UCD CENTRE FOR ECONOMIC RESEARCH

WORKING PAPER SERIES

2019

Time-Series Momentum: A Monte-Carlo Approach

Enoch Cheng, University of Colorado
Clemens Struck, University College, Dublin

WP19/06

March 2019

UCD SCHOOL OF ECONOMICS
UNIVERSITY COLLEGE DUBLIN
BELFIELD DUBLIN 4

Time-Series Momentum: A Monte-Carlo Approach*

Enoch Cheng
University of Colorado, Denver

Clemens Struck
University College Dublin

March 3, 2019

Abstract

This paper develops a Monte-Carlo backtesting procedure for risk premia strategies and employs it to study Time-Series Momentum (TSM). Relying on time-series models, empirical residual distributions and copulas we overcome two key drawbacks of conventional backtesting procedures. We create 10,000 paths of different TSM strategies based on the S&P 500 and a cross-asset class futures portfolio. The simulations reveal a probability distribution which shows that strategies that outperform Buy-and-Hold in-sample using historical backtests may out-of-sample i) exhibit sizable tail risks ii) underperform or outperform. Our results are robust to using different time-series models, time periods, asset classes, and risk measures.

JEL: C12, C52, G12, F37
Keywords: Monte-Carlo, Extreme Value Theory, Backtesting, Risk Premia, Time-Series Momentum

*Email (corresponding author): clemens.struck@ucd.ie. For helpful comments we thank Morgan Kelly, David-Michael Lincke, Karl Whelan, and seminar participants at UCD. Annika Sophia Benecke provided excellent research assistance.

1 Introduction

Many investment strategies such as risk-premia strategies are (back-)tested using historical data. Rietz (1988), Barro (2006) and Farhi & Gabaix (2016) emphasize the importance of tail events for understanding risk-premia. Historical data, however, contain few tail events, making it difficult for a backtest to judge i) the true risks involved in a strategy and ii) the chances of a strategy to outperform a benchmark in the long-term.[1]

In order to overcome these backtesting problems, several papers have employed bootstrapping procedures. Yet, existing bootstrapping procedures are unsuitable for our analysis as they i) do not preserve the time-series and cross-sectional dependencies of returns and ii) re-employ the same extreme residuals again and again. Thus, while boostrapping can replicate *past* historic tail events, it cannot generate a sufficiently large variety of possible *future* tail events to robustly examine a strategy. We develop a Monte-Carlo backtesting procedure that addresses these issues.

Monte-Carlo methods are popular in finance to price derivatives, and to stress test balance sheets, yet are rarely used to assess risk-premia strategies. Our approach involves parameterizing several plausible time-series processes by fitting them to actual data, and then empirically estimating the distributions of the residuals and their dependencies across assets using copulas to create new generic residuals. We then backward transform our generated residuals, using the empirical distributions and the estimated time-series processes, to create new returns with the same statistical properties as in the original data. This process gives us many different possible realizations on which to test investment strategies, get an overall distribution of their outcomes, and examine their characteristics.

We illustrate our approach using a set of investment strategies called Time-Series Momentum (TSM). Jegadeesh & Titman (1993) first document Cross-Sectional Momentum (CSM) returns that arise from a market-neutral portfolio of stocks that buys winners and sells losers ranked according

[1] Bailey & López de Prado (2014), among others, criticize these backtesting procedures for hiding the asymptotic properties of investment strategies.

to their past relative performance. More recently, Moskowitz *et al.* (2012) and Hurst *et al.* (2017) have documented the presence of Time-Series Momentum across asset classes. TSM takes a non-market-neutral exposure in, for example, an index based on its past own performance.

It is not well-understood how such a simple rule can earn returns in excess of Buy-and-Hold (BH) in historical backtests, making it the subject of much interest. Fama & French (1996, 2008) call Momentum the "premier anomaly" and a "main embarassment" for the standard theory. Such anomalies are usually explained either as a risk, a bias or a statistical artifact.[2] However, the characteristics of a Time-Series Momentum strategy (it involves trading based on a set of fixed rules and it combines multiple assets) that make it difficult to examine using conventional backtests, also make it suitable to be analyzed within our framework.

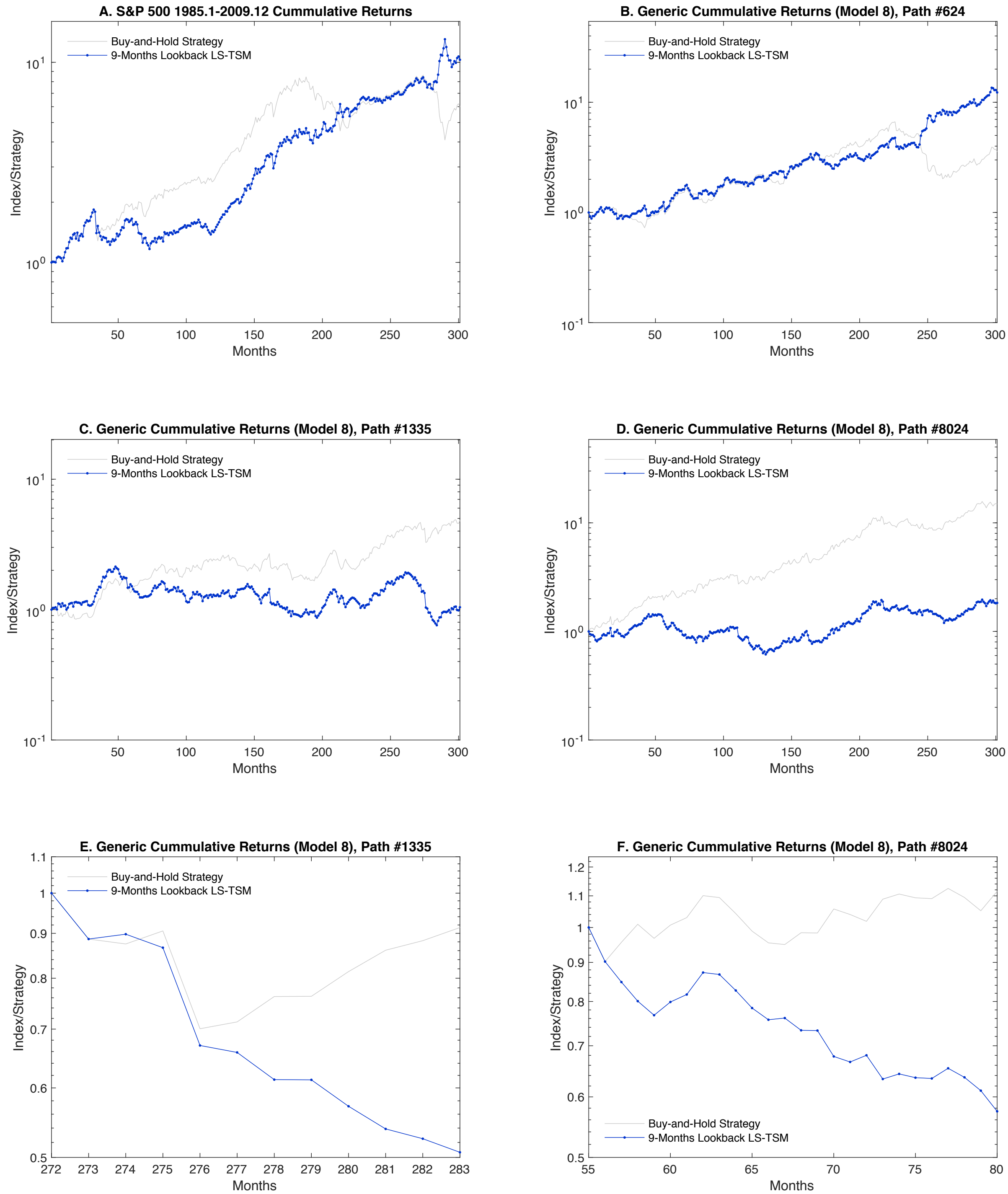
Figure 1 shows how our methodology can be used, for example, to identify hidden risks or to evaluate the chances of a strategy to perform out-of-sample: Panel A of Figure 1 shows a literature standard backtest based on the historical data from 1985.1-2009.12.[3] The data shown in this panel reveal that a Long-Short Time-Series Momentum strategy (LS-TSM) has higher risk-adjusted returns than BH. Tail events are crucial to understand this outperformance. Without the severe market decline around month 200 (during a recession in the early 2000s), the strategy underperforms BH in the overall sample.

Panels B, C, and D show three of the 10,000 simulated paths from an autoregressive model with 24 lags. While Panel B confirms the conclusions of the conventional backtest of Panel A that outperformance of this particular LS-TSM strategy can be a long-run outcome, Panels C and D reveal

[2] The first explanation views return anomalies as a risk compensation and stipulates that asset pricing is rational and can be reconciled with a multi-factor version of the capital asset pricing model (CAPM) or arbitrage pricing theory (APT). Contributors to this line of literature are, among many others, Fama & French (1993, 2016), Jagannathan & Wang (1996), Rietz (1988), Barro (2006), and Farhi & Gabaix (2016). The second views return anomalies as behavioral biases that result in mispricing. Explanations that follow this path include, among many others, De Bondt & Thaler (1990), De Long *et al.* (1990), MacKinlay (1995), Shleifer & Vishny (1997), Grinblatt & Han (2005), DellaVigna & Pollet (2009), and Hirshleifer (2015). The third views return anomalies as statistical artifacts, see e.g. Lo & MacKinlay (1990), Black (1993), Kothari *et al.* (1995), Jorion & Goetzmann (1999), Mclean & Pontiff (2016), and Harvey *et al.* (2016).

[3] Figure 1 computes a Long-Short Time-Series Momentum strategy with a 9-months lookback window on the S&P 500. Such a strategy shorts the S&P 500 if the cumulative return of the past 9 months is negative and buys the S&P500 if the cumulative return of the past 9 months is positive. A rebalancing is conducted at the last day of each month.

