3.3.1). For an investment variable, such as capital expenditures, this corresponds to the **opposite** of what Fama and French (2015) document: High investment – not low, as defined by Fama and French (2015) – is associated with larger risk premiums for a given price (via larger $X_{i,t}$ in Eq. (3)).

I create the X-value factor following the exact same procedure used to create the value factor in Fama and French (1996) with one difference: I use XM where Fama and French (1996) use BM. This creates the XHML portfolio, similar to the HML portfolio, based on six portfolios double sorted by size and XM. These portfolios have the same desirable properties reported by Fama and French (1996) for sorts by size and BM: The average returns on the portfolios and their significance increase from small to big stocks within all XM groups. The premiums also increase monotonically from low- to high-XM stocks, within both the groups of big or small stocks. Indeed, all of this is theoretically expected since the firm's XM is exactly proportional to its risk premium, $R_{i,t}$, according to Eq. (3).

It is possible to visualize the economic significance of the X-value premium in Fig. 1: The premium is substantially larger than all premiums introduced by Fama and French (1996, 2015). The premium is also more significant (has a larger Sharpe ratio) than all factors in Fama and French (2015), including the market premium, as I show in Section 4.

[Figure 1 about here.]

I run several spanning tests involving the X-value premium in Section 4. The first set consists of 10 tests regressing X-value on the factors in Fama and French (2015) (individually or not). The X-value premium is not spanned by any of the five factors individually. XHML also generates significant abnormal returns based on the standard multivariate risk adjustment (with all five factors) of Fama and French (2015). Finally, the X-value premium left unexplained by the CAPM cannot be explained with the addition of any one of the other factors in Fama and French (2015): The premium remains unspanned when regressed on the market premium in conjunction with any of the other four factors in Fama and French (2015). Hence, the risks related to XM are not subsumed by the ones supposedly related to any of the characteristics used to create the factors in Fama and French (2015), individually or not.

A second set of tests shows that, in addition to being unspanned, the X-value premium spans both the value and the investment premiums of Fama and French (2015). This suggests that the risks linked to the stock's BM and investment characteristics are just an incomplete subset of the risks linked to the XM variable. Once again, this is in line with the theoretical framework in Berk (1995) and de Oliveira Souza (2019a): XM should capture **every** risk, given the right cash flow adjustment, X, as Eq. (3) shows.

The final set of spanning tests explores XHML as a replacement for the HML portfolio in Fama and French (2015). For example, HML earns significantly negative abnormal returns based on the risk adjustment of Fama and French (2015) modified with XHML. Most importantly, replacing HML by XHML "resurrects" the value premium: Unlike HML, XHML is unspanned by the other four factors.

The tests also show that it is not necessarily true that a four-factor model that drops HML performs as well as the five-factor model of Fama and French (2015) in applications where the sole interest is abnormal returns, as claimed by Fama and French (2015). The two models do not produce the same risk adjustment. For example, at the 1% significance level, there is no evidence of abnormal returns associated with the XHML portfolio based on a version of Fama and French (2015) that drops HML: The intercept of this regression has 2.39 t-statistic. In contrast, there is evidence of abnormal returns associated with the XHML portfolio, even at the 0.1% significance level, based on a version of Fama and French (2015) that **keeps** HML: The t-statistic of the intercept is 3.35 in this case.

Finally, it is unclear whether the investment factor should (or could) be included in a possible factor model: On the one hand, XHML spans CMA in a bivariate regression, which makes it a redundant factor. On the other hand, CMA is unspanned in a multivariate regression including XHML and the other three factors in Fama and French (2015), which makes the factor not redundant.

1.1 Closest literature and contribution

The present paper is primarily about the theoretical explanation of the value premium. However, the paper also addresses the value "factor" within an empirical asset pricing framework. Hence, it intersects with both these theoretical and empirical strands of the literature.

From a theoretical perspective the paper builds on, and can be interpreted as an empirical test of, the framework in de Oliveira Souza (2019a) and Berk (1995). Within this framework, BE is **strictly** only a proxy for expected cash flows. Inevitably, the paper also relates to, but contradicts, a large heterogeneous group of theories of the value premium. In all theories in this group, for different reasons, BE carries information relevant to determine the risk of the firm, as explained in detail by de Oliveira Souza (2019a). From this perspective, the major contribution of the paper is to provide further empirical support for the theory of the value premium in de Oliveira Souza (2019a) and Berk (1995) and further remove empirical support from alternative theories. This adds and relates to Ball et al. (2018) and

Gerakos and Linnainmaa (2017), who also provide evidence against theories that declare that BM is directly linked with the value premium.

There is also a crucial theoretical distinction between the framework in the present paper, based on de Oliveira Souza (2019a), and previous models of the value premium. Within the present framework, X in XM (or BE in BM) is **not** a measure of "fundamental value" (i.e., the price of the firm's future cash flows discounted at the "correct" rate). Instead, X is just a measure of expected cash flows **without** any discounting.[3] This distinction is theoretically important because it is the main difference between the present approach and the more agnostic approach of Vuolteenaho (2002), for example. In particular, the VAR model of Vuolteenaho (2002) can include anything. It can include discount rate predictors to forecast the BM ratio, for example. In contrast, the theoretical framework in de Oliveira Souza (2019a) establishes that only cash flows should be predicted.

Finally, from an empirical asset pricing perspective, the paper is about the role of value in the factor model of Fama and French (1996, 2015). In this sense, its major contribution is to resurrect value as an unspanned factor, and raise its importance relative to the other factors: In addition to being the largest in Fama and French (2015), the X-value premium spans the investment premium.

2 A simple model of free cash flow production

Let D_i be the profit maximizing steady-state free cash flow to equity holders (net of costs and in logs) earned by firm i for producing and selling good(s) i (each firm is indexed by the goods that it produces). Importantly, "cash flow" in the present paper is not accounting cash flows, dividends, nor net income: It is the free cash flow available to be paid out to equity holders (usually, but not always, distributed as dividends and stock repurchases).

[3] Indeed, de Oliveira Souza (2019a) formally rejects the hypothesis that BM measures mispricing (which is implied by the hypothesis that BE measures "fundamental value").

The cash flows are produced according to the firm- and product-specific Cobb-Douglas production function (in logs)

$$D_i = \delta_{i_k} k + \delta_{i_{k'}} k' + \delta_{i_l} l + \delta_{i_{rd}} rd + \delta_{i_{ad}} ad + \delta_{i_{aq}} aq + \omega_{i,I}, \tag{4}$$

where the input variables are (old) capital, k, and labor, l, augmented with investment in (new) capital, k', R&D investment, rd, advertising investment, ad, or simply acquisitions of other firms, aq. These inputs are also widely mentioned in the applied corporate finance literature on discounted cash flow valuation. For example, variables that forecast cash flows are related to investment (broadly defined), such as capital expenditures, R&D, advertisement, and acquisitions, and their importance typically varies with the industry in which the firm operates (Damodaran, 2008, chapter 4). The output elasticities are all firm-specific, and $\omega_{i,I}$ is the firm's idiosyncratic productivity, unexplained by these inputs.

Intuitively, Eq. (4) is a particular case of a more general Cobb-Douglas production function,

$$Y_i = F(K, L) = A_i K^{\alpha_i} L^{\beta_i}, \tag{5}$$

where K is capital, L is labor, α_i and β_i are the respective output elasticities, and A_i is the firm-specific labor and capital augmenting technology. It is possible to interpret all the extra variables in Eq. (4) as augmenting the cash flow production (and as part of A_i).

Now let industries, I, define groups of firms with similar products and, thus, similar production functions and productivity. In particular, assume that a vector of inputs at time t, $\boldsymbol{\varphi}_{i,t}$ (boldface indicates vectors), have output elasticities, $\boldsymbol{\delta}_I$, that are constant across firms in the same industry, I. For example, advertising could be important to determine cash flows of all firms in the beverage sector, but not in the mining sector. In discrete time, this corresponds to

$$D_{i,t+1} = \delta_{I_0} + \boldsymbol{\delta}_I^{\top} \boldsymbol{\varphi}_{i,t} + \omega_{i,t+1}, \tag{6}$$

where δ_{I_0} is the industry-wide productivity unexplained by the inputs, constant for all firms in industry I, and $\omega_i \neq \omega_{i,I}$ is the firm's residual idiosyncratic productivity, unexplained by the inputs, net of the average industry productivity.

