Buzzacchi, Luigi; Ghezzi, Luca

Article

The odds of profitable market timing

Journal of Risk and Financial Management

Provided in Cooperation with:
MDPI – Multidisciplinary Digital Publishing Institute, Basel

Suggested Citation: Buzzacchi, Luigi; Ghezzi, Luca (2021) : The odds of profitable market timing, Journal of Risk and Financial Management, ISSN 1911-8074, MDPI, Basel, Vol. 14, Iss. 6, pp. 1-14, https://doi.org/10.3390/jrfm14060250

This Version is available at:
https://hdl.handle.net/10419/239666

Journal of
Risk and Financial Management

MDPI

Article

The Odds of Profitable Market Timing

Luigi Buzzacchi [1,*] **and Luca Ghezzi** [2]

[1] Interuniversity Department of Regional and Urban Studies and Planning & FULL, Politecnico di Torino, Viale Mattioli 39, 10125 Torino, Italy
[2] Department of Integrated Business Management, Università Carlo Cattaneo, Corso Matteotti 22, 21053 Castellanza, VA, Italy; lghezzi@liuc.it
* Correspondence: luigi.buzzacchi@polito.it

Abstract: This statistical study refines and updates Sharpe's empirical paper (1975, Financial Analysts Journal) on switching between US common stocks and cash equivalents. According to the original conclusion, profitable market timing relies on a representative portfolio manager who can correctly forecast the next year at least 7 times out of 10. Four changes are made to the original setting. The new data set begins and ends with similar price-earnings ratios; a more accurate approximation of commissions is given; the rationality of assumptions is examined; a prospective and basic Monte Carlo analysis is carried out so as to consider the heterogeneous performance of a number of portfolio managers with the same forecasting accuracy. Although the first three changes improve retrospectively the odds of profitable market timing, the original conclusion is corroborated once more.

Keywords: market timing; commission; forecasting accuracy; Monte Carlo analysis

Citation: Buzzacchi, Luigi, and Luca Ghezzi. 2021. The Odds of Profitable Market Timing. *Journal of Risk and Financial Management* 14: 250. https://doi.org/10.3390/jrfm14060250

Academic Editor: Shigeyuki Hamori

Received: 14 April 2021
Accepted: 1 June 2021
Published: 4 June 2021

1. Introduction

Since 2008, exchange traded funds have experienced staggering growth, especially in the US. As concluded by Pedersen (2018), even if the proportion of passive ETFs continues to grow, actively managed funds are here to stay to make prices more informative and markets more efficient. Since competition among actively managed funds is likely to be fierce, it deserves an up-to-date examination.

Active portfolio management is an investment process that includes feedback and feedforward loops. Its constituent stages are summarised by Damodaran (2012, chp. 1) and thoroughly presented by Maginn et al. (2007) and Gibson (2008, chp. 13).

A strategic asset allocation is a set of target weights that meets long-term financial goals; it goes along with a set of feasible deviations. Suppose that only two asset classes are considered, namely stocks and bonds. By way of example, let each target weight be 50% and each feasible deviation from a target weight be $\pm 10\%$. Both tactical asset allocation and security selection may be performed. Roughly speaking, tactical asset allocation is constrained market timing in line with short term expectations. According to the distinction drawn in Henriksson and Merton (1981), successful market timing rests on macroforecasting skills, whereas a successful security selection rests on microforecasting skills.

More generally, the guidelines for a contrarian tactical asset allocation are laid down in Sharpe et al. (2007). According to the Berkshire Berkshire Hathaway (1991) Annual Report, the definitive guidelines of John Maynard Keynes on discretionary (rather than systematic) portfolio management include: balanced concentration rather than diversification, only fundamentals in a medium/long term perspective, and low (and hence slow) turnover. Intuitive and systematic approaches are contrasted in Jackson (2003, chp. 11).

The challenge posed by market timing is analyzed in depth by Sharpe's seminal paper (1975). A peculiar yet insightful case is considered, where shifting from stocks to cash equivalents or vice versa can occur once a year at most. Use is made of historical data.

Stocks outperform cash equivalents in good years, whereas cash equivalents outperform stocks in bad years. According to the paper's conclusion, considerable forecasting accuracy is needed for market timing to be profitable. More precisely, portfolio managers need to distinguish in advance a good year from a bad one at least 7 times out of 10. A summary of this paper is given in Section 2 below.

Our study attempts to refine his original analysis so as to update the empirical findings on the odds of profitable market timing. In view of this:

- As advocated by Bernstein (1997), our data set begins and ends with similar price-earnings ratios;
- A more accurate approximation of commissions is given;
- A lower bound is placed on forecasting accuracy by examining the rationality of all assumptions.

These changes imply that much less forecasting accuracy is needed for market timing to be profitable. However, attention is paid only to the representative portfolio manager, who is neither consistently luckier nor consistently unluckier than all other managers who have the same forecasting accuracy. Therefore, a number of managers is also considered, all having the same forecasting accuracy. Individual performances are heterogeneous, since there are many different sequences of forecasting errors. A basic yet prospective Monte Carlo analysis is performed so as to obtain the range of performances that corresponds to each selected individual's forecasting accuracy.

Although allowance is made for heterogeneous decision making, the original conclusion is corroborated once more.

The remainder of our study is organized as follows: the relevant literature is reviewed in Section 2, the data set and conceptual framework are presented in Section 3, statistical findings are reported in Section 4, and conclusions are provided in Section 5.

2. Literature Review

Damodaran (2012, chp. 12) presents a review of the scientific literature on market timing that includes both a classification of the main approaches and a summary of the empirical evidence. According to the classification, market timing approaches can be based on:

- Several nonfinancial indicators, which may find a link between market performance and market mood. For instance, the US stock market has displayed a negative correlation with the hemlines on women's skirts;
- Several technical indicators. For instance, price-earnings ratios and dividend yields do slowly revert to their means. Remarkably, Shiller (2005, chps. 1 and 10) shows that cyclically adjusted price-earnings ratios display the same behavior. Moreover, Siegel (2014, chp. 20) also finds that an investor using a 200-day moving average would have avoided the 1929 crash. A complementary yet unusual use of moving averages is considered by Ilomäki et al. (2018), who are concerned only with the Dow Jones Industrial Average and its constituent stocks;
- Macroeconomic variables, such as short-term interest rates, yields to maturity on Treasury bonds, and GDP growth rates. For instance, the pioneering work in Harvey (1993) finds that appropriate spreads between US Treasury yields to maturity may anticipate the start (and end) of a recession. Harvey (1993) also mentions similar evidence for several developed countries.
- Relative value models, which make a comparison between two markets. For instance, Pesaran and Timmermann (1994b) take advantage of mean reversion and determine profitable investment policies that may switch, either yearly or quarterly, between either the S&P 500 index or the Dow Jones index and appropriate T-bills. Shen (2003) considers a heuristic threshold, i.e., the 10th percentile of the spread between the earnings-price ratio of the S&P 500 index and different interest rates; he obtains a profitable investment policy that may switch monthly between the S&P 500 index and a monthly loan.

