

Escaping the Backtesting Illusion

Thorsten Hens, Klaus Reiner Schenk-Hoppé, Mathis-Hendrik Woesthoff

Department of Banking and Finance, University of Zurich, Plattenstr. 32, 8032 Zurich, Switzerland

Thorsten Hens is Professor of Financial Economics at the University of Zurich, and Adjunct Professor of Finance at the Norwegian School of Economics in Bergen and at the University of Lucerne.

thorsten.hens@bf.uzh.ch

Klaus Reiner Schenk-Hoppé is Professor of Financial Economics at the University of Manchester and Adjunct Professor of Finance at the Norwegian School of Economics in Bergen.

klaus.schenk-hoppe@manchester.ac.uk

Mathis-Hendrik Woesthoff is Research Fellow at the University of Zurich.

mathis.woesthoff@bf.uzh.ch

Escaping the Backtesting Illusion

Abstract

Two tests can help asset managers to develop more robust investment strategies: an impact test and a survival test. Both tests complement the backtest where one checks how a proposed investment strategy would have performed in the past. The impact test considers the performance of the strategy when assets under management grow (crowdedness) and it checks the impact that a growth of asset under management in competing strategies have on the proposed strategy (cross impact). The survival test considers the effect of the long-term evolution of assets under management in competition for market capital. Using Shiller's S&P 500 index and bond market data we show that time-series momentum (relative strength) performs best in the backtest and in the impact test but that an expected relative cash-flow rule (relative dividend yield) has the best long-term survival properties.

Highlights/key takeaways

- Robustness checks of investment strategies require more than backtesting
- An impact test is proposed to measure the effects of crowdedness and cross impact on investment strategies
- A survival test is proposed to assess the long-term effects of investment strategies' competition for market capital

Keywords: Investment strategies, robustness, crowdedness, impact test, survival test.

JEL classification: G11, G17.

Earlier versions of this paper have been presented and benefitted from presentation at LMU, CPU, ESMT, MIT as well as at CFM and unigestion. We thank the referee (Bill Ziemba) for his helpful comments.

Every product brochure in asset management includes the disclaimer that past performance is no guarantee of future results. But it is still industry standard to develop new asset management strategies using backtesting. Moreover, some asset managers advertise their strategies by pointing out that they would have generated excess returns - indicating that they will also do so in the future. This hope is naive since in total excess returns are zero.

Although the edifice of backtesting has become more transparent about its limitations, e.g., Harvey and Liu (2015) and Fabozzi and López de Prado (2018), even the most sophisticated and proper application does not address another important issue behind the choice of investment strategies. The issues that really need consideration are: who is on the other side and how long is the other side able or willing to finance one's excess returns. Relying on backtesting suffers from the illusion that there will be no market reaction.

In backtesting one has to treat oneself – and the investment strategy one pursues – as immaterial in the sense of not having an impact on the market and the prices. This can lead to satisfactory results if the market of the future is not unlike the one of the past (and assuming no data snooping has taken place). Financial investment however is, at least in part, driven by fashion—and fashion can lead to large flows of capital. For instance, factor investing in the 2010s, long-short quantitative strategies in the 2000s, and portfolio insurance in the 1980s were all the craze of their time.¹ Each of these fashions is signified by more capital being invested in a particular investment style. That the inflow of capital was relevant became apparent when quantitative strategies led to fire sales in August 2007 and portfolio insurance failed spectacularly in October 1987. The sudden withdrawal of capital had a major impact on financial markets and prices, and one must conclude that therefore the inflow must have had an effect as well but was not noticed at the time. One may also wonder what will happen in the market if more investors start following Spare (1999)'s advice and place more weight on dividend yields of stocks, or if fundamentals, as reported in Shiller (1981), take on a larger role in investors' minds.

Our paper develops a model in which one can explore the crowdedness of investment strategies, i.e., the effect of capital flows that have permanent price impact. The basic idea behind the model is taken from evolutionary finance, Evstigneev, Hens and Schenk-Hoppé (2006, 2009, 2016). This approach views financial markets as the interaction, or rather competition, between different investment strategies. Each strategy has capital behind it, and the larger this amount, the more it can drive asset prices. In our model we consider a world in which changes in capital allocations across investment strategies have permanent impact. For instance, we ask what if more capital would have been invested in dividend yield strategies in the past. Of course this would have led to slightly different market prices if other investors would not have fully compensated this shift in demand. Indeed the idea that financial markets and the investment strategies in these markets adapt over time rather than being based on the perfect anticipation of future distributions of returns and instantaneous adjustment has become more common recently, see, e.g., Farmer and Lo (1999) and Lo (2017).

The model is applied to Robert Shiller's S&P 500 stock market index and U.S. government

¹Portfolio insurance and its failure is discussed in Shiller (1988), for the risk created by quantitative strategies and its effect see Khandani and Lo (2007, 2011), and herding in factor investing during the financial crisis is considered in Krkoska and Schenk-Hoppé (2019).

bond yields which go back to 1871. By adding an endogenous flow of capital into investment strategies based on dividend yields and on time-series momentum, the price of the index becomes semi-endogenous. Our aim is to illustrate the potential effect of an increase in the capital behind these two strategies onto themselves, each other and the market.

This approach goes beyond transaction cost models that take into account the short-term impact of trading on prices. In these models, the flow of capital from one investment strategy to some other is treated as a phenomenon that has a temporary impact on prices. In efficient markets flows are irrelevant as arbitrageurs correct any mispricing and thus any impact can at best be temporary. But as observed in Greenwood, Shleifer and You (2019) run-ups in prices (of industry sector portfolios) increase the probability of a subsequent decline. Factor investing, for instance, follows non-fundamental investment strategies. Unless there is a matching amount of arbitrage capital on the other side, it is hard to believe that a surge in this investment style will not impact prices permanently. Asness (2015) describes such a permanent impact eloquently as the transition from alpha to pure beta, i.e., from a mostly unknown, excess return generating strategy (the alpha) to a widely known strategy where the only compensation is for the actual risk taken (the beta).

