

# The Perils of Relying on Return Data When Testing Asset Pricing Models

John F. Pinfold

Vestec Limited, Auckland, New Zealand

Email: j.f.pinfold@extra.co.nz

**How to cite this paper:** Pinfold, J.F. (2022) The Perils of Relying on Return Data When Testing Asset Pricing Models. *Journal of Mathematical Finance*, 12, 71-83.  
<https://doi.org/10.4236/jmf.2022.121004>

**Received:** November 29, 2021

**Accepted:** January 17, 2022

**Published:** January 20, 2022

Copyright © 2022 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

---

## Abstract

Asset pricing models are almost always tested using stock returns over multiple time periods, and the returns of portfolios over the investment horizon determined using the arithmetic average of these portfolio returns. The arithmetic average returns of portfolios selected using the model's parameters are calculated and compared. However, investors' returns are derived from changes in the value of their portfolios. This paper shows how the use of arithmetic returns creates large biases in the magnitude and statistical significance of asset pricing models' outcomes. It argues only evaluations using the values of portfolios produce reliable results. The identified bias is created because a positive return and its equal but negative return, represent different sized price movements, and this becomes obscured when returns are analysed and averaged over multiple periods. Most existing pricing models are potentially invalid because of the biases generated by the methodology used in their development.

## Keywords

Arithmetic, Geometric, Asset, Pricing, Model

---

## 1. Introduction

The problems associated with using arithmetic returns in place of geometric returns when calculating stock returns and portfolio returns over multiple periods are well known. The problem of using arithmetic averages for portfolio performance first came to prominence in the venture capital industry where poorly performing fund managers were blatantly deceiving potential investors about their past performance when raising new funds, as detailed in Phalippou [1]. Funds did this by expressing performance as the average arithmetic annual return of the fund rather than the actual return of the fund over the entire invest-

ment period. As venture capital funds tend to suffer losses early and make their gains as investments mature, the effect can be dramatic with loss making funds appearing to be highly profitable. This problem has also been extensively discussed for other types of funds.

The problem of using arithmetic returns in testing asset pricing models was demonstrated in Roll [2] when he evaluated the size effect. He showed how both the magnitude of the size effect and its statistical significance were overstated when arithmetic averages were used in place of geometric averages to consolidate returns over time. The shorter the buy and hold period, and the longer the period sampled, the greater the magnitude of the effect.

Despite this, almost all research on asset pricing models and stock performance uses arithmetic stock returns to evaluate the returns of an investment strategy. This seems to be based on the idea that while the magnitude of the effect may be overstated, the differences in portfolio performance of the strategies are real, and the statistical significance tests are valid. The ambivalence towards substituting arithmetic returns for geometric returns has also been adopted by practitioners. A good example is when using CAPM to determine a company's cost of capital. Brunner *et al.* [3] found practitioners prefer to use arithmetic averages to calculate the market risk premium from historical data rather than the geometric return which is what investors receive. While this is theoretically incorrect it is deemed to provide more acceptable results.

The two most obvious dimensions of the problem are typically ignored or overlooked. Firstly, if stock returns are averaged to give a portfolio return for a single buy and hold period, the average stock return is not the portfolio return over the period. The second dimension is the bias generated when using returns of portfolios in place of their change in value over multiple periods. **Table 1** has been prepared to give a simple numerical example which shows the errors in returns generated in a five-stock portfolio over four monthly periods. The typical methods used in evaluating asset price models are subject to return errors generated over the total holding period in both dimensions. **Table 1** portfolio is worth \$500 at the start and at the end of the four-month period. Using arithmetic averages, the return is calculated to be 0.99 percent instead of zero percent, which is the return the investor receives. When the true value of each portfolio, each period, is used to calculate the average portfolio return over the four months, the return still comes out at 0.31 percent.

**Table 1** shows two dimensions of the problems, that is the erroneous return of the portfolio, and the error created using average arithmetic portfolio returns to measure returns over multiple periods. There is however a third dimension to the problem not readily discernible from **Table 1**. While it is easy to see percentage returns are a biased representation of price changes, it is less obvious that this bias is proportional to the volatility of the returns of each stock. If a stock price moves from \$100 to \$125 and then back to \$100, its returns are +25 percent and -20 percent giving an average return of 2.5 percent, but a real return of zero. The greater the price movement, the greater the error generated.

