The Volatility of Stock Investor Returns

Ilia D. Dichev*
Emory University

Xin Zheng
University of British Columbia

January 22, 2021

Abstract: The volatility of investor returns depends not only on the volatility of the stocks investors hold but also on their time-varying capital exposure to these holdings. We measure investor returns as dollar-weighted returns (IRRs), and provide comprehensive evidence on the volatility of investor returns using individual stocks, portfolios of stocks, and market indexes from the U.S. and major international stock markets. Our main finding is that the volatility of investor returns is higher than the corresponding volatility of stock returns in nearly all specifications. The relative magnitude of the volatility differential varies from as little as 10% and up to 75%, where this differential tends to increase with investment horizon. Probing into the drivers of this volatility differential reveals that investors tend to "chase stability" but have bad timing, adding capital to stocks after low past volatility but before high future volatility. Overall, taken together with existing evidence that investor returns tend to be lower than corresponding stock returns, this study suggests that the risk-return trade-off for stock investors is worse than previously thought.

We appreciate helpful comments from workshop participants at George Washington University, NYU Summer Camp, Santa Clara University, Georgia State University, Georgetown University, KAIST, and Seoul National University, and especially from Badri Kottimukkalur, Hank Bessembinder, and Meir Statman. We also appreciate research assistance from Jingyi Qian and Xinyi Huang. Ilia Dichev is the corresponding author (idichev@emory.edu).

The Volatility of Stock Investor Returns

1. Introduction

Stock investors experience volatility of returns because of the volatility of the stocks they hold. But they also experience return volatility because of their time-varying capital exposure to their stock holdings. For example, consider an investor who starts from scratch, and is steadily saving and investing for retirement in a stock index fund over a period of 20 years. Assume that the first 10 years the fund has low volatility of returns, while the last 10 years the fund has high volatility of returns. Intuitively, the volatility of this investor's returns over the 20-year life of the investment will be higher than the corresponding volatility of the index fund returns because his capital exposure was low when volatility was low, and high when volatility was high.

In this paper, we develop this intuition more rigorously, and provide comprehensive evidence on the volatility of investor returns vs. the volatility of corresponding stock returns. Stock returns are measured as buy-and-hold returns (BH returns). Investor returns are measured as dollar-weighted returns (DW returns), which are the IRRs from representing stocks or portfolios of stocks as sequences of signed investor capital flows. As shown in more detail later in the paper, the essential difference between these two return metrics is that BH returns assign equal weighting over each period of the time-series of returns, while dollar-weighted returns weight period returns by invested capital. Thus, dollar-weighted returns more properly reflect the actual investor experience, especially when capital exposure varies over time.

Our U.S. tests rely on NYSE/AMEX and Nasdaq data over 1925-2018. We start with a comparison of BH and DW returns at the individual stock level, using investment horizons from 5 to 30 years. The main finding is that over all horizons the volatility of investor returns is higher than the corresponding volatility of stock returns. This difference is statistically significant, and also seems economically substantial. For example, for the 20-year investment horizon, the standard deviation of annualized returns across stocks is 9.4% for the BH returns, and 13.9% for the DW returns, which indicates that the volatility of the actual investor experience is nearly 50% higher than the corresponding volatility of stock returns. International tests for individual stocks for Canada, France, Germany, Japan, and U.K. at 10 and 20-year investment horizons are largely in line with the U.S. results.

We proceed with portfolio tests, where we form random portfolios of 10 to 100 stocks, and track their returns over various investment horizons. The advantage of portfolio tests is that they reflect a more realistic representation of actual investor experiences. Specifically, existing evidence indicates that individual investors typically hold portfolios of 10 stocks or less, while mutual funds typically hold about 100 stocks (Barber, Odean, and Zhu 2009; Huang, Sialm, and Zhang 2011). Thus, our portfolios represent reasonably well the investor experience for both retail and institutional investors. In addition, the use of value-weighted portfolios tempers the effect of small stocks, which often have extreme returns and could dominate the tails of the distributions of stock returns across individual stocks. For all these reasons, we consider the portfolio results the main results of the paper.

In all U.S. and international portfolio results, the volatility of DW returns is higher than the volatility of BH returns, and the differences are statistically significant. In addition, the BH vs. DW volatility differential is economically large, and substantially increases in investment horizon. For example, for U.S. stocks and for the 10-year horizons the DW returns are about 15-20% more volatile but this differential rises to 70-75% for the 30-year horizon. These results indicate that the consideration of the right volatility metric is quite consequential.

Additional specifications include tests at the index level for U.S. and 19 international stock exchanges. A limitation of the index data is that there is no cross-section, and the available time series is usually relatively short, which necessitates using overlapping observations, with attendant problems of interpretation. Taken as a whole, the results of the index tests are largely congruent with the preceding tests. The statistical significance and the economic magnitude of the results, however, are generally weaker. Overall, the documented pattern of results provides consistent evidence that the volatility of investor returns is higher than the volatility of the corresponding stock returns.

Our final tests probe into the drivers for the documented differential volatility effect. Since the volatility differential arises because of the timing of equity capital flows, the drivers are related to the interactions of managerial and investor behavior, which determine the timing. In a nutshell, managers seem to cater to investors chasing stability but in the wrong way. Specifically, investors tend to add capital after past low volatility but before future high volatility (with the converse pattern for capital withdrawals). The result is that investors have high capital exposure when volatility is high, and accordingly investor returns that are more volatile than the corresponding stock returns, where the effects are exacerbated for settings with large capital flows and high volatility of returns.

Since return volatility is a key metric for investors, the implications of this study are potentially far-reaching. For example, combined with previous findings that DW returns are lower than BH returns (Dichev 2007; Friesen and Sapp 2007), the results of this study suggest that "the equity premium puzzle" may be less of a puzzle after the correct consideration of the actual investor experience. In addition, the bad timing of investors with respect to volatility suggests another argument for passive-style investing strategies that minimize trading, and avoid timing the market.

2. Theory and simulations

2.1 The difference between investor and stock returns

Since the difference between investor and stock returns is still a fairly new area of investigation, we start with a brief introduction to this topic. This difference is best illustrated by a simple example. Consider an investor who purchases an initial stake of $1,000 in stock X at the beginning of period 1.[1] The stock doubles during period 1, and the investor increases his stake by $1,000 at the end of period 1. The stock goes down by 50% in period 2, and the investor liquidates his entire holdings at the end of period 2. There are no dividends. It is clear that the buy-and-hold return for stock X over periods 1-2 is zero because the stock doubled, and then went down 50% to end up where it started. But the investor return is clearly not zero. The investor puts in $1,000 at the beginning of period 1, and another $1,000 at the end of period 1 for a total investment of $2,000, while the liquidation proceeds were only $1,500. Thus, the investor experience must have been negative since the liquidation proceeds were less than the total amount invested.

This intuition can be operationalized by using dollar-weighted returns, which are calculated as the internal rate of return on investor capital flows into and out of the investment. In this case, from the investor's point of view the capital flows are negative $1,000 at the

[1] This example is largely borrowed from Dichev (2007).

beginning of period 1, negative $1,000 at the end of period 1, and positive $1,500 at the end of period 2. Plugging these capital flows into the IRR calculation yields -17.7%, which correctly reflects the intuition that the investor lost money on this investment. The example also reveals that the reason for the negative return is bad timing – the investor increases his capital exposure after the initial high stock return, and before the subsequent negative return (see also Fischer and Wermers 2012 for more elaborate examples on this point). Another way to state the same conclusion is that the investor's return is low because his capital exposure is relatively low when returns are high, and his capital exposure is high when returns are low.

Summarizing, investor returns can differ from stock returns because of time-varying capital exposure. Specifically, investor returns differ from stock returns when there are correlations between investor capital flows and past or future returns. For example, if investors tend to have bad timing in the sense of pouring capital into stocks after superior stock returns and before inferior returns (with the converse for redeeming capital), investor returns will be lower than stock returns.

Appendix A develops the theory of DW returns further, including showing that DW returns are a re-weighting of period returns by beginning-of-the period invested capital. Thus, the essential difference between BH and DW returns is that DW returns are capital-weighted over time while BH returns are equally-weighted over time. In other words, DW returns are value-weighted *over time*, and thus they extend and complete the well-known advantages of using value-weighting in the cross-section of stocks. The implication is that DW returns are a better representation of the actual investor experience when capital exposure varies over time.[2]

[2] The treatment of dividends is another reason why DW returns are better than BH returns in reflecting actual investor performance. As is well-known, BH returns assume the reinvestment of dividends. While this may be a

The difference between DW and BH returns has been recognized and studied for the first moment of returns. Dichev (2007) documents that DW returns tend to be lower than BH returns for stock index returns in the U.S. and 18 international markets. Bessembinder et al. (2020) provide a comprehensive examination of individual stock performance around the world, and find that dollar-weighted returns tend to be lower than buy-and-hold returns for nearly all markets.

