The volatility of investor returns

Ilia D. Dichev
Emory University

Xin Zheng
University of British Columbia

April 17, 2019

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 magnitude of the volatility of investor returns vs. corresponding buy-and-hold stock returns using data from the U.S. and 19 major international markets. Our main finding is that the volatility of investor returns is higher than the corresponding volatility of stock returns in nearly all specifications.

Very Preliminary!

Comments welcome. Please contact:
Ilia D. Dichev
idichev@emory.edu
404-727-9353

The volatility of 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 latter 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 the stocks or portfolios of stocks as a sequences of signed investor capital flows. The essential difference between these two return metrics is that BH returns assign equal weight over each period of the time-series of returns, while dollar-weighted returns weight period returns by beginning invested capital. Thus, dollar-weighted returns more properly reflect the actual investor experience, especially when capital exposure varies over time.

Our tests rely primarily 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 returns across stocks is 0.094 for the BH returns, and 0.139 for the DR returns, which indicates that the volatility of the actual investor experience is nearly 50% higher than the corresponding volatility of stock returns. A closer look at the empirical distributions of BH and DW returns indicates that most of the heightened volatility of DW returns comes from a heavier left tail, which indicates increased downside risk.

We proceed with portfolio tests, where we form random portfolios of 10, 30, and 100 stocks, and track their returns over various investment horizons. The advantage of portfolio tests is that they represent a more realistic representation of actual investor experience. 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 closely mimic the investor experience for both retail and institutional investors. Such portfolios also reflect the diversification of unsystematic risk, which typically tops out for 30 to 50 stocks. 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 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 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 specifications. 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 international data. However, the consideration of long investment horizons and the relatively short time-series of index returns (especially for international data) necessitate some compromises, chiefly the use of overlapping observations, and corresponding problems of statistical testing and interpretation. Overall, the results of the index tests are largely in line with the preceding tests. The statistical significance and the economic magnitude of the results, however, are generally weaker.

Taken as a whole, the pattern of the results provides strong evidence that the volatility of investor returns is higher than the volatility of the corresponding stock returns. Estimates of the magnitude of this difference vary widely, from about 15% to 75%, with larger differences for longer investment horizons. While the full implications of these findings are probably not quite clear at this point, it seems safe to say that the risk-return trade-off of stock investors seems worse than previously thought. For example, combined with previous findings that DW returns are lower than BH returns, the results of this study suggest that "the equity premium puzzle" may be less puzzling after the correct consideration of the actual investor experience. Future research can explore this conjecture more carefully. Other research possibilities include investigating the drivers of the volatility differential, and how this differential varies across stock and market characteristics like size, book/market, the magnitude of capital flows, investor sentiment, and others.

2. Theory and examples

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 on this topic. This difference is best illustrated by an 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 put in $1,000 at the beginning of period 1, which doubled to $2,000, and with the addition of another $1,000 became a total value of $3,000 as of the end of period 1. Since the stock went down 50% in period 2, the investor proceeds from liquidation were only $1,500. Thus, the investor experience must have been negative because the investor put in a total of $2,000, and eventually got only $1,500 out.

Investor returns are measured by dollar-weighted returns, which are calculated as the internal rate of return on the investor capital flows into and out of the investment. In this case, from the point of view of the investor the capital flows are negative $1,000 at the 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.

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

This simple example can be extended to illustrate the driving forces of the difference between buy-and-hold (BH) and dollar-weighted (DW) returns. To trace the effect of the capital flows, initial investment always stays the same at $1,000 but the capital inflow at the end of period 1 is allowed to take the values of $500, $1,000, and $2,000. To trace the effect of the variation in returns, we consider three scenarios: i) stock goes up by 25% in period 1, and down by 20% in period 2, ii) the baseline example above, up by 100%, then down 50%, and iii) stock up 300%, then down 75%.

Notice that in all of these cases the BH return is zero. The DW returns are all different, however. After deriving the liquidating capital outflows from the applicable investment inflows and stock returns, and using the IRR calculation, the results are summarized in the following 3X3 matrix (rounded to the nearest percent).