**Table 1.** Pricing anomalies caused by using arithmetic returns.

	Start	Month 1		Month 2		Month 3		Month 4		Arithmetic Average Return
	Price	Price	Return	Price	Return	Price	Return	Price	Return	
Stock 1	\$100	\$90	-10.00%	\$105	16.67%	\$120	14.29%	\$100	-16.67%	<b>0.86%</b>
Stock 2	\$100	\$115	15.00%	\$120	4.35%	\$99	-17.50%	\$100	1.01%	<b>0.57%</b>
Stock 3	\$100	\$95	-5.00%	\$90	-5.26%	\$84	-6.67%	\$100	19.05%	<b>0.42%</b>
Stock 4	\$100	\$90	-10.00%	\$105	16.67%	\$110	4.76%	\$100	-9.09%	<b>0.47%</b>
Stock 5	\$100	\$75	-25.00%	\$109	45.33%	\$93	-14.68%	\$100	7.53%	<b>2.64%</b>
Portfolio	\$500	\$465		\$529		\$506		\$500		
Average Stock Return			-7.00%		15.55%		-3.96%		0.37%	<b>0.99%</b>
Arithmetic Portfolio Return			-7.00%		13.76%		-4.35%		-1.19%	<b>0.31%</b>
Geometric Portfolio Return										<b>0.00%</b>

Everyone knows individual stocks are more volatile than the market as a whole and investors use diversification to take advantage of this fact. As individual stock prices are much more volatile than movements in the stock index, not only actual returns, but returns relative to the market, *i.e.*, excess returns, inevitably follow the underlying pattern of positive and negative returns. This inevitably generates a bias when comparing stock returns with market returns. As stocks have different overall price volatilities, and these volatilities vary over time, the bias created differs from stock to stock and time to time. A volatile and unpredictable bias inevitably reduces the reliability of any estimate of the magnitude of any given price effect and impugns the statistical validity of any tests done to establish statistical validity. In addition, if price movements are normally distributed, returns will inevitably be skewed in the positive direction, creating problems with statistical tests, which may not be properly rectified by the methods adopted.

This explanation of the problem does not persuasively establish the nature and magnitude of the problem, so a simulation will be used to show how random price movements can generate errors which have the potential to accumulate and invalidate a pricing model's claims as to the economic and statistical significance. The simulation will start with stocks all having the same price. Each stock's price varies randomly from period to period. As the prices are purely random, it should be impossible to find a non-random price effect in the data. The use of stock return data instead of prices can allow researchers to find statistically significant relationships in random data and this will be demonstrated.

## 2. Testing a Random Simulation

The nature and magnitude of the problem can be established by analysing simulated stock prices with known characteristics and determining the distortion of results created by typical analysis techniques. CAPM was chosen as it forms the backbone of modern finance and is well entrenched in economic theory. It is an

ideal candidate as CAPM beta is a measure of volatility relative to an index and it will therefore be sensitive to errors resulting from volatility differences.

A set of random returns with similar characteristics to stock prices was generated for testing. Monthly prices of 330 stocks over a 20-year period were randomly generated. To make the prices representative of price movements of real stocks, the characteristics of 330 S&P 500 continuously listed stocks over the 20-year period ending 31 Dec 2020 were used as a reference. An equal weighted monthly index was calculated for the S&P 500 reference stocks. Each of the 330 simulated stocks was assigned a price volatility representative of a stock found in the S&P 500. The first stock in the simulation was assigned a volatility of 5.0 percent with each subsequent stock's volatility increasing at the constant amount needed to give the 330<sup>th</sup> stock a 20.0 percent volatility.