We put these findings into context by considering a market in which the price of the index is fully endogenous. The market is assumed to be populated by three distinct investment strategies: equal-weight, relative dividend yield and time-series momentum (relative strength). The dividend payments and bond yields are taken from Shiller’s data but the price is determined solely by the investment strategies and the capital behind each of these. For instance, if more capital is managed according to dividend yield, then in a situation where the S&P 500 dividend yield exceeds that of U.S. government bonds, more capital flows into the S&P 500 and its price will rise. Since capital flows actually impact market capitalizations, the model is defined in these quantities rather than prices. Of course, knowing the total number of shares issued, market capitalizations and price can easily be converted into each other.

Our paper recommends to go beyond the standard backtesting methodology by augmenting it with an impact test—analyzing the robustness if the strategy gets crowded, and a survival test—endogenising the degree of crowdedness by modeling the evolution of wealth. It will be shown that the robustness of trading strategies depends on the degree of endogeneity of prices, wealth and wealth evolution.

The software used in this paper is available from the authors upon request. We invite the reader to explore many more scenarios than space permits us to do in this paper.

Data

Our model uses Shiller’s S&P 500 data which provide monthly prices and dividends for the S&P 500 stock market index and U.S. government bond yields since 1871. We add to these exogenous data an endogenous component: The model allows for the price of the S&P 500 to change in response to capital in- or out-flows. The index will thus be driven by changes in demand and its price is (partially) endogenous. The bond yield is treated as exogenous.

Denote the market capitalization of the S&P 500 in month t by q_t , and the yield on a GS10 bond issued in month t by i_t . Holders of these bonds will receive the yield but also

experience capital gains or losses due changes in the yield from month t to $t + 1$. We account for these changes and calculate the effective yield as follows. Bond prices of newly issued bonds are set to 1. Portfolios are adjusted monthly. The bond holdings are rolled over into the newly issued bond. We calculate the price of the bond issued in month $t - 1$ with monthly yield i_{t-1} in month t as

$$P_{t-1,t} = i_t \cdot \frac{1 - (1 + i_{t-1})^{-119}}{i_{t-1}} + (1 + i_{t-1})^{-119}$$

i_t is the monthly yield derived from Shiller's GS10 annual yields.

The effective, i.e., realized, interest is therefore

$$r_t = i_{t-1} + (P_{t-1,t} - 1).$$

The two yields are close to each other before 1950 (with the highest monthly capital gain due to bond price changes of 0.51%) but their difference varies considerably during the 1980s. During essentially all years these effects are short term and cancel each other out if the bond is held for 12 months or more. At a 12-month holding period the maximum 12-month average capital gain is 1.81% over the entire time horizon. For the performance of investment strategies such short-term effects fortunately matter little. The average annual drag on bond yields, i.e., the average difference between the effective interest and the GS10 yield, amounts to just -0.089%.

Finally transaction costs can be taken into account by charging, in any one period, the value of stock turnover times a per dollar fee. We use the cost reported in Jones (2002) who estimates per dollar costs for trading Dow Jones stocks over the last 100 years. The potential caveat is that prior to the launch of index products in the early 1990s, trading of individual stocks to represent the S&P 500 returns might be underestimated.

Strategies. Investment strategies are denoted by real numbers λ_t . The value of λ_t is the fraction of the investment strategy's capital invested in the S&P 500 index in month t . The remaining fraction of wealth $1 - \lambda_t$ is invested in bonds. A negative value of λ_t means a short position in the index, and a value of λ_t larger than 1 is a leveraged position in the index (financed with a short position in the bond, i.e., by borrowing).

We consider two active and one passive investment strategies in this paper. The active investment strategies are based on signals. The passive investment strategy sets $\lambda_t \equiv 0.5$ and thus invests equal amounts in the index and in the bond. We call this the equal weight investment strategy. One could also use the term market portfolio strategy which is justified in the present setting because the market capitalization of the S&P 500 and that of U.S. government debt has historically been roughly of the same magnitude. For example, as of 31 December 2018, the market capitalization of the S&P 500 is 21.03 trillion and the US national debt stands at 21.97 trillion.

The active strategies considered here are one reversal and one momentum strategy. The reversal strategy buys assets with high dividend yield. Thus if prices increase (decrease) more than dividends it sells (buys). The momentum strategy buys (sells) the asset that has previously increased (decreased) in price. Of course there are many other reversal and momentum strategies but to keep the analysis tractable, we decided to include only one of each type.

The relative dividend yield signal is given by

$$s_t^{dividend} = \frac{d_t/q_{t-1} - i_{t-1}}{i_{t-1} + d_t/q_{t-1}}$$

and the relative strength signal is given by

$$s_t^{strength} = \frac{(q_t - q_{t-1})/q_{t-1}}{i_{t-1} + |q_t - q_{t-1}|/q_{t-1}}$$

with both signals taking values between -1 and +1 (assuming interest rates are positive).

Signals s_t are translated into an investment strategy as follows:

$$\lambda_t = \frac{1}{\pi} \arctan(a \cdot s_t) + \frac{1}{2}$$

with $a \geq 0$. This function is S-shaped with slope a at zero. Investment strategies generated in this way are non-levered, long only. It is straightforward to design more aggressive strategies, e.g., by multiplying the first term by a factor larger than one, which then can lead to short and levered positions in the index. We do not look into these strategies here but rather vary the parameter a to make the strategies respond more or less strongly to the signal.

If the strategy does not react to the signal, i.e., $a = 0$, then it invests equal amounts of capital in the S&P 500 and the bond at all times, i.e., holds the market portfolio and matches the equal weight investment strategy.

