

# Is the Value Premium Smaller Than We Thought?

Mathias Hasler<sup>1</sup>

January 12<sup>th</sup>, 2023

## Abstract

The construction of the original HML portfolio (Fama and French, 1993) includes six seemingly innocuous decisions that could easily have been replaced with alternatives that are just as reasonable. I propose such alternatives and construct HML portfolios. In sample, the average of these alternative estimates of the value premium is smaller than the original estimate of the value premium. The difference is 0.08% per month and statistically significant. Out of sample, however, this difference is statistically indistinguishable from zero. The results suggest that the original value premium estimate is upward biased because of a potential chance result in the original research decisions.

JEL Classification: G10, G11, G12, G14, C1, C2

Keywords: Value premium, HML portfolio, research decisions, chance result, statistical biases

---

<sup>1</sup>Boston College, Carroll School of Management, 140 Commonwealth Avenue, Chestnut Hill, MA 02467. My email address is [haslerm@bc.edu](mailto:haslerm@bc.edu) and my webpage is [www.mathiashasler.com](http://www.mathiashasler.com). I thank my dissertation committee, Jeffrey Pontiff, Ronnie Sadka, Alan Marcus, and Vincent Bogousslavsky, for their advice and their valuable time. I also thank the anonymous referee at the Critical Finance Review and the seminar participants at Boston College, the Asia School of Business (in collaboration with MIT Sloan School of Management), the European Financial Management Association Conference (EFMA) 2021, the French Finance Association (AFFI PhD workshop) 2021, the Future of Financial Information Conference 2022, and the Financial Management Association (FMA) Conference 2022.

The value premium compensates investors for a unit exposure to the value factor. Fama and French (1993) propose the HML portfolio as a proxy for the value factor and estimate a statistically significant value premium of 0.40% per month using data from July 1963 to December 1991.

The construction of this HML portfolio includes seemingly innocuous decisions that could easily have been replaced with alternatives that are just as reasonable. For example, the decision to sort stocks into portfolios once a year at the end of June is just as reasonable as the alternative to sort stocks into portfolios at the end of each month. If we assume that the original decisions are similar to the alternative decisions, then the original decisions and their alternatives should produce similar estimates of the value premium. Thus, the average estimate over all estimates produced from all HML portfolios constructed with all possible combinations of the original and the alternative decisions serves as an estimate of the underlying value premium. The difference between the original and this average estimate serves as an estimate for the bias in the original value premium estimate that is potentially due to chance in the original research decisions. This new approach can be extended to any empirical study.

I study the original value premium estimate because of its academic importance. The Fama and French (1993) three-factor model, which includes the original HML portfolio as a proxy for the value factor, has been serving as a benchmark asset pricing model in empirical finance. Mutual fund studies on performance evaluation, for example Fama and French (2010), include the original HML portfolio as a right-hand side variable to proxy for the value factor. Bias in the original value premium estimate may therefore affect our inference of the three-factor model's importance as an asset pricing model and on mutual fund managers' abilities.

I estimate the bias of the original value premium estimate that is potentially due to chance in the original decisions, because a chance result is always possible in empirical research. Gelman and Loken (2013), for example, explain that researchers' implicit research decisions are bifurcations that create a "Garden of Forking Paths" in which researchers may end up on a path by chance on which they find an unusual result. Moreover, many statistical biases are inherent to research and lead to biased estimates in published papers. Harvey's (2017) Presidential Address to the American Finance Association elaborates on the garden of forking paths as well as the file drawer effect, data-mining, multiple hypothesis testing, etc. McLean and Pontiff (2016) and

Linnainmaa and Roberts (2018) provide empirical evidence that statistical biases explain much of the anomaly returns in published papers.

In this paper, I focus on six seemingly innocuous decisions in the construction of the original HML portfolio. The first decision is about the timing of market equity, the second decision is about the timing of book equity, the third decision is about negative book equity, the fourth decision is about financial firms, the fifth decision is about portfolio sorting breakpoints, and the sixth decision is about the timing of market equity to account for the size effect. I propose alternatives that are just as reasonable, form all possible combinations of these six decisions and their alternatives, construct 96 HML portfolios, and collapse these 96 HML portfolios in each month into an equally weighted average portfolio. This average HML portfolio is a valuable proxy for the value factor because it reflects an average decision that mitigates a decision-specific chance result. The average return difference between the original HML portfolio and the average HML portfolio is therefore an estimate for the bias of the original value premium estimate.

I start my empirical analysis with the replication of the original HML portfolio as described in Fama and French (1993). I report that I can closely replicate the original HML portfolio. The HML portfolio has an average monthly return of 0.40% per month in the original study's sample from July 1963 to December 1991. The replicated HML portfolio has an average monthly return of 0.39% per month in the same sample period. The difference of 0.01% per month may be a result of differences in the exact construction or that of updating the CRSP and Compustat datasets as reported in Akey, Robertson and Simutin (2022).

In the baseline empirical test, I calculate the average monthly returns of the average HML portfolio and of the original HML portfolio in the original sample's study from July 1963 to December 1991. The average HML portfolio has an average monthly return of 0.31% per month ( $t$ -statistic of 2.14) while the original HML portfolio has an average monthly return of 0.39% per month ( $t$ -statistic of 2.87). The average return difference between the original HML portfolio and the average HML portfolio is 0.08% per month ( $t$ -statistic of 1.72). Also, the original estimate of the value premium is at least as large as 85% of all other value premium estimates. These findings suggest that the original value premium estimate is upward biased because of a potential chance result in the original research decisions.

I also perform a White (2000) Reality Check Bootstrap to test the null hypothesis that the value premium is zero. This is important because the  $t$ -statistic of the average HML portfolio does not account for the fact that we have more information than just the average portfolio and because the  $t$ -statistic of the original HML portfolio does not account for the fact that a researcher has the freedom to choose the HML portfolio with the highest  $t$ -statistic out of the 96 HML portfolio to proxy for the value factor. I find that the  $t$ -statistic of the original HML portfolio exceeds the critical value from the bootstrap assuming a confidence level of 95% and thus rejects the null hypothesis that the value premium is zero.

I compare the standard deviations of the original HML portfolio and the average HML portfolio to better understand whether the positive return difference is compensation for more risk. The original HML portfolio has a lower standard deviation than the average HML portfolio (2.53% versus 2.72% per month), and thus suggests that the positive return difference is not compensation for more risk.

I perform a principal component analysis of the monthly returns of the 96 HML portfolios to better understand whether the HML portfolios may be proxying for more than one underlying factor. The first principal component explains 91% of the variation of the HML portfolios, and thus suggests that they are proxying for one underlying factor.

I estimate the main empirical test using data that is out of sample. If the return difference between the original and the average HML portfolio is the result of a bias that is due to chance in research decisions, then I expect the return difference to be zero out of sample. In the pre-sample, from July 1926 to June 1963, the return difference between the original and the average HML portfolio is 0.07% per month ( $t$ -statistic 1.34). This pre-sample result is somewhat at odds with the idea of a bias. In the post-sample, from January 1992 to December 2021, the return difference between the original and the average HML portfolio is 0.05% per month ( $t$ -statistic 0.72). The post-sample result is consistent with a bias.

I also estimate the main empirical test using the full sample from July 1926 to December 2021. The average HML portfolio has an average monthly return of 0.25% per month ( $t$ -statistic of 2.19) while the original HML portfolio has an average monthly return of 0.32% per month ( $t$ -statistic of 3.02). The return difference between the original and the average HML portfolio on average is

0.07% per month ( $t$ -statistic of 2.01). The full sample results suggest that the original value premium estimate is upward biased.

In the last empirical test, I study the conditional value premium estimate using the approach in Barillas and Shanken (2017) and the Fama and French (2015) five-factor model augmented with momentum. The conditional value premium estimates using the original HML portfolio as a proxy for the underlying value factor is smaller than the estimate using the average HML portfolio in sample, but not out of sample. The result suggests that the conditional value premium estimate is downward biased due to a potential chance in seemingly innocuous research decisions. The negative bias is consistent with Asness and Frazzini (2013). The spanning regressions also report that the original HML portfolio has a less negative beta with momentum than the average HML portfolio. This suggests that the original value premium estimate may be larger in sample and unconditionally because it avoids trading against momentum. The inference whether or not the original value premium is biased therefore hinges on the unresolved debate whether or not momentum is an asset pricing factor (Fama and French, 2018; Jegadeesh and Titman, 2010).

## 1 Related Literature

### 1.1 Robustness of the Original Value Premium Estimate

This paper is related to the literature on the robustness of the original value premium estimate to address concerns put forth by Black (1993) and by Lo and MacKinlay (1990) that the value premium is not real.

The tenor of this literature is that the original value premium estimate is robust. Davis, Fama, and French (2000) document a significant value premium estimate in the US equity market in the sample from 1929 to 1963 that precedes the original study's sample. Barber and Lyon (1997) document a significant value effect in the holdout sample of financial firms in the US equity market. Chan, Hamao, and Lakonishok (1991) report a significant value premium estimate in Japan's equity market, which is the second largest equity market in 1990. Fama and French (1998) document significant value premium estimates in international equity markets. Asness, Moskowitz, and Pedersen (2013) report significant value premium estimates in other asset classes.