DW effects are not limited to stocks, and in fact several other studies find a more general pattern of investor returns that are lower than corresponding security returns. For example, Friesen and Sapp (2007) find that DW returns are about 1.5% lower than BH returns for a broad sample of U.S. mutual funds.[3] Hsu, Myers, and Whitby (2016) find that DW returns are about 1.3% lower than BH returns for value mutual funds but this gap rises to about 3% for growth funds. Madhavan and Sobczyk (2019) identify a return gap on the magnitude of 0.5 to 1% for broad samples of mutual funds and ETFs, for both equity and fixed-income instruments. Dichev and Yu (2011) document that DW returns for a broad sample of hedge funds lag BH returns by

---

reasonable assumption for any given investor, notice that the reinvestment of dividends is impossible for *investors as a class*. For example, if General Motors distributes an aggregate dividend of $100 million, investors as a class cannot reinvest the dividend because the aggregate number of shares has not changed. But DW returns handle this situation quite naturally because the $100 million appears as a capital distribution in the IRR calculation, and is therefore correctly reflected in the return experience of GM investors as a class. This difference in the treatment of dividends can be quite consequential. In GM's case, the annualized BH return on the original stock over 1925-2009 is 5.37%, while the DW return is 18.84%. What explains this huge difference? And which one is the "right" return? GM paid a lot of dividends over the years, and the buy-and-hold calculation assumes reinvestment of dividends, so essentially it implies an increasing all-in bet on GM's stock, which ultimately proves disastrous in the near-wipeout of the 2009 bankruptcy. But as discussed above, the reinvestment of dividends is impossible for GM investors as a class. In contrast, the DW return correctly reflects the fact that over 84 years of receiving generous dividends investors as a class have recovered their investment in GM stock several times, and so the ending price collapse does not skew the investor experience nearly as much (see also Bessembinder 2018 for similar insights from the GM experience).

[3] Some industry sources report much larger return gaps, e.g., DALBAR's Quantitative Analysis of Investor Behavior report (QAIB) estimates return gaps for U.S. mutual fund investors on the magnitude 3% to 4% for annualized returns. However, as Madhavan and Sobczyk (2019) explain, this large gap seems to be due to using methodology which does not adjust for the time value of money.

3% to 7%. The main takeaway is that by and large the return experience of investors is worse than the returns on the securities they hold.

The principal reason for these results seems to be a propensity to chase returns, i.e., investors tend to pour capital in after superior past returns and before average or subpar future returns, and so most investors end up earning less than the superior returns they are chasing (Fischer and Wermers 2012). There are also exceptions, however. Dvorak (2012) finds that defined-benefit pension plans have DW returns that are about 1% higher than their corresponding BH returns. The reason is that contributions to defined benefit plans tend to be counter-cyclical because legal obligations compel firms to contribute capital when returns are low and funds become underfunded.

2.2 The volatility of investor returns vs. the volatility of stock returns

The main insight of this study is that the difference between investor and security returns extends to the second moment of returns. Pursuing this insight, though, quickly reveals that conceptualizing the volatility of investor returns presents some unique challenges.[4] To illustrate, consider that any investment situation represents an ordered sequence of returns and capital flows, and this sequence is used to compute the BH and DW returns. For example, an investor buys some amount of stock ABC at time 0, receives a dividend each year, invests some more capital in stock ABC at the end of period 2, experiences a given sequence of returns over three years, and sells all holdings at the end of period 3. The sequence of returns over the three years

[4] Notice that finding IRRs in our setting involves solving high-degree polynomials. Since there are no analytical solutions for such polynomials, pursuing an analytical treatment of the volatility of DW returns is precluded as well.

is used to compute the BH return, and the sequence of initial, intermediate and final cash flows is used to compute the DW return. Notice that there is only *one* BH return, and only *one* DW return for this sequence of cash flows and returns, so our customary notion of volatility as a time-series phenomenon is hard to apply for our investigation.

Our solution is to turn to cross-sectional specifications. The hypothesized effects are best illustrated by a simulated example. We simulate a stock market that has 1,000 stocks and a total life of 20 years. For each stock, total life is split into two periods of 10 years each, where the two periods have the same expected returns but differ in simulated volatilities. Specifically, during the first period of 10 years the returns for each stock are drawn from a normal distribution with a mean of 10% and standard deviation of 10%. During the second period of 10 years the returns for each stock are drawn from a normal distribution with a mean of 10% and standard deviation of 20%. Each stock starts with an investment of $2,000, which evolves according to the pattern of simulated returns. There are no dividends, and there is a single capital contribution of $2,000 at the end of year 10, which marks the regime shift from low volatility to high volatility.

The gist of this simulation is that the stock market experiences a low-volatility period, followed by a high-volatility period, and there is a substantial capital infusion just before the start of the high-volatility period. Since the resulting capital exposure is relatively low during the low-volatility period, and relatively high during the high-volatility period, the expectation is that investor returns will be more volatile than the corresponding security returns. Panel A of Table 1 provides the baseline results for the simulation, presenting the empirical distributions of the simulated BH and DW returns across the 1,000 stocks.

An examination of Panel A reveals that the empirical distribution of DW returns across stocks is more spread-out than the empirical distribution of BH returns. The standard deviation of DW returns is 4.0% as compared to the standard deviation of BH returns of 3.5%, and this difference is highly statistically significant. A comparison of the percentiles of the empirical distributions confirms the impressions from the standard deviations.[5]

Panel B of Table 1 extends the baseline evidence from Panel A by varying the key parameters of the simulation, and tracing the resulting difference in the realized volatilities of BH and DW returns. Recall that this difference arises because of two factors, magnitude of the capital flows, and the difference in underlying stock volatility across the two periods. Panel B models the capital flows as taking five possible values at the end of the first period: $1,500, $1,000, ($1,000), ($1,500), and ($2,000). To be consistent with the empirical work later in the paper, capital flows are signed from the point of view of the investor, where positive values signify capital distributions from the stock to the investor, and negative values signify capital contributions from the investor to the stock. The difference in volatility across the two periods comes from holding the volatility of returns in the first period to be always 10%, but simulating the volatility in the second period as taking values of 15%, 20%, and 25%. The combined effect of varying these two key parameters is captured in a 5X3 matrix, where the cells of the matrix represent the difference between the standard deviation of BH returns and DW returns over the 1,000 stocks in the market. Note that the central bottom cell in the matrix corresponds to the baseline results in Panel A.

[5] We also model the null hypothesis where the simulated volatilities of returns are the same at 10% over the two 10-year periods. Since there is no correlation between the volatility of returns and the capital flows, the volatility of BH and DW returns should be nearly the same across the 1,000 simulations. The results confirm that the two standard deviations are nearly identical, and the same applies for nearly all percentiles of the empirical distribution.

The results in Panel B confirm the intuition that the difference between stock and investor returns becomes larger when there are larger capital flows, and larger differences in the time-varying volatility of stock returns. The difference between the volatility of investor and stock returns increases as we move down and right in the 5X3 matrix. While it is difficult to calibrate the expected real-world empirical magnitudes of these effects from this stylized example, the consideration of the relative magnitudes of the results in Panel B suggests strong non-linearity in the interaction of the two explanatory factors (magnitude of the capital flows and difference in volatility). Note that while the intensity of the two factors varies on the magnitude of 2-4 times along the axis of the matrix, the largest positive difference in the bottom rightmost cell (0.68%) is more than 20 times larger than the smallest positive difference in the leftmost cell on the third row (0.03%). This pattern suggests that investor timing can have substantial effects on the volatility of investor returns. Note also that the difference between the volatility of DW and BH returns is negative for capital distributions (the first two rows of Panel B), which is consistent with intuition that the volatility of investor returns is lower when capital is distributed away from the stock before the period of high volatility of returns.

3. Empirical tests and results

3.1 Individual stock returns

We start our investigation with tests at the individual stock level. At this level, we seek to answer the question "What is the dispersion of returns that stock investors experience if investing in one stock vs. another?" While most investors invest more broadly, the individual stock results are useful because they provide a natural baseline for the portfolio results later. In addition, they are descriptive for the experience of investors whose investments are dominated by a single stock (e.g., founders, entrepreneurs, executives with heavy stock compensation component).

Since dollar-weighting is essentially the re-weighting of returns over time, our main interest is in long-run investment experiences.[6] Specifically, for the U.S. tests we take all stocks from the major U.S. stock exchanges (NYSE, AMEX, and Nasdaq), and examine the empirical distribution of their BH and DW returns over plausible long-run investor horizons of 5, 10, 20, and 30 years.[7] To keep construction and interpretation simple, we present two specifications, in Table 2, Panels A and B, respectively. In the first specification, we require that included stocks have a full return history over the indicated horizons, e.g., all stocks included in the 10-year specification have at least a 10-year history of stock returns. Of course, having this requirement introduces some survivorship effects, so we also pursue a second specification that includes all stocks. All returns are annualized.

BH returns are computed as the geometric average of monthly returns over the full history of available data.[8] Delisting returns are included in the return calculations. DW returns are computed as the solution to the IRR calculation:[9]

---

[6] BH and DW returns are by definition identical at the single-period level. The gap between these two performance measures becomes pronounced only for longer periods, where there is sufficient time for meaningful interactions between the capital flows and returns.

[7] We also examined 40-year horizons but the number of available stocks drops under 1,000. Nonetheless, the results for the 40-year horizon are similar.

[8] Thus, stocks with "at least 10 years" have at least a 10-year life but the returns are computed over the full life of each stock, which means that the computed returns can be over 15 or 17 or 25 years. In Table 2, Panel B, and in the later portfolio tests, the return horizons indicate exact return spans, so a 20-year portfolio will be held for exactly 20 years (with possibly changing firm composition over time depending on delistings).