The Excel random number generator was used to randomly generate price changes to give each stock the assigned volatility. The random price adjustments were generated to produce a normal price distribution for each stock when calculated over the 240-month test period. In other words, each stock price, in each period, was set at \$100 and a random normally distributed error averaging \$0.00 was used to adjust the price so the resulting price series for the stock achieved its assigned price volatility. Each period the price generated for each stock was scaled to match the movement in the S&P 500 during the reference period. Each stock's randomly generated price was multiplied by the equal weighted index value at the date in question, divided by the index value on 31 Dec 1999. Each stock has the same intrinsic price each period, plus or minus its random adjustment. As the price movements are randomly generated, the difference in the price of any two stocks in any given period is purely random. It would of course have been possible to simply use unadjusted random movements in price so, on average, the stocks had the same price at end of the 20-year period as the beginning, but by scaling prices to match the movement of a real world index the prices better represent actual stock prices.

**Table 2** compares some key characteristics of the simulated stocks with the S&P 500 it is designed to mimic. The main differences are due to the different assumptions used, and the random variations inherent in a randomised simulation. In the simulation any differences in the geometric average return over the 240-month period are random because the intrinsic price of all stock is identical each period before the random variation is applied. S&P 500 stocks have varied returns, and superior performance is naturally associated with both geometric and arithmetic returns being higher than the average, and thus are, to a degree, correlated.

The price volatility of a simulated stock is the same over the entire period but varies over time in the S&P 500, leading to different price volatility averages. It is important to recognise it is inevitable the simulation will vary from the S&P 500 on which it is based, as the actual stocks have varying volatilities over time, whereas the simulated stocks have the same volatility, subject to random fluctuations, over the entire 20-year period. The simulation resembles real stock prices

**Table 2.** Comparison between simulation and S&P 500.

	S&P500 2000-2020		Simulation - 240 months	
	Average	Range	Average	Range
Index geometric return (%)	0.351	-	0.352	-
Index average arithmetic return (%)	0.42	-	0.47	-
Stock geometric returns (%)	0.743	-1.03 - 2.21	0.353	0.167 - 0.545
Stock arithmetic returns (%)	1.21	0.345 - 1.210	1.10	0.465 - 2.39
Price Volatility (%)	9.85	4.79 - 29.31	10.98	5.33 - 27.0
Correlation of geo. ret. & arith. ret	0.750		0.037	

which are commonly evaluated using pricing models, but all price changes are random, and any valid pricing model should generate zero excess returns. Any variation from zero should be clearly shown by tests to be statistically insignificant.

**Table 2** shows the simulation has return variation and volatility distribution of a similar magnitude to those found in actual stocks. The returns of the two indexes are similar when measured as arithmetic or geometric returns. The greatest difference between the two sets of data is the geometric return of the stock (0.743 vs 0.353). This arises because any difference between the geometric return of a stock and the index in the simulation is created by random chance, whereas the long term returns of individual real-world stocks are often markedly different from the index. Some S&P500 stocks outperformed the market by a large amount over the 20 year period and others badly underperformed the market. The expected geometric return of all simulated stocks is the market return, and any variation is due to random fluctuations which were set to be proportional to the assigned volatility.

The best test of validity of performance claims for a stock pricing model is to calculate the returns obtained by an investor holding the chosen portfolios. This involves valuing the stocks in each portfolio and monitoring the change in value over time as the portfolios are rebalanced. This can then be compared with the predictions of the model. If this is done using stocks in the simulation, this should generate a zero return over the 240-month period irrespective of the selection technique adopted.

Our test model is the Capital Asset Pricing Model, given in Equation (1). The simulation follows the return pattern of the S&P 500 and risk-free bond prices for corresponding periods are therefore available. In this case the 20 yr US Government bond was arbitrarily chosen as it matches the 20 yrs of data. Data for the 11 months before the start of the test period was collected or generated, and then CAPM betas calculated from the current month and trailing 11 months of returns. Equation (2) was used to calculate the beta used as the independent variable in our model.