Figure 1: Time-Series Momentum vs. Buy-and-Hold



Notes: Panel A shows the performances of an unleveraged Long-Short Time-Series Momentum strategy with a 9-months lookback window and of a Buy-and-Hold strategy on the S&P 500 (monthly rebalanced) during 1985.1 and 2009.12. Table 3 summarizes these findings and reveals that Momentum achieves a monthly Sharpe ratio of 0.20, Buy-and-Hold of 0.16. Panels B,C and D show three of the 10,000 paths obtained through simulations based on an MA(12) model. Panels E and F are two excerpts from Panels C and D in which Momentum significantly underperforms Buy-and-Hold.

that this strategy may also significantly underperform. Later in Section 3 we document that this particular LS-TSM strategy underperforms BH in more than 84% of the 10,000 simulated paths. Other TSM strategies do outperform in both historical backtests and Monte-Carlo simulations.

Likewise, Panels E and F reveal two hidden risks of the LS-TSM strategy. Panel E reveals that, after a drawdown of the market, LS-TSM shorts the market while BH is long when the market recovers, thus, leading to a LS-TSM underperformance of almost 40% over a period of 11 months. Panel F shows that during an innocently looking sideways movement of the market, LS-TSM underperforms by almost 50% in 25 months.

Convincing risk explanations of Momentum remain scarce. Berk *et al.* (1999), Johnson (2002), Sagi & Seasholes (2007) provide rational models based on growth options, Chordia & Shivakumar (2002) and Ahn *et al.* (2003) emphasize macroeconomic risk, Cespa & Vives (2012) and Albuquerque & Miao (2014) highlight the role of information. While many of these explanations may apply to CSM, few apply to TSM which, as a strategy, creates a portfolio exposure based on some past history of returns, and is often not-market neutral. This is why we focus on a CSM strategy that puts strategy crashes into the spotlight, see e.g. Barroso & Santa-Clara (2015) and Daniel & Moskowitz (2016). We differ from these papers in that our method can also be used to reveal out-of-sample risks, risk-return trade-offs, and possible data-snooping bias in the implementation of the strategy.

In order to examine the significance of financial anomalies, out-of-sample tests have often been suggested as another alternative to in-sample historical backtesting, see e.g. Inoue & Kilian (2005), Rapach & Wohar (2006), Hubrich & West (2010), Clark & McCracken (2012), and Rossi & Inoue (2012), and Mclean & Pontiff (2016). Since such tests typically require splitting the historical sample into two or more parts, they suffer to an even larger extent from a problem highlighted earlier, i.e. the low number of tail events, which further reduces when the sample is split.

To overcome this data limitations, researchers create new generic returns by using bootstrapping methods. However, proper bootstrapping of financial data needs to take into consideration the fact that returns are not independent and should not be directly bootstrapped. Instead, adjustments need to be made to preserve the dependency structure of the series. One of those methods, block bootstrapping, involves resampling blocks of returns. A second approach, and similar to our method, involves fitting a statistical model, resampling the residuals, and then generating a new pseudo-series by applying the estimated model to the residuals.

This latter method of bootstrapping has been employed in testing trading rules (Brock *et al.* (1992), Conrad & Kaul (1998), LeBaron (1999), Sullivan *et al.* (1999, 2001), Maillet & Michel (2000), Taylor (2000), Jegadeesh & Titman (2002), White (2000), Karolyi & Kho (2004), Qi & Wu (2006)), Jegadeesh & Titman (2015), fund manager performance (Fama & French (2010), Kosowski *et al.* (2006)), other predictive factors of returns (Harvey *et al.* (2016), Yan & Zheng (2017), and the impact of leverage Engle & Siriwardane (2017). Some papers attempt to preserve the time-series dependencies, while others preserve the cross-sectional dependencies. Yet, none preserve both, and all suffer from the same lack of variety of tail events.

In contrast, our Monte-Carlo procedure overcomes these drawbacks by using time-series models, copulas, and the empirical residual distribution to generate a large variety of new residuals (instead of resampling) while maintaining the dependency structure across residuals and time. Applying the tools of Extreme Value Theory in this way is more commonly used in the risk management literature, which frequently uses Monte-Carlo methods for stress-testing, see e.g. Nyström & Skoglund (2002a,b).[4]

The paper proceeds as follows. Section 2 explains the methodology and its implementation. Section 3 discusses the empirical findings. Section 4 interprets them while Section 5 concludes.

[4]Monte-Carlo methods have been employed by few papers in the literature on market anomalies. Evans & Lewis (1994, 1995) use Monte-Carlo methods to study risk-premia in the bond and in the forward exchange markets, Conrad & Kaul (1998) to assess Momentum (but without preserving the time-series or cross-sectional structure), Bollerslev *et al.* (2011) to study the volatility risk premium, Creal & Wu (2016) to study time variation in bond term premia. The way these papers employ Monte-Carlo methods is not directly related to what we do.

2 Methodology and Implementation

We implement our methodology in four steps: First, we estimate a variety of plausible time-series models (such as autoregressive or moving average models) to statistically describe the historical returns of different assets. Next, using our estimated models, we back out the residuals and determine their distribution using a Kernel estimator. We then apply the inverse of these estimated distributions to a series of new random numbers to obtain generic residuals that have the same statistical properties as the actual ones. Next, we reintroduce statistical features such as autocorrelation and heteroskedasticity from our estimates into the new residuals to create generic returns that have the same statistical properties as the ones in the historical data. Finally, we compute various Momentum and Buy-and-Hold strategies using the generic returns.

2.1 Data Description

We use monthly futures prices on 27 different assets including 9 equity markets futures (S&P 500, DAX, FTSE 100, CAC 40, AEX, IBEX 35, S&P/TSX 60, Nikkei 225, Hang Seng indices), 4 bonds futures (US 30y Treasury Bonds, US 10y Treasury Notes, UK 10y Treasury Bonds, Japanese 10y Treasury Bonds), 8 commodities futures (Gold, Silver, Crude Oil, Unleaded Gasoline, Heating Oil, Cotton, Coffee, Wheat), and 6 currencies futures (EUR/USD, JPY/USD, GBP/USD, CHF/USD, CAD/USD, AUD/USD). Before conducting any calculations, foreign denominated prices are all converted into USD using the USD spot exchange rate. The series are rolled at the end of each month into the front month contract. Prior to the availability of futures contracts we use excess return spot indexes instead. This procedure allows us to construct a cross-asset class portfolio from 1989.2 to 2018.12. For the S&P 500 we can go back even a bit further an obtain an overall sample ranging from 1985.1 to 2018.12.

We conduct all analysis parallel for both the i) S&P 500 and the ii) cross-asset class portfolio of 27 assets, including the S&P500. We divide our data into two periods: i) in-sample 1985.1

to 2009.12 (for the S&P500) and 1989.2 to 2018.12 (for the cross-asset class portfolio) ii) out-of-sample 2009.12-2018.12. The in-sample period is what Moskowitz *et al.* (2012) also use. The out-of-sample period serves as a robustness sample to check the predictions of both conventional and Monte-Carlo backtesting procedures. In our calculations, we rely on excess return indexes (as opposed to total return indexes) as the fundamentally different interest regimes in the two sample periods may undermine any meaningful analysis.

2.2 Step 1: Estimating Time-Series Models

In this section we separately estimate the statistical processes underlying the returns of each of the 27 assets. We rely on autoregressive and moving average time-series models since those models are most suitable to capture the dynamics behind Time-Series Momentum. As the previous literature indicates, Time-Series Momentum effects are strong with lookback windows up to 12 months. This finding determines the maximum number of lags that we choose for our analysis. Since there is no model agreed upon by the literature that perfectly captures all the statistical properties of returns, we estimate eight different basic models. As a robustness test, in Appendix C, we estimate eight additional models with more advanced dynamics. This analysis supports our main conclusions that we reach using the basic models. The sixteen models employed in this paper are not the only models that one could examine, however, we believe they capture the salient characteristics of returns that might make TSM strategies viable. We leave it to the reader to extend our analysis further by including other models. The log return of the asset j is given by

$$r_{j,t} = ln\frac{P_{j,t}}{P_{j,t-1}} \tag{1}$$

where $P_{j,t}$ denotes the asset j's excess return index value at time t. The first set of models that we employ are simple autoregressive **AR(n) Models**. We estimate these models using three, six, nine, and twelve lags, from a period of one quarter up to one year. Formally,

$$r_{j,t} = \mu_j + \sum_{k=1}^{n} \phi_{j,k} r_{j,t-k} + \epsilon_{j,t} \tag{2}$$

where μ_j is a constant; $\phi_{j,k}$ denotes the autoregressive coefficient at lag k; $z_{j,t} = \epsilon_{j,t}/\sigma_{j,t}$ i.i.d. distributed $t(\nu)$ is the standardized residual and modeled as a standardized Student's t-distribution to account for fat tails. In our models, $n = 3, 6, 9, 12$. The second set of models that we employ are simple moving average **MA(n) Models**. Again, we estimate these models using three, six, nine, and twelve lags, from a period of one quarter up to one year. Formally,

$$r_{j,t} = \mu_j + \sum_{k=1}^{n} \theta_{j,k} \epsilon_{j,t-k} + \epsilon_{j,t} \tag{3}$$

where θ denotes the moving average coefficient. To summarize, we estimate the following eight models for each asset class, using the in-sample data: **Model 1:** AR(3), **Model 2:** MA(3), **Model 3:** AR(6), **Model 4:** MA(6), **Model 5:** AR(9), **Model 6:** MA(9), **Model 7:** AR(12), **Model 8:** MA(12).

Table 1, for example, contains the estimated coefficients for these eight models based on the data for the S&P 500 during the period 1985.1-2009.12. As Table 1 shows, the lags in the AR and MA models are mostly insignificant suggesting that Momentum may not be very strong in these data.