[9] A potential problem with using IRRs is that the polynomial can have multiple real roots. We use Matlab for IRR calculations, which has the advantage of warning about the presence of multiple real roots, and providing all real roots in the output. In practice, the multiple-root problem is only a significant concern when dealing with individual stocks. For example, about 50% of U.S. individual stocks have multiple roots (where about 90% of the stocks that

$$MV_0 = \frac{MV_T}{(1+r_{dw})^T} + \sum_{t=1}^{T} \frac{\text{Capital flow}_t}{(1+r_{dw})^t} \quad (1)$$

which equates beginning market value to ending discounted market value plus the discounted investor capital flows. Similar to Dichev (2007) and Friesen and Sapp (2007), the capital flows are computed using the formula:

$$\text{Distributions}_t = MV_{t-1} \times (1+r_t) - MV_t \quad (2)$$

where r_t is the BH return for period t, and the resulting capital flow variable *Distributions* is signed from the investor's point of view, where positive *Distributions* signify capital outflows like dividends and stock repurchases and negative *Distributions* signify capital inflows like stock issues.[10]

The individual stock results for the U.S. are presented in Table 2. Panel A includes the results for the first specification where individual stocks are held "at least n years", with n between 5 and 30. Panel A includes the number of stocks available at each horizon, and the mean and standard deviation of the BH and DW returns across stocks. To assess the effect of possible outliers, the table also includes the major percentiles for the full empirical distributions of the two return metrics.

have multiple roots have only two real roots). When there are multiple real roots, we designate the root which is closest to the BH return as the DW return. Since we are benchmarking against the BH returns, this procedure ensures that our estimates of the volatility of DW returns are conservative. In additional tests, we verify that limiting our sample to stocks with a single real root does not change the tenor of the results in the paper. Finally, for portfolio and index results the IRR calculation produces a single real root in about 98%+ of the calculations, so this is close to a non-issue there. The absence of multiple root problems is another reason we consider the portfolio results the main results of the paper.

[10] The intuition behind expression (2) is that the change in MV during a given period can come from only two sources, stock returns and investor capital flows. Thus, for any given period t, capital flows can be imputed from changes in MV during that period controlling for stock returns.

Table 2, Panel A reveals that at all horizons the volatility of DW returns is substantially higher than the corresponding volatility of BH returns. Taking the sample that has at least 20 years of data, the standard deviation of DW returns is 13.9% vs. 9.4% for BH returns, where the difference is highly statistically significant using bootstrap tests of significance (results using F-tests are close to identical).[11] The difference also seems large in economic terms, suggesting that stock investors experience return volatility which is almost 50% higher than the volatility of the stocks they invest in.

Closer scrutiny of the empirical distributions of the two return metrics confirms the summary impressions from the standard deviations but also reveals that this difference is especially pronounced in the left tails. For example, for the 20-year horizon the median and the percentiles above it are almost identical across BH vs. DW returns. But for the percentiles below the median the DW percentiles quickly plunge below their equivalents for BH returns, and the difference becomes rather substantial for the extreme percentiles. The 5th percentile (1st percentile) for BH returns is -10.7% (-21.2%), while their DW counterparts are -16.6% (-56.7%). Since DW returns are capital-weighted over time, one possible explanation for this result is that equity capital that comes in later in the life of a stock tends to do worse than initial investment. This explanation can be further explored as a conjecture that stocks with poor ex post outcomes have had to raise more equity capital, e.g., to cover operational losses. A preliminary

[11] Since stock returns are statistically not well-behaved, especially for long-run returns (Kothari and Warner 1997), we opt for bootstrap tests of significance. Specifically, since we are testing whether the effect of capital flows influences the volatility of returns, the null distribution of returns is given by the distribution of the BH returns across stocks (since the computation of BH returns is not affected by the timing and magnitude of capital flows). To create the bootstrap distribution, we draw from the distribution of BH returns, with replacement, until we have the number of necessary observations, for example 3,786 for the 20-year specification in Table 2, and calculate the standard deviation of the drawn sample. We repeat this procedure 1,000 times to give us the bootstrap distribution of the standard deviations of the BH returns. The p-value is computed by comparing the actual value of the standard deviation of DW returns against the derived bootstrap distribution.

investigation reveals that indeed this is the case, with stocks with performance delistings having histories with higher capital inflows.[12]

The difference in volatility between DW and BH returns also seems to increase with the investment horizon, where in relative terms it is the smallest for the 5-year horizon, and rises to 71% for the 30-year horizon (standard deviation is 6.8% for BH returns vs. 11.6% for DW returns). This evidence is consistent with the idea that longer horizons allow more interactions between capital flows and time-varying volatility, compounding the shorter-term effects.

Table 2, Panel B provides an additional specification for individual U.S. stocks, which includes all stocks, and where starting dates and stocks are picked randomly and stocks are held exactly n years, where n varies between 5 and 30. Since the number of stocks/trials is not naturally defined in this specification, we choose 10,000 trials, roughly corresponding to the number of stocks with 10-year history in Panel A. Note that since this specification includes all stocks, a great number of trials include one or more delisting events, which we handle as the delisting stock being replaced with a stock with the closest market value as of the delisting date, and further tracking the performance of that stock until the completion of the designated investment horizon.

An inspection of the results in Table 2, Panel B reveals the same pattern as the specification in Panel A. The standard deviation of BH returns is lower than the standard deviation of DW returns, and this difference is statistically significant at all horizons. The

[12] Specifically, we construct a stock-level variable that reflects the magnitude of stock issues as the average of *Distributions* (as defined in expression 2) scaled by beginning market value over the life of the stock. As is probably clear, more negative values of this variable indicate stocks that have received more capital inflows from investors over their life, i.e., they have issued more equity. We regress this stock-level variable on a dummy variable defined as one for performance delistings (Shumway 1997), and zero otherwise. We find highly statistically significant results that stocks with performance delistings have considerably higher equity issues.

economic magnitude of the differences in Panel B is somewhat more muted than that in Panel A but just like in Panel A, the differences are more pronounced in the left tail of the return distributions, and the volatility differential is increasing in investment horizon.

We extend the evidence for individual stocks by using Datastream data for major stock markets around the world. Our focus is on the largest, most developed markets since they matter the most in terms of market capitalization, have long time-series, and most importantly the cross-sectional availability is wide enough to allow meaningful computations and comparisons of standard deviation across stocks. Operationally, our definition of major markets is the G7 countries but we drop Italy since the coverage starts only in 1987, and the cross-sectional availability is on the magnitude of only about 100 stocks for most years (before investment horizon restrictions). Thus, the tests are limited to Canada, France, Germany, Japan and the U.K. Since the pattern of results is quite similar across the two individual U.S. stock specifications in Table 2, we only retain the first specification for the international stocks.

The evidence from the international stocks is presented in Table 3. Stock coverage is quite good for Japan and U.K. where even after requiring stocks to have 10 or 20 years of data, there are thousands of stocks in the cross-section. Coverage is good for the rest of the countries at the 10-year horizon, on the magnitude of 600 to 800 stocks in the cross-section but becomes only fair for the 20-year horizon at 200 to 400 stocks. We follow the same steps as for the U.S. specification in Table 2 but for parsimony the presented results are limited to two investment horizons, 10 and 20 years, and to the standard deviations for BH and DW returns rather than presenting statistics for the full empirical distribution of returns.

An examination of Table 3 reveals that the volatility of DW returns is greater than the volatility of BH returns for all countries and for both investment horizons, and all differences are highly statistically significant. Similar to the U.S. evidence in Table 2, the differences tend to increase as investment horizon lengthens. The differences are also quite large in economic terms. Using the volatility of BH returns as the baseline, the volatility of DW returns at the 20-year horizon is 61% higher for Japan, and 74% higher for the U.K., the two largest markets, with comparable magnitudes for the rest of the countries.[13] Overall, for individual stocks the international evidence on the volatility of DW returns is quite consistent with the U.S. results.

3.2 Portfolio results

Next, we examine portfolio specifications of BH and DW returns. For the U.S. tests, we include all available stocks on the major U.S. exchanges, with no restrictions on stock longevity. We simulate the investment experience of investors who choose portfolios of 10, 30, and 100 stocks, and hold them over 10 and 30-year horizons. Using the 10-stock portfolio as an example, the specification works as follows. First, we choose a random date within the available period, which is designated as the portfolio formation date. Then, we randomly choose 10 stocks from all available stocks as of that date, and start calculating the returns with that portfolio composition. When a stock drops out before the end of the investment horizon, we randomly choose a replacement stock available as of the dropout date, i.e., the portfolio always has 10 stocks over the entire investment horizon. At the conclusion of the investment horizon, we have the beginning market value, the returns and the capital flows during the intermediate years, and

[13] Taking Japan as an example, this calculation is based on the difference between the 20-year volatilities, divided by the volatility of BH returns, (0.087-0.054)/0.054 = 61%

ending market value, which allow us to compute the BH and DW returns. Then, we repeat this procedure 1,000 times, generating an empirical distribution of portfolio BH and DW returns.

While the portfolio specifications are still stylized, we believe that they are the most representative of the actual investor experience we seek to capture. Our belief is based on existing evidence suggesting that portfolios of 10 to 100 stocks reflect most real-world investment portfolios, and also fully reflect the effects of portfolios diversification. For example, extant research of retail investors suggests that their portfolios typically comprise only 4 to 7 stocks (Barber et al. 2009; Koestner et al. 2017), while mutual funds typically hold about 100 stocks (Huang et al. 2011; Kacperczyk, Sialm, and Zheng 2009). Our choice of portfolio size also reflects evidence that the benefits of portfolio diversification top out at portfolios of 20 to 50 stocks (Elton 2014; Campbell, Lettau, Malkiel, and Xu 2001).[14]

The portfolio results are presented in Table 4 and illustrated in Figure 1, again offering summary statistics and key percentiles of the empirical distributions of returns. We present results for portfolios with 10-year and 30-year horizons, in Panels A and B respectively. The key metric for our investigation, the standard deviation of BH vs. DW returns, shows a clear pattern of higher dispersion for DW returns in both panels. The differences between the standard deviation of BH and DW returns are all highly statistically significant, and large in economic magnitude. The relative differences, however, are much larger for the 30-year horizon specifications. In Panel A, the volatility of 10-year DW returns is on the magnitude of 15%-20% higher than the corresponding BH return volatility. For example, for the 100-stock portfolio, the standard deviation of BH returns is 0.051 vs. 0.060 for the corresponding DW returns. For the

[14] For completeness, we also simulate portfolios with 300 stocks. Untabulated results remain similar to those presented in Table 4.