$$E(R_i) = R_f + \beta_i (E(R_m) - R_f) \quad (1)$$

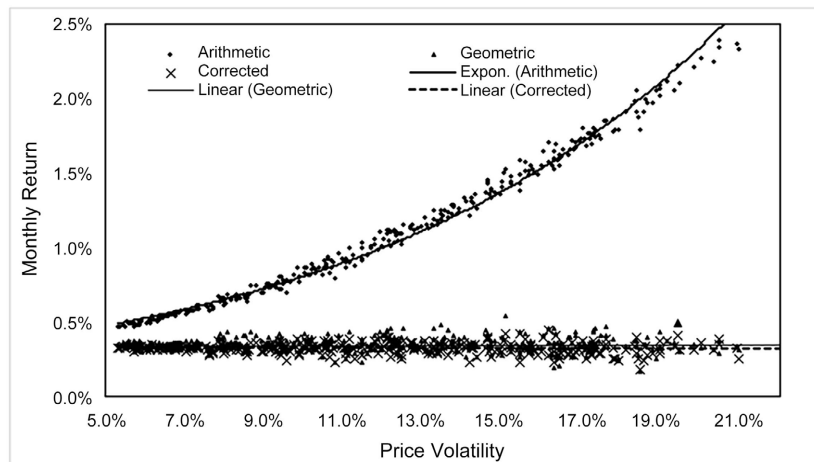
$$\beta_i = \frac{Cov(R_i, R_m)}{Var(R_m)} \tag{2}$$

We have reason to suspect this model could generate abnormal returns because of the relationship between volatility and beta, and the difference between a stock’s arithmetic and geometric returns. That is, the relationships found in the data presented in **Figure 1**. A data series has also generated using the arithmetic return data in **Figure 1** after correcting it using Equation (3). This will be used to show the relationship between arithmetic return and price volatility is created by using arithmetic returns.

$$\text{Corrected return (\%)} = \frac{P_{t+1} - P_t}{\frac{P_{t+1} + P_t}{2}} \times \frac{100}{1} \tag{3}$$

This equation calculates the percentage return from the average of the starting and finishing prices, rather than the starting price. If the price moved from \$100 to \$125 and back to \$100 it would have percentage movements of +25 percent and -20 percent. Using corrected returns, the movements would be +22.5 percent and -22.5 percent, which added together given an overall movement of 0 percent instead of the erroneous arithmetic average of +2.5 percent.

**Figure 1** shows the relationship between return and price volatility found in the simulated data. Each point on the graph represents one stock’s average returns over the 240 monthly intervals. The average arithmetic return over the 240 months increases steadily as the price volatility of the stock increases. The investor’s actual return, the geometric monthly return, does not change as volatility increases. The geometric return is calculated from a stock’s beginning and ending price, and any variation between stocks is random. Additionally, Equation (3) has been used to correct the arithmetic averages, to generate an estimation of the geometric averages. The accuracy of this approximation can be seen by comparing the positions of each cross and its corresponding triangle, which is slightly above or below it.



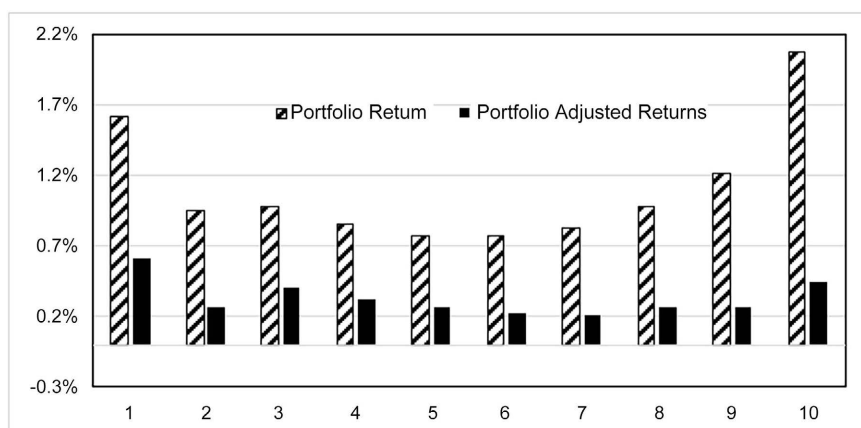
**Figure 1.** Relationship between volatility and return.

Before launching into statistical tests of the relationship, between CAPM beta and return, for the simulated data, it is useful to see how the data has been transformed. This is presented in **Figure 2** and **Figure 3**. The stocks have been sorted each period into 10 portfolios of 33 stocks according to their beta for the period. One is the lowest 10 percent of betas and 10 the highest 10 percent.