Results

The details behind the backtest and the tests for robustness, impact test and survival test are explained and the results in each test scenario are presented.

Backtest

Denote the return on the S&P 500 by

$$R_{t+1} = \frac{q_{t+1} + d_{t+1}}{q_t}$$

where q_t is the market capitalization of the S&P 500 in month t , d_{t+1} the total amount of dividends paid to those who bought the index in month t . The wealth of an investment strategy λ_t evolves as

$$w_{t+1} = \left(\frac{q_{t+1} + d_{t+1}}{q_t} \lambda_t + (1 + r_{t+1})(1 - \lambda_t) \right) w_t - cost_t$$

where r_{t+1} the effective yield on GS10 bonds bought in month t . The transaction cost incurred in month t is

$$cost_t = c_t \cdot |\lambda_t w_t - \lambda_{t-1} w_{t-1}|$$

where c_t denotes the per dollar cost of trading in month t as reported in Jones (2002).

We assume that trading the bond is costless, i.e., since trade in the model is the exchange of index and bonds we only count costs once.

The average growth rate of wealth is given by

$$g^\lambda := \frac{1}{T} \sum_{t=0}^{T-1} \log(w_{t+1}/w_t).$$

As benchmark we take the performance, measured by average annual return and the Sharpe ratio², of the equal weight investment strategy. As explained above, this investment strategy can be interpreted as holding the market portfolio.

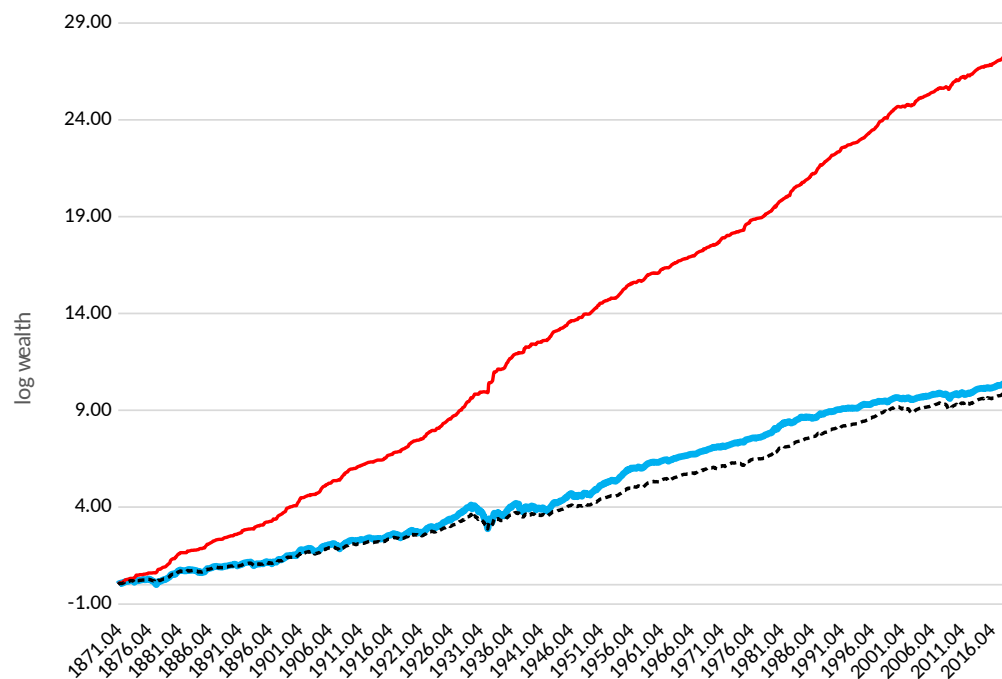


Exhibit 1: Backtest: Log wealth 1887.01-2018.12 (initial wealth is set to one). Time-series momentum relative strength (solid line, red), reversal relative dividend yield (bold solid line, blue) and equal weigh benchmark (broken line, black). $a = 5$. Wealth dynamics is net of transaction costs.

Exhibit 1 shows the logarithm of wealth in the backtest where transaction costs are taken into account. The two active investment strategies are aggressive ($a = 5$). In this backtest momentum (relative strength) performs best and reversal (relative dividend yield) does slightly better than the equal weight (our benchmark). Exhibit 2 contains more detailed results. The benchmark strategy delivers a performance of 6.72% without taking into account transaction costs and 6.66% net of transaction costs. Both the dividend yield and the relative

²Sharpe ratios are calculated as the difference of average monthly return and average riskfree rate divided by the standard deviation of monthly returns. Annualized Sharpe ratios are obtained by multiply by the square root of 12.

strength investment strategy outperform the benchmark with and without transaction costs taken into account. Aggressiveness is good for performance, i.e., the higher the parameter a the higher the excess return. But for the dividend yield investment strategy different levels of aggressiveness (parameter a) only have a minor impact on the Sharpe ratio. The Sharpe ratio is around 0.36, just a little higher than that of the benchmark.

		Benchmark (equal weight)			
		no transaction costs		with transaction costs	
Return		6.72%		6.66%	
Sharpe ratio		0.357		0.348	
a		Reversal (rel. dividend yield)	Momentum (rel. strength)	Reversal (rel. dividend yield)	Momentum (rel. strength)
0.5	Return	6.80%	11.37%	6.72%	10.40%
	Sharpe ratio	0.362	0.957	0.351	0.825
1	Return	6.87%	14.74%	6.78%	13.10%
	Sharpe ratio	0.363	1.340	0.351	1.121
2	Return	6.98%	18.33%	6.86%	15.97%
	Sharpe ratio	0.360	1.691	0.346	1.385
5	Return	7.17%	21.46%	7.02%	18.47%
	Sharpe ratio	0.351	1.952	0.336	1.575

Exhibit 2: Backtest over the full time horizon 1871.01-2018.12. Average annual log return and annualized Sharpe ratios for benchmark (equal-weight investment strategy), reversal (relative dividend yield) and time-series momentum (relative strength) investment strategies for different degrees of aggressiveness a .