Capital inflow at the end of period 1	**25% return in period 1 -20% return in period 2**	**100% return in period 1 -50% return in period 2**	**300% return in period 1 -75% return in period 2**
$500	-4%	-10%	-16%
$1,000	-7%	-18%	-28%
$2,000	-10%	-27%	-42%

The results in the matrix flesh out the intuition that DW returns differ from BH returns because of the timing of the capital flows. In all of these cases, the investor has bad timing because he increases his capital exposure after the initial high stock returns, and before the subsequent negative returns. Correspondingly, his DW returns become more negative when the additional invested capital at the end of period 1 is larger, and when the returns affecting this increased capital base are more negative. Another way to state the same conclusion is that this 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. And the investor return becomes lower when capital exposure is higher during the low stock return period, and when the low stock return becomes even lower.

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.

This difference between investor and stock returns has been recognized and studied for the first moment of returns. Dichev (2007) documents that dollar-weighted returns tend to be lower than buy-and-hold returns for stock index returns in the U.S. and 18 international markets. Dollar-weighted 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 dollar-weighted returns are about 1.5% lower than buy-and-hold returns for a broad sample of U.S. mutual funds. Dichev and Yu (2011) document that dollar-weighted returns for a broad sample of hedge funds lag buy-and-hold returns by 3% to 7%. Morningstar has been calculating dollar-weighted returns for a broad cross-section of mutual funds since 2006, and their data show reliable patterns of dollar-weighted returns falling short of buy-and-hold returns by about 1 to 2%.

The principal reason for these results is a propensity to chase returns, i.e., investors tend to pour capital in after superior past returns (Hayley 2014). The main takeaway is that the return experience of investors is worse than the widely published and followed security returns. There are also exceptions, however. Dvorak (2012) finds that defined-benefit pension plans have dollar-weighted returns that are about 1% higher than their corresponding buy-and-hold 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. Thus, the evidence on defined-benefit pension plans confirms that the correlation between capital flows and returns can create a gap between buy-and-hold and dollar-weighted returns.

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. The hypothesized effects are best illustrated by an 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 return but differ in simulated volatility. 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 mean of 10% and standard deviation of 20%. Each stock starts with an investment of $1,000, which evolves according to the pattern of simulated returns. There are no dividends, and there is a single capital contribution of $1,000 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 are substantial capital flows 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. The results from the simulation are presented in Table 1. Panel A of Table 1 provides a validity check by presenting descriptive statistics on the empirical distribution of the simulated stock returns. Since there is a large number of stocks, the parameters of the empirical distribution correspond closely to the modeled values, with means of almost exactly 10% during the two periods, and standard deviations of nearly 10% and 20% for the first and second 10-year periods, respectively.

Panel B of Table 1 provides the main results for the simulation. For each stock, we compute two returns. The buy-and-hold stock return is the geometric average of the yearly stock returns over the entire 20-year life of the stock. The investor return for each stock is computed as the internal rate of return of the stock, which is given by the simulated capital flows. Note that each stock has the same initial market value of $1,000 (negative capital flow), a capital contribution of $1,000 at the end of period 10 (negative capital flow), and a final market value (positive capital flow) that varies across stocks depending on the pattern of drawn returns for this stock. Panel B contains the empirical distributions of the simulated buy-and-hold and dollar-weighted returns over the 1,000 stocks.

An examination of Panel B reveals that the empirical distribution of dollar-weighted returns is more spread-out than the empirical distribution of buy-and-hold returns. The standard deviation of dollar-weighted returns is 4.1% as compared to the standard deviation of buy-and-hold returns of 3.7%, and this difference is highly statistically significant. A comparison of the percentiles of the empirical distributions confirms the summary evidence from the comparison of the standard deviations but also reveals that most of the wider spread in dollar-weighted returns comes from its heavier left tail.