Fama and French (1992) report a significant value effect that is robust to the return predictability of leverage, as reported in Bhandari (1988), and of earnings to price, as reported in Basu (1977, 1983). Asness and Frazzini (2013) report a significant value premium estimate for an updated HML portfolio conditional on momentum. Kessler, Scherer, and Harries (2019) find significant returns for valuation-based portfolios with different “design choices.” Fama and French (2008) report that the value effect exists in microcap stocks, small stocks, and big stocks, using portfolio sorts and cross-sectional regressions conditional on other stock return predictors, but Phalippou (2008) documents that the value effect exists only among some of the smallest stocks in the US equity market. Studies with more recent samples, however, find insignificant value premium estimates. Schwert (2003), Linnainmaa and Roberts (2018), and Fama and French (2020) report insignificant value premium estimates in the US equity market in samples that follow the original sample’s study. But Fama and French (2020) report a significant value premium estimate for the full sample from 1963 to 2019.

My contribution to the literature is the empirical finding that the value premium is smaller than we thought. The construction of the original HML portfolio (Fama and French, 1993) includes six seemingly innocuous decisions that could easily have been replaced with alternatives that are just as reasonable. In sample, the average estimate of the value premium is smaller than the original estimate of the value premium. The difference is 0.08% per month and statistically significant. Out of sample, the estimates of the value premium are similar. These results suggest that the original value premium estimate is upward biased due to a potential chance result in the original research decisions. They imply that the importance of the Fama and French (1993) three-factor model as a benchmark asset pricing model is lower than we thought and that mutual fund managers’ abilities may be larger than we thought.

## **1.2 Approaches to Address Chance or Statistical Biases**

This paper is also related to the literature on approaches to address bias in estimates of published papers that are due to chance or statistical biases.

Harvey, Liu, and Zhu (2016) propose a  $t$ -statistic of 3 as a simple critical value for future empirical asset pricing factors to account for multiple testing. They use the number of published

empirical asset pricing factors over time as a conservative estimate for the number of tested asset pricing factors and study the approaches by Bonferroni (1936), Holmes (1979), and Benjamini and Yekutieli (2001). Chordia, Goyal, and Saretto (2020) simulate 2.4 million trading strategies, use a model to infer the number of strategies that are tested by financial economists, and propose a  $t$ -statistic of 3.8 as a critical value for hypothesis tests. Harvey and Liu (2020) use a bootstrap approach and propose critical  $t$ -statistics that optimize the Type I errors in relation to Type II errors. Harvey (2017) proposes Bayesianized  $p$ -values that take prior beliefs into account.

My contribution to the literature is a new approach to estimating the bias in the original estimate of the value premium that is attributable to chance in research decisions. This new approach is simple and intuitive: Ex ante, I expect that the decisions and their reasonable alternatives produce estimates of the value premium that are the similar. If they are not similar, then the original estimate may be biased or even a false positive that occurs because of a potential chance result in the original decisions. This new approach is applicable in sample, and it can be extended to any empirical study.

My approach is related to robustness tests in general. Fama and French (1992), for example, report that the book-to-market effect is robust to replacing end of December market equity with fiscal year-end market equity. Pontiff and Singla (2019), as another example, report that the original estimate of the liquidity premium (Pastor and Stambaugh, 2003) is statistically insignificant when they introduce four modifications into the construction of the liquidity factor proxy that are expected to improve statistical power or to reduce estimation error. My approach is different, however, because it studies seemingly innocuous decisions that have alternatives that are just as reasonable and because it aggregates all decisions and alternatives into one estimate.

## 2 HML Portfolios

### 2.1 Definition of the Original HML Portfolio

The original HML portfolio is defined in Fama and French (1993). It uses data from the intersection of CRSP and Compustat. The sample is restricted to common ordinary US stocks (share code 10 or 11) that are trading on the NYSE, the NASDAQ, and the Amex (exchange code 1, 2, or 3) with non-

missing and non-negative book equity, non-missing market equity at the end of December of year  $t-1$  and at the end of June of year  $t$ , and with at least two years of available Compustat data.

The original HML portfolio is the average return of a small and a big value portfolio minus the average return of a small and a big growth portfolio in each month. Formally,  $HML = 1/2 (\text{Small Value} + \text{Big Value}) - 1/2 (\text{Small Growth} + \text{Big Growth})$ . Stocks are sorted into six portfolios by independently sorting them on market equity into small and big stocks using the median market capitalization of all stocks traded on the NYSE as breakpoint and by independently sorting them on book-to-market equity into value, neutral, and growth stocks using the 30<sup>th</sup> and 70<sup>th</sup> percentiles of book-to-market equity of all stocks traded on the NYSE as breakpoints. The portfolios are constructed at the end of June of year  $t$  and are held from July of year  $t$  to June of year  $t+1$ . Market equity observed at the end of June of year  $t$  is used to sort stocks on size. The book equity of a firm's last fiscal year with fiscal year-end before the end of December of year  $t-1$  divided by market equity at the end of December of year  $t-1$  is used to sort stocks on value. The two neutral portfolios (Small Neutral and Big Neutral) are not used. The six portfolios are value weighted.

I use the book equity definition as in Davis, Fama, and French (2000) because the description is more detailed.<sup>2</sup> Book equity is defined as stockholder book equity plus balance sheet deferred taxes and investment tax credit (txdltc), if available, minus the book value of preferred stock. Stockholder equity is the value reported by Compustat (seq). If not available, stockholders' equity is measured as the book value of common equity (ceq) plus the par value of preferred stock (pstk), or the book value of assets minus total liabilities (at-lt), in that order. The book value of preferred stock is measured as redemption (pstkrv), liquidation (pstkli), or par value (pstk), if available and in that order.

---

<sup>2</sup>I used the book equity definition as in Fama and French (1993) in a previous version of this paper. The definition in Davis, Fama and French (2000) is more detailed and it is the book equity definition that is also used in Novy-Marx (2013), Fama and French (2015), Linnainmaa and Roberts (2018), Gerakos and Linnainmaa (2018), and Ball, Gerakos, Linnainmaa, and Nikolaev (2019).



## 2.2 Replication Validation of the Original HML Portfolio

Table I reports the average monthly return of the original HML portfolio as reported in Fama and French (1993, p. 13) and the average monthly return of my replicated HML portfolio in the original study's sample from July 1963 to December 1991. I use the February 2022 vintages of CRSP and Compustat.

The original HML portfolio has an average monthly return of 0.40%. The replicated HML portfolio has an average monthly return of 0.39% over the same time period. The return difference is 0.01% per month. Table I additionally shows that the original and the replicated HML portfolios have similar standard deviations, *t*-statistics, autocorrelations, as well as similar correlations with the market risk factor and the SMB portfolio. Therefore, I report that I can closely replicate the original HML portfolio.

I additionally compare the average monthly return of my replicated HML portfolio with the average monthly return of the replicated HML portfolios of eight studies over the same time period in the eight studies. Table II reports that my replicated HML portfolio has a monthly return that is similar to the average monthly returns of the replicated HML portfolios in these eight studies.

[Table II here]

As in Akey, Robertson and Simutin (2022), I compare the monthly returns of the replicated HML portfolio with the monthly returns of the HML portfolio downloaded from Kenneth French's webpage in February 2022. In the original sample period, the replicated HML portfolio has an average return of 0.39% per month while the downloaded HML portfolio has an average return of 0.38% per month. 20 out of 342 observations have returns that deviate by a greater amount than 1% and 72 out of 342 observations deviated by a greater amount than 0.5%. In the full sample period, the replicated HML portfolio has an average return of 0.32% per month while that of the downloaded HML portfolio is 0.33% per month. 128 out of 1,145 returns deviate by more than 1% and 326 out of 1,145 returns deviate by more than 0.5%.

### 2.3 Alternatives that are Just as Reasonable

The construction of the original HML portfolio includes six seemingly innocuous decisions that could easily have been replaced with alternatives that are just as reasonable. I identified the following six decisions, and I propose alternatives that are just as reasonable.

The first decision is about the timing of book equity. The original HML portfolio uses the book equity of a firm's last fiscal year with fiscal year-end before the end of December of year  $t-1$  to sort stocks into value and growth portfolios from July of year  $t$  to June of year  $t+1$ . As an alternative that is just as reasonable, I use the book equity of a firm's last fiscal year six months after its fiscal year-end. Both specifications impose a minimum gap of six months for accounting information to become publicly available in order to address reporting issues documented in Alford, Jones, and Zmijewski (1992). Apple Inc., for example, has its last fiscal year-end on September 28<sup>th</sup>, 2019. The firm's press release on October 30<sup>th</sup>, 2019 gives relevant accounting information, including the firm's book equity. The original HML portfolio uses this book equity observation at the end of June of year 2020 for portfolio sorts, nine months after Apple Inc.'s fiscal year ended. The alternative decision uses book equity at the end of March 2020 for portfolio sorts, six months after the firm's fiscal year ended.

The second decision is about the timing of market equity. The original HML portfolio uses market equity at the end of December of year  $t-1$  as a simple way to match the timing of book equity. Specifically, market equity at the end of December of year  $t-1$  is used in the denominator of book-to-market equity to sort stocks into value and growth portfolios from July of year  $t$  to June of year  $t+1$ . As an alternative that is just as reasonable, I use market equity from the most recent month and skip one month. I skip one month to avoid the negative first-order serial correlation in monthly stock returns documented in Jegadeesh (1990). The alternative is similar to Asness and Frazzini (2013), who use market equity from the most recent month with one trading day skipped in the denominator of book-to-market equity to have a more updated value portfolio. They name this alternative a "seemingly small modification to the standard practice" (p. 3). Asness and Frazzini (2013) always report average portfolio returns conditional on momentum, but they do not report average portfolio returns unconditionally and they do not motivate momentum as a right-hand side variable. Also, Fama and French (1992, p. 430) include a short discussion about potential

issues regarding the timing of market equity. They mention that their results are robust to using market equity of a firm's fiscal year-end in the denominator of book-to-market equity, but they do not mention the robustness to using market equity of a firm's more recent month.