2.3 Step 2: Creating New Residuals

We back out the standardized residuals $z_{j,t}$ from the regressions of step 1 and estimate their distribution using a Kernel estimator. Figure 2 plots the residuals probability across all eight models for the S&P 500. Having obtained the distributions, for each of the eight models, we create 10,000 paths of uniformly distributed true random numbers as the new residuals. Each path has a length equivalent to that of the historical in-sample period. We then apply the inverse distribution function to turn the new residuals into generic residuals with distributional properties identical to those of

Table 1: Parameter Estimates for the S&P 500 Excess Return Index (1985.1-2009.12)

	parameter	value	t-stat
Model 1: AR(3)	μ	0.01	4.30
	ϕ_1	0.03	0.69
	ϕ_2	-0.06	-1.24
	ϕ_3	-0.02	-0.53
Model 2: MA(3)	μ	0.01	4.23
	θ_1	0.03	0.66
	θ_2	-0.06	-1.22
	θ_3	-0.02	-0.39
Model 3: AR(6)	μ	0.01	4.06
	ϕ_1	0.03	0.60
	ϕ_2	-0.05	-1.10
	ϕ_3	-0.02	-0.38
	ϕ_4	-0.02	-0.49
	ϕ_5	0.07	1.60
	ϕ_6	-0.01	-0.19
Model 4: MA(6)	μ	0.01	4.01
	θ_1	0.03	0.62
	θ_2	-0.06	-1.21
	θ_3	-0.01	-0.29
	θ_4	-0.03	-0.56
	θ_5	0.08	1.76
	θ_6	-0.01	-0.18
Model 5: AR(9)	μ	0.01	3.81
	ϕ_1	0.03	0.71
	ϕ_2	-0.05	-1.16
	ϕ_3	-0.01	-0.20
	ϕ_4	-0.03	-0.52
	ϕ_5	0.08	1.71
	ϕ_6	-0.01	-0.28
	ϕ_7	0.04	0.78
	ϕ_8	-0.02	-0.39
	ϕ_9	0.04	0.77
Model 6: MA(9)	μ	0.01	3.88
	θ_1	0.03	0.72
	θ_2	-0.06	-1.22
	θ_3	-0.02	-0.32
	θ_4	-0.02	-0.44
	θ_5	0.08	1.68
	θ_6	0.00	-0.07
	θ_7	0.02	0.42
	θ_8	-0.01	-0.25
	θ_9	0.03	0.59
Model 7: AR(12)	μ	0.01	3.76
	ϕ_1	0.03	0.67
	ϕ_2	-0.06	-1.18
	ϕ_3	-0.01	-0.12
	ϕ_4	-0.03	-0.56
	ϕ_5	0.08	1.77
	ϕ_6	-0.02	-0.39
	ϕ_7	0.04	0.80
	ϕ_8	-0.02	-0.45
	ϕ_9	0.04	0.78
	ϕ_{10}	-0.03	-0.69
	ϕ_{11}	0.03	0.58
	ϕ_{12}	0.02	0.31
Model 8: MA(12)	μ	0.01	3.83
	θ_1	0.03	0.71
	θ_2	-0.06	-1.18
	θ_3	-0.01	-0.26
	θ_4	-0.02	-0.41
	θ_5	0.07	1.58
	θ_6	0.00	-0.03
	θ_7	0.02	0.49
	θ_8	-0.01	-0.28
	θ_9	0.03	0.56
	θ_{10}	-0.02	-0.42
	θ_{11}	0.01	0.26
	θ_{12}	0.02	0.36

Notes: This table shows the parameter estimates obtained from the models described in Section 2.2. The underlying data are the monthly log-returns of the S&P 500 during the in-sample period 1985.1-2009.12.

the standardized residuals backed out from the historical data.

For the S&P 500 alone this process is straightforward, yet for the cross-asset class portfolio it is not. The main reason is the dependency of residuals across assets, i.e. the time-varying residual correlation. The appropriate tool that we employ to generate residuals with a time-varying dependency structure relies on copulas. An m-dimensional copula is a joint distribution function on $[0,1]^m$ with all marginal distributions being standard uniform. Let F be a joint distribution function and $F_j, j, ..., m$ be the marginal distributions. Then there exists a copula $C : [0,1]^m \to [0,1]$ such that

$$F(x_1, ..., x_m) = C(F_1(x_1), ..., F_m(x_m)) \quad (4)$$

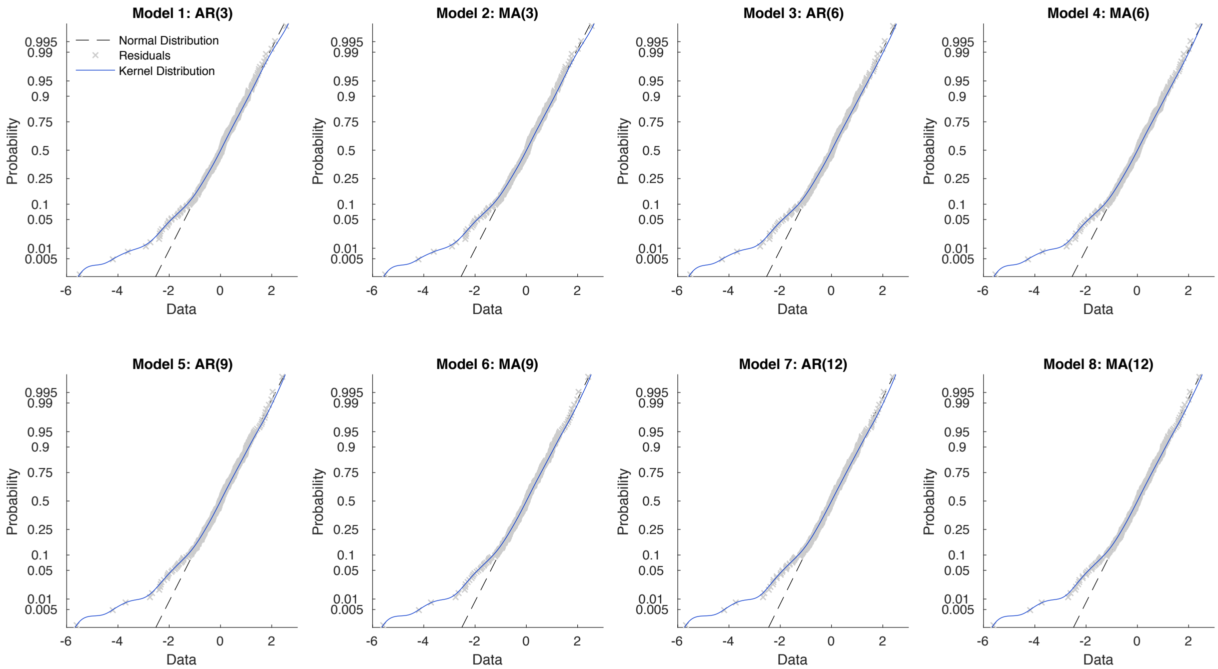
for all $x_1, x_2, ..., x_m \in [-\infty, \infty]$. Let $\Theta = \{(\nu, \Sigma) : \nu \in (1, \infty), \Sigma \in \mathbb{R}^{mxm}\}$. The Student's t-copula can be written as

$$C_\Theta(u_1, u_2, ..., u_m) = \mathbf{t}_{\nu,\Sigma}(\mathbf{t}_\nu^{-1}(u_1), \mathbf{t}_\nu^{-1}(u_2), ..., \mathbf{t}_\nu^{-1}(u_m)) \quad (5)$$

where $\mathbf{t}_{\nu,\Sigma}$ is the multivariate Student's t distribution with a correlation matrix Σ with ν degrees of freedom. When using copulas to describe dependence between asset returns, we do not need to make any assumptions about their marginal distributions. The process of using copulas also allows us to estimate the dependence structure, separately from the marginal distributions. A major benefit of this approach, as opposed to bootstrapping, is that we can generate returns for each asset, and have the aggregate series match the correlation across assets. If however, we were to bootstrap returns, each selection requires us to choose the unique set of returns across the assets at a certain date in order to maintain their cross correlation structure.

We first transform the standardized residuals to uniform residuals by the Kernel empirical CDF and then fit the t-copula to the transformed data. To estimate the parameters ν and Σ for each of the eight models we use a maximum likelihood approach. Once we have obtained the parameter estimates we can use them to generate new uniformly distributed residuals maintaining the correlation structure. We then apply the inverse distribution function to turn the new residuals into generic residuals with distributional properties identical to those of the standardized residuals backed out from the historical data. It is perhaps important to stress that this last step does not impair the correlation structure.

Figure 2: Residual Distribution of the S&P 500 Excess Return Index (1985.1-2009.12)



Notes: This figure shows the Kernel distribution estimates described in Section 2.3 obtained from the residuals backed out from the eight models described in Section 2.2. The underlying data used in the models' estimation are the monthly log-returns of the S&P 500 1985.1-2009.12.

2.4 Step 3: Creating Generic Returns

We use the generic standardized residuals obtained in step 2 as the input noise process into Models 1-8 to obtain generic returns with the autocorrelation patterns observed in the actual data. Table 2 compares the actual with the generic returns for the S&P 500 and the cross-asset class portfolio. For the S&P 500, the table shows that the mean of the monthly generic returns matches the historical data's mean return of 0.60%. For the cross-asset class portfolio the mean is computed using equal weights on all portfolio components. The table shows that the mean generic returns is slightly higher than in the historical data (0.43% vs. 0.2%). For the variance, generic and historical means are identical at second digit precision for both the S&P 500 nor the cross-asset class portfolio.

In order to verify the similarities between the generic and historic return distributions, we conducted a two sample Kolmogorov-Smirnov test with the Null that the distributions are identical. As Table 2, we fail to reject the Null in all of the cases by quite some margin — at the 45%

significance level to be precise. In addition, we also test whether the distributions of squared returns are identical. Again, we cannot reject the Null at the 45% significance level for neither the S&P 500 nor the cross-asset class portfolio. These result support the conclusion that the generic returns have properties comparable to the those of the historical returns.