Panel C of Table 1 extends the baseline evidence from Panel B by varying the key parameters of the simulation, and tracing the resulting difference in the realized volatilities of buy-and-hold and dollar-weighted returns. Recall that this difference arises because of two factors, magnitude of the capital flows, and the difference in underlying stock volatility across periods. Accordingly, Panel C models the capital contribution as taking three possible values at the end of period 10: $500, $1,000, and $2,000. 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%. Thus, the combined effect of varying these two key parameters is captured in a 3X3 matrix, where the cells of the matrix represent the difference between the standard deviation of buy-and-hold returns and dollar-weighted returns over the 1,000 stocks in the market. Note that the central cell in the matrix corresponds to the baseline results in Panel B.

The results in Panel C 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 3X3 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 C suggests strong non-linearity in the interaction of the two explanatory factors. 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 effect in the lower rightmost cell is about 100 times stronger than the weakest effect in the top leftmost cell. This pattern suggests that investor timing can have substantial effects on the volatility of investor returns.

3. Empirical results

3.1 Individual stock results

We start our investigation with a battery of tests on U.S. data, covering volatility effects at the individual stock level, for various portfolios reflecting plausible investor strategies, and for broad stock indexes. At the individual stock 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.[2] Specifically, we take all stocks from the major U.S. stock exchanges (NYSE, AMEX, and Nasdaq), and examine the empirical distribution of their buy-and-hold (BH) and dollar-weighted (DW) returns over plausible long-run investor horizons of at least 5, 10, 20, and 30 years.[3] To keep construction and interpretation simple, 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 survivorship effects but that is not much of a problem here because the short-lived stocks are small and otherwise marginal, and so they have little effect on the overall (value-weighted) investor experience we aim to capture. In any case, the portfolio tests later include all available stocks without exception. All returns are annualized.

Buy-and-hold returns are computed as the geometric average of monthly returns. Dollar-weighted returns are computed as the solution to the IRR calculation:

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

which equates ending (discounted) market value to beginning 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:

$$\text{Capital flow}_t = MV_t - MV_{t-1} \times (1+r_t) \qquad (2)$$

---

[2] Buy-and-hold and dollar-weighted returns are by construction 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.

[3] We also examined 40-year horizons but the number of available stocks drops under 1,000, so we do not include them in the table. In any case, the results for the 40-year horizon are similar.

Where r_t is the buy-and-hold return for period t, and the resulting capital flow variable is signed, where positive capital flow signifies capital inflows like stock issues, and negative capital flows signify capital outflows like dividends and stock repurchases.[4]

The individual stock results are presented in Table 2. The table includes the number of stocks available at each horizon, and the mean and standard deviation of the buy-and-hold and dollar-weighted 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. An inspection of Table 2 reveals that dollar-weighted investor returns tend to be lower than the corresponding buy-and-hold stock returns, in line with existing evidence (Dichev 2007, Hayley 2014).

Table 2 also provides evidence about our main interest in this investigation, where at all horizons the dispersion of DW returns is substantially higher than the corresponding dispersion of BH returns. Taking the 20-year horizon as an example, the standard deviation of DW returns is 13.9% vs. 9.4% for BH returns, where the difference is highly statistically significant. 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 distribution of the two return metrics confirms the summary impressions from the standard deviations. The dispersion of returns is higher for dollar-weighted returns, and 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

[4] 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.

across buy-and hold vs. dollar-weighted returns. But for the percentiles below the median the dollar-weighted percentiles quickly plunge below their equivalents for buy-and hold returns, and the difference becomes rather substantial for the extreme percentiles. The 5^{th} percentile (1^{st} percentile) for buy-and-hold returns is -10.7% (-21.2%), while their dollar-weighted counterparts are -16.6% (-56.7%). Overall, the impression is that dollar-weighted returns are considerably more volatile than buy-and-hold returns, and most of this difference comes from "downside risk".

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 buy-and-hold returns vs. 11.6% for dollar-weighted 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.