The third decision is about firms with negative book equity. The original HML portfolio excludes stocks with negative book equity observations. As an alternative, I include them. Firms can naturally and legally have negative book equity under US GAAP. Revlon Inc., for example, reported negative book equity in 2013. Hewlett-Packard, as another example, reported negative book equity in 2016.

The fourth decision is about financial firms. The original HML portfolio includes financial firms. As an alternative, I exclude financial firms (defined as firms with a one-digit SIC code of 6). This alternative is consistent with the decision to exclude financial firms in Fama and French (1992, p. 429) "... because the high leverage that is normal for these firms probably does not have the same meaning as for nonfinancial firms, where high leverage more likely indicates distress." Barber and Lyon (1997) study the book-to-market equity effect for the holdout sample of financial firms and find empirical support for the value effect. They do not reject the null hypothesis that financial firms and non-financial firms have differential value effects.

The fifth decision is about the book-to-market equity breakpoints that are used to sort stocks into value, neutral, and growth portfolios. The original HML portfolio uses the 30<sup>th</sup> and 70<sup>th</sup> percentiles of book-to-market equity of all stocks trading on the NYSE as breakpoints. As an alternative, I use the 20<sup>th</sup> and 80<sup>th</sup> percentiles and the 40<sup>th</sup> and 60<sup>th</sup> percentiles of book-to-market equity of all stocks trading on the NYSE as breakpoints.

The sixth decision is about the timing of market equity to sort stocks into small and big portfolios in order to account for the size effect. The original HML portfolio uses market equity at the end of June of year  $t$ . As an alternative, I use market equity at the end of December of year  $t-1$ . Banz (1981) is the first study on the size effect, and it uses market equity at the end of December of year  $t-1$  as a proxy for the size effect.

Many decisions in the definition of the original HML portfolio do not have an alternative that is "just as reasonable". The decision to use book equity, for example, does not have an alternative that is just as reasonable because book equity is the firm's accounting value and it matches market

equity as the firm's market value. Earnings, cashflows, or dividends are different accounting variables and thus lead to different anomaly portfolios. If the value premium estimate decreases upon replacing book equity with earnings, cashflows, or dividends, then this decrease is not about a chance result in research decisions but about different anomaly portfolios. Moreover, replacing book-to-market equity with an industry-demeaned book-to-market equity, as in Asness, Porters, and Stevens (2000), is not an alternative that is just as reasonable and instead it is economically motivated because it accounts for the heterogeneity in accounting practices across industries. Furthermore, replacing book-to-market equity with a measure that includes intangible assets, as in Park (2022) or Einfeldt and Papanikolaou (2013), is not an alternative that is just as reasonable but instead an alternative that is economically motivated because it accounts for internally created intangible assets that are not fully recognized under US GAAP. If the value premium estimate increases when such intangible assets are added to book equity, then this increase is not about a chance result in research decisions but about a better measure of value that accounts for intangible assets that are not fully recognized under US GAAP.

Excluding penny stocks is also not an alternative that is just as reasonable because this is about stock return predictability among illiquid stocks and because it leaves out information in the construction of a proxy for a systematic risk factor. Using breakpoints that are not based on stocks trading on the NYSE is also not an alternative that is just as reasonable because this is about stock return predictability among stocks that are costly to trade. See Hou, Xue, and Zhang (2018). Additionally, using quarterly book equity is not as reasonable as annual book equity because quarterly book equity is only available from 1970 while annual book equity is available from 1963. Using quarterly book equity data would leave out six years and result in a less powerful estimate than using annual book equity data.

Kessler, Scherer, and Harries (2019) study how "design choices" affect valuation-based portfolio returns. These design choices are not about alternatives that are just as reasonable. The design choices include different accounting variables (book equity, cyclically adjusted earnings, operating earnings, earnings, dividends, cash flows, earnings to growth, etc.), different sector adjustments (unadjusted, subtract the industry median, etc.), different transformations (unadjusted, z-score, percentile rank, etc.), using a different portfolio for the short leg (shorting

the growth portfolio, shorting the market portfolio, etc.), different exposures (cash neutral, beta neutral, risk neutral, etc.), and different rebalancing frequencies (monthly, quarterly, annual).

## 2.4 Snapshot of Decisions in the Literature

Table III provides a snapshot of the decisions in the empirical literature on the value effect in the 1980s and early 1990s. The heterogeneity in researchers' decisions is consistent with alternatives that are just as reasonable.

[Table III here]

The table reports all empirical studies on the value effect that are referenced in Fama and French (1992, 1993) in order to include the most relevant of a large number of empirical studies. Column (1) lists the author(s), the publication year, and the academic journal. Column (2) reports the month in which the portfolio is rebalanced. Most studies rebalance their portfolios once a year at the end of December, March, April, or June. Rosenberg, Reid, and Lanstein (1985) sort stocks into portfolios at the end of each month. Column (3) reports on the timing of the accounting information. Studies use accounting information from fiscal year-ends before the end of December or March, from the previous fiscal year, or hand-collected accounting information from fiscal year-ends from the previous month. Column (4) reports on the timing of market equity. Studies use market equity at the end of December, June, March, or the previous month. Column (5) reports if observations with negative accounting information are excluded. Some studies include negative observations, some exclude negative observations, and some sort them into a separate portfolio. Column (6) reports if financial firms are excluded. Some studies include financial firms and some exclude them. Column (7) reports on the timing of market equity to account for the size effect. Studies use market equity at the end of December, March, June, or the previous month.

### 3 Is the Original Value Premium Estimate Biased?

#### 3.1 Main Empirical Findings

Table IV reports Fama and MacBeth (1973) regressions of the monthly returns of the 96 HML portfolios on a constant and a dummy variable that is one for the original HML portfolio and zero otherwise, from July 1963 to December 1991. Equation (1) defines the regression specification. The dependent variable,  $R_{it}$ , is the return of HML portfolio  $i$  in month  $t$ .

$$R_{it} = \text{Constant} + \beta \cdot \text{Original HML Dummy}_{it} + e_{it} \quad (1)$$

The constant is the average monthly return of the average HML portfolio, and thus an estimate for the value premium. If the value premium exists, I expect to find a constant that is statistically significant. If the value premium does not exist, however, I expect to find a constant that is insignificant.

The slope on the dummy variable for the original HML portfolio is the average monthly return difference between the original HML portfolio and the average HML portfolio in each month, and thus an estimate of the bias due to chance in the original decisions. If the original value premium estimate is biased, then I expect to find a slope estimate that is statistically significant. If the original value premium estimate is not biased due to chance in the original decisions, I expect to find a slope estimate that is insignificant.

Fama and MacBeth (1973) regressions produce  $t$ -statistics that account for the contemporaneous correlation in the residuals. Petersen (2009) shows analytically and numerically that the standard errors of these estimates account for the cross-sectional correlation in the residuals.

[Table IV here]

Column (1) estimates a constant of 0.31% per month with a  $t$ -statistic of 2.14 when all HML portfolios on average are used as a proxy for the value factor. Column (2) reports the result of a

regression in which the constant is suppressed. This produces the value premium estimate using the original research decisions in Fama and French (1993). The original value premium estimate is 0.39% per month with a  $t$ -statistic of 2.87. Column (3) reports a slope estimate of 0.08% per month with a  $t$ -statistic of 1.72. These findings suggest therefore that the original estimate of the value premium is upward biased because of a potential chance result in the original research decisions.

### 3.2 Illustration of the Main Finding

Figure I shows a histogram of the monthly average returns of each of the 96 HML portfolios from July 1963 to December 1991. The original HML portfolio is marked with “HML”, and the average HML portfolio is marked with “AHML”.

The average monthly returns of the 96 HML portfolios range from 0.19% to 0.47%. I find that 82 of the 96 HML portfolios have returns that are lower or equal the original value premium estimate and that 14 HML portfolios have returns that are above it. The original value premium is thus in the 85<sup>th</sup> percentile. The average return of the average HML portfolio is 0.31% per month and is much lower than the average return of the original HML portfolio. This is consistent with the idea that the original estimate of the value premium is upward biased because of a potential chance result in the original decisions.

[Figure I here]

Also, the CAPM alpha of the original HML portfolio is in the 82<sup>nd</sup> percentile and the Sharpe ratio of the original HML portfolio is in the 92<sup>nd</sup> percentile.

### 3.3 Bootstrap

I bootstrap the distribution of the value premium under the null hypothesis that the underlying value premium is equal to the average value premium estimate to better understand whether the original estimate is within the sampling noise of the average estimate. Specifically, I draw 1,000 random samples from the monthly in-sample returns of the average HML portfolio and calculate

their average monthly returns. The bootstrap is similar to Arnott, Harvey, Kalesnik and Linnainmaa (2021).

Figure II shows the histogram of the bootstrapped distribution. I find that 90% of the bootstrapped estimates are between 0.08% and 0.56% per month. The original value premium estimate of 0.39% per month thus falls within range of sampling error. The original value premium estimate is at least as large as 673 of the 1,000 bootstrapped estimates and thus in the 67<sup>th</sup> percentile of the estimates.

[Figure II here]

### 3.4 White (2000) Reality Check Bootstrap

I additionally bootstrap  $p$ -values using the White (2000) Reality Check Bootstrap to test the null hypothesis that the value premium is zero. This is important because the  $t$ -statistic of the average value premium estimate does not account for the fact that we have more information than just the average portfolio (Table IV, column 1) and because the  $t$ -statistic on the slope estimate (Table IV, column 2) does not account for the fact that a researcher has the freedom to choose the HML portfolio with the highest  $t$ -statistic out of all 96 HML portfolio as a proxy for the value factor.