The firm's unexplained productivity can be re-written as

$$\omega_{i,t} = \left(1 - \delta_{I_0} - \boldsymbol{\delta}_I^\top \boldsymbol{\varphi}_{i,t-1}\right) D_{i,t} \tag{7}$$

$$\equiv \delta_{I_D} D_{i,t}, \tag{8}$$

where $\delta_{I_D} \equiv \left(1 - \delta_{I_0} - \boldsymbol{\delta}_I^\top \boldsymbol{\varphi}_{i,t-1}\right)$ is the fraction of cash flows unexplained by the inputs and the industry average productivity. I assume this to be constant over time and across firms in a given industry. In this formulation, idiosyncratic productivity is autocorrelated, as in Doraszelski and Jaumandreu (2013), for example, as long as output is also autocorrelated. For a firm in industry I, this finally yields the production function

$$D_{i,t+1} = \delta_{I_0} + \delta_{I_k} k_{i,t} + \delta_{I_{k'}} k'_{i,t} + \delta_{I_l} l_{i,t} + \delta_{I_{rd}} rd_{i,t} + \delta_{I_{ad}} ad_{i,t} + \delta_{I_{aq}} aq_{i,t} + \delta_{I_D} D_{i,t}. \tag{9}$$

Section 3.2 explains how I obtain proxies for the variables in Eq. (9) and discusses data availability.

3 Data and empirical methods

The security data are from the Center for Research in Security Prices (CRSP) and the accounting data are from Compustat Annual Fundamental Files, with the CRSP/Compustat Merged Database (CCM) to link the two data sets. The data end in December 2018 and start as described next.

3.1 Returns and industry classification data

The security data from CRSP must have share codes 10 or 11 and exchange codes 1, 2, or 3, which correspond to the NYSE, NYSE MKT (previously AMEX), and NASDAQ. The variables are the stock's SIC code (item siccd), th number of shares outstanding (item shrout), the price (item prc), the monthly return (item ret), and delisting returns from the CRSP Delisting Event Histories table when needed (item dlret).

From Kenneth French's data library, I collect the series of realized monthly returns on the market, size, value, profitability, and investment premiums of Fama and French (2015) starting in July 1963. They are, respectively, the returns on the MP, SMB, HML, RMW, and CMA portfolios. I also obtain a list linking SIC codes with the 48 industry classifications in Fama and French (1997). I follow the procedure in Fama and French (1997) to assign each stock to an industry based on its four-digit SIC code in each period: I use the SIC codes in Compustat or CRSP, in this order, as available.

3.2 Proxies for the variables in the production function

The proxies for the variables in the production function in Eq. (9) come from Compustat: Total assets (Compustat item at) is a proxy for "legacy" capital, k; capital expenditures (item capx or, if missing, growth in property, plant, and equipment, item ppegt) proxies for recently installed capital, k'; number of employees (item emp) proxies for labor, l; R&D expenditures (item xrd) proxies for knowledge, rd; advertisement expenditures (item xad) proxies for demand influence, ad; acquisitions (item aqc) is acquisitions, aq; and net income (item ni) proxies for cash flows, D. The industry classification is obtained from the SIC codes from Compustat or CRSP as explained in the next paragraphs. From Compustat, I obtain historical SIC codes (item sich). All Compustat variables start in 1950, except acquisitions, which starts in 1971. I require the information to be in the data set for at least two years; thus the Compustat data set effectively starts in 1951 (1972 for acquisitions).

[Figure 2 about here.]

Fig. 2 displays the evolution over time of the number of reported (non-missing) values for the cash flow-related variables in Compustat. The data on R&D and advertisement (in addition to acquisitions) do not become available for a large number of firms until after 1970. Fig. 3 complements this information showing that even after the data start to appear in Compustat, the reported values are often zero. Their average (non-missing) values tend to be very low. This especially affects the beginning of the sample.

[Figure 3 about here.]

3.3 Panel data estimates of the cash flow production function

I estimate the coefficients in the production function in Eq. (9) by industry, based on the proxies described in Section 3.2. The SIC code classification in Fama and French (1997) splits the firms into 48 industries. The resulting 48 sets of estimates, $\hat{\boldsymbol{\beta}}_I$, are based on 48 independent panel regressions of the form

$$ni_{i,t+1} = \boldsymbol{\beta}_I^\top \boldsymbol{f}_{i,t} + \varepsilon_{i,t+1}, \tag{10}$$

where $\boldsymbol{f}_{i,t}$ is a vector containing the constant and the proxies for the variables in Eq. (9) (without acquisitions in the main analysis): $ni_{i,t}$ is net income of firm i in period t (a proxy for cash flows, $D_{i,t}$), $at_{i,t}$ is total assets (a proxy for "legacy" physical capital), $capx_{i,t}$ is capital expenditures (a proxy for "recently installed" physical capital), $emp_{i,t}$ is the number of employees (a proxy for labor), $xrd_{i,t}$ is R&D expenses (a proxy for knowledge), $xad_{i,t}$ is advertising expenses (a proxy for demand influence), and $aqc_{i,t}$ is acquisitions. $\boldsymbol{\beta}_I$ is the vector with all respective industry-specific coefficients, and $\varepsilon_{i,t}$ is the error term. Excluding or including acquisitions,

$$\boldsymbol{f}_{i,t} = \begin{cases} (1, at_{i,t}, capx_{i,t}, emp_{i,t}, xrd_{i,t}, xad_{i,t}, ni_{i,t})^{\top}, \\ (1, at_{i,t}, capx_{i,t}, emp_{i,t}, xrd_{i,t}, xad_{i,t}, aqc_{i,t}, ni_{i,t})^{\top}, \end{cases} \tag{11}$$

respectively.

The reason for the special treatment of acquisitions is that acquisition data do not start until 1972.[4] Therefore, out-of-sample cash flow predictions based on this predictor are only available from 1974 (compared to 1952 for all other variables). I report the results that include acquisitions in the cash flow production function as a robustness check in Section 5. The exercise serves both as a robustness check with respect to sample period and a model of cash flow production.

Importantly, every econometric or theoretical misspecification regarding the cash flow prediction in Eq. (10) results in less accurate cash flow forecasts. In terms of Eq. (3), X becomes a worse proxy for cash flows, the XM variable becomes less correlated with risks, and the X-value premium becomes smaller. This is particularly true because the cash flow predictions are done out of sample. Therefore, eventual improvements in the cash flow prediction should make the results in the paper even stronger (not weaker).

3.3.1 Overview of the in-sample predictive relations

Table 1 provides an in-sample summary of the predictive relation between the variables in Eq. (10). The estimates are based on all data available in Compustat, regardless of being available in CRSP, for example. It ignores the industry information, so that everything can be displayed in a single table.

[4] The first year with reported acquisition data is 1971, but I require the information to be in the data set for at least two years.

Table 1 shows that increases in all regressors significantly predict one-year ahead increases in net income jointly or individually. The multivariate regression in Eq. (10) is displayed in column (8), and the regressors, $\boldsymbol{f}_{i,t}$ in Eq. (11), include acquisitions. Univariate regressions of net income on each of the seven regressors individually, $f_{i,t}$, (and only if the regressor is different from zero) are displayed in columns (1) through (7),

$$ni_{i,t+1} = \beta_{I_0 f} + \beta_{If} f_{i,t}, \tag{12}$$
$$f_{i,t} = \left\{ni_{i,t},\, at_{i,t},\, capx_{i,t},\, emp_{i,t},\, xrd_{i,t},\, xad_{i,t},\, aqc_{i,t}\right\}.$$

The number of firm-years in each of these cases shows that data availability is substantially lower for R&D, advertisement and acquisitions. Net income, total assets, and number of employees remain absent when the information is missing. However, I replace eventual missing observations for capital expenditures, R&D, advertisement, or acquisitions with zeros. This is equivalent to the assumption that firms that do not report these expenses do not incur them.