2.5 Step 4: Computing Investment Strategies

We use both historic and generic returns to compute and compare a variety of Time-Series Momentum investment strategies. Since volatility weighting (risk-parity) may increase investment returns (see e.g. Kim *et al.* (2016)) we compute all our strategies in a volatility adjusted and unadjusted variant. Volatility weighting usually has the goal of making returns across asset classes with different risk levels more comparable (see e.g. Maillard *et al.* (2010)). The adjustment entails weighting the return of asset j in period t by the asset's own past volatility. Formally, this exponentially weighted moving average volatility is given by

$$\phi_{j,t-1} = \frac{\sqrt{12\mu_{S\&P500}}}{\sqrt{\frac{12}{\sum_{i=1}^{12} e^{(1-i)\alpha}} \sum_{k=1}^{12} e^{(1-k)\alpha}(r_{j,t-k} - \mu_{j,t-1:t-12})^2}} \tag{6}$$

where $\mu_{S\&P500}$ is the monthly variance of the S&P 500 log-returns during the in-sample period 1985.1-2009.12 which, as Table 2 shows, amounts to 0.21%. The effect of this term is that the leverage factor for each asset is scaled relative the S&P 500. Thereby, assets with lower volatility than the S&P 500 become more leveraged and assets with higher volatility are less leveraged. The exponentially weighted moving average volatility. $\alpha \in (0, 1)$ is a constant governing the degree of decay of the exponential weighting. $\mu_{j,t-1:t-12}$ denotes the exponentially weighted return variance during the 12 months period.

As benchmark we consider a simple **Buy-and-Hold** strategy. The period t return to this strategy is given by:

Table 2: Actual vs. generic Data

	Actual Data	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
Data: S&P 500, 1985.1-2009.12									
$H0: D(\hat{r}) = D(r)$ (p-value)	-	0.88	0.87	0.80	0.78	0.74	0.77	0.73	0.77
$H0: D(\hat{r}^2) = D(r^2)$ (p-value)	-	0.62	0.61	0.61	0.65	0.73	0.74	0.74	0.78
mean	0.60%	0.60%	0.60%	0.60%	0.60%	0.60%	0.60%	0.60%	0.60%
variance (mean)	0.21%	0.22%	0.22%	0.22%	0.22%	0.22%	0.22%	0.22%	0.22%
Data: Cross-Asset Class Portfolio 1989.2-2009.12									
$H0: D(\hat{r}) = D(r)$ (p-value)	-	0.50	0.51	0.51	0.50	0.50	0.50	0.50	0.46
$H0: D(\hat{r}^2) = D(r^2)$ (p-value)	-	0.4785	0.4902	0.4922	0.50	0.51	0.52	0.52	0.46
mean	0.20%	0.43%	0.42%	0.43%	0.43%	0.43%	0.43%	0.44%	0.44%
variance (mean)	0.08%	0.08%	0.08%	0.08%	0.08%	0.08%	0.08%	0.08%	0.08%

Notes: $^{**},^{*}$ denote 1% and 5% significance, respectively. This table compares actual and generic returns. $H0: F_1(\hat{r}) = F_2(r)$ is the Null Hypothesis of a two sample Kolmogorov-Smirnov test that the empirical distribution of the generic returns,$F_1(\hat{r})$, is equal to the distribution of the actual returns, $F_2(r)$. Similarly, $H0: F_1(\hat{r}^2) = F_2(r^2)$ tests the equivalence of the distributions of the actual squared returns and the generic squared returns. If the p-value is less than 5%, we reject the Null hypothesis that the distributions are equal. *mean* denotes the average return during the sample period. *variance* denotes the variance during the sample period.

$$q_{BH,t} = \frac{1}{|J|}\sum_{j\in J}[\exp(r_{j,t}) - 1]. \tag{7}$$

where $j = 1, 2, ..., 27 \in J$. The volatility weighted benchmark equivalent is given by $q_{BH,vol-adj,t} = \frac{1}{|J|}\sum_{j\in J}\phi_{j,t-1}[\exp(r_{j,t}) - 1]$. As the baseline Momentum strategy, we consider a **Long-Short Momentum** strategy. This strategy considers a signal within a lookback window that determines whether to go long or short the next period's return. The signal to go long or short is the sign of the return that occurred during the lookback window. For example, using a nine-month lookback window, one would short the asset for next period if the returns over those nine months were negative. Various papers in the literature also use different holding-periods, periods to remain long or short after receiving the signal, and they report strong Momentum effects for holding periods less than a year. In our paper, we use a holding period of one month which is common practice.

Our Long-Short strategy will thus be equivalent to the Buy-and-Hold strategy if the cumulative return over the lookback window $t - k - 1 \to t - 1$ is positive, and shorts Buy-and-Hold otherwise. More specifically, the period t return to this strategy is given by

$$q_{LS,t} = \frac{1}{|J|}\sum_{j\in J}[\exp(r_{j,t}) - 1] \times \frac{P_{j,t-1}/P_{j,t-1-k}}{|P_{j,t-1}/P_{j,t-1-k}|} \quad \text{where} \quad k = 1, 3, 6, 9, 12. \tag{8}$$

The volatility adjusted variant is given by $q_{LS,vol-adj,t} = \frac{1}{|J|}\sum_{j\in J}\phi_{j,t-1}\times[\exp(r_{j,t})-1]\times\frac{P_{j,t-1}/P_{j,t-1-k}}{|P_{j,t-1}/P_{j,t-1-k}|}$. As an additional Momentum strategy, we consider a **Long-Cash Momentum**, which yields returns equivalent to Buy-and-Hold if the cumulative return over the lookback window $t - k - 1 \to t - 1$ is positive, and is 0 (not invested in the market) otherwise. More specifically, the return to this strategy is given by

$$q_{LC,t} = \frac{1}{|J|}\sum_{j\in J}[\exp(r_{j,t}) - 1] \times \max\left\{\frac{P_{j,t-1}/P_{j,t-1-k}}{|P_{j,t-1}/P_{j,t-1-k}|}, 0\right\} \quad \text{where} \quad k = 1, 3, 6, 9, 12. \tag{9}$$

The volatility adjusted variant is $q_{LC,vol-adj,t} = \frac{1}{|J|}\sum_{j\in J}\phi_{j,t-1}\times[\exp(r_{j,t})-1]\times\max\left\{\frac{P_{j,t-1}/P_{j,t-1-k}}{|P_{j,t-1}/P_{j,t-1-k}|}, 0\right\}$.

In addition we restrict our Momentum strategies when the cumulative returns reach zero, they will stay at zero instead of potentially becoming infinitely negative. Any real-life implementation of these strategies would require such limited liability. Having defined our core strategies, we will now examine them using a range of parameter values. Appendix A offers, as a robustness test, an alternative Time-Series Momentum implementation that makes use of exponential moving averages. The results obtained under this alternative strategy are similar to those obtained under our main TSM strategy.

3 Empirical Analysis

To evaluate the performance of various strategies, we use two main risk criteria: i) the sample period Sharpe ratio and ii) the maximum drawdown during the sample period. The Sharpe ratio includes a risk-free rate of zero. We do this because we intend to compare a momentum strategy with a Buy-and-Hold strategy - and the risk free rate is the same between the two (given that we implement our strategies using futures). Any conclusion reached about the comparison would be the same regardless of the interest rate. We also consider the maximum drawdown because it allows us to see the asymptotic risks on the downside.

3.1 Historical Backtest

Table 3 presents a summary of the performance in the historical data, as in a conventional backtest. By examining the Sharpe ratio, we can see that a Long-Short Momentum strategy on the S&P 500 outperforms Buy-and-Hold in-sample (1985.1-2009.12) with a lookback window of 9 or 12 months. With shorter lookback windows it does worse. The Long-Cash Momentum on the S&P 500 outperforms Buy-and-Hold in-sample when the lookback window is selected to be 6 months or longer. These results hold independent of any volatility-adjustment. The in-sample peak performance of Momentum occurs at a lookback-window of 9 months. Thus, based on the conventional backtests one should expect out-of sample outperformance particular at this lookback horizon.

By examining the drawdown behavior, we can instantly observe one of the drawbacks of volatility adjustment. The volatility adjusted strategies on the S&P 500 have a worse drawdown behavior compared to the unadjusted variants. Given that these variants leverage it is not surprising. What is a bit surprising is that this behavior is not also captured when examining the Sharpe ratio. Even though the 9M and 12M volatility-adjusted Long-Short strategies perform better in terms at Sharpe Ratio than both their corresponding unadjusted counterparts, they perform significantly worse in terms of their drawdown behavior in-sample. Relative to the LS-TSM strategies on the S&P 500, the LC-TSM strategy achieves both higher Sharpe ratios and a better drawdown behavior in-sample.

Out-of-Sample (2009.12-2018.12), all 9M and 12M strategies on the S&P 500 (vol-adjusted and unadjusted, Long-Short and Long-Cash) underperform their corresponding Buy-and-Hold benchmarks both in terms of Sharpe ratio and drawdown behavior. Long-Cash strategies maintain Sharpe ratios close to the ones observed in-sample, but Long-Short Sharpe ratios collapse across different lookback horizons. Some Long-Cash strategies without volatility adjustment outperform Buy-and-Hold (1m and 6m) in terms of drawdown behavior.

The out-of-sample underperformance is also reflected in the overall sample where the Sharpe ratios of the 9M and 12M unadjusted LS-TSM strategies are 0.15 lower than Buy-and-Hold (0.18). Only the vol-adj. Long-Cash Sharpe ratios are marginally higher than their Buy-and-Hold counterpart (0.21 vs. 0.20). To summarize, for the S&P 500, at a 9M horizon, all Momentum strategies outperform Buy-and-Hold in-sample in terms of their Sharpe ratios but underperform out-of-sample. In the next subsection, we will see that in 2009.12, our Monte-Carlo procedure would have indicated that underperformance is a highly probably outcome in subsequent years.

Turning to the cross-asset class portfolio, we observe Momentum outperformance (relative to Buy-and-Hold) in terms of Sharpe ratios across all strategies in-sample (1988.2-2009.12). Particularly, strategies with a 12M horizon are characterized by much higher Sharpe ratios. In terms of drawdown behavior, all Momentum strategies outperform Buy-and-Hold. Without volatility-adjustment, the the LC strategies beat the LS strategies. With volatility adjustment, at 12M, the

LS strategy beats the LC strategy.