The White (2000) Reality Check Bootstrap provides a way to calculate a critical value for the hypothesis that the value premium is zero that accounts for the researcher's freedom to choose the HML portfolio with the highest  $t$ -statistic as a proxy for the value factor out of the 96 HML portfolios. The bootstrap procedure is defined as: I first demean each of the 96 HML portfolios so that no value effect exists by construction. I then bootstrap a random sample of months with replacement, calculate the average returns and the  $t$ -statistics of each of the 96 HML portfolios, and take the highest  $t$ -statistic as the  $t_{max}$ -statistic. I repeat this bootstrap 1,000 times. Eventually, I calculate the 95<sup>th</sup> percentile of all the  $t_{max}$ -statistics and use it as the critical value for the hypothesis test.

Figure III shows the histogram of the bootstrapped  $t_{max}$ -statistics under the null hypothesis that no value premium exists from July 1963 to December 2021. The 95<sup>th</sup> percentile of the  $t_{max}$ -



statistics is 2.09 and serves as the critical value of the hypothesis test using a confidence level of 95%. The original HML portfolio has a  $t$ -statistic of 2.87. This  $t$ -statistic is larger than the critical value and therefore leads to the rejection of the null hypothesis that the value premium is zero.

[Figure III]

### 3.5 Compensation for Risk

To address the concern that the higher returns of the original HML portfolio relative to the average HML portfolio may be compensation for more risk, I compare the return standard deviation and the Sharpe ratio of the original HML portfolio with the average HML portfolio (AHML) for the sample from July 1963 to December 1991. If the higher returns of the original HML portfolio are compensation for more risk, I expect to find a higher return standard deviation and a similar Sharpe ratio. Table V reports that the standard deviation of the HML portfolio is lower than that of the average HML portfolio (2.53% versus 2.72% per month) and that the Sharpe ratio of the original HML portfolio is larger than that of the average HML portfolio (0.15 versus 0.11). These findings are difficult to reconcile with a risk explanation, but they are consistent with the explanation of a result occurring due to chance in research decisions.

[Table V here]

### 3.6 One Common Factor

I study the common variation in the monthly returns of the 96 HML portfolios to better understand whether the 96 HML portfolios are proxying for one or for many risk factors.

Specifically, I perform a principal component analysis of the monthly returns of the 96 HML portfolios from July 1963 to December 1991. Table VI reports that the first principal component explains 91% of the variation in the monthly returns, the second principal component explains 6%, the third principal component explains 1%, and each of the remaining 93 principal components

explain less than 1% of the variation. These findings suggest that the 96 HML portfolios are proxying for one underlying risk factor and not for multiple risk factors.

[Table VI here]

I also find an average pairwise correlation of 0.91 for the monthly returns of the 96 HML portfolios from July 1963 to December 1991. This correlation is high compared to the average pairwise correlations among different anomaly portfolios, and it suggests that the 96 HML portfolios are all proxies for one common factor. McLean and Pontiff (2016) report an average correlation of 0.03 across 97 different anomaly portfolios, and Green, Hand, and Zhang (2017) report an average correlation of 0.09 among 60 different predictor portfolios.

### **3.7 Out-Of-Sample Evidence**

I estimate the main regression specification in (1) using data that is outside of the original study's sample period. If the value premium exists, I expect to find that the average estimate of the value premium is statistically significant out of sample. If the value premium does not exist, however, I expect to find that the average estimate of the value premium is statistically insignificant out of sample. If the original value premium estimate is biased due to a potential chance result in the original decisions, I expect to find an insignificant difference between the original and the average estimate of the value premium out of sample. Otherwise, I expect to find a statistically significant difference between the original and the average value premium estimate out of sample.

Table VII reports Fama and MacBeth (1973) regressions of the monthly returns of the 96 HML portfolios on a constant and a dummy variable that is one for the original HML portfolio and zero otherwise. The pre-sample is July 1926 to June 1963, and the post-sample is January 1992 to December 2021. The pre-Compustat book equity data is from Kenneth French's webpage, and I hand-collected fiscal year-end months from the Moody's Manuals that are available to me.

[Table VII here]

Columns (1) and (4) estimates a value premium of 0.33% per month with a *t*-statistic of 1.45 in the pre-sample and a value premium of 0.09% per month with a *t*-statistic of 0.48 in the post-sample.

Columns (2) and (5) report the result for a specification in which the constant is suppressed. This produces the value premium estimate using the original research decisions in Fama and French (1993). The estimate of the value premium is 0.40% per month with a *t*-statistic of 1.95 in the pre-sample and it is 0.14% per month with a *t*-statistic of 0.81 in the post-sample when the original HML portfolio is used as a proxy for the underlying value factor. The pre-sample finding is consistent with Davis, Fama, and French (2000). The post-sample result is consistent with Fama and French (2020). They report an insignificant value premium of 0.10% per month for the post-sample period using the difference between a market value and a market growth portfolio as a proxy for the value factor (Table 1 on p. 15,  $MV-MG = 0.11-0.01 = 0.10$ ). The post-sample findings are also consistent with Linnainmaa and Roberts (2018), who report that the original HML portfolio has an average monthly return of 0.31% per month with a *t*-statistic of 1.71 in the post-sample, and they are consistent with Schwert (2003), who reports an insignificant value effect for the sample from 1994 to 2002.

Columns (3) and (6) estimate a bias of 0.07% per month with a *t*-statistic of 1.34 in the pre-sample and a bias of 0.05% per month with a *t*-statistic of 0.72 in the post-sample. The pre-sample slope estimate is similar in magnitude but statistically less significant compared to the in-sample slope estimate, which is somewhat at odds with an estimation bias in the original value premium estimate. The post-sample estimate is smaller and statistically insignificant, and thus suggests that the return difference between the original and the average HML portfolios is an estimate for the bias due to chance in the original decisions.

I also study the robustness of my findings to using different start dates. Davis, Fama, and French (2000) use July 1929 as a start date because they need the first three years to estimate betas. My results are robust to using July 1929 as start date. Cohen, Polk, and Vuolteenaho (2003) report that the Securities Exchange Act of 1934 was enacted to ensure accurate accounting information. They characterize the first two years after that act as an enforcement period and

determine that accounting information from 1937 is of sufficiently high quality for empirical research. My results are robust to using July 1938 as the start date.

I also acknowledge that the positive return difference between the original and the average HML portfolio in the pre- and in the original sample period, and the lower return difference in the post-sample period are consistent with McLean and Pontiff's (2016) assertion that sophisticated investors learn about stock return predictability due to mispricing and start to trade on it. Fama and French (2020) argue that the post-sample decline of value minus growth portfolios is economically large but not statistically significant.

### **3.8 Individual Decisions**

I study how much each of the six decisions accounts for the return difference between the original and the average HML portfolio. Table VIII reports Fama and MacBeth (1973) regressions of the monthly returns of the 96 HML portfolios on a constant and six dummy variables for each of the six decisions, from July 1963 to December 1991.

[Table VIII here]

Columns (1) to (5) report estimates of the six dummy variables individually, and column (6) reports estimates for the six dummy variables jointly. Four out of six slope estimates are close to zero, which is consistent with researchers' expectation that the alternative decisions lead to similar empirical finding as the original decisions. The slope on the second dummy variable about the timing of market equity in the denominator of book-to-market equity is 0.11% per month with a *t*-statistic of 1.50. This effect is economically large but is not statistically significant at the 95 percent confidence level. The slope on the sixth dummy variable about the timing of market equity to account for the size effect is 0.03% per month with a *t*-statistic of 2.29. This effect is statistically significant at the 5% significance level. These finding are related to Gerakos and Linnainmaa (2018), who report that most of the value premium estimate is driven by changes in market equity. The empirical findings also suggest that decisions with more "degrees of freedom", for example

decisions that use monthly stock market data compared to annual accounting data, are more likely to pick up a result due to chance in research decisions.

### 3.9 Robustness to Cross-Sectional Slope Estimates

The empirical tests use monthly portfolio returns as the dependent variable. As a robustness test, I use the monthly slope estimates from cross-sectional regressions of one-month ahead returns on each of the individual value measures as the dependent variable. Fama (1976) shows that these slope estimates are returns from long-short portfolios with an exposure of one to the underlying value factor.

Specifically, I estimate monthly cross-sectional regressions of one-month ahead returns on the natural logarithm of each of 32 value measures and on the natural logarithm of market equity using all but microcap stocks from July 1963 to December 1991. Microcap stocks are defined as stocks with market equity below the 20<sup>th</sup> percentile of market equity using all stocks trading on the NYSE. I have 32 value measures compared to the 96 HML portfolios, because the fifth decision on breakpoints to sort stocks into value, neutral, and growth portfolios is not used in cross-sectional regressions. I exclude microcap stocks as a simple way to account for the unique weighting that is used in the construction of the original HML portfolio. The original HML portfolio is constructed by equal weighting two long and two short portfolios (Small Value and Big Value, Small Growth and Big Growth), and these four portfolios are constructed by value weighting stocks.

Table IX reports Fama and MacBeth (1973) regressions of the monthly slope estimates on a constant and a dummy variable that is one for the slope estimates that are consistent with the original HML portfolio returns. The *t*-statistics are reported in parentheses.

[Table IX here]

Column (1) estimates a value premium of 0.19% per month with a *t*-statistic of 2.06. Column (2) estimates a value premium of 0.27% per month with a *t*-statistic of 2.95 for the value measure that is consistent with the original HML portfolio. Column (3) estimates a bias of 0.08%

per month with a  $t$ -statistic of 3.76. My main results are therefore robust to using monthly slope estimates from cross-sectional regressions instead of portfolios returns.