3.2 Portfolio results

Next, we examine portfolio specifications of BH and DW returns. For these tests, we include all available stocks on the major U.S. exchanges without restrictions on stock longevity. We simulate the investment experience of investors who choose a portfolio 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 that 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 (including the value of stock leaving or entering the portfolio), and 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, Odean, and Zhu 2009; Koestner, Loos, Meyer, and Hackethal 2017), while mutual funds typically hold about 100 stocks (Huang, Sialm, and Zhang 2011; Kacperczyk, Sialm, and Zheng 2009). In any case, there is reliable evidence that the benefits of portfolio diversification top out at portfolios of 20-50 stocks (Elton 2014; Campbell, Lettau, Malkiel, and Xu 2001), so there isn't much benefit to looking at much larger portfolios.

The portfolio results are presented in Table 3, 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 30-year horizons 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 3, there is also evidence of stratification in the right tail as well.

3.3 Broad market index for the U.S.

We also present 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 adjusting for the resulting statistical dependence, and 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 20-year and 40-year investment horizons.

The evidence is presented in Table 4, constructed along the lines of the previous tables, presenting key statistics and percentiles for the empirical distributions of the resulting buy-and-hold and dollar-weighted returns. The mean returns in Table 3 are on the magnitude of 10% percent, consistent with much existing evidence on long-run U.S. returns. Similar to the previous tables, our focus in on the difference between the dispersion of BH and DW returns. 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. 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%. This difference is statistically significant at the 0.01 level, using a F-test which adjusts for the influence of overlapping observations. It is also 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 relative gap between the volatility of buy-and-hold and investor returns tends to grow 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.

3.4. International results

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.

Seeking to maximize the information content of international data, we present two sets of results. The first set uses the full time-series of available data for each country, and presents the standard deviation of buy-and-hold and dollar-weighted returns *across countries*. This evidence essentially answers the question about the variation of investor experience from investing in one country vs. another. An examination of the results in Panel A of Table 5 confirms existing evidence that DW returns tend to be lower than BH returns, with this pattern obtaining in 18 out of the 19 available countries. While the standard deviation of DW returns across countries is higher than the corresponding standard deviation of BH returns, the difference is not statistically significant.

The second set of results looks at the variation of investor experience *within countries*. 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 as for the broad U.S. index in Table 4. Using a fairly typical case as an example, Japan's stock index data starts in 1973. We form a portfolio in 1974, hold it through 1993, and calculate the requisite 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. For parsimony, we only present results on the standard deviations of BH and DW returns in Table 5; untabulated results based on the percentiles of the empirical distributions have the same tenor.

An inspection of the results in Panel B of Table 5 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 20% 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. Overall, international results mostly confirm the tenor of the U.S. results but they are statistically and economically weaker.

4. Conclusion

The volatility of 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 investor returns, and contrast it to the volatility of buy-and-hold stock returns. 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 magnitudes of this differential vary from as little as 10% to as high as 75%, substantially increasing in investment horizon. More work is needed to fully map out the driving forces and the implications of these findings. At this point, however, it seems safe to say that the risk-return trade-off of stock investors seems worse than previously thought.

References:

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

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.

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

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

Dvorak, Tomas, 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.

Hayley, Simon, 2014, Hindsight effects in dollar-weighted returns, *Journal of Financial and Quantitative Analysis* 49 (1), 249-262.

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.

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

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 the life span of 20 years which are split into two period 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 $1,000. We assume there are no dividends and there is a single capital contribution of $1,000 at the end of year 10 (i.e., end of period 1). Panel A presents the distribution of simulated returns for periods 1 and 2. Panel B reports the distributions of buy-and-hold returns and dollar-weighted returns for the 1,000 stocks. Panel C reports the difference in volatilities between buy-and hold and dollar-weighted 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 the simulated returns

	Period 1	Period 2
N	10,000	10,000
Mean	0.101	0.100
Std Deviation	0.102	0.197
100% Max	0.497	0.878
99%	0.338	0.553
95%	0.266	0.428
90%	0.231	0.352
75% Q3	0.169	0.232
50% Median	0.103	0.101
25% Q1	0.031	-0.031
10%	-0.029	-0.158
5%	-0.068	-0.232
1%	-0.134	-0.361
0% Min	-0.275	-0.586