For relative strength a different picture emerges. Returns, even for the doveish $a = 0.1$, are higher than in the benchmark case and increase with more aggressiveness. The most remarkable feature of relative strength however that the Sharpe ratio is considerably higher than in the benchmark case and increasing with aggressiveness. A Sharpe ratio of 1.952 at $a = 5$ is impressive. Transaction costs do take their toll but overall the picture is unchanged. After taking account of these costs, relative strength still delivers a Sharpe ratio of 1.575. Transaction costs incurred are highest for the relative strength investment strategy which indeed has the highest turnover. Turnovers are reported in Exhibit 3. When aggressive, $a = 5$, relative strength has a turnover in excess of 415% per year. Dividend yield, in contrast, has low turnover even when it is aggressive.

The result of the backtest is that the momentum strategy relative strength outperforms the reversal strategy relative dividend yield and also the index. Hurst, Ooi and Pedersen (2017) establish that in backtests time-series momentum strategies perform well in global equity markets in every decade since 1880 and identify their performance during 8 of the largest 10 crises in the last 100 years as a distinctive feature. Further Lempérière et al. (2014) have found that time-series momentum strategies have an excellent historical record in backtesting in many stock indices as well as in commodities, currencies and bonds in the past two centuries.

In the next section we check the crowdedness of the strategies. What would have hap-

	Benchmark (equal weight) 9.12%	
a	Reversal (relative dividend yield)	Momentum (relative strength)
0.5	11.2%	134.0%
1	13.2%	226.8%
2	16.5%	327.0%
5	20.7%	415.9%

Exhibit 3: Backtest. Average annual turnover in S&P 500 index for benchmark (equal-weight investment strategy), the reversal (relative dividend yield) and time series momentum (relative strength) investment strategies for different degrees of aggressiveness a .

pened if all investors knew about the superior performance of the relative strength strategy?

Impact test

The impact test considers a situation where a passive strategy becomes active. Suppose the investment strategy λ_t holds capital equal to the fraction $\epsilon \geq 0$ of the index and the same amount in the bond market, i.e., its total assets under management are

$$2\epsilon q_t.$$

If this strategy changes from holding the market portfolio, i.e. $\lambda_t = 0.5$ for all t , to some other allocation, then, if the impact is permanent, the market capitalization of the S&P 500 will become

$$q_t^\lambda := (1 - \epsilon)q_t + \lambda_t 2\epsilon q_t = q_t + \epsilon(2\lambda_t - 1)q_t.$$

Thus if $\lambda_t < 0.5$, the index will be lower than in the historical data in month t , and if $\lambda_t > 0.5$ it will be higher. The market capitalization-neutral strategy is $\lambda_t \equiv 0.5$, i.e., the equal-weight investment strategy.

To test the impact of this change on the performance of some strategy μ_t , one can define the new returns

$$R_{t+1}^\lambda := \frac{q_{t+1}^\lambda + d_{t+1}}{q_t^\lambda}$$

where d_{t+1} is the S&P 500 dividend paid in month $t + 1$ to investors who took a position in the index in month t .

The wealth of some strategy μ evolves as

$$w_{t+1}^\mu = (R_{t+1}^\lambda \mu_t + (1 + r_{t+1})(1 - \mu_t)) w_t^\mu - cost_t$$

with $cost_t$ denoting the transaction costs incurred in month t .

As above, one can calculate the average growth rate of wealth (gross of transaction costs) as

$$g_\lambda^\mu := \frac{1}{T} \sum_{t=0}^{T-1} \log(w_{t+1}^\mu / w_t^\mu).$$

Whether and, if yes to what extent, an investment strategy is susceptible to crowding can be tested by setting $\mu = \lambda$ and calculating the average growth rate g_λ^λ for varying values of ϵ . The larger ϵ , the larger the amount of assets under management of the investment strategy λ relative to the overall market.

The cross impact, i.e., the impact of an increase in the assets under management of investment strategy λ on some other investment strategy μ is g_λ^μ , again calculated for varying values of ϵ .

In the impact test capital is moved from the benchmark investment strategy to either the relative dividend yield or the relative strength investment strategy. This impact changes the market capitalization of the index and thus leads to different returns. We then rerun the backtests at these, new returns. Exhibit 4 reports the results.

The results give a clear message. The impact of the relative dividend yield (reversal) investment strategy (RDY) is negative for both strategies' performance. This is true for the results without and with transaction costs. Net of transaction costs the impact is slightly more pronounced. Overall the impact of the relative dividend yield investment strategy never exceeds 6 basis points and is therefore less than one-hundredth of the performance in the backtest. The impact is marginally more negative for relative strength than for relative dividend yield itself.

The momentum (relative strength) investment strategy has a much stronger impact on performance. Both strategies' performance improves in all scenarios considered but relative strength is benefiting to a much larger extent. For instance with $a = 2$, $\epsilon = 400bp$ and net of transaction costs, the annual return of relative strength improves by 7.7 percentage points and that of dividend yield by 0.5 percentage points. The time series momentum strategy is mostly boosting its own performance.

With the change in the market capitalization of the index due to a strategy's impact the original strategies are no longer correct in the sense that the information on which they are based has changed as well. For instance if the dividend yield of the index decreases, the relative dividend yield investment strategy would allocate less capital to the index compared to the original data.