### 3.10 Full Sample Evidence

I estimate the main regression specification using data from the full sample from July 1926 to December 2021. Table X reports the empirical results. Column (1) estimates a value premium of 0.25% per month with a  $t$ -statistic of 2.19. Column (2) suppresses the constant. This produces the value premium estimate using the original research decisions in Fama and French (1993). The estimate of the value premium is 0.32% per month with a  $t$ -statistic of 3.02 when the original HML portfolio is used as a proxy for the value factor. Column (3) estimates a bias of 0.07% per month with a  $t$ -statistic of 2.01. These findings suggest that the value premium exists and that the original estimate of the value premium is upward biased due to chance in the original decisions.

[Table X]

## 4 Spanning Regressions

I use the approach in Barillas and Shanken (2017) and the Fama and French (2015) five-factor model augmented with momentum (Jegadeesh and Titman, 1993; Carhart, 1997) to assess whether the value premium estimate conditional on the factors of this model is biased due to a potential chance result in seemingly innocuous research decisions. If the value premium estimate conditional on other factors is unbiased, then I expect to find that the original and the average HML portfolios produce similar estimates. If the value premium estimate conditional on other factors is biased, however, then I expect to find that the original and the average HML portfolios produce estimates that are different from each other.

Table XI reports the value premium estimate conditional on the Fama and French (2015) five-factor model augmented with momentum (Jegadeesh and Titman, 1993; Carhart, 1997). Columns (1) and (2) report the value premium estimate conditional on the five factors when the original

HML portfolio and when the average HML portfolio is used as proxy for the underlying value factor using data from the original study's sample period from July 1963 to December 1991. The original HML portfolio estimates a value premium of 0.28% per month with a *t*-statistic of 2.82, and the average HML portfolio estimates a value premium of 0.36% per month with a *t*-statistic of 3.68. Column (3) reports that the difference between the estimates is  $-0.08\%$  per month and with a *t*-statistic of  $-2.45$ . The statistically significant difference between the value premium estimates conditional on the five factors suggests that the value premium estimate using the original HML portfolio is biased due to chance in the original study's research decisions, and the negative sign suggests that the value premium estimate using the original HML portfolio conditional on the augmented five-factor model is downward biased. This is consistent with Asness and Frazzini (2013) that "... seemingly small modifications to standard practice..." in the construction of the original HML portfolio lead to a larger value premium estimates conditional on the augmented five-factor model.

[Table XI here]

If the difference between the value premium estimates using the original and the average HML portfolio and conditional on the five factors is a bias due to chance in seemingly innocuous decisions, then I expect to find no difference between the estimates out of sample. If the difference is not a bias, however, then I expect to find that the value premium estimates conditional on the five factors are different from each other out of sample. Columns (4) to (6) in Table XI report the spanning regressions using data prior to the original study's sample period (July 1926 to June 1963). The original HML portfolio estimates a value premium of 0.39% per month with a *t*-statistic of 2.07, and the average HML portfolio estimates a conditional value premium estimate of 0.40% per month with a *t*-statistic of 2.24. The difference is  $-0.01\%$  per month with a *t*-statistic of  $-0.26$ . The statistically insignificant difference in the pre sample is consistent with a bias in seemingly innocuous research decisions.

Columns (7) to (9) in Table XI report the spanning regressions using data post the original study's sample period (January 1993 to December 2021). The original HML portfolio estimates a

value premium of  $-0.22\%$  per month with a  $t$ -statistic of  $-1.52$ , and the average HML portfolio estimates a conditional value premium of  $-0.14\%$  per month with a  $t$ -statistic of  $-1.13$ . The difference is  $-0.08\%$  per month with a  $t$ -statistic of  $-1.39$ . The statistically insignificant difference in the post sample is also consistent with a bias in sample that is the result of chance in seemingly innocuous research decisions.

Table XI additionally reports on the betas of the original and the average HML portfolios with the factors of the augmented five-factor model. Column (3) reports that the return difference between the original and the average HML portfolios has a beta with the UMD portfolio of  $0.18$  with a  $t$ -statistic of  $16.14$  in the original study's sample period. Columns (6) and (9) report that this return difference has a beta with the UMD portfolio of  $0.13$  ( $t$ -statistic of  $5.98$ ) in the pre sample period and a beta with the UMD portfolio of  $0.23$  ( $t$ -statistic of  $11.42$ ) in the post sample period. The finding that the original HML portfolio has a less negative beta with the UMD portfolio suggests that the original value premium estimate has a higher return than the average HML portfolio unconditionally and in sample because the original HML portfolio avoids training against momentum. The inference of whether or not the original value premium is biased unconditionally therefore hinges on the unresolved debate of whether or not momentum is an asset pricing factor: On one hand, Fama and French (2018, p. 237) are reluctant to accept momentum as an asset pricing factor, arguing that momentum lacks economic motivation as an asset pricing factor and that recognizing momentum as an asset pricing factor marks the beginning of mining the data for factors. Griffin, Ji, and Martin (2003) argue that international momentum portfolios are not related to macroeconomic risk factors. On the other hand, Barberis, Jin, and Wang (2019) argue that momentum can be motivated as an asset pricing factor in a model in which investors evaluate risk according to prospect theory. Jegadeesh and Titman (2010) document that momentum returns are also statistically significant after the publication of their 1993 paper.



## 5 Conclusion

Fama and French (1993) propose the HML portfolio as a proxy for the value factor and estimate a value premium of 0.40% per month using data from July 1963 to December 1991. The construction of this HML portfolio includes six seemingly innocuous research decisions that could easily have been replaced with alternatives that are just as reasonable. I propose such alternatives, construct HML portfolios using all possible combinations of the six original and the alternative decisions, and study their value premium estimates.

If we assume that the original decisions are similar to the alternative decisions, then the original and the alternative decisions should produce similar estimates of the value premium. Thus, the average estimate over all estimates produced from all HML portfolios constructed with all possible combinations of the original and the alternative decisions serves as an estimate of the underlying value premium. The difference between the original and this average estimate serves as an estimate for the bias in the original value premium estimate that is potentially due to chance in the original research decisions.

In sample, the original value premium estimate is 0.39% per month and the average value premium estimate is 0.31% per month. The difference is 0.08% per month and statistically significant. Out of sample, this difference is statistically indistinguishable from zero however. These findings suggest that the original value premium estimate in Fama and French (1993) is upward biased because of a potential chance result in the original research decisions.

These findings contribute another piece of puzzle about the validity of the original value premium estimate to Akey, Robertson and Simutin's (2022) puzzling finding about the value premium estimate's sensitivity to different data vintages.

## References

- Alford, A., J. J. Jones, and M. E. Zmijewski, 1994, Extensions and violations of the statutory SEC form 10-K filing requirements, *Journal of Accounting and Economics*, 17, 229-254.
- Akey, P., A. Z. Robertson, and M. Simutin, 2022, Noisy factors, Working paper, University of Pennsylvania.
- Ang, A., and J. Chen, 2007, CAPM over the long run: 1926–2001. *Journal of Empirical Finance*, 14, 1-40.
- Arnott, R. D., C. R. Harvey, V. Kalesnik, and J. T. Linnainmaa, 2021, Reports of value's death may be greatly exaggerated. *Financial Analysts Journal*, 77, 44-67
- Asness, C. S., and A. Frazzini, 2013, The Devil in HML's Details. *The Journal of Portfolio Management*, 39, 49-68.
- Asness, C. S., T. Moskowitz, and L. Pedersen, 2013, Value and Momentum Everywhere. *Journal of Finance*, 68, pp. 929-985.
- Asness, C. S., R. B. Porters, and R. L. Stevens, 2000, Predicting Stock Returns Using Industry-Relative Firm Characteristics, Working Paper.
- Banz, R. W., 1981, The relationship between return and market value of common stocks. *Journal of Financial Economics*, 9, 3-18.
- Banz, R. W., and W. J. Breen, 1986, Sample-Dependent Results Using Accounting and Market Data: Some Evidence, *The Journal of Finance*, 41, 779-793.
- Barber, B., and J. Lyon, 1997, Firm Size, Book-to-Market Ratio, and Security Returns: A Holdout Sample of Financial Firms, *The Journal of Finance*, 52, 875-883.
- Barberis, N., L. Jin, and B. Wang, 2019, Prospect Theory and Stock Market Anomalies, Working paper.
- Barillas, F., and J. Shanken, 2017, Which Alpha?, *Review of Financial Studies*, 30, 1116-1338.
- Basu, S., 1977, Investment performance of common stocks in relation to their price-earnings ratios: a test of the efficient market hypothesis, *Journal of Finance*, 32: 663-682.
- Basu, S., 1983, The relationship between earnings' yield, market value and return for NYSE common stocks: Further evidence. *Journal of Financial Economics*, 12, 129-156.

Benjamini, Y., and D. Yekutieli, 2001, The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics*, 29, 1165–88.

Bhandari, L. C., 1988, Debt/Equity Ratio and Expected Common Stock Returns: Empirical Evidence, *The Journal of Finance*, 43, 507-528.

Black, F., 1993, Beta and Return, *Journal of Portfolio Management*, 20, 8-18.

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

Chan, K. C., and N.-F. Chen, 1991, Structural and Return Characteristics of Small and Large Firms, *The Journal of Finance*, 46, 1467-1484.

Chan, K. C., N.-F. Chen, and D. A. Hsieh, 1985, An exploratory investigation of the firm size effect, *Journal of Financial Economics*, 14, 451-471.