30-year horizon and 100-stock portfolios, though, the volatility of BH returns is 0.020 vs. 0.035 for the DW returns, which translates to a very large relative difference of 75%. A closer look at the percentiles of the empirical distribution of returns confirms the message from the standard deviations. Similar to Table 2, most of the higher dispersion for DW returns is due to a heavier left tail, i.e., DW returns have more extreme poor return realizations. But in Table 4, there is also evidence of stratification in the right tail as well.

In Table 5, we expand the portfolio results using international stock data. For the same reasons as in the tests at the individual stock level, we limit our attention to markets in Canada, France, Germany, Japan, and the U.K. Similar to the U.S. tests, we construct 1,000 portfolios, picking randomly the date of construction and the stocks in the portfolio. For parsimony, we limit the presented results to one specification which relies on constructing portfolios of 50 stocks, and holding them over 20-year horizons. We also depict the results in Figure 2.

The volatility of DW returns is higher than the volatility of BH returns for all examined countries in Table 5, and all differences are highly statistically significant. The last column of Panel B presents the ratio of the difference between the two standard deviations divided by the standard deviation of BH returns, i.e., the ratio indicates how much larger in relative terms is the standard deviation of DW returns as compared to BH returns. The smallest ratio in the last column is 28.3% for France, and for three of the five countries the ratio exceeds 50%, reaching 65% for Germany. Thus, the economic magnitude of this difference seems rather substantial, and is largely in line with the corresponding evidence for U.S. portfolios.

3.3 Evidence from broad market indexes

We start with U.S. broad-market evidence based on the value-weighted CRSP index combining all stocks from NYSE, AMEX, and Nasdaq, and spanning years 1925-2018. An advantage of this specification is that it allows insight into the experience of investors in the whole market. A disadvantage is that there is only a single time-series, and thus if one is looking into the investment experience over 20-30 year horizons, there are only 3-4 non-overlapping observations. Our compromise solution is to look at overlapping windows, while keeping in mind the different interpretation of the results. For example, looking at 30-year investment horizons, we start with 1925-1955, and then roll forward by one year for each successive horizon, 1926-1956, 1927-1957, and so on. This specification essentially looks into how much of a difference does chance make in long-run index returns, if one starts in a given year vs. another. We also present results for 10-year, 20-year, and 40-year investment horizons.

The evidence is presented in Table 6. As is probably clear, since these are long-term index returns, and there is a heavy overlap between adjacent return intervals, the resulting variation in returns is relatively small. Nevertheless, the same basic pattern of results is clear in this specification as well. For example, the standard deviation of BH returns is 1.4% for the 30-year horizon, while the standard deviation of DW returns is 1.8%. The indicated difference is economically significant, where the volatility of DW returns is about a third higher than the volatility of the corresponding BH returns. Similar to the previous tables, the volatility differential grows larger over longer horizons. Overall, the results for the broad CRSP index are largely in line with the preceding results, keeping in mind the caveats about the differing interpretation.

In Table 7, we extend the broad-index evidence to international markets. We include all international markets that have a sufficiently long time-series of data, defined here as at least 30 years of annual index data on Datastream. The obvious advantage of international data is providing evidence on the generalizability of U.S. results around the world. The disadvantage is much shorter time-series than U.S. data, which is a key concern for our investigation seeking to capture long-run investor experience. Since the data series for international data are relatively short, we limit our attention to 20-year investment horizons, using the same overlapping-periods specification for the broad U.S. index in Table 6.

An inspection of the results in Table 7 reveals that the volatility of DW returns is higher than the corresponding volatility of BH returns in 14 out of 19 countries. To provide some summary measure of these differences, we provide the mean and median standard deviation of BH and DW returns across countries. The mean volatility of BH returns is 3.52% vs. 3.84% for DW returns, while medians are 3.16% vs. 3.69%, respectively. In economic terms, these results imply that the volatility of DW returns is on the magnitude of about 10% to 15% higher than the corresponding volatility of BH returns. In interpreting these economic magnitudes, it is useful to keep in mind the U.S. evidence above that dollar-weighted effects become stronger for longer horizons. Thus, international results are likely to be somewhat muted by the short time-series of Datastream data.

3.4 Evidence about the drivers of the volatility of stock investor returns

The evidence in all preceding specifications paints a consistent picture of the volatility of DW returns being considerably higher than the corresponding volatility of BH returns. In this section, we delve into the possible drivers of this differential. The theory and simulations earlier in the paper suggest some possible drivers, and we investigate them further including calibrating their effect in our data. For parsimony, we limit our analysis to portfolio specifications in U.S.

data. The reason is that, as explained earlier, we consider the portfolio results as closest to actual investor experiences, and U.S. data on stock returns has much wider coverage, and especially a long time-series, which is important for our investigation. Specifically, we use 1,000 portfolios of 30 stocks that are held for 30 years, representing the experience of diversified investors over long horizons. Using the exact same procedures as for the earlier portfolio tests, portfolio formation dates are picked randomly, portfolio stocks are picked randomly from all available stocks on that date, and stocks are held throughout the 30-year period unless they delist, in which case they are replaced by another random pick from the stocks available as of the delisting date.

Panel A in Table 8 provides the basic BH and DW return statistics for the 1,000 portfolios. Not surprisingly, the results in Panel A are quite similar to comparable portfolio results earlier in the paper; we provide them here for calibration, and to ensure that this sample replicates the earlier results in a reasonable manner. Most importantly, DW returns are much more volatile than corresponding BH returns in Panel B, and that volatility differential is substantial, with across-portfolios standard deviations of returns of 3.1% vs. 1.9%.

The theory and simulations in Section 2 suggest that the interaction between capital flows and time-varying volatility is the principal driver of the DW volatility differential. For example, if investors "chase stability" by withdrawing capital after stock volatility is high, and injecting capital after volatility is low, the resulting DW return volatility will be higher than the corresponding BH volatility. To investigate for such effects, we examine for possible correlations between capital flows and the volatility of past and future returns. Since the horizons of the possible interactions are not clear, we examine four return volatility windows, stretching from one to seven years. Specifically, in Table 8, Panel B, we examine the correlations between monthly *Distributions* and the standard deviation of past monthly returns over 1-year, 3-year, 5-

year, and 7-year windows. We repeat the same procedure for the correlation between monthly *Distributions* and the volatility of future returns.

The results in Panel B reveal that the correlation between current *Distributions* and the volatility of past returns is small but reliably positive for all four horizons, ranging in magnitudes from one to five percent. Since *Distributions* has a positive sign when capital is distributed to stock investors, the positive correlations with past volatilities mean that capital tends to flee stocks after episodes of high volatility. In other words, investors "flee volatility". In addition, the correlations between current *Distributions* and future volatilities of returns are negative and significant for all horizons except the 1-year horizon (the shortest). These results suggest that investors tend to pour capital into the market before future volatility turns out to be high, and tend to flee the market before future volatility turns out to be low.

The combination of these results implies that investors tend to have bad timing with respect to volatility. They tend to leave the market following high past volatility but before future low volatility. And they tend to pour capital into the market after past low volatility but before future high volatility. The implication is that stock investors have high capital exposure when stock volatility is high, and low capital exposure when volatility is low. As a result, (capital-weighted) DW returns will have higher volatility than (time-weighted) BH returns, which is exactly the main result of this paper. Thus, the propensity of investors to "chase stability" explains the volatility differential between DW and BH returns. (In untabulated additional results we find that the well-documented propensity to “chase returns” is not an explanation for the volatility differential identified in this paper).[15]

---

[15] Specifically, we repeat the tests in Table 8, Panel C using a “return-chasing“ stratifying variable defined as the correlation between Distributions and past 3-year returns. If return chasing is related to the identified volatility

We provide further corroborating evidence on this key result in Panels C and D. In Panel C, we stratify the 1,000-portfolio sample by the magnitude of the correlation between *Distributions* and the volatility of past 3-year returns. We expect that the effects identified in Panel B will be magnified when the correlation between capital flows and past volatility is the highest (most positive). Specifically, In Panel C we present the differential volatility results for deciles 1 and 10, where the 1,000 portfolios are ranked on the magnitude of their correlation between capital flows and past volatility of returns. Consistent with expectations, the difference between DW and BH volatility is much higher when the correlation is the most positive. The standard deviation of DW returns is only 38% higher than that for BH returns for decile 1 but this differential almost triples to 108% for decile 10. Thus, the volatility differential is much larger when the capital flows are more sensitive to past return volatility. Panel D provides analogous results for the correlation between *Distributions* and future volatility of returns, where the expectation is that the return differential is larger for decile 1, where the correlations are the lowest (most negative). Indeed, the pattern of results in Panel D confirms this expectation, with the volatility differential for decile 1 (68%) nearly double that for decile 10 (38%).

Finally, Panels E and F provide results for stratifying the sample on the magnitude of stock volatility, and on the magnitude of capital flows. Stock volatility is the volatility of monthly portfolio returns over the life of the portfolio, and magnitude of capital flows is operationally defined as the average absolute value of *Distributions* for each portfolio over portfolio life. The motivation is that Panels C and D inform on the strength of the interaction between the capital flows and past and future volatilities. But for a given interaction effect, the

---

differential, we expect that the sample comprising firms in the bottom 10% of the return chasing variable (most negative correlation between high positive returns and Distributions) will have higher volatility differential than the sample of firms in the top 10% of the return chasing variable. The empirical results actually show the opposite pattern.

economic magnitude of the volatility differential should be an increasing function of the magnitude of the two interacting variables, stock volatility and capital flows (see Madhavan and Sobczyk 2019 for a similar motivation on volatility). In other words, Panels C and D confirm that investors have bad timing with respect to volatility. But the effect of bad timing will be magnified by the magnitude of the volatility and capital flows in play with this bad timing.