If allowance is made for futures on stock indexes, their interplay with futures on commodities, currencies, and Treasury bonds must be considered. Therefore, reference can be made to:

- The sentiments of investors and traders. For instance, Wang (2003) examines the futures on the S&P 500 index by making reference to COT reports, which distinguish large speculators from large hedgers and small traders. On isolating extreme values from their sentiments, he finds that combining extremely bullish speculator sentiment and extremely bearish hedger sentiment results in a price continuation indicator. Wang (2003) also mentions similar empirical evidence on commodity futures;
- The VIX (futures) term structure. For instance, Fassas and Hourvouliades (2019) find that a downward slope is a "contrarian market timing indicator".

Consistent market timing is very hard. Combining forecasts is essential, as found once by Ambachtsheer and Farrell (1979), and recently by Mascio et al. (2020). According to the summary of Damodaran (2012):

- There is little empirical evidence that mutual funds have market timing skills;
- There is empirical evidence that hedge funds may time bond and currency markets;
- There is empirical evidence that professional advisers may have market timing skills.

In contrast, mixed empirical evidence on mutual funds is reported by Elton et al. (2014, chp. 26).

Mutual funds can attempt to time markets by changing their cash balances and their portfolio beta. According to a parametric test, when portfolio excess returns are regressed against market excess returns, two different characteristic lines can be obtained by making use of a dummy variable. One characteristic line matches the periods when the market does better than the riskless security; another characteristic line matches the periods when the market does worse than the riskless security. Therefore, two top-down portfolio betas are estimated. Whenever market timing is successful, positive and statistically significant parameters can be estimated, with the former portfolio beta being greater than the latter. Moreover, a positive and statistically significant Jensen's alpha indicates skillful stock picking.

Theoretical analysis of this parametric test is performed by Henriksson and Merton (1981) under the joint null hypothesis that there is no market timing skill and equilibrium security returns are consistent with the CAPM. The empirical analysis of Fung et al. (2002, p. 19) is concerned with global equity-based hedge funds; it finds accordingly that "*managers do not show positive market-timing ability but do demonstrate superior security-selection ability*".

However, bottom-up portfolio betas lead to a better prediction of portfolio performance. Therefore, the time pattern of a bottom-up portfolio beta can be compared with the time pattern of the market. To estimate a bottom-up portfolio beta, security betas are possibly estimated from security returns, actual weights are retrieved, and a weighted sum of all relevant security betas is computed. Such a procedure ought to be applied repeatedly at different points in time.

The odds of profitable market timing are examined by three empirical papers below. Altogether, it is found that market timing is challenging, with successful performances being possibly achievable only with suitable forecasting accuracy. According to Pedersen (2018, p. 29), "*Naturally, large institutional investors can better afford to spend resources on a manager selection team. So, it is not a surprise that institutional investors have been more successful in their active management than smaller investors*".

More specifically, Sharpe (1975) focuses on the link between forecasting accuracy and superior performance. His data set spans the historical period of 1929–1972; it includes the total returns of two asset classes: either prime bankers' acceptances (until 1942) or US Treasury bills (after 1942), as well as the Standard and Poor's Composite index.

Years can be either good or bad; stocks outperform cash equivalents in good years, whereas cash equivalents outperform stocks in bad years. Markets can be timed only

at the beginning of each year; every shift from stocks to cash equivalents or vice versa brings about commissions, which amount to 2% of portfolio value. Profitable market timing follows from a suitable forecasting accuracy, i.e., a suitable proportion of right forecasts. The same accuracy is assumed when forecasting a good or bad year; stocks or cash equivalents are held accordingly.

Each asset class is represented by two pairs (arithmetic mean return, standard deviation); a pair is attached to good years, another pair to bad years. The timing policy is compared with two different policies, either buy and hold, or constant mix. The former invests only in stocks, whereas the latter requires yearly rebalancing to restore the weights of stocks and cash equivalents. Since cash equivalents are less volatile than stocks, the timing policy has a lower standard deviation than the buy and hold one. By construction, the timing policy has roughly the same standard deviation as the constant mix one. All policies are compared in terms of their mean returns. According to the original findings, the timing policy has a greater geometric mean than the buy and hold policy whenever the proportion of right forecasts is 83% or more; in contrast, it has a greater arithmetic mean than the constant mix policy whenever the proportion of right forecasts is 74% or more. This explains the conclusion that a portfolio manager should not try to time the stock market unless he/she can correctly forecast the next year at least 7 times out of 10.

However, the conclusion applies only to the average manager, who is neither consistently luckier nor consistently unluckier than all other managers with the same forecasting accuracy. Unfortunately, overall performance is heterogeneous, since each manager with the same forecasting accuracy makes his/her own peculiar sequence of decisions on asset allocation.

Such an original conclusion is corroborated by Chua et al. (1987), who apply a Monte Carlo analysis to Canadian common stocks and Treasury bills under the assumption that their logarithmic returns have a bivariate normal distribution. Moreover, a different accuracy is used when forecasting good or bad years. On realizing whether a year is good or bad, a random number is drawn from a (0,1) uniform distribution so as to check whether the attendant forecast is right or wrong.

According to Monte Carlo simulations, profitable market timing relies on a portfolio manager who can correctly forecast good years at least eight times out of ten, and bad years at least 6 times out of 10. However, each simulated policy is compared only with a buy and hold one. Use is made of the win/loss ratio, i.e., the ratio of right switches to wrong switches. Unfortunately, its financial implications are unclear when both a 1% commission and high forecasting accuracies are considered. According to the resulting win/loss ratio, perfect timing is matched by a modest chance of outperforming a buy and hold policy.

Additional references on market timing are available in Hallerbach (2014), where a Monte Carlo analysis is performed under the assumptions that the probability density function of monthly excess returns has fat tails and volatility varies over time in accordance with a set EWMA. Neither Chua et al. (1987) nor Hallerbach (2014) deal with the heterogeneous performance of portfolio managers.