Exhibit 5 compares the performance of these recalculated investment strategies, i.e., each investment strategy is formed at the new, adjusted market capitalization of the index and then a backtest is performed for these strategies at the new returns. We report only scenarios where the difference in performance exceeds 10 basis points. The main message is that neglecting the impact of the change in market capitalization on the strategies has mostly a minor effect. The largest change in performance is observed for aggressive, relative strength. But even then the performance is underestimated by only 1 percentage point (1.3 when transaction costs are ignored). However this is only a contribution of about 6% of the average annual return which is in excess of 18.47% (21.46% in the absence of transaction costs). This is surely a consequence of considering an index rather than individual stocks. Here one would expect a much larger effect on changes to the investment strategies' portfolios.

The impact test shows that reversal strategies suffer from crowding while the momentum strategy thrives the more assets are invested accordingly. Yet, the next test will reveal that the second result is illusionary. With fully endogenous prices we see that momentum drives prices away from fundamentals increasing the returns of the reversal strategy which will then flourish and itself impact prices. The interesting question is which strategy will impact

		no transaction costs		with transaction costs		
		impact strategy				
ϵ (in bp)		RDY	MOM	RDY	MOM	
$a = 0.5$	50	RDY	-0.0004	0.0115	-0.0004	0.0135
		MOM	-0.0011	0.1452	-0.0011	0.1466
	100	RDY	-0.0009	0.0237	-0.0009	0.0277
		MOM	-0.0023	0.2911	-0.0023	0.2938
	200	RDY	-0.0017	0.0502	-0.0018	0.0573
		MOM	-0.0045	0.5848	-0.0046	0.5900
	400	RDY	-0.0035	0.1115	-0.0035	0.1227
		MOM	-0.0090	1.1798	-0.0091	1.1893
$a = 1$	50	RDY	-0.0009	0.0252	-0.0009	0.0287
		MOM	-0.0031	0.4349	-0.0032	0.4377
	100	RDY	-0.0017	0.0524	-0.0018	0.0590
		MOM	-0.0063	0.8715	-0.0063	0.8767
	200	RDY	-0.0034	0.1132	-0.0035	0.1241
		MOM	-0.0125	1.7498	-0.0126	1.7590
	400	RDY	-0.0067	0.2595	-0.0069	0.2716
		MOM	-0.0249	3.5270	-0.0251	3.5412
$a = 2$	50	RDY	-0.0016	0.0472	-0.0017	0.0524
		MOM	-0.0074	0.9476	-0.0074	0.9524
	100	RDY	-0.0032	0.0986	-0.0033	0.1077
		MOM	-0.0147	1.8979	-0.0149	1.9063
	200	RDY	-0.0063	0.2143	-0.0065	0.2265
		MOM	-0.0294	3.8068	-0.0296	3.8195
	400	RDY	-0.0124	0.4967	-0.0128	0.5022
		MOM	-0.0583	7.6581	-0.0589	7.6750
$a = 5$	50	RDY	-0.0028	0.0758	-0.0029	0.0826
		MOM	-0.0152	1.6251	-0.0154	1.6315
	100	RDY	-0.0055	0.1574	-0.0057	0.1662
		MOM	-0.0304	3.2528	-0.0307	3.2620
	200	RDY	-0.0107	0.3378	-0.0111	0.3427
		MOM	-0.0605	6.5159	-0.0612	6.5284
	400	RDY	-0.0207	0.7678	-0.0215	0.7488
		MOM	-0.1200	13.0745	-0.1215	13.0915

Exhibit 4: Excess annual return (in percentage points) of original strategies after market impact. The impact is caused by the column strategy with capital amount determined by the value of ϵ in the respective row. Performance is reported for the investment strategy in the corresponding row. We measure performance as the excess annualized return over the performance of that strategy in the original market (Exhibit 2, log returns for the same value of a). Reversal (relative dividend yield, RDY) and time-series momentum (relative strength, MOM). Different degrees of aggressiveness a .

		no transaction costs		with transaction costs			
		impact strategy					
$a = 0.5$	ϵ (in bp)	RDY	MOM	RDY	MOM		
$a = 0.5$	200	RDY	-0.0002	<i>-0.0017</i>	-0.0002	<i>-0.0195</i>	
		MOM	<i>0.0014</i>	0.1327	-0.0002	<i>0.0356</i>	
	400	RDY	-0.0005	-0.0003	-0.0003	<i>-0.0330</i>	
		MOM	<i>0.0031</i>	0.2847	-0.0001	0.1292	
$a = 1$	100	RDY	-0.0004	<i>-0.0031</i>	-0.0004	<i>-0.0206</i>	
		MOM	<i>0.0024</i>	0.1771	-0.0004	0.0418	
	200	RDY	-0.0008	-0.0016	-0.0007	<i>-0.0348</i>	
		MOM	<i>0.0053</i>	0.3756	-0.0003	0.1670	
	400	RDY	-0.0017	<i>0.0154</i>	-0.0014	<i>-0.0421</i>	
		RS	<i>0.0112</i>	0.7967	0.0003	0.4912	
	$a = 2$	50	RDY	-0.0006	<i>-0.0046</i>	-0.0006	<i>-0.0191</i>
			RS	0.0033	0.1453	-0.0006	0.0049
100		RDY	-0.0012	-0.0048	-0.0012	<i>-0.0331</i>	
		RS	0.0073	0.3058	-0.0005	<i>0.0982</i>	
200		RDY	-0.0024	0.0079	-0.0023	<i>-0.0440</i>	
		RS	0.0151	0.6331	0.0001	0.3403	
400		RDY	-0.0046	<i>0.0861</i>	-0.0045	-0.0010	
		RS	<i>0.0320</i>	1.3099	0.0032	0.9139	
$a = 5$		50	RDY	-0.0017	<i>-0.0084</i>	-0.0020	<i>-0.0288</i>
			RS	0.0127	0.1440	0.0020	-0.0119
	100	RDY	-0.0034	-0.0049	-0.0039	<i>-0.0424</i>	
		RS	0.0224	0.3049	0.0026	0.0949	
	200	RDY	-0.0065	<i>0.0375</i>	-0.0076	<i>-0.0271</i>	
		RS	0.0345	0.6303	0.0007	0.3681	
	400	RDY	-0.0120	0.2642	-0.0142	0.1598	
		RS	0.1328	1.2623	0.0684	0.9230	

Exhibit 5: Increase in excess annual return (in percentage points) of recalculated strategies after market impact over the original strategies, all else as in Exhibit 4. Significant changes in performance at a 1% level in a 2-sided paired t-test of more resp. less than 10 basis points are highlighted in bold resp. italic. Cases with lower ϵ are not reported if the change in performance is less than 10 basis points.

prices such that no other strategy flourishes more than the dominant strategy. The answer will be given in the survival test.