[Table 1 about here.]

3.3.2 The out-of-sample expected cash flows

For each year t (with expanding recursive windows), I run 48 independent panel regressions, as given by Eq. (10). There is one regression for each group of firms in the 48 industries in Fama and French (1997), based only on the data available until that year. This generates 48 sets of estimated coefficients (one per industry, I) valid for year t, $\hat{\boldsymbol{\beta}}_{I,t}$. These sets of coefficients produce the out-of-sample prediction of the cash flows in year $t+1$ for each individual firm. The significance of the coefficients is Bonferroni adjusted to account for the $48 \times 6 = 288$ multiple tests in each period (or $48 \times 7 = 336$ with acquisitions) and, after that, I set the insignificant coefficients to zero. The resulting one year ahead out-of-sample expected cash flow for firm i that belongs to industry I in period t is given by

$$\mathrm{E}_t[ni_{i,t+1}] = \hat{\boldsymbol{\beta}}_{I,t}^{\top} \boldsymbol{f}_{i,t}, \tag{13}$$

where $\boldsymbol{f}_{i,t}$ in Eq. (11) excludes acquisitions in most of the analyses, except for the robustness tests in Section 5.

The BM of Fama and French (1996), for example, is a particular case of Eq. (13), in which the firm's expected cash flow at time $t+1$ (and the variable used to normalize ME) is simply its BE at time t. For the sake of comparison, Fama and French (1996) assume

$$\mathrm{E}_t[ni_{i,t+1}] = BE_{i,t}. \tag{14}$$

3.4 Construction of the X-factor portfolios and summary statistics

There is only one difference in the construction of the X-value and X-size factors compared to the value and size factors in Fama and French (1996): I replace the firm's BE (for fiscal year t) with its out-of-sample net income expectation (also for fiscal year t), given by Eq. (13). So I calculate the stock's XM (instead of BM) to sort the stocks into (X-)value portfolios. The rest, as I describe next, is exactly the same.

At the end of June of each year t, the stocks are allocated to two size groups (Small or Big) based on their June market cap (ME) being below or above the median NYSE market cap. The stocks are also independently allocated to three XM groups according to their XM being either below the NYSE XM 30 percentile (Low), above the 70 percentile (High), or between the two (Medium). This generates six double sorted portfolios with value weighted returns. The return on the XSMB portfolio is the difference, each month, between the average returns on the three small-stock portfolios and the three big-stock portfolios.

The return on XHML is the difference between the average returns on the two high-XM portfolios and the two low-XM portfolios.

The XM ratio used to form portfolios in June of year t is the out-of-sample expected net income for the fiscal year ending in calendar year $t-1$, divided by market equity at the end of December of year $t-1$. I ignore firms with negative expected cash flows to follow the treatment of firms with negative BM in Fama and French (1996).

Table 2 shows that the six portfolios double sorted by size and XM have the same desirable properties reported by Fama and French (1996) for sorts by size and BM: The excess returns on the portfolios and their significance increase from small to big stocks within all XM groups. The premiums also increase monotonically from low- to high-XM stocks within either the group of big or small stocks.

[Table 2 about here.]

4 The X-value and the Fama and French (2015) factors

Table 3 shows how the X-value premium compares to the premiums on the five factors in Fama and French (2015) in terms of economic and statistical significance: The X-value premium is more significant (has larger Sharpe ratio) than all factors in Fama and French (2015), including the market premium. The mean X-value premium is also larger than the mean premium on all factors created by Fama and French (1996, 2015) (only the mean market premium is larger). In particular, the X-value premium is substantially larger and more significant than the value premium of Fama and French (1996).

[Table 3 about here.]

4.1 The X-value factor is unspanned

Table 4 displays the results of 10 spanning tests of the X-value premium between July 1963 and December 2018 based on the factors in Fama and French (2015). It shows that the X-value factor is unspanned by these factors (either individually or in different combinations). The general form of these spanning regressions is

$$R_{xhml,t} = \alpha + \boldsymbol{\theta}^{\top} \boldsymbol{R}_{b,t} + \varepsilon_t, \tag{15}$$

where $R_{xhml,t}$ is the return on the XHML portfolio described in Section 3.4, α and $\boldsymbol{\theta}$ are the coefficients to be estimated, ε_t is an error term, and each of the 10 tests uses a different set of regressors, $\boldsymbol{R}_{b,t}$, which are the returns on the five Fama and French (2015) benchmark portfolios either in groups or individually.

[Table 4 about here.]

Regressions (1) through (5) test whether any of the existing factors individually spans the X-value premium. Hence, each (bivariate) regression is based on $\boldsymbol{R}_{b,t}$ being, respectively, the market premium, MP, or the returns on the SMB, HML, RMW, or CMA portfolios of Fama and French (2015).

Regressions (6) through (9) test whether the X-value premium left unexplained by the CAPM can be explained with the help of any one of the other factors in Fama and French (2015) individually. Thus, these (multiple) regressions are based on $\boldsymbol{R}_{b,t}$ containing the market premium, MP, in addition to, respectively, either the returns on the SMB, HML, RMW, or CMA portfolios,

$$\boldsymbol{R}_{b,t} = \left(MP_t, FF_{i,t}\right)^{\top}, \tag{16}$$

with $FF_{i,t} \in \{SMB_t, HML_t, RMW_t, CMA_t\}$.

Finally, in column (10) I test whether the X-value premium is an unspanned factor based on the risk adjustment in Fama and French (2015). Therefore, $\boldsymbol{R}_{b,t}$ contains all factors in Fama and French (2015) jointly in this case,

$$\boldsymbol{R}_{b,t} = (MP_t, SMB_t, HML_t, RMW_t, CMA_t)^\top, \tag{17}$$

The most important result in Table 4 is that the intercepts in all tests are highly significant: The X-value premium remains unexplained by the factors in Fama and French (2015) in all these combinations. Hence, the risks related to XM are not subsumed by the ones related individually to any of the characteristics used to create the factors in Fama and French (2015), based on models (1) through (5), nor are they subsumed by the risks related to combinations of these characteristics, based on models (6) through (10).

Models (1) through (5) show positive unconditional correlations between the X-value premium and the value, profitability, and investment premiums of Fama and French (2015), and negative ones with the market and size premiums. However, in a multivariate sense in model (10), the only significantly negative correlation is with the market premium, the correlations are positive with profitability, and especially with value. In fact, the point estimates of the coefficients on the SMB and CMA portfolios change from the bivariate to the (insignificant) multivariate case in model (10). Fama and French (2015) also mention (relatively puzzling) changes from univariate to multivariate "risk loadings" regarding their factors. Models (6) through (9) show that adding the market premium to the factors in Fama and French (2015) does not change the results of estimations (2) through (5) substantially. The biggest difference is that the correlation with the RMW becomes marginally insignificant. Finally, the variation in the X-value premium is mostly explained by variation in the value and, next, in the investment premium, as we see when comparing the R^2 values of models (1) through (5) or (6) through (9).

4.2 X-value spans the value and investment factors

Having established that X-value is not spanned by the other factors in Fama and French (2015), the next question is whether X-value spans these factors. Table 5 displays the results of spanning regressions for each of the five factor premiums in Fama and French (2015) on the X-value premium between July 1963 and December 2018. The equations are of the form

$$R_{FF,t} = \alpha + \theta R_{xhml,t} + \varepsilon_t, \tag{18}$$

where $R_{FF,t}$ is either the market premium, MP, or the returns on the SMB, HML, RMW, or CMA portfolios,

$$R_{FF,t} \in \{MP_t, SMB_t, HML_t, RMW_t, CMA_t\}, \tag{19}$$

and the remaining variables are the same as described in Eq. (15).

[Table 5 about here.]