Our study refines the choice of data as well as the treatment of commissions and heterogeneity so as to fill a gap in the scientific literature on the odds of profitable market timing.

3. Historical Data and Conceptual Framework

3.1. Historical Data

Use is made of annual total returns. The data set includes two major US asset classes:

- Cash equivalents with a maturity of 1 year[1];
- The S&P Composite.

Original data were downloaded in spring 2021 from the webpage of Professor Robert Shiller, Yale University. Our data set spans the historical period of 1929–2018. As advocated by Bernstein (1997), the data set begins and ends with similar price-earnings ratios. The whole historical period is divided into the subperiods of 1929–1972 and 1973–2018, which

begin and end with similar price-earnings ratios[2]. The two subperiods are made up of 44 and 46 years, respectively, whereas the key subperiod of 1934–1972 in Sharpe (1975) is made up of 39 years.

The sample moments of the two major US asset classes are reported in Table 1; annual rates are considered. Table 1 allows us to compare the subperiods of 1929–1972 and 1973–2018.

Table 1. Sample Moments of Two Major US Asset Classes: Annual Rates.

	1929–2018		
	Cash Equivalents	**S&P Composite**	**Inflation**
Arithmetic mean	4.18%	10.92%	3.11%
Standard deviation	3.47%	18.64%	4.04%
Worst rate	0.29%	−43.12%	−10.06%
Best rate	17.63%	54.87%	18.13%
	1929–1972		
	Cash Equivalents	**S&P Composite**	**Inflation**
Arithmetic mean	2.85%	10.40%	2.20%
Standard deviation	2.20%	20.35%	4.68%
Worst rate	0.53%	−43.12%	−10.06%
Best rate	9.11%	54.87%	18.13%
	1973–2018		
	Cash Equivalents	**S&P Composite**	**Inflation**
Arithmetic mean	5.45%	11.42%	3.98%
Standard deviation	3.96%	16.83%	3.08%
Worst rate	0.29%	−35.19%	−0.09%
Best rate	17.63%	38.56%	13.91%

Raw data downloaded from www.econ.yale.edu/~shiller/data.htm (accessed on 20 May 2021).

Notice that in the former subperiod:

- Inflation had a lower mean but a higher standard deviation and a wider range;
- Cash equivalents had a lower mean, a lower standard deviation, and a narrower range;
- The S&P Composite had a lower mean, a higher standard deviation, and a wider range.

As recalled in Section 2, years can be either good or bad; stocks outperform cash equivalents in good years, whereas cash equivalents outperform stocks in bad years.

The breakdown into good and bad years is reported in Table 2, which allows us to compare good and bad years. Notice that in the good years of each historical period the S&P Composite had a higher mean by assumption. However, it had a lower standard deviation only in the subperiod 1973–2018.

In Section 4.1, all arithmetic means and standard deviations as well as all geometric means are determined by assuming that population moments are the same as the available historical moments.

The best possible performance in our setting is ideal, since it requires perfect timing. In other words, the proportion of right forecasts is 100% so that only the asset class with the higher total return is held each year.

The sample moments of perfect timing are reported in Table 3, assuming that:

- Either no commission is charged;
- Or a commission is charged on transaction value.

Table 2. Sample Moments of Two Major US Asset Classes: Annual Rates in Good and Bad Years.

	1929–2018	1929–1972	1973–2018
Proportion of good years	68.89%	70.45%	67.39%
Cash equivalents in good years			
Arithmetic mean	4.11%	2.89%	5.33%
Standard deviation	3.42%	2.12%	3.99%
Cash equivalents in bad years			
Arithmetic mean	4.33%	2.76%	5.70%
Standard deviation	3.57%	2.37%	3.87%
S&P Composite in good years			
Arithmetic mean	20.73%	20.36%	21.10%
Standard deviation	11.82%	14.11%	8.96%
S&P Composite in bad years			
Arithmetic mean	−10.80%	−13.36%	−8.58%
Standard deviation	11.09%	11.25%	10.47%

Raw data downloaded from www.econ.yale.edu/~shiller/data.htm (accessed on 20 May 2021).

Table 3. Sample Moments of Perfect Timing: Annual Rates.

		1929–2018	
	S&P Composite	**Perfect Timing No Commission**	**Perfect Timing 1% Commission**
Arithmetic mean	10.92%	15.63%	14.89%
Standard deviation	18.64%	12.56%	12.86%
Worst rate	−43.12%	0.53%	−1.47%
Best rate	54.87%	54.87%	53.87%
		1929–1972	
	S&P Composite	**Perfect Timing No Commission**	**Perfect Timing 1% Commission**
Arithmetic mean	10.40%	15.16%	14.48%
Standard deviation	20.35%	14.36%	14.64%
Worst rate	−43.12%	0.53%	−1.47%
Best rate	54.87%	54.87%	53.87%
		1973–2018	
	S&P Composite	**Perfect Timing No Commission**	**Perfect Timing 1% Commission**
Arithmetic mean	11.42%	16.08%	15.29%
Standard deviation	16.83%	10.54%	10.86%
Worst rate	−35.19%	0.86%	−1.14%
Best rate	38.56%	38.56%	37.56%

Raw data downloaded from www.econ.yale.edu/~shiller/data.htm (accessed on 20 May 2021).

More precisely, a commission $c = 1\%$ is paid whenever stocks are bought or sold and whenever cash equivalents are bought; no commission is paid when cash equivalents expire. Table 3 allows us to compare perfect timing (with and without commissions) with a buy and hold policy.

Take the payment of commissions into consideration and notice that:

- Perfect timing always has a greater arithmetic mean than the S&P Composite. The difference is large and ranges from 3.87% to 4.08%;
- The S&P Composite always has a greater standard deviation than perfect timing. The difference is large and ranges from 5.71% to 5.97%;

- Perfect timing always has a much better worst rate than the S&P Composite by assumption.

3.2. Conceptual Framework: Expected Commissions

The aim of this statistical study is to investigate market timing when portfolio managers have different forecasting accuracies. First, we consider a sufficiently large population of homogeneous managers and analyze the properties of their representative portfolio by using the historical mean returns and standard deviations reported in Section 3.1. Next, we move to a sufficiently large population of heterogeneous managers and carry out a basic Monte Carlo analysis of the effects exerted by individual forecasting errors on the related portfolios.