Survival test

The survival test is an application of evolutionary finance theory as described in Evstigneev, Hens and Schenk-Hoppé (2006, 2009, 2016). In the survival test the price of the S&P 500 index will be fully determined by the investment strategies and their wealth under management. The price of the index and the wealth of investment strategies all become

endogenous. Only the dividend process remains exogenously given by the data. The price of the S&P 500 index will be determined by the strategies in the market and their wealth under management through short-term equilibrium of demand and supply. The return of the index is then given by the endogenous prices and the exogenous dividends, and the wealth of all investment strategies changes according to these returns.

Suppose there are I investment strategies λ_t^i , $i = 1, \dots, I$ with wealth w_t^i at time t . Assuming the supply of the asset is normalized to one unit, the price is given by³

$$q_t = \lambda_t^1 w_t^1 + \dots + \lambda_t^I w_t^I \quad (1)$$

and the dynamics of wealth, if all investors have a reinvestment rate $0 < \rho \leq 1$ is

$$w_{t+1}^i = \rho \left(\frac{q_{t+1} + d_{t+1}}{q_t} \lambda_t^i + (1 + r_{t+1})(1 - \lambda_t^i) \right) w_t^i. \quad (2)$$

The $t + 1$ -period wealth appears on both sides of the last equation due to the fact that the index is endogenous. The online appendix shows how to solve this equation.

In the survival test, one can set the initial amount of capital to reflect their prevalence in the market. In the example below most capital is held by the equal-weight strategy.

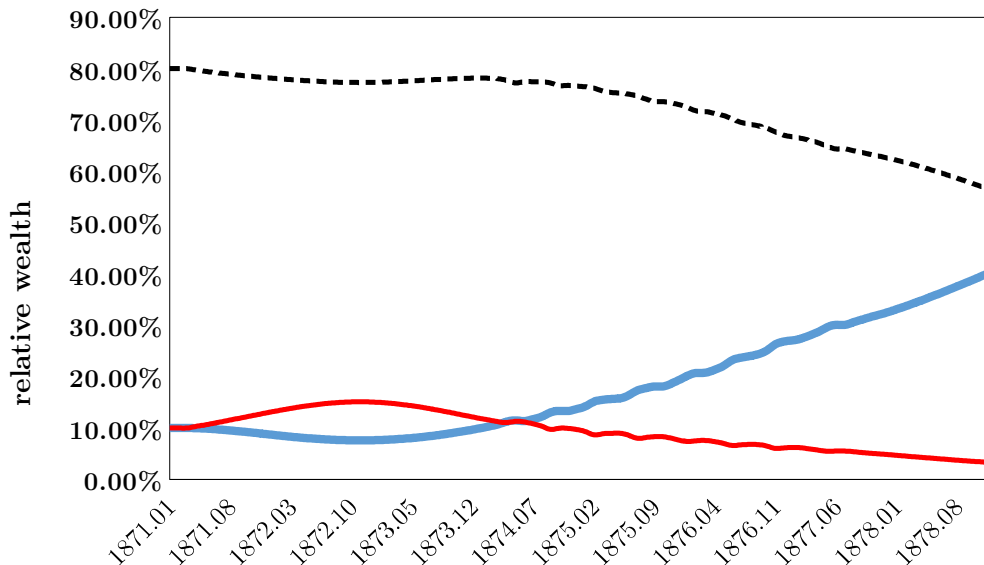


Exhibit 6: Survival test for time period 1871.01-1889.07. $\rho = 0.92$, $a = 1$, $w_{1871.01} = (8, 1, 1)$. Share of total wealth of time-series momentum relative strength (solid line, red), reversal relative dividend yield (bold solid line, blue) and equal weight benchmark (broken line, black). Performance gross of transaction costs.

Exhibit 6 illustrates the endogenous dynamics of wealth for the three investment strategies. The initial capital allocated is $w_0 = (8, 1, 1)$ (in month 1871.01), i.e., the equal-weight strategy holds 80% of the total capital initially whereas the two active strategies each own

³Dividing equation (1) by q_t reveals that this relation states the equality of supply and demand.

10% of the total capital. In the first period, the market capitalization of the index matches roughly that of the Shiller S&P 500 value in January 1871. The exhibit shows market shares, defined as the capital of a strategy divided by the total capital of all the strategies.

In the short-term, relative strength is the only gainer in the market. The effect is as predicted by the impact test. However after about one and a half years, dividend yield starts to gain at the expense of relative strength. This effect is due to the endogenous wealth dynamics. After an initial phase, the growth of capital generated by the relative strength strategy is insufficient to further maintain the increase in the index momentum. The index becomes overpriced which benefits relative dividend yield. There is no qualitative difference between the transaction cost and the no transaction cost case.