Table 5 shows that the X-value premium spans both the value and investment premiums of Fama and French (2015): The intercepts, α , in the spanning regressions of both the HML and the CMA portfolios are insignificant. Given that X-value is unspanned in tests (3) and (5) in Table 4, this implies that the risks linked individually to the stock's BM and investment characteristics are just part of the risks captured by XM, which subsume them. The X-value factor, on the other hand, does not span the market, size, or profitability factors, as shown by their significant intercepts, α.

The other results in Table 5 have already been discussed: The interpretation of the estimates of θ in Eq. (18) and corresponding R^2 values is exactly the same as the ones obtained from Eq. (15) in tests (1) through (5) in Table 4 for each respective premium.

4.3 Does replacing value by X-value fix Fama and French (2015)?

The short answer is: Partially. One of the surprises in Fama and French (2015) is that value becomes redundant after controlling for the other four factors, so by dropping HML the five-factor model seems to be effectively a four-factor model. However, Fama and French (2015) claim that "in applications where the sole interest is abnormal returns (measured by regression intercepts), [...] a four-factor model that drops HML performs as well as the five-factor model" (p. 19). Thus, we should be able to keep the redundant factor in the model. This section basically answers two questions related to this statement: The first is whether X-value (as a replacement for value) would also be redundant in the presence of the other factors in Fama and French (2015), or if the model would effectively become a five-factor model. The second is whether keeping a redundant factor, such as HML, in Fama and French (2015) really has no effect on the estimated regression intercepts and related conclusions about abnormal returns.

The results of the spanning regressions in Table 6 help to answer these questions and analyze the properties of X-value as a substitute for the value factor in Fama and French (2015). In particular, I run spanning regressions, between July 1963 and December 2018, of each of the five factors in Fama and French (2015) on the other factors, but with the XHML portfolio replacing the HML portfolio. The estimated equations are of the form

$$R_{i,t} = \alpha + \boldsymbol{\theta}^{\top} \boldsymbol{R}_{x5,t} + \varepsilon_t, \tag{20}$$

where $^{\top}$ is the transposition sign, $R_{i,t}$ is either the market premium, MP, or the returns on the SMB, HML, RMW, CMA, or XHML portfolios,

$$R_{i,t} \in \{HML_t, MP_t, SMB_t, RMW_t, CMA_t, XHML_t\}, \tag{21}$$

and $\boldsymbol{R}_{x5,t}$ is (potentially, except for the factor eventually being spanned, R_i) the group of factors in Fama and French (2015), with XHML replacing HML,

$$\boldsymbol{R}_{x5,t} = (MP_t, SMB_t, RMW_t, CMA_t, XHML_t)^\top . \tag{22}$$

[Table 6 about here.]

The first column in Table 6 shows, for example, that HML earns significantly negative abnormal returns after the risk adjustment based on the modified Fama and French (2015) model that replaces HML by XHML: The regression intercept is −0.14 (−2.35 t-statistics).

Regarding the first question in this section, Table 6 shows that the regression of XHML on the other four factors in Fama and French (2015) has an estimated intercept of 0.19 (2.39 t-statistic). Therefore, replacing HML by XHML could effectively turn Fama and French (2015) into a five-factor model by resurrecting the "value" factor, given that XHML is not redundant in the presence of the other factors, unlike HML.

Regarding the second question, it is not true that a four-factor model that drops HML performs as well as the five-factor model of Fama and French (2015) in applications where the sole interest is abnormal returns. The two versions produce different risk adjustments. For example, at a 1% significance level, there is no evidence of abnormal returns associated with the XHML portfolio based on a version of Fama and French (2015) that drops HML. The regression's intercept has 2.39 t-statistic (Table 6, XHML column). In contrast, there is evidence of abnormal returns associated with the XHML portfolio, even at the 0.1% significance level, based on a version of Fama and French (2015) that **keeps** HML. In this case, the intercept of the regression has 3.35 t-statistic (Table 4, column (10)).

The results in this table also raise a related (and open) question about including or not including the investment factor in the model: According to Table 5, XHML spans CMA, so CMA is a redundant factor. According to Table 6, however, CMA is unspanned by the XHML and the other three factors in Fama and French (2015) jointly. In this case, the intercept is 0.23 (3.65 t-statistic) and CMA is not redundant. Given that Fama and French (2015) do not provide strong theoretical grounds to include or remove factors, it is unclear whether CMA should be included.

5 Acquisitions and a different sample period

This section provides a robustness check of the previous results across two dimensions: (i) It adds acquisitions to the list of inputs in the production function in Eq. (9) and in the estimated panel regressions in Eq. (10). This generates the estimate of expected cash flows needed to create the X-value factor. (ii) Acquisition data are only available from 1971, so the sample period over which the method can be evaluated starts in 1974 instead of 1963. In this section, the vector $\boldsymbol{f}_{i,t}$ in Eq. (11) includes acquisitions,

$$\boldsymbol{f}_{i,t} = (1, at_{i,t}, capx_{i,t}, emp_{i,t}, xrd_{i,t}, xad_{i,t}, aqc_{i,t}, ni_{i,t})^{\top}, \tag{23}$$

but all empirical procedures are the same.

The first quantitative (but not qualitative) difference appears in Fig. 4 (compared to Fig. 1): The X-value premium is larger than all premiums introduced by Fama and French (1996, 2015) in this shorter sample period in addition to using the different regressors in Eq. (23). Table 7 (compared to Table 3) shows that the X-value premium is still the largest and most significant (with the largest Sharpe ratio) among all factors in Fama and French (2015). But this happens by a smaller margin and excluding the market premium.

[Figure 4 about here.]

[Table 7 about here.]

Table 8 and Table 2 show the same properties for the six portfolios double sorted by size and XM under both formulations and sample periods: The excess returns on the portfolios and their significance increase from small to big stocks within all XM groups. The premiums also increase monotonically from low- to high-XM stocks within either the group of big or small stocks, as expected.

[Table 8 about here.]

The spanning tests in Table 9 (compared to Table 4) lead to the same conclusion: The X-value factor is unspanned by the Fama and French (2015) factors (either individually or in different combinations): The intercepts in all tests are significant (albeit less than in the longer sample).

[Table 9 about here.]

In terms of spanning the other factors, the results in Table 10 are even stronger than the ones in Table 5: Again, the X-value premium spans both the value and the investment premiums of Fama and French (2015). However, the intercepts, α, are even less significant in this sample (and the unspanned profitability premium is also marginally less significant).

[Table 10 about here.]

Finally, Table 11 confirms the main result in Table 6: Replacing HML by XHML "resurrects" the value premium. Unlike HML, XHML is not redundant in the presence of the other factors and it earns abnormal returns with significance given by the 2.06 t-statistic of this regression's intercept (compared to 2.39 in the longer sample in Table 6). Table 11 also shows that the negative abnormal returns associated with HML are even more significant in this sample: The regression intercept is -0.17 (-2.64 t-statistics) compared to -0.14 (-2.35) in Table 6.

[Table 11 about here.]

6 Conclusion

In this paper, we learn about further empirical support for the theoretical explanation of the value premium in Berk (1995) and de Oliveira Souza (2019a) in general, and the hypothesis that BE is simply a proxy for expected cash flows, in particular. The XM variable is – based on this theory – unconditionally more correlated with risk premiums than size, BM, investment, and profitability. Indeed, the ranking by XM is particularly successful at capturing (and subsumes) the risks associated with rankings by BM and investment.

We also learn that the value factor only becomes irrelevant in the five-factor model of Fama and French (2015) because BE is a relatively bad proxy for expected cash flows. As a consequence, the HML portfolio does not capture risks very well. Replacing BE by properly modeled cash flow expectations for each firm not only resurrects the unspanned value premium within the factor model of Fama and French (2015), but the resulting X-value factor in fact spans the investment factor instead.

References

Ball, Ray, Joseph J. Gerakos, Juhani T. Linnainmaa, and Valeri V. Nikolaev, 2018, Earnings, retained earnings, and book-to-market in the cross section of expected returns, *Journal of Financial Economics* forthcoming.