Financial markets are represented by a Bernoulli random variable r such that $r = 1$ (or $r = 0$) when a year is good (or bad). Roughly speaking, stocks are expected to perform sufficiently better than cash equivalents in a good year, whereas cash equivalents are expected to perform sufficiently better than stocks in a bad year. Forecasts are represented by a Bernoulli random variable w such that $w = 1$ (or $w = 0$) when the next year is called good (or bad) and stocks (or cash equivalents) are held accordingly by an operator who is a price taker.

The table below represents the joint probability distribution

	$r = 1$	$r = 0$
$w = 1$	$p\rho$	$(1-p)(1-\rho)$
$w = 0$	$p(1-\rho)$	$(1-p)\rho$

where p is the marginal probability of a good year, $(1 - p)$ is the marginal probability of a bad year, and ρ is the proportion of right forecasts. More precisely, ρ is a conditional probability: it is the probability of forecasting a good year when a good year is coming, as well as the probability of forecasting a bad year when a bad year is coming; the same accuracy is thus assumed when forecasting good or bad years. Notice that $w = 1$ is matched by a correct prediction for $r = 1$ and an incorrect prediction for $r = 0$.

It can be readily ascertained that:

- $1 - p - (1 - 2p)\rho$ is the marginal probability of holding stocks, whereas $p + (1 - 2p)\rho$ is the marginal probability of holding cash equivalents. For any given $p > 0.5$, the marginal probability of holding stocks increases linearly with ρ, whereas the marginal probability of holding cash equivalents decreases linearly with ρ. Both probabilities come in useful when determining the probability of incurring commissions;
- Summing the probabilities on the main diagonal gives ρ, which is the probability of a right call. In other words, the higher ρ, the higher is the forecasting accuracy;
- The covariance between r and w is

$$p(1-p)(2\rho-1)$$

- For $\rho = 0.5$, forecasting is like flipping a coin so that the Bernoulli random variables r and w are independent. As explained in Section 3.3, no rational operator can do so. The marginal probabilities of holding stocks and cash equivalents are the same;
- For $\rho = 1$, the Bernoulli random variables r and w are perfectly correlated; in other words, forecasting is clairvoyant. The marginal probability of holding stocks is equal to p, whereas the marginal probability of holding cash equivalents is equal to $1 - p$;
- For any given p and $0.5 \leq \rho \leq 1$, Pearson's correlation coefficient φ ranges between 0 and 1, being higher the higher ρ is. Indeed, the variance $p(1 - p)$ is constant. In contrast, the variance $(1 - p - \rho + 2p\rho)(p + \rho - 2p\rho)$ decreases with ρ for $p \neq 0.5$, whereas it is equal to 0.25 for $p = 0.5$.

Let c be the commission charged on a transaction that is worth \$1. Expected commissions depend on the four different events that can take place at the end of each year:

- If stocks are held for two years in a row, no commissions are paid. The probability is $(1-p-\rho+2p\rho)^2$;
- If stocks are replaced with cash equivalents, commissions are paid twice. The probability is $(1-p-\rho+2p\rho)(p+\rho-2p\rho)$;
- If cash equivalents are replaced with stocks, commissions are paid once. The probability is $(p+\rho-2p\rho)(1-p-\rho+2p\rho)$;
- If cash equivalents are held for two years in a row, commissions are paid once. The probability is $(p+\rho-2p\rho)^2$.

As a consequence, expected commissions per unit of capital amount to

$$3c(1-p-\rho+2p\rho)(p+\rho-2p\rho)+c(p+\rho-2p\rho)^2 \quad (1)$$

Remarkably, expected commissions per unit of capital take a different form in Sharpe (1975):

$$4c(1-p-\rho+2p\rho)(p+\rho-2p\rho) \quad (2)$$

since only switches are considered and commissions are always paid twice. It can be readily ascertained that:

- For any given $p > 0.5$ and $\rho = 0.5$, both expected values are equal to c;
- For any given $p > 0.5$ and $0.5 < \rho \leq 1$, the expected value (1) is lower than the expected value (2).

3.3. Conceptual Framework: Rational Behaviour

Since operators don't know the marginal probability of a good year p, rationality requires

$$\rho\mu_{SG}+(1-\rho)\mu_{SB} > \rho\,\mu_{CG}+(1-\rho)\mu_{CB} \quad (3)$$

$$(1-\rho)\mu_{CG}+\rho\mu_{CB} > (1-\rho)\,\mu_{SG}+\rho\mu_{SB} \quad (4)$$

provided that operators are risk-neutral. The right-hand side of (3) is the expected rate of return conditional on $w = 1$; since stocks are held, cash equivalents are considered in the left-hand side of (3). The right-hand side of (4) is the expected rate of return conditional on $w = 0$; since cash equivalents are held, stocks are considered in the left-hand side of (4).

The inequalities (3) and (4) respectively imply that

$$\rho > \overline{\rho} = \frac{\mu_{CB}-\mu_{SB}}{\mu_{SG}-\mu_{SB}+\mu_{CB}-\mu_{CG}}$$

$$\rho > 1-\overline{\rho} = \frac{\mu_{SG}-\mu_{CG}}{\mu_{SG}-\mu_{SB}+\mu_{CB}-\mu_{CG}}$$

where $0 < \overline{\rho} < 1$. Since we need that

$$\rho > \max(\overline{\rho}, 1-\overline{\rho}) \geq 0.5$$

We have

$$\rho > \max(\overline{\rho}, 1-\overline{\rho}) = 0.5201 \quad \text{i.e.,} \quad \varphi > 0.08$$
$$> \max(\overline{\rho}, 1-\overline{\rho}) = 0.5248 \quad \text{i.e.,} \quad \varphi > 0.02$$

in the subperiod of 1929–1972 and in the subperiod of 1973–2018, respectively. Therefore, the lowest feasible value of the forecasting accuracy is slightly larger than 0.5, whereas the lowest feasible value of Pearson's correlation coefficient φ is slightly larger than 0.

4. Statistical Findings on Profitable Market Timing

4.1. Retrospective Approach: The Representative Portfolio

Both the dichotomous classification of Section 3.2 and the historical returns of Section 3.1 come in useful when focusing on the representative portfolio, in order to assess the requisites for profitable market timing. More precisely, we assume that:

- The forecasting accuracy ρ is homogeneous;
- The historical moments in Table 2 are the relevant parameters of the investment process;
- A commission $c = 1\%$ is charged on transaction value so that Table 3 above, as well as Tables 4 and 5 below, can be contrasted with the scientific works mentioned in Section 2.

Table 4. Market-Timing Performance versus Forecasting Accuracy, 1929–1972.