		Gains/losses of total capital shares		
		Reversal (rel. div. yield)	Momentum (rel. strength)	Equal weight (benchmark)
		Panel I		
a	1871.01-1878.12, $\rho = 0.9$			
0.5	256.01%	-53.26%	-25.34%	
1	476.34%	-79.40%	-49.62%	
2	627.16%	-91.96%	-66.90%	
5	698.03%	-97.20%	-75.10%	
		Panel II		
ρ	1871.01-1878.12, $a = 1$			
0.92	304.51%	-67.26%	-29.66%	
0.95	91.00%	-40.42%	-6.32%	
0.98	-5.11%	14.16%	-1.13%	
		Panel III		
ρ	1871.01-1886.12, $a = 1$			
0.92	811.41%	-93.01%	-89.80%	
0.95	508.48%	-68.88%	-54.95%	
0.98	76.11%	-49.18%	-3.37%	

Exhibit 7: Survival test, $w_{1871.01} = (8, 1, 1)$. Gains/losses of share of total wealth relative to initial capital share for different degrees of aggressiveness a (during 1871.01-1878.12), and different values of the re-investment rate ρ for fixed $a = 1$ (during 1871.01-1878.12 and 1871.01-1886.12). Gross of transaction costs.

Exhibit 7 provides a more detailed look at the outcome of the survival test and reports the relative gains and losses of capital shares, i.e., the capital of a strategy divided by the sum of all strategies' capital, over a certain time horizon. Denote the share of the total wealth in month t by r_t . Then the gains/losses are given by $(r_T - r_0)/r_0$ with 0 the initial month and T the final month of the time horizon.

Panel I of the Exhibit 7 explores the role of aggressiveness. The value of the parameter a is lower or raised for both relative dividend yield and relative strength. We observe that in a market with more aggressive strategies, the dividend yield strategies performs better.

Or, in other words, market selection works faster.

The role of the re-investment rate is explored in Panel II of Exhibit 7. The lower the re-investment rate, the more a gap in performance matters as more capital has to be spent on consumption. Small difference in performance are amplified by a low re-investment rate as consumption has to be financed by selling assets.

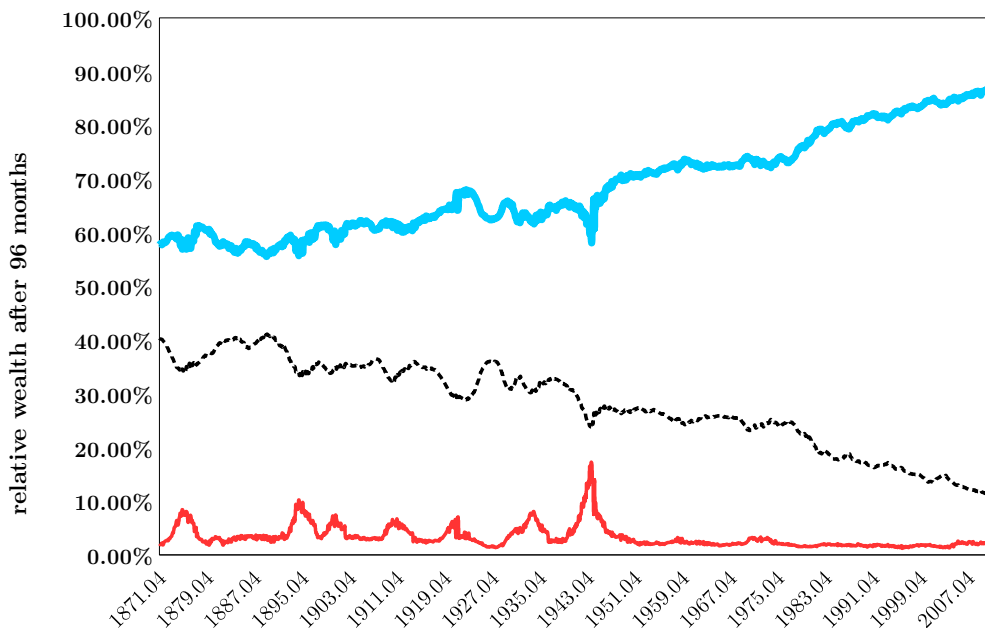


Exhibit 8: Survival test. Wealth share after 96 months for different initial time periods. $\rho = 0.9$, $a = 1$, $w_{initial} = (8, 1, 1)$. Share of total wealth of time-series momentum relative strength (solid line, red), reversal relative dividend yield (bold solid line, blue) and equal weigh benchmark (broken line, black). Performance gross of transaction costs.

Indeed with higher values of ρ market selection works slower. For instance at the end of the shorter time period with $\rho = 0.98$ the market is in the phase that relative strength gains at the expense of the other strategies (the hump appearing in Exhibit 6). But for the longer time period, we see that relative strength eventually loses to the dividend yield strategy. When doubling the time horizon from 8 (Panel I, Exhibit 7) to 16 years (Panel III, Exhibit 7) a ρ of 0.95 roughly produces the same result after 16 years than a ρ of 0.90 over 8 years.

Exhibit 8 depicts the wealth shares of the investment strategies after 8 years (96 month) for different initial time periods, i.e., each data point is the wealth share of the corresponding investment strategy 8 years after the initial period with wealth distribution (8, 1, 1) where the equal weight strategy has capital eight times that of either competing investment strategy. We observe that the results are qualitatively similar to those reported in Exhibit 7 (Panel I). In more recent decades, the dividend yield strategy is doing even better over 8 year horizons than for initial periods in 1871–1940. The same test can be carried out for the scenarios reported in Panel III of Exhibit 7. Again the performance of the investment strategies in different time periods are qualitative with the dividend yield strategy doing even better on

data from more recent decades, see Exhibit 9.

It is instructive to test different distributions of capital, different time periods and investment durations, all of which can be done with the software used (available from the authors upon request).