Chan, K. C., Y. Hamao, and J. Lakonishok, 1991, Fundamentals and Stock Returns in Japan. *The Journal of Finance*, 46, 1739-1764.

Chordia, T., A. Goyal, and A. Saretto, 2020, Anomalies and False Rejections, *The Review of Financial Studies*, 33, 2134–2179.

Cohen, R. B., C. Polk, and T. Vuolteenaho, 2003, The Value Spread. *The Journal of Finance*, 58, 609-641.

Davis, J., E. F. Fama, and K. R. French, 2000, Characteristics, covariances, and average returns, *Journal of Finance* 46, 389-406.

Eisfeldt, Andrea L., and Dimitris Papanikolaou, 2013, Organization capital and the cross-section of expected returns, *The Journal of Finance*, 68, 1365–1406.

Fama, E. F., 1976, *Foundations of Finance*, Basic Books, New York.

Fama, E. F., 1991, Efficient capital markets: II, *Journal of Finance* 46, 1575–1617.

Fama, E. F., and K. R. French, 1992, The cross-section of expected stock returns, *Journal of Finance*, 47, 427-465.

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

Fama, E. F., and K. R. French, 1998, Value versus growth: The international evidence. *Journal of Finance*, 53, 1975–1999.

Fama, E. F., and K. R. French, 2010, Luck versus Skill in the Cross-Section of Mutual Fund Returns. *The Journal of Finance*, 5, 1915-1947.

Fama, E. F., and K. R. French, 2008, Dissecting Anomalies. *The Journal of Finance*, 63, 1653-1678.

Fama, E. F., and K. R. French, 2015, A five-factor asset pricing model. *Journal of Financial Economic*, 116, 1–22.

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

Fama, E. F., and K. R. French, 2020, The Value Premium. Working Paper.

Fama, E. F., and J. D. MacBeth, 1973, Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy*, 81, 607–636.

Gelman, A., and E. Loken, 2013, The garden of forking paths: Why multiple comparisons can be a problem, even when there is no “fishing expedition” or “p-hacking” and the research hypothesis was posited ahead of time. Working paper.

Gelman, A., and E. Loken, 2014, The statistical crisis in science, *American Scientist* 102, 460–465.

Gerakos, J., and J. Linnainmaa, 2018. Decomposing value. *Review of Financial Studies*, 31, 1825–1854.

Gibbons, M. R., S. Ross, and J. Shanken, 1989, A Test of the Efficiency of a Given Portfolio, *Econometrica*, 57, 1121-52.

Green, J., J. R. Hand, and X. F. Zhang, 2017, The Characteristics that provide independent information about average U.S. monthly stock returns, *Review of Financial Studies*, 30, 4389-4436.

Griffin, J.M., X. Ji, and J. S. Martin, 2003, Momentum Investing and Business Cycle Risk: Evidence from Pole to Pole, *The Journal of Finance*, 58, 2515-2547.

Harvey, C. R., 2017, Presidential Address: The scientific outlook in financial economics. *The Journal of Finance*, 72, 1399-1440.

Harvey, C. R., and Y. Liu, 2020, False (and Missed) Discoveries in Financial Economics. *The Journal of Finance*, 75, 2503-2553.

Harvey, C. R., Y. Liu, and H. Zhu, 2016. ... and the cross-section of expected returns. *The Review of Financial Studies*, 29, 5-68.

Holm, S., 1979, A simple sequentially rejective multiple test procedure, *Scandinavian Journal of Statistics*, 6, 65-70.

Hou, K., C. Xue, and L. Zhang, 2018, Replicating Anomalies, *The Review of Financial Studies*, 33, 2019-2133.

Jaffe, J., D. B. Keim, and R. Westerfield, 1989, Earnings Yields, Market Values, and Stock Returns, *The Journal of Finance*, 44, 135-148.

Jegadeesh, N., 1990, Evidence of Predictable Behavior of Security Returns, *Journal of Finance*, 45, 881-898.

Jegadeesh, N., and S. Titman, 1993, Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency, *Journal of Finance*, 48, 65-91.

Jegadeesh, N., and S. Titman, 2001, Profitability of Momentum Strategies: An Evaluation of Alternative Explanations. *The Journal of Finance*, 56, 699-720.

Keim, D. B., 1983, Size-related anomalies and stock return seasonality: Further empirical evidence, *Journal of Financial Economics*, 12, 13-32.

Kessler, S., B. Scherer, and J. P. Harries, 2020, Value by Design. *The Journal of Portfolio Management*, 46.

Linnainmaa, J. T., and M. R. Roberts, 2018, The history of the cross-section of stock returns, *Review of Financial Studies* 31, 2606-2649.

Lo, A. W., and A. C. MacKinlay, 1990, Data-snooping biases in tests of financial asset pricing models. *Review of Financial Studies*, 3, 431-67.

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

Novy-Marx, R., 2013, The other side of value: The gross profitability premium. *Journal of Financial Economics*, 108, 1-28.

Park, H., An Intangible-Adjusted Book-to-Market Ratio Still Predicts Stock Returns. *Critical Finance Review*, 11, 265-297.

Pastor, L. and R. Stambaugh, 2003, Liquidity Risk. *Journal of Political Economy*, 111, 642-685.

Pedersen, M. A., 2009, Estimating Standard Errors in Finance Panel Data Sets: Comparing Approaches. *The Review of Financial Studies*, 22, 435-480.

- Phalippou, L., 2008, Where is the value premium? *Financial Analysts Journal*, 64, 41-48.
- Pontiff, J., 1996, Costly arbitrage: Evidence from closed end funds, *Quarterly Journal of Economics*, 111, 1135–1151.
- Pontiff, J., and R. Singla, 2019, Liquidity Risk? *Critical Finance Review*, 8, 257-276.
- Rosenberg, R., K. Reid, and R. Lanstein, 1985, Persuasive Evidence of Market Inefficiency, *Journal of Portfolio Management*, 11, 9-16.
- Rosenthal, R., 1979, The “file drawer problem” and tolerance for null results. *Psychological Bulletin* 86, 638–641.
- Schwert, G. W., 2003, Anomalies and market efficiency, In: Constantinides, G., M. Harris, and R. Stulz, *Handbook of the Economics of Finance*, North Holland, Amsterdam, 937-972.
- Zhang, L., 2005, The Value Premium. *The Journal of Finance*, 60, 67-103.

**Table I**  
**Replication Validation, 1963 to 1991**

This table compares the summary statistics of the reported and the replicated HML portfolio. Column (1) reports summary statistics of the HML portfolio as reported in Fama and French (1993, p. 13, Table 2) and column (2) reports the summary statistics of the replicated HML portfolio. The sample is from July 1963 to December 1991. I use the book equity definition as in Davis, Fama, and French (2000). Book equity is defined as stockholder book equity plus balance sheet deferred taxes and investment tax credit (txditc), if available, minus the book value of preferred stock. Stockholder equity is the value reported by Compustat (seq). If not available, stockholders' equity is measured as the book value of common equity (ceq) plus the par value of preferred stock (pstk), or the book value of assets minus total liabilities (at-lt), in that order. The book value of preferred stock is measured as redemption (pstkrv), liquidation (pstk), or par value (pstk), if available and in that order. **Interpretation:** The replicated HML portfolio looks similar to the original HML portfolio in Fama and French (1993).

	(1)	(2)
	Original HML portfolio as Reported in Table 2 in Fama and French (1993, p. 13)	Replication of the original HML portfolio
Mean	0.40	0.39
St. Dev.	2.54	2.53
<i>t</i> -statistic	2.91	2.87
Autocorr. Lag 1	0.18	0.17
Autocorr. Lag 2	0.06	0.06
Autocorr. Lag 12	0.07	0.08
Corr. with SMB	-0.08	-0.11
Corr. with MKTRF	-0.38	-0.37
Observations	342	342

**Table II**  
**Replication Validation**

This table reports the average monthly return (in percentages) of the original HML portfolio reported in Fama and French (1993, p. 13) and compares it with the average monthly return of the replicated HML portfolio, as well as with the average monthly returns of HML portfolios reported in eight additional studies. Column (1) names the authors, the publication year, and the academic journal of the study. Column (2) reports the time horizon that is used in each study. Column (3) reports the average monthly return of the HML portfolio over the time horizon of the study as reported in the study. Column (4) reports the average monthly return of the replicated HML portfolio over the same time horizon as in each study. **Interpretation:** The replicated HML portfolio looks similar to the original HML portfolio.