Since more volatility allows for more timing effects, we expect that the volatility differential will be more pronounced for portfolios with high volatility. Indeed, in Panel E the volatility differential for decile 10 is nearly double that for decile 1. Panel F compares the results for stocks with the smallest capital flows in decile 1 vs. the stocks with the largest capital flows in decile 10, where the expectation is that the volatility differential is larger in decile 10. An examination of Panel F reveals a strong confirmation of this expectation, where the volatility differential is only 28% for decile 1 but rises more than eight-fold to 237% for decile 10. In fact, the volatility differential for decile 10 in Panel F is the single largest volatility differential in this paper. Thus, the magnitude of stock capital flows seems to be a decisive driver of the identified volatility differentials.

4. Discussion of the results and implications

Summing up the empirical results, the volatility of stock investor returns is much higher than the corresponding volatility of the stocks they hold. The explanation for this differential is timing of the capital flows with respect to volatility, specifically capital tends to flee stocks after high past volatility but before low future volatility, with the converse for capital inflows. Note that the timing of capital flows is a function of both managerial and investor actions. Essentially, managers seem to time market volatility with their net issuances while catering to corresponding volatility-sensitive investor demand for capital contributions and distributions. In that sense, the findings of this paper are reminiscent of the pseudo market timing effects of equity issues on the level of returns in Schultz (2003). Note also that because of the role of managers in timing equity issues the patterns of dollar-weighted results for stocks are meaningfully different from those for open-end mutual funds and ETFs, for example, where investor actions dominate by design.[16]

Keeping the dual manager/investor role in mind, our analysis and discussion mostly lean on the investor side, as that side more naturally aligns with existing findings on the first moment of dollar-weighted returns. In a nutshell, the findings of this paper imply that investors tend to "chase stability" but in the wrong way, ending up worse off as compared to a passive investor. These results for the second moment of stock returns closely mirror previous findings for the first moment of returns. Existing research has found ample evidence that stock investors tend to chase returns, and end up worse off as compared to passive investors, e.g., Dichev (2007) finds that DW returns are lower than BH returns on the magnitude of 1.5% over long horizons for U.S. and international stock markets. Thus, the combined effect is that stock investors have both lower and more volatile returns than those for the corresponding stocks.

Since the level and volatility of stock returns are obviously key metrics for investors, the implications of our findings are potentially far-reaching. At a minimum, it looks worthwhile to extend the volatility analysis to other asset classes and investors, e.g., mutual funds, hedge funds, pension funds, bonds, ETFs, and others. Looking deeper into some of the documented key results is also worthwhile, e.g., why the volatility differential grows with investment horizon.

[16] Further analysis of the relative strength of these two sides can be a useful extension of our results.

But for the purposes of this study, we would like to offer some observations on at least two possible implications, where the conclusions seem more clear-cut.

One is the so-called equity premium puzzle, i.e., stock returns seem too high as compared to their risk (e.g., Mehra and Prescott 1985; Constantinides 1990; Benartzi and Thaler 1995). Since the main result of this study is that the volatility of investor returns is much higher than the corresponding volatility of stock returns, the equity premium puzzle is perhaps less puzzling after a consideration of the true volatility of investor returns. A caveat for this implication is that the arguments for the equity premium puzzle are based on the volatility of the whole stock market, while the evidence in this paper is based on cross-sectional volatility. The suggestion seems warranted, nevertheless, as existing research documents a reliable positive relation between cross-sectional and aggregate stock volatility (Stivers 2003 and Connolly and Stivers 2006). Notice also that, as discussed above, investor returns tend to be lower than corresponding stock returns. The implication is that, after the consideration of DW returns, the equity premium puzzle is chipped away both on the returns side, and on the risk side.

The other fairly straightforward implication is on the investor behavior side. Investor skills and experiences vary widely, so generalizations have to be made with caution. In addition, it is not clear whether investors are already aware of the patterns of volatility reflected in our results, and if so, whether they (or at least some subsets) are already taking actions to counter their effect. But with these qualifications in mind, the results for DW returns suggest that on average investors have bad timing with respect to stock volatility. They tend to "chase stability" but end up being more invested during periods of high volatility, and so their returns end up being more variable than those for corresponding stock returns. It also looks unlikely that this is a "rational allocation" in the sense of being more invested during volatile times hoping to earn higher returns; if that was the case, DW returns would be higher than BH returns while existing results show it is exactly the opposite. The upshot then is that investors need to be careful about adjusting their capital allocation over time. Simple strategies like passive buy-and-hold would by definition earn the BH returns, which tend to be higher and less variable than those for the average investor. If capital allocation needs to be adjusted, for example for the purpose of increasing or decreasing retirement funds, the adjustments are best done gradually, and with no attempt to time the market.

5. Conclusion

The volatility of stock investor returns depends not only on the volatility of the stocks they hold but also on their capital exposure to their holdings over time. We use dollar-weighted returns to investigate the volatility of stock investor returns, and contrast it to the volatility of the corresponding buy-and-hold stock returns. We provide a comprehensive investigation of investor experiences, including settings for single stocks, portfolios of stocks, and indexes for U.S. and international data. Our main finding is that the volatility of investor returns is considerably higher than the corresponding volatility of stock returns for nearly all specifications. The relative magnitudes of this differential vary from as little as 10% to as high as 75%, substantially increasing in investment horizon. Additional tests reveal that investor propensity to chase stability is the principal driver of this volatility differential, where the effects are exacerbated for settings with large capital flows and high volatility of returns. One implication of these results is that existing metrics understate the volatility of investor returns, and so the equity premium puzzle may be less puzzling than previously thought. On the investor behavior side, the findings of this study are another argument for investing styles that minimize trading, and avoid timing the market.

Appendix A

As shown in Dichev and Yu (2011) one can take the expression for capital flows from equation (2), plug it into the DW returns calculation in equation (1), and after re-arranging, obtain:

$$\sum_{t=1}^{T} \frac{MV_{t-1}}{(1+r_{dw})^{t-1}} \times r_{dw} = \sum_{t=1}^{T} \frac{MV_{t-1}}{(1+r_{dw})^{t-1}} \times r_t . \tag{3}$$

Notice that r_{dw} is a constant in the left-hand summation of expression (3), and can therefore be taken out of the summation term, which allows one to divide the right-hand side of expression (3) by the summation term on the left-hand side. After that, re-arranging and introducing the new variable (weights) w_t produces:

$$r_{dw} = \sum_{t=1}^{T} w_t \times r_t$$

where $w_t = \dfrac{MV_{t-1}}{(1+r_{dw})^{t-1}} \Big/ \sum_{t=1}^{T} \dfrac{MV_{t-1}}{(1+r_{dw})^{t-1}}$.

Notice that the weights w_t for period t are equal to the beginning of the period t discounted market value scaled by the sum of discounted market values over the life of the stock. In other words, dollar-weighting returns are essentially value-weighted returns over time, where the only difference with cross-sectional value-weighting is that market values are discounted to account for the fact that they occur at different points in time, and need to be re-scaled to account for the time value of money. It is exactly this value-weighting over time which implies that DW returns are the proper metric for measuring the actual investor experience since they apply more weight to the period returns when there is more invested capital.

References:

Barber, B.M., Odean, T., Zhu, N., 2009. Systematic noise. *Journal of Financial Markets* 12, 547–569.

Benartzi, S. and Thaler, R.H., 1995. Myopic loss aversion and the equity premium puzzle. *The Quarterly Journal of Economics* 110 (1), 73-92.

Bessembinder, H., 2018, Do stocks outperform Treasury bills? *Journal of Financial Economics* 129 (3), 440-457.

Bessembinder, H., T.F Chen, G. Choi, and K.C. John Wei, 2020, Long-term shareholder returns: Evidence from 64,000 global stocks, Working paper, Arizona State University.

Campbell, J.Y., Lettau, M., Malkiel, B.G., Xu, Y., 2001. Have individual stocks become more volatile? An empirical exploration of idiosyncratic risk. *The Journal of Finance* 56, 1–43.

Connolly, R., and C. Stivers, 2006, Information content and other characteristics of the daily cross-sectional dispersion in stock returns, *Journal of Empirical Finance* 13, 79-112.

Constantinides George M., 1990. Habit formation: A resolution of the equity premium puzzle. *Journal of Political Economy* 98(3), 519-543.

Dichev, I.D., 2007, What are stock investors' actual historical returns? Evidence from dollar-weighted returns, *American Economic Review* 97, 386-401.

Dichev, I.D., and G. Yu, 2011, Higher risk, lower returns: What hedge fund investors really earn, *Journal of Financial Economics* 100, 248-263.

Dvorak, T., 2012, Timing of Retirement Plan Contributions and Investment Returns: The Case of Defined Benefit versus Defined Contribution, *The B.E. Journal of Economic Analysis & Policy* 12 (1), 1-26.

Elton, E.J., 2014. Modern portfolio theory and investment analysis, Ninth edition. ed. Wiley, Hoboken, NJ.

Friesen, G., and Sapp, 2007, Mutual fund flows and investor returns: an empirical examination of fund investor timing ability, *Journal of Banking & Finance* 31, 2796-2816.

Fischer, B., and R. Wermers, 2012, Performance evaluation and attribution of security portfolios, in *Handbook of Economics*, Academic Press, 1-715.

Hsu, J., B. Myers, and R. Whitby, 2016, Timing poorly: A guide to generating poor returns while investing in successful strategies, *Journal of Portfolio Management* 42 (2), 90-98.