Berk, Jonathan, 1996, An empirical re-examination of the relation between firm size and return, *Unpublished manuscript, University of British Columbia* .

Berk, Jonathan B., 1995, A critique of size-related anomalies, *Review of Financial Studies* 8, 275–286.

Damodaran, Aswath, 2008, *Damodaran on valuation: Security analysis for investment and corporate finance* (John Wiley & Sons, Hoboken, NJ).

de Oliveira Souza, Thiago, 2019a, Correlation, causation, and stock returns, Unpublished working paper. University of Southern Denmark.

de Oliveira Souza, Thiago, 2019b, Price of risk fluctuations and the size premium, Unpublished working paper. University of Southern Denmark.

Doraszelski, Ulrich, and Jordi Jaumandreu, 2013, R&d and productivity: Estimating endogenous productivity, *Review of Economic Studies* 80, 1338–1383.

Fama, Eugene F., and Kenneth R. French, 1996, Multifactor explanations of asset pricing anomalies, *Journal of Finance* 51, 55–84.

Fama, Eugene F., and Kenneth R. French, 1997, Industry costs of equity, *Journal of Financial Economics* 43, 153–193.

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

Gerakos, Joseph, and Juhani T. Linnainmaa, 2017, Decomposing Value, *Review of Financial Studies* 31, 1825–1854.

Vuolteenaho, Tuomo, 2002, What drives firm-level stock returns?, *Journal of Finance* 57, 233–264.

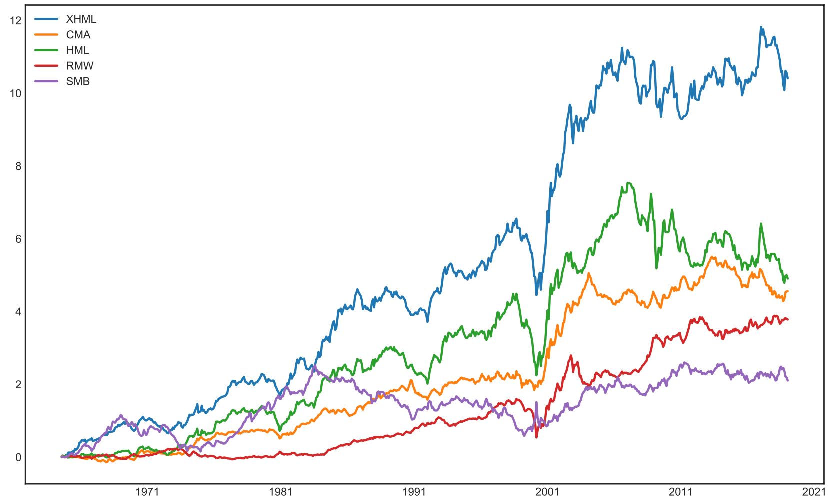

Figure 1: **Cumulative returns for each factor portfolio between July 1963 and December 2018.** The factor portfolios correspond to the X-value premium (XHML) and the Fama/French size (SMB), value (HML), profitability (RMW), and investment (CMA) premiums. The cumulative returns each month are calculated as

$$\pi_{i,t} = \prod_{t_0}^{t} \left(1 + R_{i,t}\right) - 1,$$

where t_0 is July 1963. The expected cash flow that normalizes the market value of each firm (to construct the XHML portfolio) is the expected net income as predicted by its 1-year lagged net income, total assets, number of employees, capital expenditures, R&D expenses, and advertisement expenses, with coefficients estimated by industry (based on the 48 industry classification of Fama and French, 1997). The X-value premium tends to be consistently and substantially larger than the others.

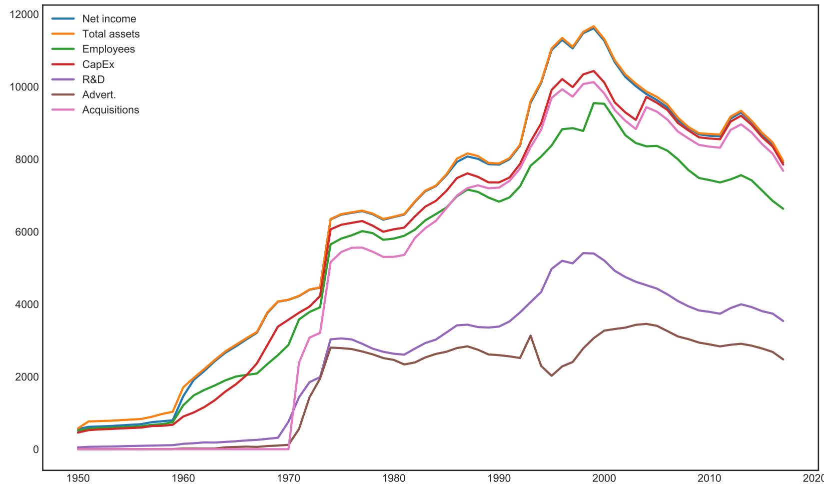

Figure 2: **Number of non-missing values for the cash flow-related variables in Compustat each year.** The variables are: Net income, total assets, number of employees, capital expenditures, R&D expenses, advertisement expenses, and acquisitions. The data on R&D, advertisement, and especially acquisitions do not become readily available until 1970.

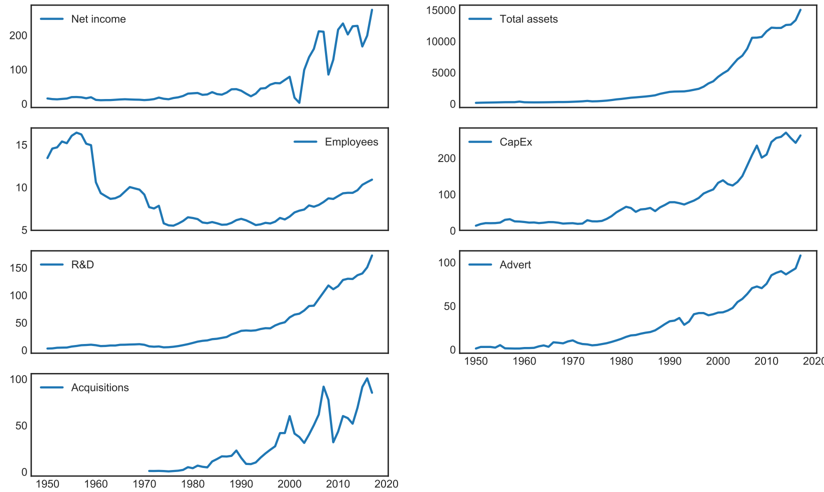

Figure 3: Cross-sectional average of non-missing values for the cash flow-related variables in Compustat per year. The variables, f, are: Net income (millions), total assets (millions), number of employees (thousands), capital expenditures (millions), R&D expenses (millions), advertisement expenses (millions), and acquisitions (millions). Each respective cross-sectional mean is given by

$$\overline{f}_t = \frac{1}{N_t}\sum_{i=1}^{N_t} f_{i,t},$$

where N_t is the total number of non-missing values for variable f in year t. Although acquisitions start to be reported in 1971, a large part of the firms still report very low values until close to 1980, when data quality seems to improve.