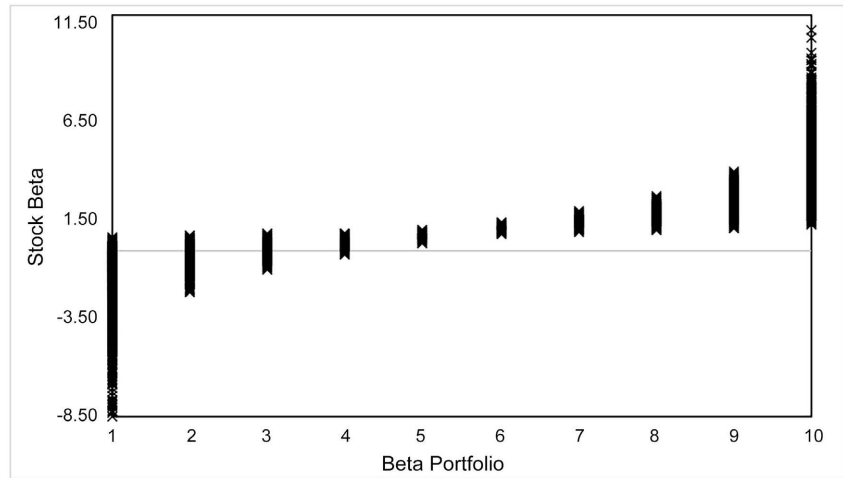
Two things are immediately obvious. Firstly, the difference between the average monthly arithmetic returns of the lowest and highest portfolios is more than one percent per month. Secondly the relationship is not a linear one. Instead, we see a concave curve. This makes it obvious that a conventional linear regression test is not suitable for determining the statistical significance of the differences. **Figure 3** gives us an insight how this pattern of returns came about. Portfolios 1 and 10 contain the outlying negative and positive outliers from the beta calculation, and as the tails in the returns are from a normal distribution, they are thin. The discrepancy between arithmetic and geometric returns is greater when there is a higher the percentage price movement, and this becomes exaggerated as the tails of the normal distribution thin out.

Also, in **Figure 2** the returns for each beta portfolio have been calculated using Equation (3) and included as a comparison. This correction goes a long way towards eliminating the exaggerated return tails created when arithmetic averages are used.

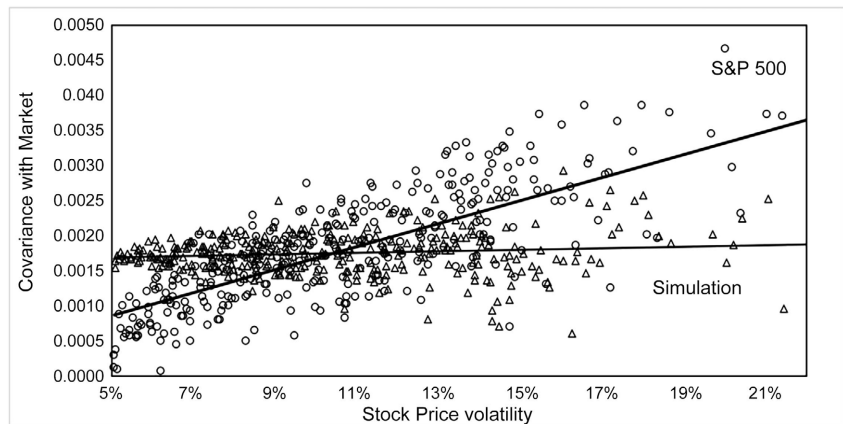
The beta calculation given in Equation (2) uses the covariance of the stock return with the market return to determine beta. The differences in covariances between the simulation and the S&P 500 are presented in **Figure 4**. Regression lines have been added to help interpretation. The S&P500 stocks, upon which the simulation is based, show the covariance of a stock with the market increases with the price volatility of the stock. The simulated stocks show no such relationship as price variation is purely random. For this reason, we would not expect the relationship between beta and return to be the same for the simulation as the actual stock market. The uncharacteristic relationship between beta and return found in **Figure 2** must, a least in part, be due to this difference.