Huang, J., Sialm, C., Zhang, H., 2011. Risk Shifting and Mutual Fund Performance. *Review of Financial Studies* 24, 2575–2616.

Kacperczyk, M., Sialm, C., Zheng, L., 2005. On the Industry Concentration of Actively Managed Equity Mutual Funds. *The Journal of Finance* 60, 1983–2011.

Kothari, S.P., and J.B. Warner, 1997, Measuring long-horizon security performance, *Journal of Financial Economics* 43 (3), 301-339.

Koestner, M., Loos, B., Meyer, S., Hackethal, A., 2017. Do individual investors learn from their mistakes? *Journal of Business Economics* 87, 669–703.

Madhavan, A., and A. Sobczyk, 2019, Does trading by ETF and mutual fund investors hurt performance? Evidence from time- and dollar-weighted returns, *Journal of Investment Management* 17 (3), 1-17.

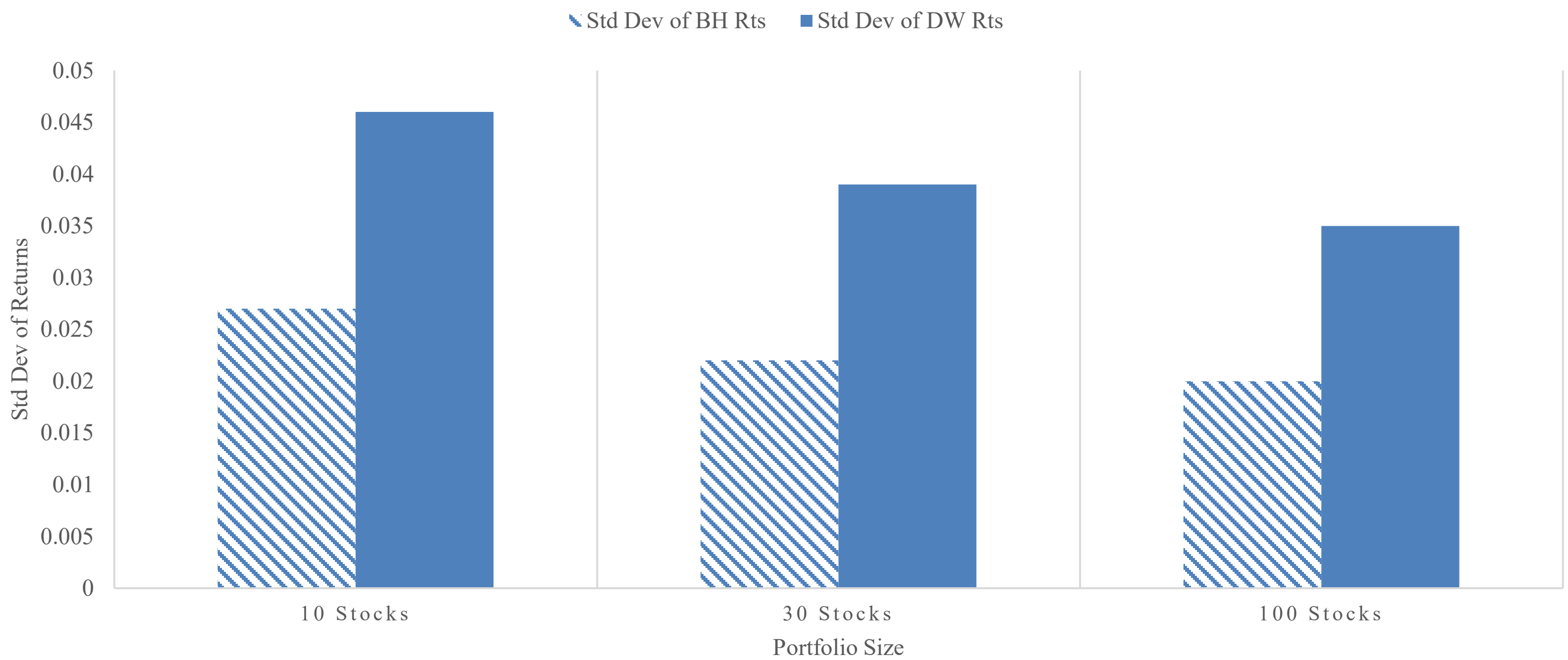
Mehra, R. and Prescott, E.C., 1985. The equity premium: A puzzle. *Journal of Monetary Economics*, 15(2), 145-161.

Schultz, P., 2003, Pseudo market timing and the long-run underperformance of IPOs, *The Journal of Finance* 58 (2), 483-517.

Shumway, T, 1997, The delisting bias in CRSP data, *The Journal of Finance* 56 (3), 983-1009.

Stivers, C.T., 2003, Firm-level return dispersion and the future volatility of aggregate stock market returns, *Journal of Financial Markets* 6, 389-411.

Figure 1: Std Dev of BH returns and DW returns for 1,000 Simulated 30-year Portfolios with U.S. Stocks



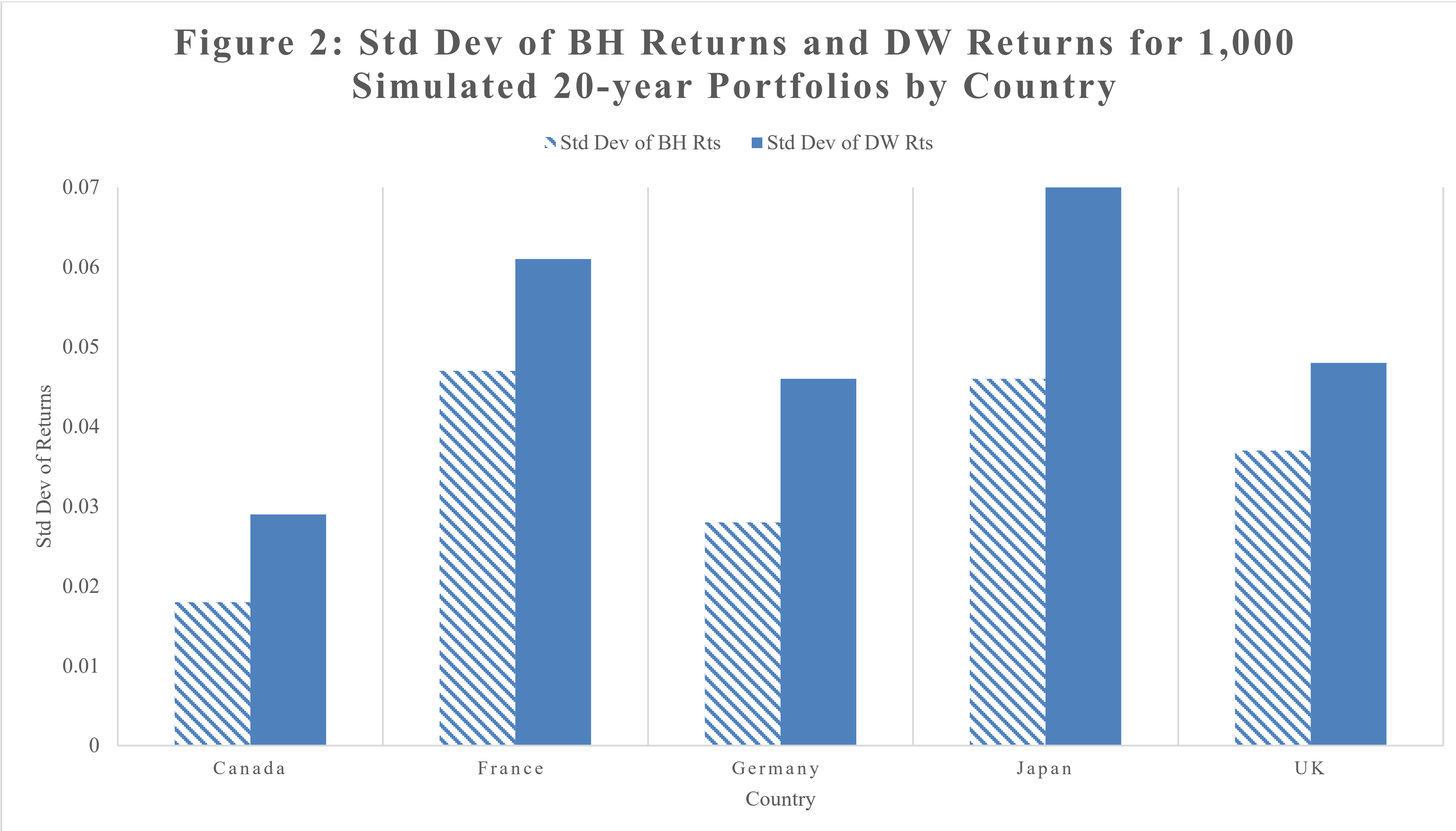


Figure 2: Std Dev of BH Returns and DW Returns for 1,000 Simulated 20-year Portfolios by Country
Std Dev of BH Rts
Std Dev of DW Rts
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
Std Dev of Returns
Canada
France
Germany
Japan
UK
Country

Table 1: Statistics for the Motivation Example

This table presents the results of the motivation example. We simulate a stock market that has 1,000 stocks. Each stock has a life span of 20 years, which is split into two periods of 10 years each. For the first period, the returns for each stock are drawn from a normal distribution with a mean of 10% and standard deviation of 10%. For the second period, the returns are drawn from a normal distribution with mean of 10% and standard deviation of 20%. Each stock starts with an investment of $2,000. We assume there are no dividends and there is a single capital contribution of $2,000 at the end of year 10 (i.e., end of period 1). Panel A reports the distributions of BH returns and DW returns for the 1,000 stocks. Panel B reports the difference in volatilities between buy-and hold and DW returns by varying two parameters. The two parameters are differences in variances in periods 1 and 2, and capital contribution at the end of the first period.