	Market Timing Policy				Buy and Hold Policy	Constant Mix Policy	
Forecasting Accuracy	**Gross Arithmetic Mean Return**	**Standard Deviation**	**Net Arithmetic Mean Return**	**Net Geometric Mean Return**	**Difference in Net Geometric Mean Returns**	**Net Arithmetic Mean Return**	**Difference in Net Arithmetic Mean Returns**
ρ	μ	σ	$\bar{\mu}$	g	$g-\hat{g}$	$\bar{\bar{\mu}}$	$\bar{\mu}-\bar{\bar{\mu}}$
0.50	0.0662	0.1496	0.0562	0.0457	−0.0383	0.0781	−0.0219
0.55	0.0748	0.1512	0.0650	0.0543	−0.0297	0.0788	−0.0138
0.60	0.0833	0.1523	0.0737	0.0630	−0.0210	0.0792	−0.0055
0.65	0.0918	0.1529	0.0825	0.0718	−0.0122	0.0795	0.0030
0.70	0.1004	0.1531	0.0913	0.0806	−0.0034	0.0795	0.0118
0.75	0.1089	0.1527	0.1001	0.0896	0.0056	0.0794	0.0207
0.80	0.1174	0.1519	0.1090	0.0986	0.0146	0.0791	0.0299
0.85	0.1260	0.1506	0.1178	0.1077	0.0237	0.0785	0.0393
0.90	0.1345	0.1488	0.1267	0.1169	0.0329	0.0778	0.0489
0.95	0.1430	0.1465	0.1356	0.1262	0.0422	0.0768	0.0587
1.00	0.1516	0.1436	0.1445	0.1355	0.0515	0.0757	0.0688

Table 5. Market Timing Performance versus Forecasting Accuracy, 1973–2018.

	Market Timing Policy				Buy and Hold Policy	Constant Mix Policy	
Forecasting Accuracy	**Gross Arithmetic Mean Return**	**Standard Deviation**	**Net Arithmetic Mean Return**	**Net Geometric Mean Return**	**Difference in Net Geometric Mean Returns**	**Net Arithmetic Mean Return**	**Difference in Net Arithmetic Mean Returns**
ρ	μ	σ	$\bar{\mu}$	g	$g-\hat{g}$	$\bar{\bar{\mu}}$	$\bar{\mu}-\bar{\bar{\mu}}$
0.50	0.0843	0.1259	0.0743	0.0670	−0.0170	0.0947	−0.0204
0.55	0.0920	0.1261	0.0822	0.0748	−0.0092	0.0948	−0.0126
0.60	0.0996	0.1258	0.0900	0.0828	−0.0012	0.0947	−0.0047
0.65	0.1073	0.1251	0.0978	0.0907	0.0067	0.0943	0.0035
0.70	0.1149	0.1239	0.1057	0.0988	0.0148	0.0938	0.0119
0.75	0.1226	0.1222	0.1136	0.1069	0.0229	0.0930	0.0205
0.80	0.1302	0.1200	0.1215	0.1151	0.0311	0.0920	0.0294
0.85	0.1378	0.1173	0.1293	0.1233	0.0393	0.0908	0.0386
0.90	0.1455	0.1140	0.1373	0.1316	0.0476	0.0893	0.0480
0.95	0.1531	0.1100	0.1452	0.1399	0.0559	0.0874	0.0577
1.00	0.1608	0.1054	0.1531	0.1483	0.0643	0.0853	0.0678

The dichotomous classification can go along with a contingency table; if forecasts are observable, the latter may be used to validate a specific timing policy. Notice that the

attendant nonparametric tests do not depend on any assumption about the distribution of portfolio returns. This is the case of Shen (2003), mentioned above in Section 2. This is also the case of Brown et al. (1998), where both nonparametric and bootstrapping tests are performed. The nonparametric test of Henriksson and Merton (1981) is extended by Cumby and Modest (1987). As remarked by Pesaran and Timmermann (1994a, p. 1), the former test "*is asymptotically equal to the* χ^2 *test of independence in the context of a* 2×2 *contingency table*".

Our statistical findings are reported in Tables 4 and 5, which are based on the historical periods of 1929–1972 and 1973–2018, respectively. Use is made of linear rates of return so that approximate geometric means are computed. Outcomes include the difference between two net geometric means as well as the difference between two net arithmetic means. The geometric means belong to either the timing policy or the buy and hold one, whereas the arithmetic means belong to either the timing policy or the constant mix one.

First, the above-mentioned market timing policy is examined; its arithmetic means μ and standard deviations σ are displayed in column 2 and column 3 of Tables 4 and 5. Arithmetic means are computed in accordance with the equation

$$\mu = p\rho\mu_{SG} + (1-p)(1-\rho)\mu_{SB} + p(1-\rho)\mu_{CG} + (1-p)\rho\mu_{CB} \tag{5}$$

with ρ taking the values in column 1 of Tables 4 and 5. Moreover, μ_{SG} and μ_{SB} are the arithmetic means of stocks in good and bad years, respectively, whereas μ_{CG} and μ_{CB} are the arithmetic means of cash equivalents in good and bad years, respectively. Both p and all historical arithmetic means are the entries from Table 2.

Standard deviations are computed in accordance with the equation

$$\begin{aligned}\sigma^2 = {} & p\rho\sigma^2_{SG} + (1-p)(1-\rho)\sigma^2_{SB} + p(1-\rho)\sigma^2_{CG} + (1-p)\rho\sigma^2_{CB} + p\rho(\mu_{SG}-\mu)^2 \\ & +(1-p)(1-\rho)(\mu_{SB}-\mu)^2 + p(1-\rho)(\mu_{CG}-\mu)^2 + (1-p)\rho(\mu_{CB}-\mu)^2\end{aligned} \tag{6}$$

where σ^2_{SG} and σ^2_{SB} are the historical variances of stocks in good and bad years, respectively, whereas σ^2_{CG} and σ^2_{CB} are the historical variances of cash equivalents in good and bad years, respectively. All attendant standard deviations are the entries from Table 2.

When a commission is charged on transaction value, net arithmetic means $\overline{\mu}$ and net geometric means g are computed and displayed in column 4[3] and in column 5 of Tables 4 and 5[4].

Next, a buy and hold policy is examined and its net geometric mean $\hat{g}$ is computed[5]. The timing policy and the buy and hold policy are compared in column 6 of Tables 4 and 5, where the differences $g - \hat{g}$ between their net geometric means are shown.