Study	Time Period of Each Study	HML Return in Each Study	Replicated HML Return	Diff.
Fama and French (1993, JFE)	07/63-12/91	0.40	0.39	0.01
Fama and French (1995, JF)	07/63-12/92	0.44	0.44	0.00
Fama and French (1996, JF)	07/63-12/93	0.46	0.46	0.00
Fama and French (1997, JFE)	07/63-12/94	0.45	0.45	0.00
Davis, Fama, and French (2000, JF)	07/63-06/97	0.43	0.43	0.00
Fama and French (2015, JFE)	07/63-12/13	0.37	0.37	0.00
Linnainmaa and Roberts (2018, RFS)	07/63-12/16	0.36	0.36	0.00
Gerakos and Linnainmaa (2018, RFS)	07/63-12/16	0.37	0.36	0.01
Ball, Gerakos, Linnainmaa, and Nikolaev (2019, JFE)	07/64-12/17	0.31	0.32	-0.01
Average		0.40	0.40	0.00



**Table III**  
**Snapshot of Decisions in the Related Studies, 1981 to 1993**

This table provides a snapshot of the decisions in the empirical literature. I report all empirical studies on valuation anomalies that are referenced in Fama and French (1992, 1993) to include the most relevant empirical studies as of the beginning of the early 1990s. Column (1) names the author(s), the publication year, and the academic journal. Column (2) reports the month in which stocks are sorted into portfolios. Column (3) reports the month from which the accounting variable is that is used for the portfolio sorts. Column (4) reports the month from which month market equity is that is used for the portfolio sorts. Column (5) reports if observations with negative accounting information are excluded. Column (6) reports if financial firms are excluded. Column (7) reports the month from which market equity is to control for the size effect. **Interpretation:** The variation in research decisions across value papers suggests that the original research decisions in Fama and French (1993) have alternative decisions that are just as reasonable

(1)	(2)	(3)	(4)	(5)	(6)	(7)
Study	Rebalancing Month	Timing of Accounting	Timing of ME	Negative Accounting	Financial Firms	Timing of Size Effect
Fama and French (1993, JFE)	Jun t	Dec t-1	Dec t-1	Exclude	Include	Jun t
Fama and French (1992, JF)	Jun t	Dec t-1	Dec t-1	Exclude	Exclude	Jun t
Chan, Hamao, and Lakonishok (1991, JF)	Jun t	Mar t	Jun t	Separate	Include	Jun t
Jaffe, Keim, and Westerfield (1989, JF)	Mar t	Dec t-1	Mar t	Separate	Include	Mar t
Bhandari (1988, JF)	Dec t	Prev. FYE	Prev. FYE	Include	Include	Dec t
Banz and Breen (1986, JF)	Dec t	Prev. FYE	Prev. FYE	Separate	Include	Dec t
Rosenberg, Reid, and Lanstein (1985, JPM)	Monthly	Prev. Month	Prev. Month	Include	Include	Prev. Month
Basu (1983, JFE)	Apr t	Dec t-1	Dec t-1	Exclude	Exclude	Dec t-1

**Table IV**  
**Seemingly Innocuous Decisions in the Construction of the Original HML Portfolio, 1963 to 1991**

This table reports Fama and MacBeth (1973) regressions of the monthly returns (in percentage) of 96 HML portfolios on a constant and on a dummy variable that is one for the original HML portfolio, from July 1963 to December 1991.

$$R_{it} = \text{Constant} + \beta \cdot \text{Original HML Dummy}_{it} + e_{it}$$

The construction of the original HML portfolio (Fama and French, 1993) includes seemingly innocuous decisions that could easily have been replaced with alternatives that are just as reasonable. I propose such alternatives, form all possible combinations of these decisions and their alternatives, and construct 96 HML portfolios. Note that column (2) suppresses the constant in the regression. This produces the value premium estimate using the original research decisions in Fama and French (1993). *t*-statistics are reported in parentheses. **Interpretation:** The value premium estimate is 0.08 percent per month larger in sample when the HML portfolio is constructed with the original research decisions in Fama and French (1993) compared to when the HML portfolio is constructed with reasonable alternative decisions on average.

	(1)	(2)	(3)
Constant	0.31 (2.14)		0.31 (2.13)
Original HML Dummy		0.39 (2.87)	0.08 (1.72)
Months	342	342	342

**Table V**  
**Compensation for Risk Explanation, 1963 to 1991**

This table reports the average monthly return, standard deviation, and the Sharpe ratio of the original HML portfolio and of the average HML portfolio in the original study's sample from July 1963 to December 1991. The construction of the original HML portfolio includes seemingly innocuous decisions that could easily have been replaced with alternatives that are just as reasonable. I propose such alternatives, form all possible combinations of these decisions and their alternatives, construct 96 HML portfolios, and finally collapse these 96 HML portfolios in each month into an equally weighted average HML portfolio. **Interpretation:** The higher return of the original HML portfolio in sample is not the result of more risk.

Portfolio	Average	CAPM Alpha	Standard Deviation	Sharpe Ratio
Original HML Portfolio	0.39	0.48	2.53	0.15
Average HML Portfolio	0.30	0.41	2.72	0.11

**Table VI**  
**Principal Component Analysis of 96 HML Portfolios, 1963 to 1991**

This table reports estimation results of a principal component analysis of the monthly returns of 96 HML portfolio in the original study's sample from July 1963 to December 1991. The construction of the original HML portfolio includes seemingly innocuous decisions that could easily have been replaced with alternatives that are just as reasonable. I propose such alternatives, form all possible combinations of these decisions and their alternatives, construct 96 HML portfolios. Note that I only report the first five principal components, because the other principal components account for less than half a percent of the variation. **Interpretation:** The HML portfolios that are constructed using all possible combinations of research decisions are proxying for one underlying risk factor.

Principal Component	Proportion of Variance Explained
First	0.91
Second	0.06
Third	0.01
Fourth	0.00
Fifth	0.00

**Table VII**

**Pre- and Post-Sample Evidence, 1926 to 1963 and 1992 to 2021**

This table reports Fama and MacBeth (1973) regressions of the monthly returns (in percentage) of 96 HML portfolios on a constant and a dummy variable that is one for the original HML portfolio using out of sample data. The pre-sample is July 1926 to June 1963, and the post-sample is January 1992 to December 2021.

$$R_{it} = \text{Constant} + \beta \cdot \text{Original HML Dummy}_{it} + e_{it}$$

The construction of the original HML portfolio includes seemingly innocuous decisions that could easily have been replaced with alternatives that are just as reasonable. I propose such alternatives, form all possible combinations of the decisions and their alternatives, and construct 96 HML portfolios. Note that columns (2, 5) suppress the constant in the regression. This produces the value premium estimate using the original research decisions in Fama and French (1993). *t*-statistics are reported in parentheses. **Interpretation:** The incremental return of the original HML portfolio is statistically insignificant out of sample consistent with a statistical bias.

	(1)	(2)	(3)	(4)	(5)	(6)
<u>Sample</u>	<u>July 1926 to June 1963</u>			<u>January 1992 to December 2021</u>		
Constant	0.33 (1.45)		0.33 (1.45)	0.09 (0.48)		0.09 (0.48)
Original HML Dummy		0.40 (1.95)	0.07 (1.34)		0.14 (0.81)	0.05 (0.72)
Months	444	444	444	360	360	360

**Table VIII**  
**Individual Decisions, 1963 to 1991**

This table reports Fama and MacBeth (1973) regressions of the monthly returns (in percent) of 96 HML portfolios on a constant and six dummy variables for each of the six decisions of the original HML portfolio from July 1963 to December 1991. The construction of the original HML portfolio includes seemingly innocuous decisions that could easily have been replaced with alternatives that are just as reasonable. I propose such alternatives, form all possible combinations of the decisions and their alternatives, and construct 96 HML portfolios. *t*-statistics are reported in parentheses. **Interpretation:** Two decisions, both about the timing of market equity, account for the bias in the original value premium estimate.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Constant	0.31 (2.16)	0.26 (1.58)	0.31 (2.15)	0.32 (2.11)	0.31 (2.12)	0.30 (2.01)	0.24 (1.46)
Dummy for Timing of BE	-0.00 (-0.10)						-0.00 (-0.10)
Dummy for Timing of ME		0.11 (1.50)					0.11 (1.50)
Dummy for Neg. BE			-0.00 (-0.42)				-0.00 (-0.42)
Dummy for Financials				-0.00 (-0.15)			-0.00 (-0.15)
Dummy for Breakpoints					0.00 (0.18)		0.00 (0.18)
Dummy for Timing of ME						0.03 (2.29)	0.03 (2.29)
Months	342	342	342	342	342	342	343

**Table IX**  
**Robustness to Cross-sectional Slope Estimates, 1963 to 1991**

This table reports Fama and MacBeth (1973) regressions of monthly slope estimates on a constant and a dummy variable that is one for the value measure that is consistent with the original HML portfolio and zero otherwise.

$$\hat{\gamma}_{it} = \text{Constant} + \beta \cdot \text{Original HML Dummy}_{it} + e_{it}$$

The dependent variable,  $\hat{\gamma}_{it}$ , is the monthly slope estimates of monthly cross-sectional regressions in which one-month ahead returns are regressed on a constant, on the natural logarithm of each of 32 value measures, and on the natural logarithm of market equity, excluding microcap stocks. Microcap stocks are defined as stocks with end of June market capitalization below the 20 percent breakpoint of all stocks trading on the NYSE. Fama (1976) shows that these slope coefficients are returns of long-short portfolios with an exposure of one to the value measure. I exclude microcaps as a simple way to account for the equal and value weighting of stocks that is used in the construction of the original HML portfolio. I have 32 value measures (instead of 96 HML portfolios) because the fifth decision about the breakpoints to sort stocks into value and growth portfolios is not applicable. Column (2) suppresses the constant in the regression. This produces the value premium estimate using the original research decisions in Fama and French (1993). *t*-statistics are reported in parentheses. **Interpretation:** The results are robust to using cross-sectional regression slopes

	(1)	(2)	(3)
Constant	0.19 (2.06)		0.18 (2.03)
Original HML Dummy		0.27 (2.95)	0.08 (3.76)
Months	342	342	342

**Table X**  
**Full Sample Evidence, 1926 to 2021**

This table reports Fama and MacBeth (1973) regressions of the monthly returns (in percentage) of the HML portfolios on a constant and on a dummy variable that is one for the original HML portfolio, from July 1926 to December 2021.

$$R_{it} = \text{Constant} + \beta \cdot \text{Original HML Dummy}_{it} + e_{it}$$

The construction of the original HML portfolio (Fama and French, 1993) includes seemingly innocuous decisions that could easily have been replaced with alternatives that are just as reasonable. I propose such alternatives, form all possible combinations of the decisions and their alternatives, and construct 96 HML portfolios. Column (2) suppresses the constant in the regression. This produces the value premium estimate using the original research decisions in Fama and French (1993). *t*-statistics are reported in parentheses. **Interpretation:** In the full sample, the value premium estimate is biased.