Panel A: Distributions of annualized BH returns and DW returns for the 1,000 stocks

	BH Returns	DW Returns
N	1,000	1,000
Mean	0.089	0.088
Standard Deviation	**0.035**	**0.040**
99%	0.172	0.176
95%	0.149	0.154
90%	0.134	0.137
75% Q3	0.114	0.116
50% Median	0.089	0.087
25% Q1	0.064	0.061
10%	0.044	0.035
5%	0.031	0.020
1%	0.010	-0.006
P-value of F-test of the equality of variances between BH returns and DW returns	<0.001	

Panel B: Difference in volatility between DW returns and BH returns by varying parameters

Capital distributions to investor at the end of the first period	*Differences in standard deviation across the two periods*		
	5%	*10%*	*15%*
$1,500	-0.0049	-0.0055	-0.0065
$1,000	-0.0039	-0.0040	-0.0046
($1,000)	0.0003	0.0017	0.0032
($1,500)	0.0013	0.0031	0.0050
($2,000)	0.0023	0.0044	0.0068

Table 2

Panel A: Distributions of BH Returns and DW Returns at the U.S. Individual Stock Level, only long-lived stocks

This table reports the distributions of BH returns and DW returns at the individual U.S. stock level, only using stocks with lives exceeding thresholds between 5 and 30 years. We obtain monthly stock information from the CRSP monthly files. We include stocks listed on NYSE, AMEX, and NASDAQ from 1925 to 2018. The first row of the table indicates the investment horizons. At least 5, 10, 20, 30 years indicates that stocks are required to have at least 5, 10, 20, 30 years of consecutive monthly non-missing records in CRSP. All returns are annualized. We test the difference in standard deviations between BH returns and DW returns using a bootstrap test with 1,000 trials and report the associated p-value.

	At least 5 Years		*At least 10 Years*		*At least 20 Years*		*At least 30 Years*	
	BH Rts	DW Rts	BH Rts	DW Rts	BH Rts	DW Rts	BH Rts	DW Rts
N	14,786	14,786	9,068	9,068	3,786	3,786	1,634	1,634
Mean	-0.013	-0.032	0.031	0.012	0.071	0.056	0.086	0.073
Std Dev	**0.219**	**0.253**	**0.154**	**0.193**	**0.094**	**0.139**	**0.068**	**0.116**
p-value	**<0.001**		**<0.001**		**<0.001**		**<0.001**	
99%	0.371	0.382	0.274	0.291	0.221	0.239	0.210	0.230
95%	0.235	0.233	0.199	0.198	0.182	0.181	0.171	0.177
90%	0.185	0.181	0.169	0.166	0.159	0.158	0.149	0.153
75%	0.121	0.117	0.122	0.118	0.126	0.123	0.125	0.123
Median	0.052	0.047	0.068	0.063	0.089	0.082	0.097	0.090
25%	-0.111	-0.123	-0.021	-0.027	0.041	0.033	0.059	0.053
10%	-0.328	-0.398	-0.174	-0.216	-0.041	-0.061	0.006	-0.001
5%	-0.442	-0.559	-0.259	-0.389	-0.107	-0.166	-0.035	-0.065
1%	-0.685	-0.853	-0.444	-0.766	-0.212	-0.567	-0.136	-0.536

Table 2 (continued)
Panel B: Distributions of BH Returns and DW Returns at the U.S. Individual Stock Level, all stocks

This table reports the distributions of BH returns and DW returns at the individual U.S. stock level, using all available stocks. We obtain monthly stock information from the CRSP monthly files. We include stocks listed on NYSE, AMEX, and NASDAQ from 1925 to 2018. The first row of the table indicates the investment horizons, where 5, 10, 20, or 30 years indicate that stocks are held exactly 5, 10, 20 or 30 years. We simulate the investment experience at each investment horizon, by using 10,000 picks of a random date and a random stock, and tracking their performance over the indicated horizon. If a stock delists during the examined period, we replace it with the stock that has the nearest market value as of the date of delisting, and continue tracking performance until the end of the indicated horizon. All returns are annualized. We test the difference in standard deviations between BH returns and DW returns using a bootstrap test with 1,000 trials and report the associated p-value.

	5 Years		*10 Years*		*20 Years*		*30 Years*	
	BH Rts	DW Rts	BH Rts	DW Rts	BH Rts	DW Rts	BH Rts	DW Rts
N	10,000	10,000	10,000	10,000	10,000	10,000	10,000	10,000
Mean	0.077	0.075	0.079	0.079	0.084	0.084	0.080	0.079
Std Dev	**0.208**	**0.218**	**0.137**	**0.149**	**0.097**	**0.126**	**0.085**	**0.123**
p-value	**<0.001**		**<0.001**		**<0.001**		**<0.001**	
99%	0.598	0.625	0.370	0.392	0.266	0.320	0.218	0.288
95%	0.400	0.406	0.273	0.282	0.212	0.228	0.182	0.200
90%	0.309	0.318	0.230	0.234	0.187	0.194	0.163	0.172
75%	0.194	0.195	0.163	0.164	0.144	0.146	0.133	0.135
Median	0.091	0.089	0.093	0.092	0.098	0.098	0.099	0.099
25%	-0.032	-0.038	0.017	0.012	0.045	0.042	0.052	0.051
10%	-0.192	-0.197	-0.093	-0.099	-0.038	-0.030	-0.032	-0.024
5%	-0.293	-0.301	-0.184	-0.188	-0.108	-0.102	-0.096	-0.117
1%	-0.487	-0.506	-0.344	-0.351	-0.233	-0.396	-0.215	-0.407

Table 3: Distributions of std(BH Returns) and std(DW Returns) at the Individual Stock Level for Selected Countries

This table reports the distributions of standard deviation of BH returns and DW returns at the individual stock level for Canada, France, Germany, Japan and the UK. We obtain monthly stock information between 1973 and 2018 from Datastream. The first row of the table indicates the investment horizons. 10 years and 20 years indicate that stocks are required to have at least 10 or 20 years of consecutive monthly non-missing records in Datastream. We report standard deviation of annualized returns. We test the difference in standard deviations between BH returns and DW returns using a bootstrap test with 1,000 trials and report the associated p-value.

	Canada				**France**			
	10 Yrs		*20 Yrs*		*10 Yrs*		*20 Yrs*	
	BH Rts	DW Rts	BH Rts	DW Rts	BH Rts	DW Rts	BH Rts	DW Rts
N	636	636	246	246	797	797	351	351
Std Dev	0.102	0.155	0.083	0.151	0.086	0.108	0.067	0.100
Difference in Std Deviation	**0.053**		**0.068**		**0.022**		**0.033**	
p-value	<0.001		<0.001		<0.001		<0.001	

	Germany				**Japan**			
	10 Yrs		*20 Yrs*		*10 Yrs*		*20 Yrs*	
	BH Rts	DW Rts	BH Rts	DW Rts	BH Rts	DW Rts	BH Rts	DW Rts
N	752	752	372	372	4187	4187	2549	2549
Std Dev	0.085	0.110	0.053	0.098	0.084	0.117	0.054	0.087
Difference in Std Deviation	**0.025**		**0.046**		**0.033**		**0.033**	
p-value	<0.001		<0.001		<0.001		<0.001	

	UK			
	10 Yrs		*20 Yrs*	
	BH Rts	DW Rts	BH Rts	DW Rts
N	2536	2536	1280	1280
Std Dev	0.128	0.171	0.081	0.141
Difference in Std Deviation	**0.043**		**0.060**	
p-value	<0.001		<0.001	

Table 4: Portfolio Tests Using U.S. Stocks

We include all stocks on the major U.S. exchanges between 1925 and 2018 without restrictions on stock longevity. The data is obtained from CRSP monthly data set. We simulate portfolio sizes of 10, 30, and 100 stocks, and hold them over 10-year and 30-year horizons, results of which are presented in panels A and B respectively. Using the 10-stock portfolio as an example, the specification works as follows. First, we choose a random date within the available period, which is designated as the portfolio formation date. Then, we randomly choose 10 stocks from all available stocks as of that date, and start calculating the returns with that portfolio composition. When a stock drops out before the end of the investment horizon, we randomly choose a replacement stock available as of that date, i.e., the portfolio always has 10 stocks over the entire investment horizon. Then, we repeat this procedure 1,000 times, generating an empirical distribution of portfolio BH and DW returns. All returns are annualized. We test the difference in standard deviations between BH returns and DW returns using a bootstrap test with 1,000 trials and report the associated p-value.