Finally, a constant mix policy is considered and its net arithmetic mean is computed[6]. Each attendant portfolio is matched by a specific and constant mix of stocks and cash so that its specific volatility is the same as the volatility σ of the timing policy in the same row[7]. The timing policy and the constant mix policy are contrasted in column 8 of Tables 4 and 5, where the differences between their net arithmetic means $\overline{\mu} - \overline{\overline{\mu}}$ are displayed.

According to Tables 4 and 5:

- The higher the forecasting accuracy ρ, the greater is the difference $g - \hat{g}$ between net geometric means. Therefore, the timing policy outperforms the buy and hold one for $\rho \geq 71\%$ in the subperiod of 1929–1972, and for $\rho \geq 61\%$ in the subperiod of 1973–2018;
- The higher the forecasting accuracy ρ, the greater is the difference $\overline{\mu} - \overline{\overline{\mu}}$ between net arithmetic means. Therefore, the timing policy outperforms the constant mix one for $\rho \geq 63\%$ in both subperiods.

All thresholds are strikingly lower than in Sharpe (1975), especially in the latter subperiod. Keep in mind that his Table 5 is based on the subperiod of 1934–1972 which doesn't begin and end with similar price-earnings ratios; moreover, his approximations are different from ours, as regards commissions (see notes 3 and 6).

4.2. Prospective Approach: The Distribution of Individual Portfolios

The representative portfolio of Section 4.1 is attached to the average manager, who is neither consistently luckier nor consistently unluckier than all other managers who have the same forecasting accuracy ρ. Only the performance of the average manager is matched by the net arithmetic mean $\overline{\mu}$ and the net geometric mean g, respectively, given in notes 5 and 6.

However, overall performance is heterogeneous, since each manager with the same forecasting accuracy ρ makes his/her own peculiar sequence of decisions on asset allocation. Therefore, there are many different sequences of decisions and forecasting errors as well as many different accumulation patterns. Accordingly, we consider a number of managers who are price takers and have the same forecasting accuracy ρ. Each year, each manager acts in accordance with his/her forecast so that individual forecasting errors result in a range of net arithmetic means.

For each selected forecasting accuracy ρ, a Monte Carlo analysis is carried out in order to obtain a 90% confidence interval for net arithmetic means. Whatever the selected value of ρ, 10,000 portfolio managers are considered. Each individual performance is at first represented by a sequence of 90 forecasting attempts and then turned into a sequence of 90 net annual rates of return. For simplicity's sake, use is made of the annual rates of return on cash equivalents and the S&P Composite that span the historical period of 1929–2018. Forecasting errors are assumed to be independent and identically distributed Bernoulli random variables. When no forecasting error is made, the right asset class is held; when a forecasting error is made, the wrong asset class is held. A 1% commission is taken into account.

Figure 1 displays the 90% confidence interval for net medians; forecasting accuracy ρ varies from 0.5 to 1. For each selected value of ρ, the cross-sectional median as well as the 5th and 95th percentiles are plotted. Remarkably, the net arithmetic mean turns out to be hardly distinguishable from the net median. Notice that:

- The higher the forecasting accuracy ρ, the narrower is the cross-sectional range, i.e., the lower is the heterogeneity of managers' net annual return;
- For $\rho = 1$, the cross-sectional range includes a single point. Therefore, if forecasting is clairvoyant, all investors are like the average portfolio manager of Tables 4 and 5.

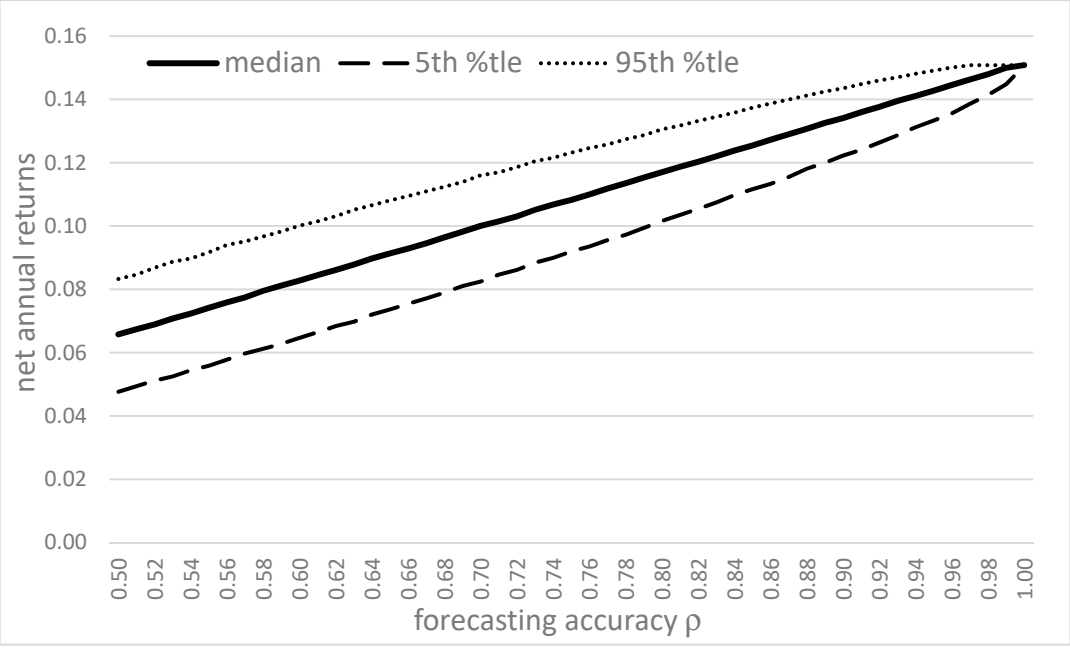

Figure 1. Net Annual Return from Heterogeneous Market Timing. Note: For each specific forecasting accuracy ρ, 10,000 simulations are run, which result in the net annual returns of 10,000 different portfolios.

Figure 2 compares the heterogeneous timing policies with a constant mix policy; forecasting accuracy ρ varies from 0.5 to 1. For each selected value of ρ, the same net arithmetic mean $\bar{\bar{\mu}}$ is taken away from the cross-sectional values plotted in Figure 1; the resulting cross-sectional median as well as the resulting 5th and 95th percentiles are plotted. Notice that:

- The cross-sectional median of all timing policies is greater than 0 for $\rho \geq 63\%$;
- All timing policies outperform the constant mix one for $\rho \geq 72\%$.