Out-of-Sample (2009.12-2018.12), the Sharpe ratios of all strategies including Buy-and-Hold are significantly lower. At the 12M horizon, the volatility adjusted strategies slightly outperform, the unadjusted strategies do not. Some of the other strategies underperform, some outperform. The drawdown behavior of Momentum is better than that of Buy-and-Hold in all cases. LS and LC perform similarly at all horizons both in terms of Sharpe ratio and drawdown behavior.

In the overall sample (2009.12-2018.12), all Momentum strategies outperform Buy-and-Hold in terms of their drawdown behavior. At the 12M horizon the Sharpe ratios of Momentum is higher than Buy-and-Hold. In summary, conventional backtests predict outperformance out-of-sample of all 12M strategies which is unobserved in subsequent years. Our Monte-Carlo analysis in the next subsection reveal the probabilities of outperformance for individual strategies. It thus offers a rationale for why some of these 12M strategies which outperform in sample, subsequently underperform out-of-sample, while others outperform both in- and out-of-sample.

3.2 Monte-Carlo Backtest

Table 4 presents the summary results of our Monte-Carlo analysis. We only focus on strategies that outperform Buy-and-Hold in-sample using the conventional backtest, in order to emphasize one of our main arguments — that in-sample outperformance does not imply out-of-sample outperformance or the opposite. This is because conventional backtests cannot (by construction) speak to the probability distribution of outperformance. Hence, our focus is on the 9M strategy for the S&P 500 and the 12M strategy for the cross-asset class portfolio.

A benefit of our methodology is by examining the distribution of returns, we can assign probabilities of outperformance. The upper half of Table 4 compares the best Momentum strategy to a Buy-and-Hold strategy on the S&P 500 using simulations created from the 1985.1-2009.12 in-sample, to examine these *(*probabilities). The probabilities shown in the table are computed as the fraction of outperforming outcomes relative to the total 10,000 simulated paths. Despite using the

Table 3: Risk Behavior of TSM Momentum Strategies (Historical Data)

	S&P500						Cross-Asset Class Portfolio					
	In-Sample 1985.1-2009.12		**Out-of–Sample** 2009.12-2018.12		**Overall Sample** 1985.1-2018.12		**In-Sample** 1989.2-2009.12		**Out-of–Sample** 2009.12-2018.12		**Overall Sample** 1989.2-2018.12	
	SR	**DD**	**SR**	**DD**	**SR**	**DD**	**SR**	**DD**	**SR**	**DD**	**SR**	**DD**
Buy-and-Hold												
Buy-and-Hold	0.16	-52.6%	0.24	-17.0%	0.18	-52.6%	0.14	-36.7%	0.01	-28.4%	0.10	-36.7%
Buy-and-Hold (vol.-adj.)	0.17	-54.4%	0.26	-22.9%	0.20	-54.4%	0.19	-25.7%	0.01	-27.9%	0.13	-27.9%
LS-TSM												
1M Momentum	0.04	-44.0%	-0.03	-31.1%	0.03	-44.0%	0.15	-15.0%	-0.08	-18.4%	0.09	-22.5%
3M Momentum	0.05	-49.6%	0.05	-32.6%	0.05	-49.6%	0.17	-9.2%	0.02	-16.1%	0.13	-16.1%
6M Momentum	0.12	-58.4%	0.04	-29.0%	0.11	-58.4%	0.16	-21.6%	0.07	-16.0%	0.13	-21.6%
9M Momentum	0.20	-36.4%	-0.02	-45.4%	0.15	-52.4%	0.19	-21.4%	0.04	-14.5%	0.15	-24.5%
12M Momentum	0.18	-39.6%	0.06	-31.9%	0.15	-39.6%	0.23	-22.2%	0.00	-17.8%	0.16	-32.6%
1M Mom (vol.-adj.)	0.03	-63.2%	0.03	-52.3%	0.03	-63.2%	0.22	-9.0%	-0.07	-15.0%	0.13	-18.1%
3M Mom (vol.-adj.)	0.06	-55.2%	0.10	-50.4%	0.07	-55.2%	0.23	-8.1%	0.08	-15.2%	0.18	-15.2%
6M Mom (vol.-adj.)	0.14	-63.8%	0.13	-34.5%	0.14	-63.8%	0.24	-12.1%	0.11	-11.0%	0.20	-12.1%
9M Mom (vol.-adj.)	0.21	-37.4%	0.09	-51.8%	0.18	-52.6%	0.29	-11.3%	0.06	-11.9%	0.22	-16.5%
12M Mom (vol.-adj.)	0.20	-39.0%	0.14	-38.4%	0.19	-39.0%	0.33	-11.2%	0.03	-14.0%	0.23	-21.6%
LC-TSM												
1M Momentum	0.15	-30.3%	0.15	-15.5%	0.15	-30.3%	0.23	-13.1%	-0.04	-18.2%	0.15	-18.2%
3M Momentum	0.14	-27.0%	0.20	-17.3%	0.16	-27.0%	0.24	-10.1%	0.02	-17.5%	0.17	-17.5%
6M Momentum	0.19	-23.7%	0.18	-14.0%	0.19	-23.7%	0.23	-10.2%	0.05	-14.4%	0.17	-14.4%
9M Momentum	0.23	-26.6%	0.14	-21.3%	0.21	-26.6%	0.26	-9.6%	0.03	-14.2%	0.19	-14.2%
12M Momentum	0.21	-30.2%	0.17	-17.5%	0.20	-30.2%	0.28	-11.8%	0.00	-16.4%	0.19	-16.4%
1M Mom (vol.-adj.)	0.14	-32.0%	0.18	-24.1%	0.15	-32.0%	0.28	-9.6%	-0.03	-17.8%	0.19	-17.8%
3M Mom (vol.-adj.)	0.14	-33.3%	0.21	-28.5%	0.16	-33.3%	0.28	-9.3%	0.05	-14.0%	0.21	-14.0%
6M Mom (vol.-adj.)	0.19	-30.5%	0.22	-24.4%	0.20	-30.5%	0.28	-6.7%	0.07	-11.6%	0.21	-11.6%
9M Mom (vol.-adj.)	0.22	-32.0%	0.19	-30.9%	0.21	-33.6%	0.31	-6.5%	0.04	-13.0%	0.22	-13.0%
12M Mom (vol.-adj.)	0.21	-33.0%	0.21	-23.7%	0.21	-33.0%	0.32	-8.2%	0.03	-14.1%	0.23	-14.1%

Notes: This table summarizes the risk-adjusted performance of the different Momentum strategies in the historical data. There are three periods: in-sample 1985.1-2009.12, out-of-sample 2009.12-2018.12, and overall sample 1985.1-2018.12 for the S&P 500. For the cross-asset class portfolio the in-sample and overall sample periods begin in 1989.2 due to a lack of data availability. 1M means 1 month. The definition of each strategy are given in Section 2.5. 'SR' denotes the the monthly Sharpe ratio calculated under the assumption of a risk free rate of 0%. 'DD' denotes the maximum drawdown during the sample period.

Table 4: Risk Behavior of TSM Momentum Strategies (Monte-Carlo)

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
Data: S&P 500, 1985.1-2009.12								
Sharpe(LS-TSM 9M) > Sharpe(Buy-Hold)	2.6%	2.8%	4.0%	4.3%	7.4%	7.0%	7.2%	7.1%
Sharpe(LS-TSM 9M vol. adj.) > Sharpe(Buy-Hold)	3.8%	3.9%	5.5%	5.7%	8.5%	8.0%	8.3%	8.2%
Sharpe(LC-TSM 9M) > Sharpe(Buy-Hold)	13.8%	14.2%	17.9%	19.1%	25.8%	24.4%	24.6%	24.8%
Sharpe(LC-TSM 9M vol. adj.) > Sharpe(Buy-Hold)	12.8%	13.2%	16.6%	17.6%	23.6%	22.8%	22.7%	22.7%
Data: S&P 500, 1985.1-2009.12								
Max DD(LS-TSM 9M) < Max DD(Buy-Hold)	88.1%	87.7%	84.8%	84.0%	78.9%	80.0%	79.3%	79.1%
Max DD(LS-TSM 9M vol. adj.) < Max DD(Buy-Hold)	87.0%	86.7%	84.0%	83.4%	78.4%	79.1%	78.%	78.7%
Max DD(LC-TSM 9M) < Max DD(Buy-Hold)	43.0%	42.5%	38.9%	37.7%	33.0%	33.5%	33.2%	33.1%
Max DD(LC-TSM 9M vol. adj.) < Max DD(Buy-Hold)	48.7%	48.0%	44.2%	43.2%	38.1%	38.5%	38.7%	38.3%
Data: Cross-Asset Class Portfolio 1989.2-2009.12								
Sharpe(LS-TSM 12M) > Sharpe(Buy-Hold)	16.0%	13.6%	10.0%	11.6%	17.9%	14.2%	39.9%	52.9%
Sharpe(LS-TSM 12M vol. adj.) > Sharpe(Buy-Hold)	19.2%	18.2%	9.3%	11.1%	22.8%	17.3%	48.1%	52.6%
Sharpe(LC-TSM 12M) > Sharpe(Buy-Hold)	51.5%	49.0%	40.7%	43.9%	56.1%	49.4%	82.0%	89.9%
Sharpe(LC-TSM 12M vol. adj.) > Sharpe(Buy-Hold)	54.5%	52.1%	36.5%	40.3%	60.9%	52.2%	85.2%	88.2%
Data: Cross-Asset Class Portfolio, 1989.2-2009.12								
Max DD(LS-TSM 12M) < Max DD(Buy-Hold)	43.5%	44.5%	50.4%	47.6%	40.1%	43.5%	23.9%	16.8%
Max DD(LS-TSM 12M vol. adj.) < Max DD(Buy-Hold)	36.2%	36.7%	48.1%	44.5%	31.9%	35.6%	17.3%	14.1%
Max DD(LC-TSM 12M) < Max DD(Buy-Hold)	4.8%	5.2%	6.2%	5.6%	4.5%	4.8%	1.8%	1.1%
Max DD(LC-TSM 12M vol. adj.) < Max DD(Buy-Hold)	4.2%	4.5%	7.0%	6.1%	3.5%	4.4%	1.4%	1.0%

Notes: Sharpe(LS-TSM 9M) > Sharpe(Buy-and-Hold) denotes the fraction of the 10,000 simulated paths (in %) in which the Sharpe Ratio of a volatility unadjusted 9M Long-Short Momentum strategy is higher than that of a Buy-and-Hold strategy. Similarly, Max DD(LS-TSM 9M) < Max DD(Buy-Hold) denotes the fraction of the 10,000 simulated paths (in %) in which the drawdown of a volatility unadjusted 9M Long-Short Momentum strategy is worse than that of a Buy-and-Hold strategy. The definition of each strategy are given in Section 2.5.

best lookback window (9M), the methodology indicates poor probabilities for outperformance.