	(1)	(2)	(3)
Constant	0.25 (2.19)		0.25 (2.18)
Original HML Dummy		0.32 (3.02)	0.07 (2.01)
Months	1,146	1,146	1,146

**Table XI**  
**Spanning Regressions, 1926 to 2021**

This table reports spanning regressions of the original and the average HML portfolios using the approach in Barillas and Shanken (2017) and the factors in Fama and French (2015) augmented with momentum over different time periods from January 1926 to December 2021. The construction of the original HML portfolio includes seemingly innocuous decisions that could easily have been replaced with alternatives that are just as reasonable. I propose such alternatives, form all possible combinations of these decisions and their alternatives, construct 96 HML portfolios, and finally collapse these 96 HML portfolios in each month into an average HML portfolio (“AHML”). The factor returns are from Kenneth French’s webpage. *t*-statistics are reported in parentheses. **Interpretation:** The original value premium estimate is downward biased conditional on the Fama and French (2015) factors including momentum.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
<u>Sample</u>		<u>1963 - 1991</u>			<u>1926 - 1962</u>			<u>1992 - 2021</u>	
<u>Portfolio</u>	<u>HML</u>	<u>AHML</u>	<u>Diff</u>	<u>HML</u>	<u>AHML</u>	<u>Diff</u>	<u>HML</u>	<u>AHML</u>	<u>Diff</u>
Constant	0.28 (2.82)	0.36 (3.68)	-0.08 (-2.45)	0.39 (2.07)	0.40 (2.24)	-0.01 (-0.26)	-0.22 (-1.52)	-0.14 (-1.13)	-0.08 (-1.39)
Mkt-RF	-0.06 (-1.83)	-0.09 (-2.91)	0.03 (3.62)	0.20 (5.12)	0.20 (4.91)	0.01 (0.50)	0.09 (1.82)	0.04 (1.05)	0.05 (2.57)
SMB	0.02 (0.53)	0.06 (1.30)	-0.03 (-2.60)	0.22 (3.08)	0.26 (3.87)	-0.04 (-1.76)	-0.02 (-0.40)	0.01 (0.19)	-0.03 (-1.31)
CMA	0.80 (11.66)	0.78 (11.88)	0.02 (0.86)				0.98 (13.64)	0.95 (16.33)	0.03 (0.95)
RMW	-0.32 (-4.33)	-0.32 (-4.50)	-0.00 (-0.19)				0.40 (4.89)	0.40 (5.65)	0.00 (0.07)
UMD	-0.06 (-1.48)	-0.24 (-5.83)	0.18 (16.14)	-0.31 (-4.11)	-0.44 (-7.02)	0.13 (5.98)	-0.13 (-4.49)	-0.36 (-15.33)	0.23 (11.42)
Months	342	342	342	438	438	438	359	359	359



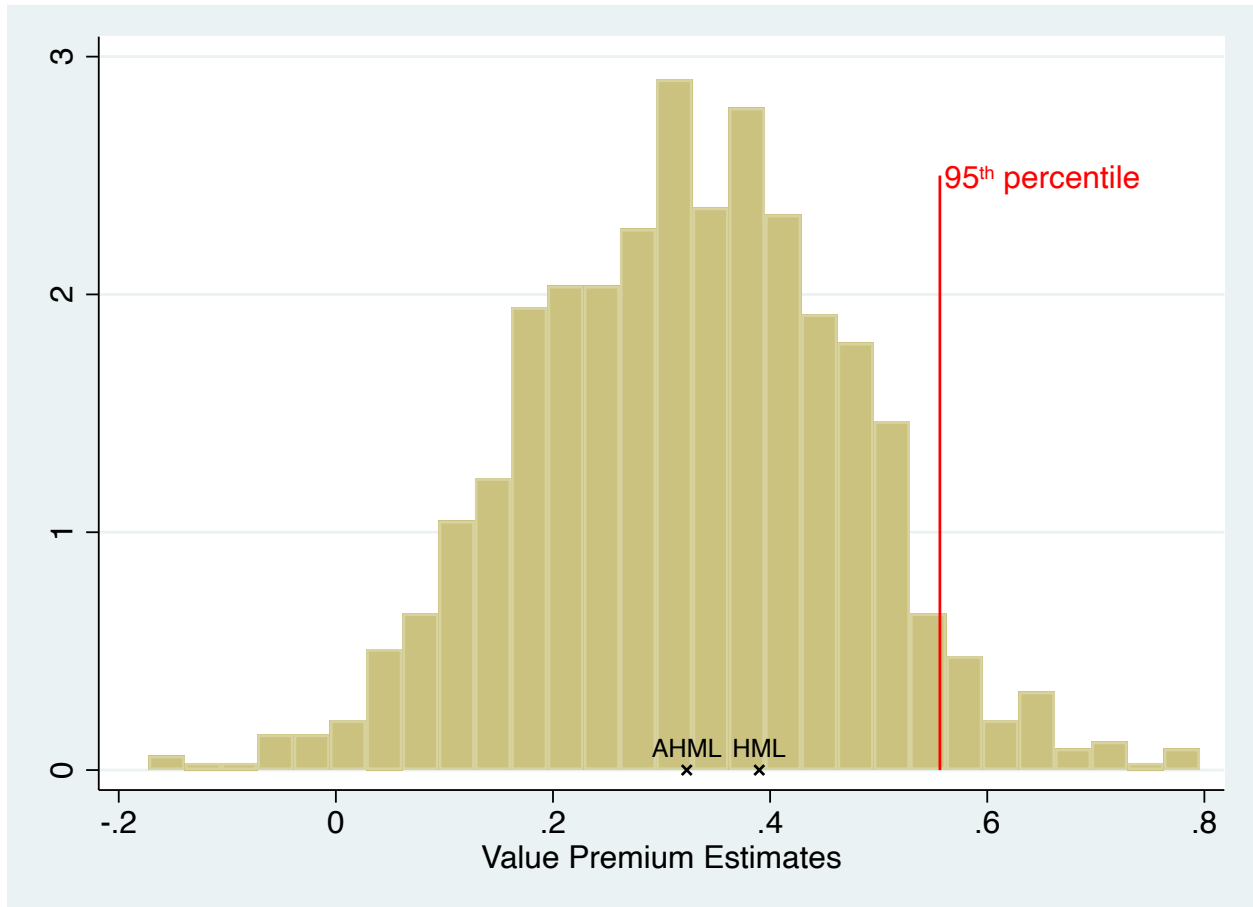
**Figure I**  
**Seemingly Innocuous Decisions in the Construction of the Original HML Portfolio, 1963 to 1991**

This histogram shows the average monthly returns of 96 HML portfolios from July 1963 to December 1991. The construction of the original HML portfolio (Fama and French, 1993) includes six seemingly innocuous decisions that could easily have been replaced with alternatives that are just as reasonable. I propose such alternatives, form all possible combinations of the six decisions and their alternatives, and construct 96 HML portfolios. The original HML portfolio (marked as “HML”) has an average monthly return of 0.39% per month (*t*-statistic of 2.87). The average HML portfolio (marked as “AHML”) has an average monthly return of 0.31% per month (*t*-statistic of 2.14). The average return difference between the original and the average HML portfolio is 0.09% per month (*t*-statistic of 1.72). The original value premium estimate is at least as large as 82 of the 96 value premium estimates and thus in the 85<sup>th</sup> percentile of the distribution. **Interpretation:** Among the reasonable value premium estimates in sample, the original estimate of the value premium is among the largest.



**Figure II**  
**Bootstrapping the Average HML Portfolio, 1963 to 1991**

This histogram shows bootstrapped of the value premium under the null hypothesis that the value premium is equal to the average value premium estimate. The average HML portfolio (marked as “AHML”) has an average monthly return of 0.31% per month and the original HML portfolio (marked as “HML”) has an average monthly return of 0.39% per month. 90% of all bootstrapped estimates are between 0.08% and 0.56% per month. The original estimate is at least as large as 673 of the 1,000 estimates and thus in the 67<sup>th</sup> percentile. The bootstrap procedure is defined as: I bootstrap monthly returns with replacement from the average HML portfolio and calculate the average monthly return. I repeat the bootstrap 1,000 times. **Interpretation:** The test suggests that the original value premium estimate falls within the sampling error of the average value premium estimate.



**Figure III**

**Reality Check Bootstrap for the Original HML Portfolio, 1963 to 1991**

This histogram shows bootstrapped  $t_{max}$ -statistics of a White (2000) Reality Check Bootstrap for the 96 HML portfolios under the null hypothesis that the value premium is zero. The original HML portfolio (marked as "HML") has  $t$ -statistic of 2.87. The 95<sup>th</sup> percentile of the  $t_{max}$ -statistics is 2.09 and marked with a red line. It serves as the critical value of a hypothesis test that considers multiple testing and that assumes a confidence level of 95%. The White (2000) reality check bootstrap procedure is defined as: I first demean each of the 96 HML portfolios. I then bootstrap months with replacement, calculate the mean return and its  $t$ -statistic for each of the 96 portfolios, and use the largest  $t$ -statistics as the  $t_{max}$ -statistic. I repeat the bootstrap 1,000 times. **Interpretation:** The test rejects the null hypothesis that the value premium is zero in sample under the assumption that the research decisions are made in a way that maximizes the  $t$ -statistic.