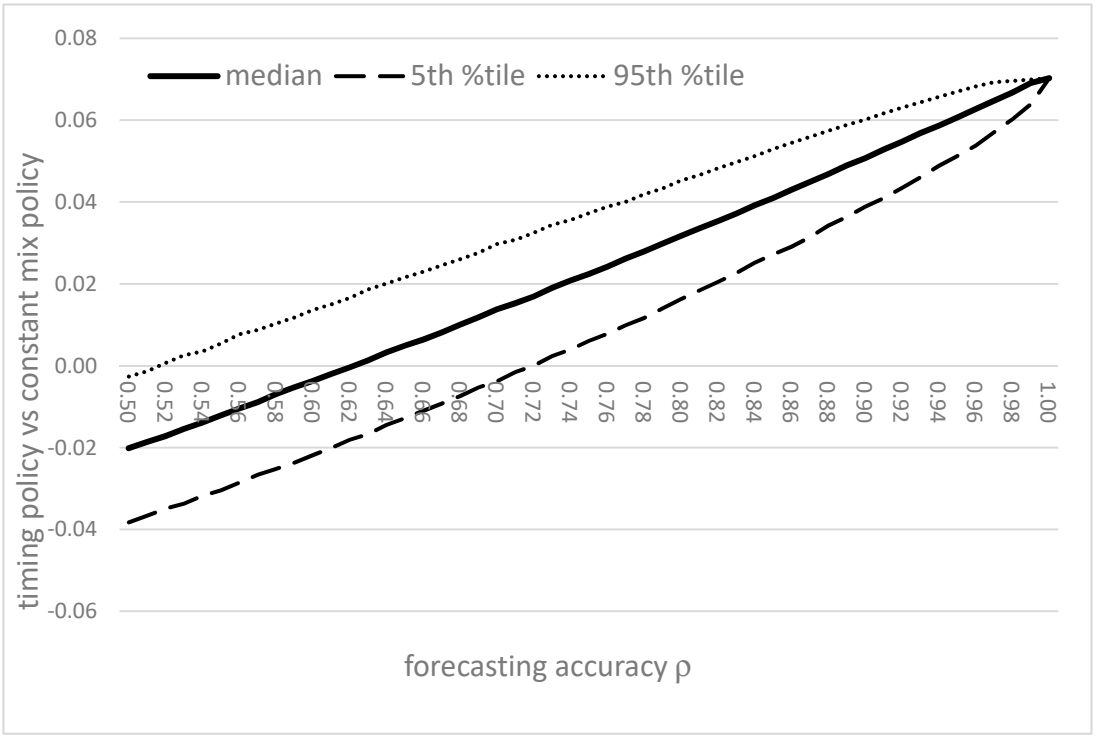

Figure 2. Net Annual Advantage from Market Timing over Constant Mix. Note: For each specific forecasting accuracy ρ, 10,000 simulations are run, which eventually result in the annual advantages of 10,000 different portfolios, each net annual advantage being the difference between a net annual return of Figure 1 and the same net arithmetic mean of the constant mix policy.

The former threshold is the same as in Tables 4 and 5 above, whereas it is much lower than in Sharpe (1975). In contrast, the latter threshold is like Sharpe's. However, if the subperiods of 1929–1972 and 1973–2018 are separately considered, a Monte Carlo analysis arrives at two slightly larger thresholds; the difference from $\rho = 72\%$ is smaller in the latter subperiod. Altogether, the seminal findings of Sharpe (1975) are corroborated[8].

5. Conclusions

This statistical study corroborates once more the original conclusion of Sharpe (1975), whereby profitable market timing depends on a portfolio manager who can correctly forecast the next year at least 7 times out of 10.

The conclusion of Damodaran (2012, p. 523) is slightly different: "To be a successful market timer, you have to be right about two thirds of the time". However, as recalled in Section 2, he remarks that hedge funds may time bond and currency markets rather than equity markets. His conclusion is also drawn from successful fund managers; for instance, Sir John Templeton (1912–2008) claimed that even the best investors cannot outperform a market more than 2 times out of 3.

It can be readily realized that Hallerbach (2014) is likely to reach a similar conclusion, since his Monte Carlo analysis is carried out under the tacit assumption that no commission is paid on transaction value.

As usual, our approach has limitations:

- Two different percent commissions could have been used, a lower one for cash equivalents and a higher one for stocks;
- Two different forecasting accuracies could have been used as in the Monte Carlo analysis of Chua et al. (1987);
- A better use could have been made of historical returns. Expanding on Hallerbach (2014), we could have computed the relative return of stocks versus cash equivalents and examined its statistical properties. Next, we could have checked whether the Bernoulli random variables r of Section 4 are independent and identically distributed.
- Risk-aversion has been disregarded, although it also depends on a country's cultural heritage, as found by Arosa et al. (2014) when considering the corporate decisions on capital structure. Needless to say, the stronger the risk-aversion, the larger are both gaps $\rho - \max(\overline{\rho}, 1 - \overline{\rho})$.

Author Contributions: Conceptualization, L.B. and L.G.; methodology, L.B. and L.G.; software, L.B.; data curation, L.G.; writing, L.B. and L.G. Both authors have read and agreed to the published version of the manuscript.

Funding: No research grant has been received in support of this scientific work.

Data Availability Statement: All the data the authors has used can be freely downloaded from www.econ.yale.edu/~shiller/data.htm (accessed on 20 May 2021).

Acknowledgments: We are grateful to Elena Conti for her editorial support and to Alberto Re Fraschini for his bibliographical assistance. We also thank three anonymous referees for their really helpful comments and suggestions. As usual, we are solely responsible for any remaining errors.

Conflicts of Interest: The authors declare no conflict of interest.

Notes

1 One-year interest rates are available at www.econ.yale.edu/~shiller/data.htm (accessed on 20 May 2021). Reference is made to the 6-month commercial paper until 1997, the 6-month Certificate of Deposit (secondary market) from 1997 to 2010, and the 2-month commercial paper from 2011 to 2018. Needless to say, periodic rates were turned into an annual yield.

2 As reported in Shiller's dataset, price-earnings ratios at the end of 1928, 1972 and 2018 were 17.77, 18.09 and 19.60, respectively. 2019 and 2020 were disregarded owing to a sharp increase in the P/E ratio.

3 Net arithmetic means are provided by the approximation where expected commissions are taken away from arithmetic means. Roughly speaking, a similar and small error is made when computing a real rate of return as the difference between a nominal rate of return and an inflation rate. Although commissions affect both means and standard deviations, the latter effect is disregarded because it is less important.