The probability of outperforming Buy-and-Hold in terms of the Sharpe ratio is less than 10% for the Long-Short Momentum volatility adjusted and unadjusted strategies. For the Long-Cash strategies, the probability is better, but still less than 25%. In terms of drawdown behavior, the probability that the two Long-Short Momentum strategies outperform is less than 25%. The probability that the Long-Cash strategies outperform is more than 50%.

It is important to notice that the probabilities are similar in terms of their qualitative predictions across the eight models.[5] For example, the Long-Short volatility un-adjusted Momentum strategy has a probability of realizing a better Sharpe ratio of 2.6% under Model 1 and 7.2% under Model 7. While the numbers are quantitatively distinct, both models have the same message: outperformance is unlikely, only occurring in less than 10% of the 10,000 paths.

How do these findings relate to the findings in the historical backtests of Table 3? The simulations predict that, in particular, the Long-Short strategies on the S&P 500 have a low probability of outperforming Buy-and-Hold. This prediction also became clear in the light of the evidence in Panel A of Figure 1 which essentially showed that in-sample outperformance was driven by a single crisis. The Monte-Carlo predictions are realized by the out-of-sample underperformance shown in Table 3. They also are realized by the relative out-of-sample outperformance of the Long-Cash over the Long-Short TSM strategy.

Turning to the simulations of the cross-asset class portfolio, the lower half of Table 4 shows the performances of various strategies with a 12M lookback window. They indicate, at best, an approximately 50-50 chance for the Long-Short strategies to outperform Buy-and-Hold in terms of the Sharpe ratio (Models 7 and 8). According to these models the strategies should outperform with 75% probability in terms of their drawdown behavior. The range of probabilities across models lie between 50% and 85% though. We do stress however, that given the 12M lookback window, we should focus on the MA and AR models that include all the lags within the window, namely Models

[5] The additional 8 models that we examine in Appendix C also make similar predictions.

7 and 8.

The Long-Cash strategies again have a higher probability of outperformance compared to the Long-Short strategies. That is both in terms of the Sharpe ratio and drawdown behavior. In Models 7 and 8, the outperformance probability for the Sharpe ratio is above 80%, and the probability of drawdown outperformance is greater than 95%. In the historical backtests of Table 3, outperformance of all 12M Momentum strategies in terms of drawdown behavior occurs both in- and out-of-sample, consistent with the Monte-Carlo simulated probabilities, and implying that this was a highly likely outcome.

With respect to the Sharpe ratio, all portfolios deliver weak out-of-sample performance which is mainly due to the decline in commodity prices during this period. Of the Long-Short strategies, one strategy underperforms, while another strategy outperforms, broadly consistent with the Monte-Carlo probabilities of 50-50. Of the Long-Cash strategies, one outperforms while another performs similar to the Buy-and-Hold out-of-sample in Table 3. These two latter outcomes are again broadly consistent with our Monte-Carlo probabilities that indicate an outperformance probability between 35% and 90%.

We now turn our focus on the two strategies that Monte-Carlo identifies as the best and weakest strategies. The 9M Long-Short volatility unadjusted Momentum on the S&P 500 and the 12M Long-Cash volatility unadjusted Momentum on the cross-asset class portfolio. Again, the in-sample historical backtests of Table 3 suggest that both of these strategies should outperform out-of-sample in terms of their Sharpe ratios. Figure 3 visualizes the distribution of the 10,000 cumulative returns at the end of the simulated paths.

The Panels of this figure reveal two key problems: first, the lowest cumulative returns of the Momentum strategy are lower than those of Buy-and-Hold, implying that the drawdown behavior can be worse and that there are asymptotic risks attached to this strategy which the historical backtest of Panel A of Figure 1 does not show; secondly, the peak of the cumulative return distribution for a momentum strategy is left of that of Buy-and-Hold, implying that underperformance is more likely than outperformance.

The latter of these two findings is not surprising. The insignificant AR and MA parameter estimates of Table 1 already suggested that evidence for structures in the data that allow for effective Momentum strategies during 1985.1-2009.12 in the S&P 500 data is rather weak. Given the weak significance of the estimates, and the fact that the stock market tends to increase over time, it is not surprising to see a strategy that shorts the market sometimes underperform a strategy that is long-only.

Figure 4 plots the cumulative return distributions for the 12M Long-Cash cross-asset class portfolio. Two things are noteworthy. First, the Figure shows that the cumulative return of Momentum does not fall as low as the cumulative return of Buy-and-Hold across all eight models. Secondly, the Buy-and-Hold distribution is significantly more right-tailed than the Momentum distribution. The former observation implies that this Momentum strategy has better drawdown behavior. The latter observation implies that the better risk behavior (relative to Buy-and-Hold) comes at the expense of giving up upside potential.

To summarize, by employing our Monte-Carlo approach, we can identify probabilities of outperformance for TSM strategies relative to a benchmark. We then employed these probabilities to rationalize findings observed in the historical data. In that data, strategies that outperformed the benchmark in-sample tended to consequently out- or underperform out-of-sample. The Monte-Carlo probabilities also identified the worst performing strategy (9M LS-TSM volatility unadjusted on the S&P 500) as the one least likely to outperform.

3.3 Bootstrap Backtest

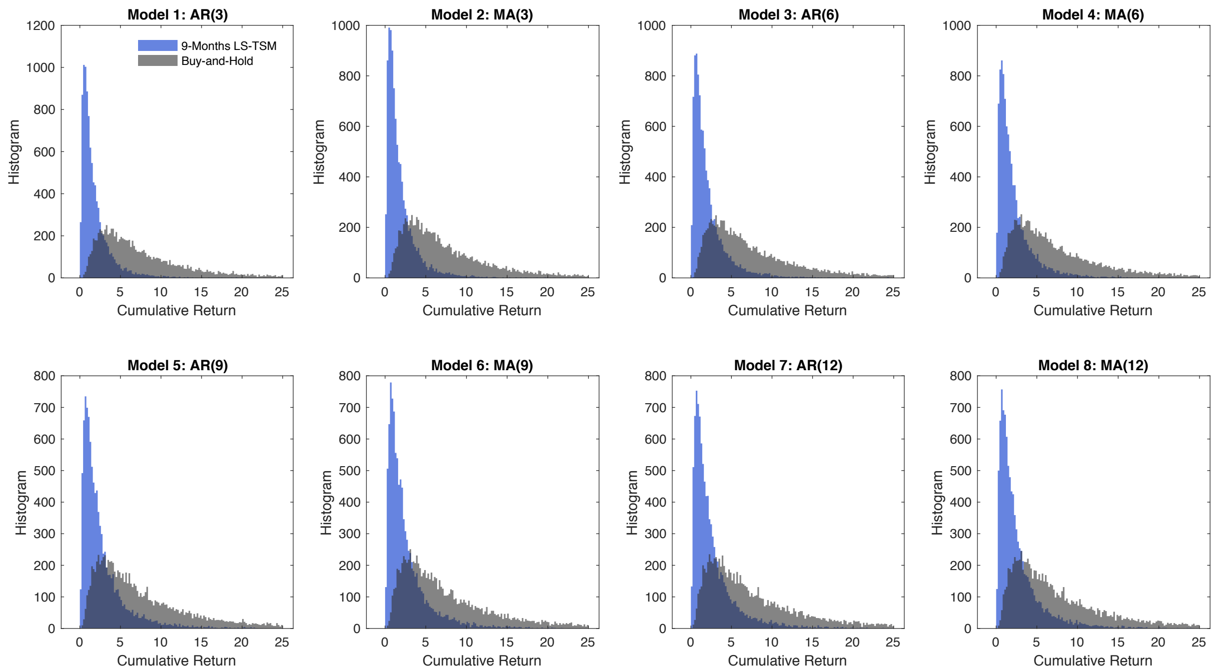
We previously listed some of the theoretical reasons why Monte-Carlo is a more appropriate procedure than bootstrapping when used in examining investment strategies particularly in the context of multi-asset portfolios. In this section, we empirically show that a bootstrapping procedure leads to vastly different predictions compared to the Monte-Carlo simulations.