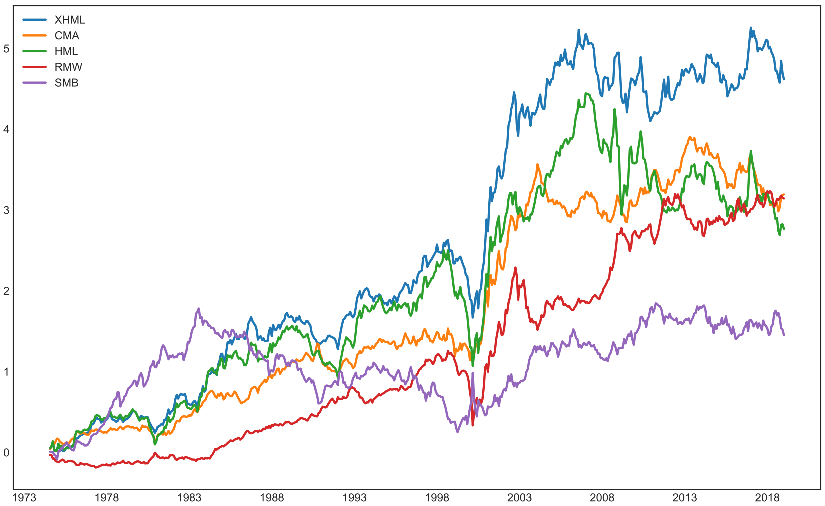

Figure 4: **Cumulative returns for each factor portfolio between July 1974 and December 2018.** The factor portfolios correspond to the X-value premium (XHML) and the Fama/French size (SMB), value (HML), profitability (RMW), and investment (CMA) premiums. The cumulative returns each month are calculated as

$$\pi_{i,t} = \prod_{t_0}^{t} \left(1 + R_{i,t}\right) - 1,$$

where t_0 is July 1974. The expected cash flow that normalizes the market value of each firm (to construct the XHML portfolio) is the expected net income as predicted by its 1-year lagged net income, total assets, number of employees, capital expenditures, R&D expenses, advertisement expenses, and acquisitions, with coefficients estimated by industry (based on the 48 industry classification of Fama and French, 1997). The X-value premium tends to be larger than the others in this shorter sample in which acquisitions can be used to predict cash flows. X-value improves relative to value, especially towards the end of the sample.

Table 1: Predictive panel regressions of net income between January 1950 and December 2018. The estimated equations are of the form

$$ni_{i,t+1} = \beta_0 + \boldsymbol{\beta}^\top \boldsymbol{f}_{i,t} + \varepsilon_{i,t+1},$$

where $^\top$ is the transposition sign, $\varepsilon_{i,t}$ is an error term, and $ni_{i,t}$ is net income of firm i at time t. In models (1) through (7), the (single) regressor $\boldsymbol{f}_{i,t}$ for firm i at time t is, respectively, net income, total assets, number of employees, capital expenditures, R&D expenditures, advertisement expenditures, or acquisitions. In model (8), the (multiple) regressors, $\boldsymbol{f}_{i,t}$, are all of these seven variables at once. I exclude observations that have zeros for the regressor in (1) through (7), but not in (8). The time unit is one year and the estimates are based on all Compustat firms. The table shows the number of firm-years depending on the regressor, the R^2, and coefficients with t statistics in parentheses, $^{*}\ p < 0.05$, $^{**}\ p < 0.01$, $^{***}\ p < 0.001$. Increases in all regressors (jointly or individually) forecast increases in cash flows.

	Net income							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Net income	0.7***							0.6***
	(454.05)							(307.09)
Total assets		0.005***						0.001***
		(155.88)						(40.85)
Employees			9.0***					1.1***
			(153.44)					(21.94)
CapEx				0.5***				0.1***
				(244.28)				(49.26)
R&D					1.1***			0.2***
					(135.43)			(34.38)
Advert.						1.8***		0.2***
						(132.73)		(14.67)
Acquis.							0.4***	0.06***
							(58.81)	(17.11)
Constant	35.6***	83.2***	34.8***	40.5***	51.6***	42.3***	161.7***	6.4***
	(23.15)	(41.82)	(15.78)	(20.84)	(12.37)	(11.15)	(26.07)	(3.94)
Firm-years	238862	239431	220018	216506	86556	76777	56525	223054
R-squared	0.463	0.092	0.097	0.216	0.175	0.187	0.058	0.491

Table 2: Average monthly excess percent returns for portfolios formed on Size and X-value between July 1963 and December 2018. At the end of each June, stocks are allocated to two Size groups (Small, Big) using NYSE market cap (median) breakpoints. Stocks are allocated independently to three X-value groups (Low, Medium, High), again using NYSE breakpoints (30th and 70th percentiles). The intersections of the two sorts produce six value-weight Size/X-value portfolios. In the sort for June of year t, X-value is the recursively estimated expected net income for the fiscal year ending in year $t-1$ divided by the market cap at the end of December of year $t-1$. The net income forecasters are (lagged) net income, total assets, number of employees, capital expenditures, R&D expenses, and advertisement expenses estimated by industry (using the 48 classification in Fama and French, 1997). The table shows averages of monthly returns in excess of the one-month Treasury bill rate, number of months, and t statistics in parentheses, $^{*}\ p < 0.05$, $^{**}\ p < 0.01$, $^{***}\ p < 0.001$. Conditioned on the other variable, the excess returns on the portfolios increase from low to high X-value and from big to small market cap.

	Small			Big		
	Low	Med.	High	Low	Med.	High
Mean	0.57^{*}	0.79^{***}	0.97^{***}	0.39^{*}	0.55^{***}	0.79^{***}
	(2.33)	(3.72)	(4.57)	(2.10)	(3.32)	(4.37)
Months	648	648	648	648	648	648

Table 3: Summary statistics for monthly factor percent returns between July 1963 and December 2018. MP is the market premium, and SMB, HML, RMW, and CMA are the other four factors in Fama and French (2015), size, value, profitability, and investment, respectively. XHML is the difference in returns between the two high and the two low X-value portfolios: At the end of each June, stocks are allocated to two Size groups (Small, Big) using NYSE market cap (median) breakpoints. Stocks are allocated independently to three X-value groups (Low, Medium, High), again using NYSE breakpoints (30th and 70th percentiles). The intersections of the two sorts produce six value-weight Size/X-value portfolios. In the sort for June of year t, X-value is the recursively estimated expected net income for the fiscal year ending in year $t-1$ divided by the market cap at the end of December of year $t-1$. The net income forecasters are (lagged) net income, total assets, number of employees, capital expenditures, R&D expenses, and advertisement expenses estimated by industry (using the 48 classification in Fama and French, 1997). The table shows the factor Sharpe ratios, sample size (in Months), and average monthly returns (Mean) with t statistics in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. The X-value factor has the largest significance/Sharpe ratio of all premiums and the difference with respect to the value premium is especially large.

	MP	SMB	HML	RMW	CMA	XHML
Mean	0.50^{**}	0.22	0.32^{**}	0.26^{**}	0.28^{***}	0.40^{***}
	(2.86)	(1.82)	(2.84)	(3.07)	(3.58)	(4.20)
Sharpe	0.39	0.25	0.39	0.42	0.49	0.57
Months	648	648	648	648	648	648

Table 4: Spanning tests of the X-value premium between July 1963 and December 2018. The estimated equations are of the form

$$R_{xhml,t} = \alpha + \boldsymbol{\theta}^{\top}\boldsymbol{R}_{b,t} + \varepsilon_t,$$

where $^{\top}$ is the transposition sign, $R_{xhml,t}$ is the return on the XHML portfolio, and $\boldsymbol{R}_{b,t}$ are the returns on the five Fama/French benchmark portfolios either in groups or individually. In columns (1) through (5), $R_{b,t}$ is, respectively, the market premium, MP, or the returns on the SMB, HML, RMW, or CMA portfolios. In columns (6) through (9), $\boldsymbol{R}_{b,t}$ contains the market premium, MP, in addition to, respectively, the returns on the SMB, HML, RMW, or CMA portfolios. In column (10), $\boldsymbol{R}_{b,t}$ contains all five Fama/French premiums. $R_{xhml,t}$ is the difference in returns between the two high and the two low X-value portfolios: At the end of each June, stocks are allocated to two Size groups (Small, Big) using NYSE market cap (median) breakpoints. Stocks are allocated independently to three X-value groups (Low, Medium, High), again using NYSE breakpoints (30th and 70th percentiles). The intersections of the two sorts produce six value-weight Size/X-value portfolios. In the sort for June of year t, X-value is the recursively estimated expected net income for the fiscal year ending in year $t-1$ divided by the market cap at the end of December of year $t-1$. The net income forecasters are (lagged) net income, total assets, number of employees, capital expenditures, R&D expenses, and advertisement expenses estimated by industry (using the 48 classification in Fama and French, 1997). The table shows the number of months in each sample, the R^2, and coefficients with t statistics in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. The X-value factor is unspanned by the existing factors, individually or in different combinations.