**Figure 2.** Comparison of arithmetic and adjusted arithmetic returns of beta decile portfolios.



**Figure 3.** Simulated beta distribution across decile portfolios.



**Figure 4.** Relationship between stock covariance and price volatility.

It is important to notice how the covariance of S&P500 stocks increases with increases in stock price volatility. This means stocks with high volatilities will tend to have higher betas. This will lead to a relationship where the stocks which are the most overpriced because of using arithmetic returns will tend to have higher betas, exactly the relationship CAPM predicts.

### 3. Are the Relationships in the Simulation Statistically Significant?

The existence of a relationship proves nothing. It is only if the relationship is statistically significant that we can draw valid conclusions. Before proceeding to statistical testing, it is worth discussing the methodologies typically used to evaluate CAPM. The existence of a relationship between beta and return can be evaluated at three distinct levels. Firstly, a single regression on the entire data set can be conducted on the raw data with return as the dependent variable and beta as the independent variable. The return used can either be for the same period as the beta calculation, or for the following period to determine its value at predicting future returns. Traditionally, asset pricing models fail to find relation-



ships which are economically and statistically significant when tested in this way. It is more common to form portfolios to enhance the relationship and reduce the random noise from the sample before analysing the predictive ability of the independent variable. Testing portfolios in this way was first used to evaluate CAPM in Black, Jensen and Scholes [4] and involved sorting stocks into 10 portfolios on the basis on their beta calculated using 5 years of monthly returns. This process removes the noise in the data by diversifying away the effect of other factors affecting return, leaving only the effect of beta in the data. This is only valid if beta is independent of all the other factors affecting return. If beta is not independent, sorting by beta can provide information associated with other factors affecting return. **Figure 4** shows the relationship between price volatility and beta in S&P500 stocks so, knowing the relationship between the over statement of arithmetic returns and price volatility shown in **Figure 1**, it is fair to conclude for CAPM, the process of sorting into decile portfolios should produce an overstatement of both the economic and statistical significance of the model.

**Figure 4** shows this is not a problem for our random simulation of stock prices. This means testing decile portfolios is still statistically valid.

The relationships represented by asset pricing models seldom produce consistent results for individual investment rebalancing periods. They require multiple investment periods to produce statistically significant results, even when stocks have been sorted into portfolios. The returns over multiple investment periods are assessed by comparing average arithmetic returns over the investment horizon tested. This introduces a further layer of potential overestimation of returns due to using multi period arithmetic averages instead of geometric averages. Such overestimations are subject to bias, particularly when there is a relationship between the chosen independent variable or variables and return volatility resulting from the relationship.

For our tests, each month the 330 simulation stocks were sorted into 10 portfolios according to beta. Beta was calculated over the 12-month period ending in the return period. The market return was the calculated using an equal weighted price index of the simulated stocks. The risk-free rate used was the 20-year US Government bond for the corresponding period of the S&P 500 reference portfolio. As the results are not expected to yield significant results from linear regression, regression against dummy variables has been used. For each of the 79201 data points, a beta portfolio was determined for the period in question. The data was then divided into two groups. One of these groups was assigned the dummy variable 1 and the other the dummy variable 0, according to which beta group they occupied. Regressions were then conducted using the dummy variable as the independent variable.

In **Table 3** we can see the returns generated by Portfolios 1 and 10 have a positive coefficient (are higher) than the returns of any of the other portfolios. This difference is statistically significant at the 0.1 percent level. When Portfolios 1 and 10 are combined, the difference is even more statistically significant. The

**Table 3.** Regression relationships found in the random simulation of CAPM.

Dependent variable	Independent variable dummies assigned		Coeff	t-stat	Adj R <sup>2</sup>
	Dummy Variable 1	Dummy variable 0			
$R_i$	Portfolio 1	Portfolios 2-10	0.0057	3.817***	0.0001
$R_i$	Portfolio 10	Portfolios 1-9	0.0107	7.148***	0.0006
$R_i$	Portfolios 1&10	Portfolios 2-9	0.0093	8.221***	0.0008
<i>Adjusted R<sub>i</sub></i>	Portfolio 1	Portfolios 2-10	0.0026	1.901*	0.0000
<i>Adjusted R<sub>i</sub></i>	Portfolio 10	Portfolios 1-9	0.0011	0.771	0.0000
<i>Adjusted R<sub>i</sub></i>	Portfolios 1&10	Portfolios 2-9	0.0012	1.133	0.0000

\*and \*\*\*significant at the 10.0% and 1.0% levels respectively.