Panel B: Distributions of BH returns and DW returns for the 1000 stocks

	BH Returns	DW Returns
N	1,000	1,000
Mean	0.090	0.088
Std Deviation	**0.037**	**0.041**
100% Max	0.201	0.202
99%	0.179	0.182
95%	0.149	0.152
90%	0.137	0.139
75% Q3	0.115	0.116
50% Median	0.090	0.089
25% Q1	0.065	0.062
10%	0.041	0.034
5%	0.027	0.018
1%	0.007	-0.010
0% Min	-0.047	-0.060
P-value of t-test for the differences in the mean returns	<0.0001	
P-value of F-test for the equality of variances between BH returns and DW returns	**0.000**	

Panel C: Differences in volatility between buy-and-hold returns and dollar-weighted returns by varying parameters

	Differences in standard deviation in the two periods		
	5%	10%	15%
Capital contributions at the end of the 1st period			
$500	0.0001	0.0015	0.0032
$1,000	0.0032	0.0045	0.0071
$1,500	0.0039	0.0066	0.0098

Table 2: Distributions of Buy-and-Hold Returns and Dollar-Weighted Returns at the U.S. Individual Stock Level

This table reports the distributions of buy-and-hold returns and dollar-weighted returns at the individual stock level. 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. 5 years, 10 years, 20 years, and 30 years indicate that stocks are required to have at least 5, 10, 20 or 30 years of consecutive monthly non-missing records in CRSP. All returns are annualized. We use one-tailed F-test to test the difference in standard deviations between buy-and-hold returns and dollar-weighted returns.

	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	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**
F-test 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 3: Portfolio tests for U.S. stock

This table presents the results of portfolio tests of the difference between the volatility of buy-and-hold and dollar-weighted returns. We include all stocks on the major U.S. exchanges 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 use F-test of equality of variances to test the difference in standard deviations between buy-and-hold returns and dollar-weighted returns.

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**
F-test 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**
F-test 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 4: Distributions of Buy-and-Hold Returns and Dollar-Weighted Returns for 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 use F-test of equality of variances to test the difference in standard deviations between buy-and-hold returns and dollar-weighted returns.

Panel A: distribution of BH returns and DW returns for 10, 20, and 30-year portfolio

	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
Mean	0.099	0.102	0.108	0.104	0.110	0.107	0.108	0.105
Std Deviation	0.057	0.052	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 of variance equality F test	**0.185**		**0.0994**		**0.0086**		**0.0083**	

Table 5: International Results

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. Panel A presents country level results, where we calculate buy-and-hold and dollar-weighted returns for each country, and then compare the volatility of investor and stock returns across countries. . Panel B presents country-level portfolio results. 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.

Panel A: Country level BH and DW Returns

Country	Start Year	(2) BH Rts	(3) DW Rts	(4): (3)-(2)
Australia	1973	0.0697	0.0559	-0.0138
Austria	1973	0.0583	0.0431	-0.0152
Belgium	1973	0.0624	0.0581	-0.0043
Canada	1974	0.0689	0.0702	0.0013
Denmark	1974	0.1022	0.1018	-0.0004
France	1973	0.0786	0.0525	-0.0261
Germany	1973	0.0563	0.0539	-0.0024
Hong Kong	1973	0.0899	0.0700	-0.0199
Ireland	1973	0.0835	0.0515	-0.0320
Italy	1973	0.0696	0.0171	-0.0524
Japan	1973	0.0392	0.0273	-0.0120
Netherlands	1973	0.0611	0.0566	-0.0045
Norway	1981	0.0939	0.0772	-0.0167
Singapore	1973	0.0383	0.0344	-0.0039
South Africa	1973	0.1388	0.1384	-0.0004
Sweden	1983	0.1060	0.0937	-0.0123
Switzerland	1973	0.0580	0.0533	-0.0047
U.K.	1965	0.0795	0.0571	-0.0224
U.S.	1974	0.0806	0.0765	-0.0041
Std Dev		0.0241	0.0277	
F stat		1.3157		
P value of F stat		0.2833		

Panel B: 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**