Table 5: Risk Behavior of TSM Momentum Strategies (Bootstrapping)

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
Data: S&P 500, 1985.1-2009.12								
Sharpe(LS-TSM 9M) > Sharpe(Buy-Hold)	2.6%	2.4%	3.8%	4.3%	6.7%	6.5%	7.1%	6.7%
Sharpe(LS-TSM 9M vol. adj.) > Sharpe(Buy-Hold)	3.8%	3.7%	5.3%	5.4%	7.6%	7.7%	8.0%	7.7%
Sharpe(LC-TSM 9M) > Sharpe(Buy-Hold)	13.8%	13.3%	17.6%	18.4%	24.9%	23.6%	24.7%	24.0%
Sharpe(LC-TSM 9M vol. adj.) > Sharpe(Buy-Hold)	13.3%	12.6%	16.8%	17.3%	22.7%	21.9%	22.8%	22.2%
Data: S&P 500, 1985.1-2009.12								
Max DD(LS-TSM 9M) < Max DD(Buy-Hold)	87.3%	87.9%	84.5%	84.2%	79.3%	80.0%	79.4%	79.4%
Max DD(LS-TSM 9M vol. adj.) < Max DD(Buy-Hold)	86.3%	87.2%	84.3%	83.3%	78.6%	80.2%	79.4%	78.8%
Max DD(LC-TSM 9M) < Max DD(Buy-Hold)	43.7%	43.0%	39.4%	39.1%	33.9%	35.7%	34.5%	34.3%
Max DD(LC-TSM 9M vol. adj.) < Max DD(Buy-Hold)	49.9%	49.8%	45.6%	45.1%	40.5%	41.7%	40.7%	41.1%
Data: Cross-Asset Class Portfolio 1989.2-2009.12								
Sharpe(LS-TSM 12M) > Sharpe(Buy-Hold)	0.7%	0.5%	0.1%	0.3%	0.7%	0.4%	7.0%	14.0%
Sharpe(LS-TSM 12M vol. adj.) > Sharpe(Buy-Hold)	0.4%	0.5%	0.0%	0.2%	0.9%	0.3%	7.4%	8.7%
Sharpe(LC-TSM 12M) > Sharpe(Buy-Hold)	31.7%	29.7%	19.7%	22.9%	37.0%	28.6%	74.0%	86.7%
Sharpe(LC-TSM 12M vol. adj.) > Sharpe(Buy-Hold)	29.4%	28.4%	11.6%	14.1%	38.5%	26.2%	75.1%	80.3%
Data: Cross-Asset Class Portfolio, 1989.2-2009.12								
Max DD(LS-TSM 12M) < Max DD(Buy-Hold)	90.4%	90.7%	93.9%	93.5%	89.7%	91.5%	77.5%	68.0%
Max DD(LS-TSM 12M vol. adj.) < Max DD(Buy-Hold)	89.0%	88.8%	94.5%	93.9%	87.0%	91.0%	74.5%	71.7%
Max DD(LC-TSM 12M) < Max DD(Buy-Hold)	20.3%	20.7%	24.9%	22.4%	18.5%	21.0%	11.5%	8.4%
Max DD(LC-TSM 12M vol. adj.) < Max DD(Buy-Hold)	23.6%	23.8%	30.5%	29.0%	21.0%	24.2%	13.6%	12.4%

Notes: Sharpe(LS-TSM 9M) > Sharpe(Buy-and-Hold) denotes the fraction of the 10,000 simulated paths (in %) in which the Sharpe Ratio of a volatility unadjusted 9M Long-Short Momentum strategy is higher than that of a Buy-and-Hold strategy. Similarly, Max DD(LS-TSM 9M) < Max DD(Buy-Hold) denotes the fraction of the 10,000 simulated paths (in %) in which the drawdown of a volatility unadjusted 9M Long-Short Momentum strategy is worse than that of a Buy-and-Hold strategy. The definition of each strategy are given in Section 2.5.

Figure 3: Cumulative Returns (Monte-Carlo, S&P 500)



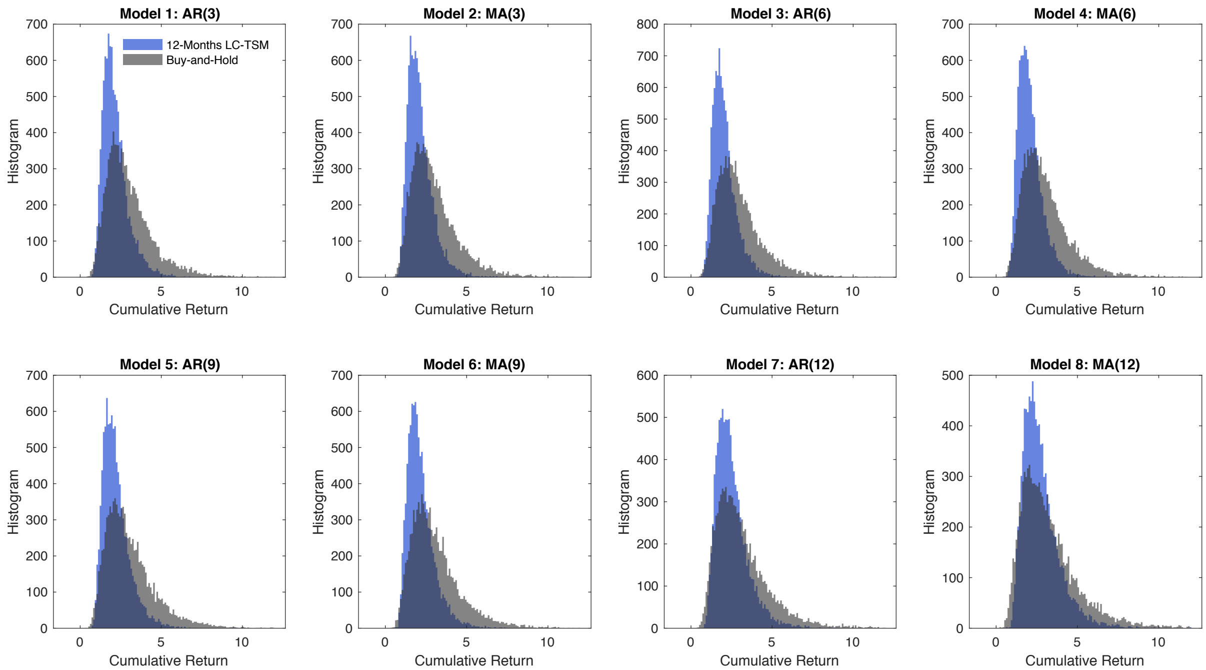
Notes: The figure shows the distribution of the cumulative returns at the ends of the 10,000 simulated paths for both the volatility unadjusted 9M LS-TSM strategy on the S&P 500 as well as the Buy-and-Hold strategy. The definition of each strategy are given in Section 2.5.

Table 5 shows the results of a bootstrapping analysis in which the residuals (that were backed out after the estimation in step 1) are randomly drawn to create 10,000 paths for each of the eight models. For the univariate time-series (the analysis of the S&P 500), this approach leads to probabilities similar to those of Table 3. This is to be expected since there is no cross-asset dependency structure for a single asset.[6] It is worth mentioning that, although the probabilities are similar, the simulated tail events will not offer as much variety of events as under Monte-Carlo. A rare event study such as the one conducted in Figure 1 may therefore miss the more interesting events that can happen to a strategy.

Focusing on the cross-asset class portfolio, we observe vast differences between 5 and 3. While the Monte-Carlo simulations imply that the 12M LS-TSM strategies have a reasonably *high* chance of outperforming the benchmark, particularly, in terms of drawdown, the bootstrapping simula-

[6] A potential solution is cross-sectional bootstrapping whereby one draws the whole cross-section of residuals at a certain point in time. This procedure has been employed by Kosowski *et al.* (2006), Fama & French (2010) and Yan & Zheng (2017) for example, yet these approaches do not attempt to preserve the time-series structure.

Figure 4: Cumulative Returns (Monte-Carlo, Cross-Asset Class Portfolio)



Notes: The figure shows the distribution of the cumulative returns at the ends of the 10,000 simulated paths for both the volatility unadjusted 12M LC-TSM strategy on the the cross-asset class portfolio as well as the Buy-and-Hold strategy. The definition of each strategy are given in Section 2.5.

tions imply that such strategies will have a very *low* chance of outperforming out-of-sample both in terms of Sharpe-ratio and drawdown behavior. In the light of the historical in- and out-of-sample outperformance in terms of the drawdown behavior of these strategies, the bootstrapping results are difficult to reconcile with the actual data.

Likewise, the 12M LC-TSM cross-asset class portfolio analysis seems distorted by bootstrapping. In Model 6, for example, the Sharpe-ratio outperformance probability of the volatility adjusted strategy is 52.2% under Monte-Carlo, but only 26.2% under bootstrapping. Similarly, the drawdown behavior predictions, although qualitatively broadly consistent, substantially differ quantitatively. Under Model 3, for example, the probability of drawdown underperformance of the volatility adjusted strategy is 6.8% using Monte-Carlo, but 30.5% using bootstrapping.

4 Discussion

Conventional backtests of risk-premia strategies are mainly based on historical data. Few papers employ simulations. Those who do rely on bootstrapping rather than Monte-Carlo methods. Bootstrapping is problematic for at least two key reasons: i) it corrupts the time-series and cross-sectional dependencies of a series; ii) it re-employs the same extreme residuals again and again. It is plausible, and conservative, to assume that future extreme events will look different compared to the past and simulations should therefore employ a variety of extreme residuals. We develop a Monte-Carlo procedure that addresses these two issues by using a combination of time-series models, empirical residual distributions and copulas to generate new residuals with a realistic dependency structure. We then apply this procedure to analyze a variety of Time-Series Momentum Strategies.

By analyzing such strategies first using conventional backtests, we find that some of them outperform a benchmark in-sample yet underperform out-of-sample, while others also outperform out-of-sample. By construction, conventional backtests fail to tell us much about the chances to outperform out-of-sample. In fact, a naive interpretation of a back-test would be to assume that in-sample outperformance (underperformance) also implies out-of-sample outperformance (underperformance). Our Monte-Carlo procedure shows the dangers of this logic by explicitly revealing a probability distribution of strategy outcomes.

This distribution can be used to not only reconcile in-sample outperformance (underperformance) with out-of-sample underperformance (outperformance), but also in-sample outperformance (underperformance) with out-of-sample outperformance (underperformance). Furthermore, we can employ the probability distribution to separate robust and weaker strategies. To separate strategies, we focused only on those strategies that outperform in-sample using conventional backtests. Our Monte-Carlo procedure then identified one particularly weak strategy that subsequently underperformed out-of-sample in the historical backtest.[7]

[7] It is beyond the scope of his paper to explore this latter channel further. In the future we plan to investigate the procedure's ability to separate strategies more systematically.