Adjusted R<sup>2</sup>s show the differences are small in economic terms. Despite the high degree of statistical significance, we know the prices are random and hence cannot be statistically different. To show the effect is caused using arithmetic returns instead of geometric returns, the returns were adjusted using Equation 3 and the regressions repeated. Now, only Portfolio 1 shows any statistical significance, and this is only at the 10 percent level, which could easily have occurred by chance. This shows the statistical significance arises from the bias created by using portfolio arithmetic averages instead of analysing the changes in value of the stocks comprising the portfolios.

#### 4. Discussion

It is a mathematical certainty that the arithmetic average return of a portfolio is greater than the percentage change in the value of the stocks held in a portfolio over any given period, except when all the returns are equal. Similarly, the arithmetic average return of a given portfolio over time is greater than the actual return received by investors over multiple periods. In this paper it has been shown there is an additional problem. The statistical significance of the return results from an investment strategy cannot be relied upon because the price anomaly created when using arithmetic returns may easily result in a hidden variable with high statistical significance.

This hidden variable arises because the arithmetic return anomaly is proportional to the volatility of a stock in any given period. Volatility varies between stocks, and if, over time, it is correlated to any of the independent variables in the model, it will contribute to the reported statistical significance of the model's predictions.

The correlation between volatility and CAPM beta is obvious. Would we expect this correlation in other asset pricing models? There are good theoretical reasons for assuming so. The most fundamental ideas of modern finance are the concepts of efficient markets and the relationship between risk and return. The value of a stock is the present value of its future cash flows. If markets are effi-

cient, market participants will seek out all publicly available information and use it to estimate the effect it has on future cash flows and revise price estimates accordingly. Price volatility will therefore be a measure of the uncertainty of future cash flows for the stock in question. The more estimates of future cash flows change the greater the price volatility. The uncertainty of the cash flows is converted into risk, that is price volatility. There are two main measures of risk used in assessing stocks, CAPM beta, used in estimating a company's cost of capital, and the Sharpe ratio which is used more generally. The Sharpe ratio is of course excess return divided by the standard deviation of excess returns. Thus, risk is predominately measured in using some form of price volatility.

If investors are risk averse, as we generally assume they are, they will require a higher return in exchange for accepting higher risk. Any asset pricing model used to predict stock performance will, of necessity, be incorporating risk into its pricing factors. If markets are efficient, the investment strategy it recommends will involve detecting stocks with high risk and diversifying away this risk by forming a suitable portfolio of stocks. Market efficiency also predicts that once sufficient participants use the model, the excess returns will be driven down by competition, as markets are assumed not to reward an investor for assuming a diversifiable risk. Nevertheless, it is reasonable to assume that any asset pricing model predicting stock returns will incorporate risk, and hence stock price volatility.

If this is true, every asset pricing model derived by analysing returns over multiple periods must be considered suspect. Until proven otherwise, the statistical significance, and the magnitude of the price advantage it is claimed to impart, must be questioned.

The simplest way of doing this is to rerun the analysis, substituting returns adjusted using Equation (3), and find the effect this has on the result. Where multiple regression is used, the difference between the return and the adjusted return can be included in the regression as an additional factor to see if it is statistically significant, and how it affects the results of the other factors in the model. Equation (3) is, however, just an estimate, even though it appears to be adequate for the task. The true test is to analyse the model by forming portfolios and adding up the value of the stocks each period then rebalancing the portfolio and determining the next periods return in the same way. At the end of the evaluation period the geometric return of each test portfolio over time can be calculated. The reference standard for excess return is normally an index, and the geometric return of the index is easily calculated from its start and end value.

Most financial research is done using SAS or a similar statistics package. Commercial statistics packages have developed to serve demand, and in the case of financial analysis demand is for analysing returns. They are not conducive to the easy analysis of portfolios of assets. In addition, when returns are analysed at the portfolio level the number of data points are greatly reduced and it is more difficult to establish statistical significance.