	XHML									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
MP	-0.17***					-0.15***	-0.06***	-0.16***	-0.06**	-0.06***
	(-8.34)					(-7.29)	(-4.53)	(-7.71)	(-3.13)	(-4.08)
SMB		-0.14***				-0.07*				0.03
		(-4.59)				(-2.41)				(1.41)
HML			0.69***				0.67***			0.69***
			(36.08)				(34.15)			(26.04)
RMW				0.15***				0.08		0.08**
				(3.57)				(1.84)		(3.07)
CMA					0.68***				0.63***	-0.04
					(17.53)				(15.08)	(-0.90)
Constant	0.48***	0.43***	0.18**	0.35***	0.20**	0.49***	0.21***	0.45***	0.25**	0.19***
	(5.32)	(4.58)	(3.23)	(3.78)	(2.60)	(5.43)	(3.89)	(5.00)	(3.13)	(3.35)
Months	648	648	648	648	648	648	648	648	648	648
R-squared	0.097	0.032	0.668	0.019	0.322	0.105	0.679	0.102	0.333	0.684

Table 5: Spanning regressions of each of the five individual Fama/French factor premiums on the X-value premium between July 1963 and December 2018. The estimated equations are of the form

$$R_{FF,t} = \alpha + \theta R_{xhml,t} + \varepsilon_t,$$

where $R_{xhml,t}$ is the return on the XHML portfolio and $R_{FF,t}$ is either the market premium, MP, or the returns on the SMB, HML, RMW, or CMA portfolios of Fama/French. $R_{xhml,t}$ is the difference in returns between the two high and the two low X-value portfolios: At the end of each June, stocks are allocated to two Size groups (Small, Big) using NYSE market cap (median) breakpoints. Stocks are allocated independently to three X-value groups (Low, Medium, High), again using NYSE breakpoints (30th and 70th percentiles). The intersections of the two sorts produce six value-weight Size/X-value portfolios. In the sort for June of year t, X-value is the recursively estimated expected net income for the fiscal year ending in year $t-1$ divided by the market cap at the end of December of year $t-1$. The net income forecasters are (lagged) net income, total assets, number of employees, capital expenditures, R&D expenses, and advertisement expenses estimated by industry (using the 48 classification in Fama and French, 1997). The table shows the sample size (in months), the R^2, and coefficients with t statistics in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. The X-value factor spans the value and investment factors.

	MP	SMB	HML	RMW	CMA
XHML	-0.58***	-0.23***	0.96***	0.13***	0.48***
	(-8.34)	(-4.59)	(36.08)	(3.57)	(17.53)
Constant	0.73***	0.31*	-0.07	0.21*	0.09
	(4.32)	(2.57)	(-1.00)	(2.47)	(1.43)
Months	648	648	648	648	648
R-squared	0.097	0.032	0.668	0.019	0.322

Table 6: **Spanning regressions with the XHML portfolio replacing the HML portfolio as the fifth Fama/French factor between July 1963 and December 2018.** The estimated equations are of the form

$$R_{i,t} = \alpha + \boldsymbol{\theta}^{\top} \boldsymbol{R}_{x5,t} + \varepsilon_t,$$

where $^{\top}$ is the transposition sign, $R_{i,t}$ is either the market premium, MP, the returns on the SMB, HML, RMW, CMA portfolios of Fama/French, or the return on the XHML portfolio. $\boldsymbol{R}_{x5,t}$ is the group of returns, MP, SMB, RMW, CMA and XHML (except the one eventually being spanned, as R_i). The return on the XHML portfolio is the difference in returns between the two high and the two low X-value portfolios: At the end of each June, stocks are allocated to two Size groups (Small, Big) using NYSE market cap (median) breakpoints. Stocks are allocated independently to three X-value groups (Low, Medium, High), again using NYSE breakpoints (30th and 70th percentiles). The intersections of the two sorts produce six value-weight Size/X-value portfolios. In the sort for June of year t, X-value is the recursively estimated expected net income for the fiscal year ending in year $t-1$ divided by the market cap at the end of December of year $t-1$. The net income forecasters are (lagged) net income, total assets, number of employees, capital expenditures, R&D expenses, and advertisement expenses estimated by industry (using the 48 classification in Fama and French, 1997). The table shows the number of months in each sample, the R^2, and coefficients with t statistics in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. The X-value factor is unspanned in a multivariate regression including the MP, SMB, CMA, and RMW, while the value premium earns negative excess returns when we add XHML to these regressors.

	HML	MP	SMB	RMW	CMA	XHML
MP	0.05***		0.11***	-0.08***	-0.11***	-0.04
	(3.87)		(4.13)	(-4.37)	(-7.42)	(-1.87)
SMB	-0.04*	0.23***		-0.26***	-0.06**	-0.01
	(-2.16)	(4.13)		(-9.84)	(-2.68)	(-0.19)
RMW	-0.01	-0.34***	-0.51***		-0.18***	0.15***
	(-0.49)	(-4.37)	(-9.84)		(-5.86)	(3.91)
CMA	0.51***	-0.70***	-0.19**	-0.28***		0.65***
	(14.33)	(-7.42)	(-2.68)	(-5.86)		(15.55)
XHML	0.74***	-0.15	-0.01	0.15***	0.42***	
	(26.04)	(-1.87)	(-0.19)	(3.91)	(15.55)	
Constant	-0.14*	0.80***	0.36**	0.38***	0.23***	0.19*
	(-2.35)	(5.02)	(3.15)	(4.82)	(3.65)	(2.39)
Months	648	648	648	648	648	648
R-squared	0.755	0.233	0.214	0.215	0.403	0.351

Table 7: Summary statistics for monthly factor percent returns between July 1974 and December 2018. MP is the market premium, and SMB, HML, RMW, and CMA are the other four factors in Fama and French (2015), size, value, profitability, and investment, respectively. XHML is the difference in returns between the two high and the two low X-value portfolios: At the end of each June, stocks are allocated to two Size groups (Small, Big) using NYSE market cap (median) breakpoints. Stocks are allocated independently to three X-value groups (Low, Medium, High), again using NYSE breakpoints (30th and 70th percentiles). The intersections of the two sorts produce six value-weight Size/X-value portfolios. In the sort for June of year t, X-value is the recursively estimated expected net income for the fiscal year ending in year $t-1$ divided by the market cap at the end of December of year $t-1$. The net income forecasters are (lagged) net income, total assets, number of employees, capital expenditures, R&D expenses, advertisement expenses, and acquisitions by industry (using the 48 classification in Fama and French, 1997). The table shows the factor Sharpe ratios, sample size (in Months), and average monthly returns (Mean) with t statistics in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. In this smaller sample, the X-value premium is still the largest and most significant (with the largest Sharpe ratio) among all factors, but now by a smaller margin and excluding the market premium.

	MP	SMB	HML	RMW	CMA	XHML
Mean	0.67***	0.23	0.29*	0.32**	0.26**	0.35***
	(3.52)	(1.77)	(2.30)	(3.23)	(3.09)	(3.46)
Sharpe	0.53	0.27	0.35	0.49	0.47	0.52
Months	528	528	528	528	528	528

Table 8: Average monthly excess percent returns for portfolios formed on Size and X-value between July 1974 and December 2018. At the end of each June, stocks are allocated to two Size groups (Small, Big) using NYSE market cap (median) breakpoints. Stocks are allocated independently to three X-value groups (Low, Medium, High), again using NYSE breakpoints (30th and 70th percentiles). The intersections of the two sorts produce six value-weight Size/X-value portfolios. In the sort for June of year t, X-value is the recursively estimated expected net income for the fiscal year ending in year $t-1$ divided by the market cap at the end of December of year $t-1$. The net income forecasters are (lagged) net income, total assets, number of employees, capital expenditures, R&D expenses, advertisement expenses, and acquisitions estimated by industry (using the 48 classification in Fama and French, 1997). The table shows averages of monthly returns in excess of the one-month Treasury bill rate, number of months, and t statistics in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Conditioned on the other variable, the excess returns on the portfolios increase from low to high X-value and from big to small market cap.