Panel A: 10-year horizon

	10 Stocks		*30 Stocks*		*100 Stocks*	
	BH Rts	DW Rts	BH Rts	DW Rts	BH Rts	DW Rts
N	1000	1000	1000	1000	1000	1000
Mean	0.147	0.144	0.150	0.148	0.155	0.153
Std Dev	**0.066**	**0.077**	**0.056**	**0.065**	**0.051**	**0.060**
p-value	**<0.001**		**<0.001**		**<0.001**	
99%	0.285	0.318	0.277	0.287	0.254	0.293
95%	0.252	0.266	0.242	0.249	0.235	0.247
90%	0.234	0.237	0.224	0.230	0.223	0.228
75%	0.192	0.192	0.190	0.194	0.195	0.193
Median	0.147	0.143	0.150	0.147	0.152	0.151
25%	0.107	0.097	0.113	0.105	0.120	0.114
10%	0.066	0.056	0.082	0.070	0.091	0.081
5%	0.034	0.014	0.056	0.038	0.073	0.058
1%	-0.029	-0.062	0.018	-0.015	0.040	0.004

Panel B: 30-year horizon

	10 Stocks		*30 Stocks*		*100 Stocks*	
	BH Rts	DW Rts	BH Rts	DW Rts	BH Rts	DW Rts
N	1,000	n	1,000	1,000	1,000	1,000
Mean	0.154	0.151	0.159	0.155	0.163	0.160
Std Dev	**0.027**	**0.046**	**0.022**	**0.039**	**0.020**	**0.035**
p-value	**<0.001**		**<0.001**		**<0.001**	
99%	0.226	0.274	0.211	0.254	0.209	0.242
95%	0.201	0.224	0.196	0.216	0.200	0.220
90%	0.187	0.200	0.188	0.201	0.194	0.207
75%	0.170	0.171	0.174	0.174	0.178	0.179
Median	0.153	0.148	0.157	0.153	0.160	0.155
25%	0.136	0.127	0.143	0.134	0.148	0.140
10%	0.121	0.103	0.132	0.115	0.139	0.123
5%	0.113	0.088	0.126	0.096	0.135	0.109
1%	0.099	0.052	0.117	0.071	0.125	0.077

Table 5: Portfolio Tests for Selected International Countries

This table presents the results of 1,000 simulations of 20-year portfolios that contain 50 stocks each by country. We obtain monthly stock returns for all equities traded in Canada, France, Germany, Japan, and UK through Datastream. Panel A reports the number of unique firms in our data by year and by country. Panel B shows the difference in standard deviations between BH returns and DW returns by country. We impose two restrictions. First, we require each stock to have at least 12 months of stock returns. Second, we require stock price to be at least 1 U.S. dollar or equivalent in local currency. We simulate portfolio sizes of 50 stocks, and hold them over 20-year horizon. First, we choose a random date within the available time period, which is designated as the portfolio formation date. Then, we randomly choose 50 stocks from all available stocks as of that date, and start calculating the returns with that portfolio composition. When a stock drops out before the end of the investment horizon, we randomly choose a replacement stock available as of that date, i.e., the portfolio always has 50 stocks over the entire investment horizon. Then, we repeat this procedure 1,000 times, generating an empirical distribution of portfolio BH and DW returns. All returns are annualized. We then calculate the standard deviation of annualized BH returns and DW returns. We test the difference in standard deviations between BH returns and DW returns using a bootstrap test with 1,000 trials and report the associated p-value.

Portfolio results by country

Country	**(1): std(BH Rts)**	**(2): std(DW Rts)**	**Difference in std (3): (2)-(1)**	**(4): p-value for (3)**	**Relative diff. in std (5): (3)/(1)**
Canada	0.018	0.029	0.011	<0.001	0.618
France	0.047	0.061	0.013	<0.001	0.283
Germany	0.028	0.046	0.018	<0.001	0.649
Japan	0.046	0.070	0.024	<0.001	0.519
UK	0.037	0.048	0.011	<0.001	0.308

Table 6: Distributions of BH Returns and DW Returns for the broad U.S. index

This table presents the results for a broad U.S. index. We use the ASIX dataset in CRSP, which is a NYSE, AMEX, and NASDAQ annual rebalanced index. We form rolling portfolios at 10 years, 20 years, 30 years, or 40 years horizons. For example, looking at 30-year investment horizons, we start with 1925-1955, and then roll forward by one year for each successive horizon, 1926-1956, 1927-1957, and so on. We test the difference in standard deviations between BH returns and DW returns using a bootstrap test with 1,000 trials and report the associated p-value.

	10 Years		*20 Years*		*30 Years*		*40 Years*	
	BH Rts	DW Rts	BH Rts	DW Rts	BH Rts	DW Rts	BH Rts	DW Rts
N	84	84	74	74	64	64	54	54
Std Deviation	0.052	0.057	0.032	0.037	0.014	0.018	0.010	0.014
Difference in Std Deviation	**0.005**		**0.0052**		**0.0048**		**0.0040**	
p-value	**0.035**		**0.007**		**<0.001**		**<0.001**	

Table 7: Results for International Index Returns

This table reports international results for sizable stock markets around the world. We obtain monthly country-level index from Datastream. All data end in 2018. For each country, we form rolling 20-year portfolios at the starting year. Using Japan as an example, Japan's stock index data starts in 1973. We form a portfolio in 1974, hold it through 1993, and calculate the BH and DW returns. The next portfolio starts in 1975 and hold it through 1994, and so on, with the end result of 25 investment experiences for Japan.

Country-level portfolio results

Country	Start Year	(2) Std Dev of BH Rts	(3) Std Dev of DW Rts	(4): (3)-(2)
Australia	1973	0.0267	0.0274	0.0006
Austria	1973	0.0316	0.0385	0.0069
Belgium	1973	0.0296	0.0335	0.0039
Canada	1974	0.0155	0.0200	0.0046
Denmark	1974	0.0309	0.0305	-0.0004
France	1973	0.0359	0.0436	0.0077
Germany	1973	0.0218	0.0287	0.0068
Hong Kong	1973	0.0515	0.0596	0.0081
Ireland	1973	0.0479	0.0621	0.0142
Italy	1973	0.0560	0.0448	-0.0112
Japan	1973	0.0428	0.0351	-0.0078
Netherlands	1973	0.0374	0.0438	0.0064
Norway	1981	0.0556	0.0538	-0.0019
Singapore	1973	0.0308	0.0336	0.0028
South Africa	1973	0.0299	0.0287	-0.0012
Sweden	1983	0.0234	0.0264	0.0030
Switzerland	1973	0.0304	0.0401	0.0097
U.K.	1965	0.0385	0.0429	0.0044
U.S.	1974	0.0318	0.0369	0.0051
Mean		**0.0352**	**0.0384**	**0.0033**
Median		0.0316	0.0369	0.0044
P-value of t-test for the difference between average standard deviations of BH Rts and DW Rts				**0.014**

Table 8: The Drivers of the Volatility Differential

In this table, we simulate 1,000 portfolios using all stocks on major U.S. stock exchanges between 1925 and 2018, and compare the properties of the resulting BH and DW returns. First, we pick a random portfolio formation date, and then each portfolio consists of 30 stocks randomly selected as of that date. We hold the portfolios for 30 years. When a stock drops out before the end of the investment horizon, we randomly choose a replacement stock available as of that date, i.e., the portfolio always has 30 stocks over the entire investment horizon. Then, we repeat this procedure 1,000 times, generating an empirical distribution of portfolio BH and DW returns. All returns are annualized. Panel A reports the distributions of BH returns and DW returns for the 1,000 portfolios. Panel B presents the Pearson correlations between Distribution and the standard deviation of returns. Panel C includes the differential volatility results for deciles 1 and 10, where the 1,000 portfolios are ranked on the magnitude of their correlation between Distributions and past volatility of returns. Similarly, Panel D provides the differential volatility results for deciles 1 and 10, where the 1,000 portfolios are ranked on the magnitude of their correlation between Distributions and future volatility of returns. Distributions are calculated at the portfolio level. Panels E and F provide results for stratifying the sample on the magnitude of stock volatility, and the magnitude of Distributions. Stock volatility is the volatility of monthly portfolio returns over the life of the portfolio, and magnitude of Distributions is defined as the average absolute value of the Distributions for each portfolio over portfolio life. Results in the table are from Pearson correlations. Results remain similar using Spearman correlations.

Panel A: Descriptive stats of BH Rts and DW Rts

	BH Rt	DW Rt
N	1,000	1,000
Mean	0.112	0.090
Std Deviation	0.019	0.031
1%	0.068	0.008
5%	0.080	0.041
10%	0.089	0.054
25% Q1	0.100	0.075
50% Median	0.111	0.091
75% Q3	0.124	0.109
90%	0.136	0.126
95%	0.144	0.135
99%	0.161	0.154

Panel B: Pearson correlations between Distribution and the Std(Rts)

	Distributions and Std(Past Yr Rts)		**Distributions and Std(Future Yr Rts)**	
	Correlation	P-value	Correlation	P-value
1 Year	0.009	<0.001	-0.002	0.283
3 Years	0.024	<0.001	-0.026	<0.001
5 Years	0.045	<0.001	-0.026	<0.001
7 Years	0.044	<0.001	-0.031	<0.001

Panel C: Stratifying on the correlation of Distributions and Std(Past 3 Yr Rts)

	Decile 1 (Lowest correlation/ most negative correlation)		Decile 10 (Highest)	
	std(BH Rt)	std(DW Rt)	std(BH Rt)	std(DW Rt)
	0.0231	0.0320	0.0161	0.0336
Difference	38%		108%	
P-value of f-test	<0.001		<0.001	

Panel D: Stratifying on the correlation of Distributions and Std(Future 3 Yr Rts)

	Decile 1 (Lowest correlation/ most negative correlation)		Decile 10 (Highest)	
	std(BH Rt)	std(DW Rt)	std(BH Rt)	std(DW Rt)
	0.0207	0.0338	0.0183	0.0253
Difference	64%		38%	
P-value of f-test	<0.001		<0.001	

Panel E: Stratifying on monthly stock volatility

	Decile 1 (Lowest stock volatility)		Decile 10 (Highest stock volatility)	
	std(BH Rt)	std(DW Rt)	std(BH Rt)	std(DW Rt)
	0.0152	0.0240	0.0276	0.0531
Difference	58%		92%	
P-value of f-test	<0.001		<0.001	

Panel F: Stratifying on Avg(abs(Distributions))

	Decile 1 (Lowest Avg(abs(Distributions)))		Decile 10 (Highest Avg(abs(Distributions)))	
	std(BH Rt)	std(DW Rt)	std(BH Rt)	std(DW Rt)
	0.0167	0.0214	0.0231	0.0780
Difference	28%		237%	
P-value of f-test	<0.001		<0.001	