	Mean wealth share and 5-95 percentile range (all in %)		
	Reversal (rel. div. yield)	Momentum (rel. strength)	Equal weight (benchmark)
ρ	8 years, $a = 1$		
0.92	54.31 [40.32, 75.26]	4.13 [2.32, 8.72]	41.56 [22.28, 55.70]
0.95	32.02 [18.77, 54.85]	5.40 [3.69, 8.43]	62.58 [41.02, 74.63]
0.98	17.35 [9.27, 34.35]	6.78 [4.56, 11.03]	75.88 [59.66, 81.81]
ρ	16 years, $a = 1$		
0.92	93.72 [90.94, 97.50]	0.81 [0.28, 1.70]	5.47 [2.23, 8.21]
0.95	71.12 [60.35, 86.36]	2.21 [1.05, 4.14]	26.67 [12.64, 37.36]
0.98	28.18 [17.03, 50.29]	3.68 [2.63, 5.06]	68.14 [46.83, 78.70]

Exhibit 9: Survival test. Wealth share of investment strategies after 8 years and after 16 years for all possible initial months, $w_{initial} = (8, 1, 1)$. Initial months run from 1871.01 to 2011.01 (8 years and 1681 overlapping periods) resp. from 1871.01 to 2003.01 (16 years and 1585 overlapping periods). Gross of transaction costs.

To summarize, the results of the survival test give a clear message: Among the investment strategies considered, relative dividend yield is the one that survives. This result is supported by Gergaud and Ziemba (2012) who find that among well-known, successful investors most are fundamental investors and not momentum investors.

Conclusion

Thorough testing of the robustness of investment strategies should be on every asset manager's to-do list. While the backtest is in the standard toolbox, our impact and survival tests are not. We illustrate the approaches using Shiller's S&P 500 dataset. The paper shows that relative strength, a time series momentum strategy that does well in the backtest, has a positive impact on itself in the impact test while the relative dividend yield strategy suffers under crowdedness. In the survival test however the fundamental dividend yield strategy shows its true potential and resilience.

While the backtest suggests to enjoy the music as long as it plays, the impact test suggests to momentum strategies to beat the drum and to reversal strategies to keep mum. However,

only the survival test is free from ignoring important feedback effects. It shows that in the long run only the fundamental strategy is sustainable.

The application of our methodology is by no means exhaustive. Asset managers who seek to develop more robust investment strategies should apply these ideas to more sophisticated strategies (such as factors) and larger asset universes.

References

- Asness, C. S. “How Can a Strategy Still Work If Everyone Knows About It?” *AQR Capital Management*, 2015.
- Evstigneev, I. V., T. Hens, and K. R. Schenk-Hoppé. “Evolutionary Stable Stock Markets.” *Economic Theory*, Vol. 27, No. 2 (2006), pp. 449–468.
- . “Evolutionary Finance.” In T. Hens and K. R. Schenk-Hoppé, eds., *Handbook of Financial Markets: Dynamics and Evolution*, pp. 507–566. Amsterdam: North Holland, 2009.
- . “Evolutionary Behavioural Finance.” In E. Haven, P. Molyneux, J.O.S. Wilson, S. Fedotov and M. Duygun, eds., *The Handbook of Post Crisis Financial Modelling*, pp. 214–234. London: Palgrave MacMillan, 2016.
- Fabozzi, F. J., and M. López de Prado. “Being Honest in Backtest Reporting: A Template for Disclosing Multiple Tests.” *The Journal of Portfolio Management*, Vol. 45, No. 1 (2018), pp. 141–147.
- Farmer, D., and A. Lo. “Frontiers of Finance: Evolution and Efficient Markets.” *PNAS*, Vol. 96, No. 18 (1999), pp. 9991–9992.
- Gergaud, O., and W. T. Ziemba. “Great Investors: Their Methods, Results, and Evaluation.” *The Journal of Portfolio Management*, Vol. 38, No. 4 (2012), pp. 128–147.
- Greenwood, R., A. Shleifer, and Y. You. “Bubbles for Fama.” *Journal of Financial Economics*, Vol. 131, No. 1 (2019), pp. 20–43.
- Harvey, C. R., and Y. Liu. “Backtesting.” *The Journal of Portfolio Management*, Vol. 42, No. 1 (2015), pp. 13–28.
- Hurst, B., Y. H. Ooi, and L. H. Pedersen. “A Century of Evidence on Trend-Following Investing.” *The Journal of Portfolio Management*, Vol. 44, No. 1 (2017), pp. 15–29.
- Jones, C. M. “A Century of Stock Market Liquidity and Trading Costs.” Working Paper, 2002.
- Khandani, A., and A. Lo. “What Happened to the Quants in August 2007?” *Journal of Investment Management*, Vol. 5, No. 4 (2007), pp. 5–54.

- . “What Happened to the Quants in August 2007? Evidence from Factors and Transactions Data.” *Journal of Financial Markets*, Vol. 14, No. 1 (2011), pp. 1–46.
- Krkoska, E., and K. R. Schenk-Hoppé. “Herding in Smart-Beta Investment Products.” *Journal of Risk and Financial Management*, Vol. 12, No. 1 (2019), 47.
- Lempérière, Y., C. Deremble, P. Seager, M. Potters, and J. P. Bouchaud. “Two Centuries of Trend Following.” *Journal of Investment Strategies*, Vol. 3, No. 3 (2014), pp. 41–61.
- Lo, A. “Adaptive Markets Financial Evolution at the Speed of Thought.” New Jersey: Princeton University Press, 2017.
- Shiller, R. J. “Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends?” *American Economic Review*, Vol. 71, No. 3 (1981), pp. 421–436.
- . “Portfolio Insurance and Other Investor Fashions as Factors in the 1987 Stock Market Crash.” In S. Fischer, ed., *NBER Macroeconomics Annual 1988, Volume 3*, pp. 287–297. Cambridge, MA: MIT Press, 1988.
- Spare, A. E. “Relative Dividend Yield: Common Stock Investing for Income and Appreciation.” (2nd ed), New York: John Wiley & Sons, 1999.