	Small			Big		
	Low	Med.	High	Low	Med.	High
Mean	0.82^{**}	1.01^{***}	1.15^{***}	0.57^{**}	0.72^{***}	0.94^{***}
	(3.10)	(4.51)	(5.07)	(2.78)	(4.00)	(4.82)
Months	528	528	528	528	528	528

Table 9: Spanning tests of the X-value premium between July 1974 and December 2018. The estimated equations are of the form

$$R_{xhml,t} = \alpha + \boldsymbol{\theta}^{\top} \boldsymbol{R}_{b,t} + \varepsilon_t,$$

where $^{\top}$ is the transposition sign, $R_{xhml,t}$ is the return on the XHML portfolio, and $\boldsymbol{R}_{b,t}$ is the vector of returns on the five Fama/French benchmark portfolios either in groups or individually. In columns (1) through (5), $R_{b,t}$ is, respectively, the market premium, MP, or the returns on the SMB, HML, RMW, or CMA portfolios. In columns (6) through (9), $\boldsymbol{R}_{b,t}$ contains the market premium, MP, in addition to, respectively, the returns on the SMB, HML, RMW, or CMA portfolios. In column (10), $\boldsymbol{R}_{b,t}$ contains all five Fama/French premiums. $R_{xhml,t}$ is the difference in returns between the two high and the two low X-value portfolios: At the end of each June, stocks are allocated to two Size groups (Small, Big) using NYSE market cap (median) breakpoints. Stocks are allocated independently to three X-value groups (Low, Medium, High), again using NYSE breakpoints (30th and 70th percentiles). The intersections of the two sorts produce six value-weight Size/X-value portfolios. In the sort for June of year t, X-value is the recursively estimated expected net income for the fiscal year ending in year $t-1$ divided by the market cap at the end of December of year $t-1$. The net income forecasters are (lagged) net income, total assets, number of employees, capital expenditures, R&D expenses, advertisement expenses, and acquisitions estimated by industry (using the 48 classification in Fama and French, 1997). The table shows the number of months in each sample, the R^2, and coefficients with t statistics in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. The X-value factor is unspanned by the existing factors, individually or in different combinations.

	XHML									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
MP	-0.18***					-0.15***	-0.07***	-0.15***	-0.08***	-0.07***
	(-8.16)					(-7.01)	(-5.94)	(-6.60)	(-3.84)	(-4.89)
SMB		-0.18***				-0.13***				-0.00
		(-5.65)				(-3.92)				(-0.12)
HML			0.67***				0.64***			0.66***
			(35.01)				(33.63)			(25.78)
RMW				0.28***				0.20***		0.10***
				(6.62)				(4.65)		(3.63)
CMA					0.68***				0.61***	-0.06
					(15.89)				(13.56)	(-1.45)
Constant	0.47***	0.39***	0.15**	0.26**	0.17*	0.48***	0.21***	0.38***	0.24**	0.19***
	(4.87)	(3.99)	(2.79)	(2.65)	(2.05)	(5.08)	(3.89)	(4.00)	(2.84)	(3.35)
Months	528	528	528	528	528	528	528	528	528	528
R-squared	0.112	0.057	0.700	0.077	0.324	0.138	0.719	0.148	0.343	0.729

Table 10: Spanning regressions of each of the five individual Fama/French factor premiums on the X-value premium between July 1974 and December 2018. The estimated equations are of the form

$$R_{FF,t} = \alpha + \theta R_{xhml,t} + \varepsilon_t,$$

where $R_{xhml,t}$ is the return on the XHML portfolio and $R_{FF,t}$ is either the market premium, MP, or the returns on the SMB, HML, RMW, or CMA portfolios of Fama/French. $R_{xhml,t}$ is the difference in returns between the two high and the two low X-value portfolios: At the end of each June, stocks are allocated to two Size groups (Small, Big) using NYSE market cap (median) breakpoints. Stocks are allocated independently to three X-value groups (Low, Medium, High), again using NYSE breakpoints (30th and 70th percentiles). The intersections of the two sorts produce six value-weight Size/X-value portfolios. In the sort for June of year t, X-value is the recursively estimated expected net income for the fiscal year ending in year $t-1$ divided by the market cap at the end of December of year $t-1$. The net income forecasters are (lagged) net income, total assets, number of employees, capital expenditures, R&D expenses, advertisement expenses, and acquisitions estimated by industry (using the 48 classification in Fama and French, 1997). The table shows the sample size (in months), the R^2, and coefficients with t statistics in parentheses, $^{*}\ p < 0.05$, $^{**}\ p < 0.01$, $^{***}\ p < 0.001$. The X-value factor spans the value and investment factors.

	MP	SMB	HML	RMW	CMA
XHML	-0.64***	-0.31***	1.04***	0.27***	0.48***
	(-8.16)	(-5.65)	(35.01)	(6.62)	(15.89)
Constant	0.90***	0.34**	-0.07	0.22*	0.09
	(4.91)	(2.65)	(-1.07)	(2.33)	(1.34)
Months	528	528	528	528	528
R-squared	0.112	0.057	0.700	0.077	0.324

Table 11: Spanning regressions with the XHML portfolio replacing the HML portfolio as the fifth Fama/French factor between July 1974 and December 2018. The estimated equations are of the form

$$R_{i,t} = \alpha + \boldsymbol{\theta}^{\top}\boldsymbol{R}_{x5,t} + \varepsilon_t,$$

where $^{\top}$ is the transposition sign, $R_{i,t}$ is either the market premium, MP, the returns on the SMB, HML, RMW, CMA portfolios of Fama/French, or the return on the XHML portfolio. $\boldsymbol{R}_{x5,t}$ is the group of returns MP, SMB, RMW, CMA and XHML (except the one eventually being spanned, as R_i). The return on the XHML portfolio is the difference in returns between the two high and the two low X-value portfolios: At the end of each June, stocks are allocated to two Size groups (Small, Big) using NYSE market cap (median) breakpoints. Stocks are allocated independently to three X-value groups (Low, Medium, High), again using NYSE breakpoints (30th and 70th percentiles). The intersections of the two sorts produce six value-weight Size/X-value portfolios. In the sort for June of year t, X-value is the recursively estimated expected net income for the fiscal year ending in year $t-1$ divided by the market cap at the end of December of year $t-1$. The net income forecasters are (lagged) net income, total assets, number of employees, capital expenditures, R&D expenses, advertisement expenses, and acquisitions estimated by industry (using the 48 classification in Fama and French, 1997). The table shows the number of months in each sample, the R^2, and coefficients with t statistics in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. The X-value factor is unspanned in a multivariate regression including the MP, SMB, CMA, and RMW, while the value premium earns negative excess returns when we add XHML to these regressors.

	HML	MP	SMB	RMW	CMA	XHML
MP	0.07***		0.07*	-0.09***	-0.10***	-0.04*
	(4.75)		(2.40)	(-4.22)	(-5.92)	(-2.01)
SMB	-0.03	0.16*		-0.30***	-0.00	-0.06
	(-1.47)	(2.40)		(-10.09)	(-0.01)	(-1.85)
RMW	-0.03	-0.37***	-0.55***		-0.10**	0.17***
	(-1.05)	(-4.22)	(-10.09)		(-2.86)	(4.31)
CMA	0.50***	-0.64***	-0.00	-0.16**		0.61***
	(12.76)	(-5.92)	(-0.01)	(-2.86)		(13.95)
XHML	0.85***	-0.19*	-0.12	0.20***	0.44***	
	(25.78)	(-2.01)	(-1.85)	(4.31)	(13.95)	
Constant	-0.17**	0.99***	0.40***	0.42***	0.20**	0.17*
	(-2.64)	(5.60)	(3.32)	(4.79)	(2.89)	(2.06)
Months	528	528	528	528	528	528
R-squared	0.775	0.218	0.242	0.279	0.371	0.384