The difficulty of doing the analysis should not be accepted as an excuse for ignoring it. Every asset pricing model which claims to outperform its reference index should be duplicated by forming actual portfolios of the stocks in the original testing of the model and tracking the value of this portfolio as it rebalances over time. If the change in value of the test portfolio over the evaluation period does not exceed the market index, the model has failed. Testing it using the Equation (3) adjusted returns is a quick way of screening the model for problems of overstatement of excess returns and over statement of statistical significance, but only using portfolios will be definitive.

Without running such tests, it is hard to estimate how large a problem exists. Explanations of how asset pricing models produce the results they do are often in short supply in the papers proposing their use. If markets are efficient, competition between investors should quickly drive profits to the point where they equal transaction costs. If this does not happen there is a convenient explanation, the excess return is a reward for additional risk. There is an alternative explanation. The excess profits are imaginary. This explanation is consistent with the difficulties active investment managers have in beating the market return, a phenomenon which has been researched in detail. Investment managers are not averse to taking on additional risk in order to produce superior results. Why do they leave the extra money sitting on the table?

Reschenhofer *et al.* [5] investigate the prediction performance of pricing models and conclude, “apparently good forecasting performance does not translate into profitability once realistic transaction costs and the effect of data snooping are taken into account.” However, others have replicated studies and found models retain profitability. Mclean and Pontiff [6] looked at 72 characteristics providing return predictability and estimated, on average, returns dropped 32% due to publication informed trading. Most of the studies used Fama-MacBeth [7] slope coefficients or long short portfolio returns. Both methods are based on regression of stock returns, hence are subject to the overstatement bias inherent in using arithmetic returns.

There will not be ready acceptance of the proposition that using arithmetic returns when testing pricing models leads to overstatement of both returns and statistical significance, as this has far reaching implications. Most financial researchers have a large investment of time and reputation in the studies they have produced. However, this paper must surely have made a compelling enough case to raise doubts in the minds of many researchers.

There are of course major limitations to the conclusions which can be drawn from this research as it does not actually test pricing models. While the difference in portfolio returns calculated using average arithmetic returns instead of actual portfolio values is mathematically certain, the overstatements of statistical significance, and the performance of model portfolios relative to market returns are not. The statistical anomalies presented in the paper arise from the relationship between price volatility and the overstatement of returns. The volatility

characteristics of any given model's parameters will differ from the simulation and may not lead to unreliable calculations of statistical significance. It does however provide a question needing an answer. Every model which relies on averaging stock returns to calculate a portfolio return or averages sub-period returns to determine long term performance is worthy of further investigation.

### Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

### References

- [1] Phalippou, L. (2008) The Hazards of Using IRR to Measure Performance: The Case of Private Equity. *SSRN Electronic Journal*, **12**, 55-66. <https://doi.org/10.2139/ssrn.1111796>
- [2] Roll, R. (1983) On Computing Mean Returns and the Small Firm Premium. *Journal of Financial Economics*, **12**, 371-386. [https://doi.org/10.1016/0304-405X\(83\)90055-7](https://doi.org/10.1016/0304-405X(83)90055-7)
- [3] Brunner, R.F., Eades, K.M, Harris, R.S. and Higgins, R.C. (1998) Best Practices in Estimating the Cost of Capital: Survey and Synthesis. *Finance Practice and Education*, **8**, 13-28.
- [4] Black, F., Jensen, M.C. and Scholes, M. (1972) The Capital Asset Pricing Model: Some Empirical Tests. In: Jensen, M.C., Ed., *Studies in the Theory of Capital Markets*, Praeger, New York.
- [5] Reschenhofer, E., Mangat, M.K., Zwatz, C. and Guzmics, S. (2020) Evaluation of Current Research on Stock Return Predictability. *Journal of Forecasting*, **39**, 334-351. <https://doi.org/10.1002/for.2629>
- [6] McLean, R.D. and Pontiff, J. (2016) Does Academic Research Destroy Stock Return Predictability? *The Journal of Finance*, **71**, 5-32. <https://doi.org/10.1111/jofi.12365>
- [7] Fama, E.F. and MacBeth, J.D. (1973) Risk, Return, and Equilibrium: Empirical Tests. *Journal of Political Economy*, **81**, 607-636. <https://doi.org/10.1086/260061>