4 Net geometric means are provided by the approximation that depends on net arithmetic means (5) and the variance (6). As shown by Booth and Fama (1992), the approximation above can be derived from a Taylor expansion truncated at 2nd order.

5 The net geometric mean of the buy and hold policy is historical and exact. It is worth $\hat{g} = 8.40\%$ in the subperiod of 1929–1972 and $\hat{g} = 10.02\%$ in the subperiod of 1973–2018.

6 Owing to annual rebalancing, an appropriate amount of stocks are sold at the end of good years, whereas an appropriate amount are bought at the end of bad years. Moreover, new cash equivalents are subscribed at the end of each year. Consequently, net expected commissions per unit of capital are worth approximately.

7 The proper weight of stocks is numerically determined; it ranges from 67.9% to 73.4% in the subperiod of 1929–1972, and from 60.3% to 73.7% in the subperiod of 1973–2018.

8 As explained by Rutterford (2012), analytical and prospective appraisal methods based on DCFs came into use in the 1980s and 1990s among US corporate finance analysts.

References

Ambachtsheer, Keith P., and James L. Farrell. 1979. Can active management add value? *Financial Analysts Journal* 35: 39–47. [CrossRef]

Arosa, Clara Maria Verduch, Nivine Richie, and Peter W. Schuhmann. 2014. The impact of culture on market timing in capital structure choices. *Research in International Business and Finance* 31: 178–92. [CrossRef]

Berkshire Hathaway. 1991. Annual Report. Available online: https://www.berkshirehathaway.com/ (accessed on 10 April 2021).

Bernstein, Peter. L. 1997. What rate of return can you reasonably expect . . . or what can the long run tell us about the short run? *Financial Analysts Journal* 53: 20–28. [CrossRef]
Booth, David G., and Eugene F. Fama. 1992. Diversification returns and asset contributions. *Financial Analysts Journal* 48: 26–32. [CrossRef]
Brown, Stephen. J., William N. Goetzmann, and Alok Kumar. 1998. The Dow theory: William Peter Hamilton's track record reconsidered. *Journal of Finance* 53: 1311–33. [CrossRef]
Chua, Jess H., Richard S. Woodward, and Eric C. To. 1987. Potentials gains from stock market timing in Canada. *Financial Analysts Journal* 43: 50–56. [CrossRef]
Cumby, Robert E., and David M. Modest. 1987. Testing for market timing ability: A framework for forecast evaluation. *Journal of Financial Economics* 19: 169–89. [CrossRef]
Damodaran, Aswath. 2012. *Investment Philosophies*, 2nd ed. Hoboken: Wiley.
Elton, Edwin J., Martin J. Gruber, Stephen J. Brown, and William N. Goetzmann. 2014. *Modern Portfolio Theory and Investment Analysis*, 9th ed. New York: Wiley.
Fassas, Athanasios P., and Nikolas Hourvouliades. 2019. VIX futures as market timing indicator. *Journal of Risk and Financial Management* 12: 113. [CrossRef]
Fung, Hung Gay, Xiaoqing Eleanor Xu, and Jot Yau. 2002. Global hedge funds: Risk, return, and market timing. *Financial Analysts Journal* 58: 19–30. [CrossRef]
Gibson, Roger C. 2008. *Asset Allocation. Balancing Financial Risk*, 4th ed. New York: McGraw-Hill.
Hallerbach, Winfried G. 2014. On the expected performance of market timing strategies. *Journal of Portfolio Management* 40: 42–51. [CrossRef]
Harvey, Campbell R. 1993. Term structure forecasts economic growth. *Financial Analysts Journal* 49: 6–8. [CrossRef]
Henriksson, Roy D., and Robert C. Merton. 1981. On market timing and investment performance. II. Statistical procedures for evaluating forecasting skills. *Journal of Business* 54: 513–33. [CrossRef]
Ilomäki, Jukka, Hannu Laurila, and Michael McAleer. 2018. Market timing with moving averages. *Sustainability* 10: 2125. [CrossRef]
Jackson, Charles. 2003. *Active Investment Management*. Chichester: Wiley.
Maginn, John L., Donald L. Tuttle, Dennis W. McLeavey, and Jerald E. Pinto. 2007. The portfolio management process and the investment policy statement. In *Managing Investment Portfolios. A Dynamic Process*, 3rd ed. Edited by John L. Maginn, Donald L. Tuttle, Dennis W. McLeavey and Jerald E. Pinto. Hoboken: Wiley, pp. 1–19.
Mascio, David A., Frank J. Fabozzi, and J. Kenton Zumwalt. 2020. Market timing using combined forecasts and machine learning. *Journal of Forecasting* 40: 1–16. [CrossRef]
Pedersen, Lasse Heje. 2018. Sharpening the Arithmetic of Active Management. *Financial Analysts Journal* 74: 21–36. [CrossRef]
Pesaran, M. Hashem, and Allan G. Timmermann. 1994a. A generalization of the non-parametric Henriksson-Merton test of market timing. *Economics Letters* 44: 1–7. [CrossRef]
Pesaran, M. Hashem, and Allan G. Timmermann. 1994b. Forecasting stock returns: An examination of stock market trading in the presence of transaction costs. *Journal of Forecasting* 13: 335–68. [CrossRef]
Rutterford, Janette. 2012. Valuing equities in the UK and the US: Fashions and trend. In *Handbook of Research on Stock Market Globalization*. Edited by Geoffrey Poitras. Cheltenham: Edward Elgar.
Sharpe, William F. 1975. Are gains likely from market timing? *Financial Analysts Journal* 31: 60–69. [CrossRef]
Sharpe, William F., Peng Chen, Jerald E. Pinto, and Dennis W. McLeavey. 2007. Asset allocation. In *Managing Investment Portfolios. A Dynamic Process*, 3rd ed. Edited by John L. Maginn, Donald L. Tuttle, Dennis W. McLeavey and Jerald E. Pinto. Hoboken: Wiley, pp. 230–27.
Shen, Pu. 2003. Market timing strategies that worked. *Journal of Portfolio Management* 29: 57–68. [CrossRef]
Shiller, Robert J. 2005. *Irrational Exuberance*, 2nd ed. Princeton: Princeton University Press.
Siegel, Jeremy J. 2014. *Stocks for the Long Run*, 5th ed. New York: McGraw-Hill.
Wang, Changyun. 2003. Investor sentiment, market timing, and futures returns. *Applied Financial Economics* 13: 891–98. [CrossRef]