In addition to revealing the probability distribution, the Monte-Carlo procedure also allows us to study possible future outcomes during tail events. As alluded to earlier, this feature of our approach is important because historical backtests contain few tail events. In the historical backtests it is often a single tail event that determines a strategy's overall performance. Figure 1, for example, illustrates how a strategy (9M LS-TSM vol.-unadjusted on the S&P 500) might behave during possible future tail events. The figure reveals possible Momentum Crashes in economically bad times (when the overall market declines or moves sideways).

The simulations help us interpret whether a strategy's outperformance is a risk premium, a behavioral bias or a statistical artifact. For example, in the case of the 9M LS-TSM vol.-unadjusted strategy on the S&P 500, we find that the strategy embodies sizable crash risks and also that the observed in-sample outperformance during 1985.1-2009.12 is rather unlikely to repeat. In the case of the 12M LC-TSM vol.-unadjusted strategy on the cross-asset class portfolio, the simulations suggest that the crash risk is rather low, but that the improved risk behavior is paid for by giving up a good part of the right tail of the cumulative return distribution. The Sharpe-ratio, a simplified risk-return measure improves, but due to its simplicity, masks this trade-off.

What is clear is that performance can vary widely between lookback windows and assets, and the reasons for a lookback window/asset combination to outperform or underperform, may not be the same reasons for another. However, by looking at their empirical distributions, we see that some strategies exhibit outcomes that have characteristics of having a risk-premium - paying to give up crash risks (12M LC-TSM vol.-unadjusted); and some strategies seem to have outcomes that are statistical artifacts (9M LS-TSM vol.-unadjusted). This uncertainty creates real risks in the choice of which window/asset combination to use

We also want to highlight a potential drawback of our approach (and other simulation approaches such as bootstrapping in general). We make the assumption that the past structure of Time-Series Momentum will be similar to its future structure. More concretely, the assumption is that the the estimated parameters of Table 1, the empirical distributions of Figure 2, and the Copula parameters will be similar in the future. While the latter two assumption are less stringent- as the empirical distributions and copula parameters are empirically observed to be relatively constant even over long-periods of time, the former assumption is much stronger.

Appendix B provides a robustness test of the Monte-Carlo simulations that addresses this issue. The parameters, distributions and copula are estimated from the overall sample (-2018.12) instead of the in-sample period (-2009.12). The Monte-Carlo implications are very consistent with those estimated from a shorter period. The simulations again identify the two 9M LS-TSM strategies on the S&P 500 as the weakest strategies most likely to underperform the benchmark, and the two 12M LC-TSM strategies on the cross-asset class portfolio as the most robust strategies.

Another potential drawback of our analysis is the limited choice of time-series models that we have employed. While the eight AR and MA models that we use are a natural fit in the context of Time-Series Momentum, it is not clear whether Momentum may also be connected to heteroskedasticity of error terms. Appendix C addresses this issue by employing a set of eight additional models that use asymmetric GARCH models. The Monte-Carlo computations shown in Table 9 are again very consistent with the main computations shown in Table 9. The outperformance probability of Momentum is slightly higher across strategies, which confirms our prior that heteroskedasticity may also be a factor behind Momentum.

Finally, we need to keep in mind, that absence of evidence is not evidence of absence. In the analysis, we use a variety of statistical time-series models to describe historical asset returns. Based on these models, and using simulations, we observe that some of these Momentum strategies such as the 9M LS-TSM strategies on the S&P 500 underperform in the majority of cases. Thus, we find absence of evidence for Momentum outperformance in these strategies - yet we cannot confidently reject Momentum outperformance as the time-series models that we use potentially do not capture the deeper reason for why Momentum works. Unfortunately, there may be a statistical dimension important for Momentum that these models simply miss.

5 Conclusions

Tail events are important drivers for the performance of many investment strategies. Historical data contain few tail events. Conventional backtests that rely on such data therefore have difficulties informing us about i) the true risks involved in a strategy and ii) the long-run chance of a strategy outperforming a benchmark.

We develop a Monte-Carlo backtesting procedure amenable to analyzing risk-premia strategies. Our approach combines elements from the risk management literature with elements from a small literature that uses bootstrap procedures to test risk-premia strategies. It overcomes two key limitations of standard bootstrapping procedures that i) corrupt the time-series and cross-sectional dependencies of returns and ii) re-employ the same extreme residuals again and again.

Our method involves parameterizing several plausible time-series processes by fitting them to actual data, empirically estimating the distributions of the residuals and the residual dependencies across assets using copulas to create new generic residuals. We then backward transform our generated residuals, using the empirical distributions and time-series processes, to create returns with statistical properties similar to those of the original data.

We illustrate our approach by analyzing a set of investment strategies called Time-Series Momentum (TSM). We show that our approach reveals both the out-of-sample risks of such strategies and the probability of their outperformance in the long-term. We find that strategies that outperform in-sample using historical backtests may, out-of-sample, underperform or outperform consistent with the predictions of our simulations. We also use this additional information to categorize different TSM strategies as a risk-premia strategy or a statistical artifact. In the future we plan to examine other risk-premia strategies using this approach.

References

Ahn, Dong-Hyun, Conrad, Jennifer, & Dittmar, Robert F. 2003. Risk Adjustment and Trading Strategies. *The Review of Financial Studies*, **16**(2), 459–485.

Albuquerque, Rui, & Miao, Jianjun. 2014. Advance information and asset prices. *Journal of Economic Theory*, **149**, 236–275.

Bailey, David H., & López de Prado, Marcos. 2014. The Deflated Sharpe Ratio: Correcting for Selection Bias, Backtest Overfitting, and Non-Normality. *The Journal of Portfolio Management*, **40**(5), 94–107.

Barro, Robert J. 2006. Rare Disasters and Asset Markets in the Twentieth Century*. *The Quarterly Journal of Economics*, **121**(3), 823–866.

Barroso, Pedro, & Santa-Clara, Pedro. 2015. Momentum has its moments. *Journal of Financial Economics*, **116**(1), 111 – 120.

Berk, Jonathan B., Green, Richard C., & Naik, Vasant. 1999. Optimal Investment, Growth Options, and Security Returns. *The Journal of Finance*, **54**(5), 1553–1607.

Black, F. 1976. Studies of Stock Price Volatility Changes. *Proceedings of the Business and Economics Section of the American Statistical Association*, 177–181.

Black, Fischer. 1993. Beta and Return. *The Journal of Portfolio Management*, **20**(1), 8.

Bollerslev, Tim, Gibson, Michael, & Zhou, Hao. 2011. Dynamic estimation of volatility risk premia and investor risk aversion from option-implied and realized volatilities. *Journal of Econometrics*, **160**(1), 235 – 245. Realized Volatility.

Brock, William, Lakonishok, Josef, & Lebaron, Blake. 1992. Simple Technical Trading Rules and the Stochastic Properties of Stock Returns. *Journal of Finance*, **47**(5), 1731–64.

Cespa, Giovanni, & Vives, Xavier. 2012. Dynamic Trading and Asset Prices: Keynes vs. Hayek. *The Review of Economic Studies*, **79**(2), 539–580.

Chordia, Tarun, & Shivakumar, Lakshmanan. 2002. Momentum, Business Cycle, and Time-varying Expected Returns. *The Journal of Finance*, **57**(2), 985–1019.

Clark, Todd E., & McCracken, Michael W. 2012. Reality Checks and Comparisons of Nested Predictive Models. *Journal of Business & Economic Statistics*, **30**(1), 53–66.

Conrad, Jennifer, & Kaul, Gautam. 1998. An Anatomy of Trading Strategies. *Review of Financial Studies*, **11**(3), 489–519.

Creal, Drew D, & Wu, Jing Cynthia. 2016. Bond Risk Premia in Consumption-based Models. *National Bureau of Economic Research Working Paper Series*, **No. 22183**(April 2016).

Daniel, Kent, & Moskowitz, Tobias J. 2016. Momentum crashes. *Journal of Financial Economics*, **122**(2), 221 – 247.

De Bondt, Werner F M, & Thaler, Richard. 1990. Do Security Analysts Overreact? *American Economic Review*, **80**(2), 52–57.

De Long, J. Bradford, Shleifer, Andrei, Summers, Lawrence H., & Waldmann, Robert J. 1990. Noise Trader Risk in Financial Markets. *Journal of Political Economy*, **98**(4), 703–738.

DellaVigna, Stefano, & Pollet, Joshua M. 2009. Investor Inattention and Friday Earnings Announcements. *The Journal of Finance*, **64**(2), 709–749.

Engle, Robert F., & Siriwardane, Emil N. 2017. Structural GARCH: The Volatility-Leverage Connection. *The Review of Financial Studies*, **31**(2), 449–492.

Evans, Martin D.D, & Lewis, Karen K. 1994. Do stationary risk premia explain it all?: Evidence from the term structure. *Journal of Monetary Economics*, **33**(2), 285 – 318.

Evans, Martin D.D., & Lewis, Karen K. 1995. Do Long-Term Swings in the Dollar Affect Estimates of the Risk Premia? *The Review of Financial Studies*, **8**(3), 709–742.

Fama, Eugene F., & French, Kenneth R. 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, **33**(1), 3–56.

Fama, Eugene F., & French, Kenneth R. 1996. Multifactor Explanations of Asset Pricing Anomalies. *The Journal of Finance*, **51**(1), 55–84.

Fama, Eugene F., & French, Kenneth R. 2008. Dissecting Anomalies. *The Journal of Finance*, **63**(4), 1653–1678.

Fama, Eugene F., & French, Kenneth R. 2010. Luck versus Skill in the Cross-Section of Mutual Fund Returns. *The Journal of Finance*, **65**(5), 1915–1947.

Fama, Eugene F., & French, Kenneth R. 2016. Dissecting Anomalies with a Five-Factor Model. *The Review of Financial Studies*, **29**(1), 69–103.

Farhi, Emmanuel, & Gabaix, Xavier. 2016. Rare Disasters and Exchange Rates. *Quarterly Journal of Economics*, **131**(1), 1–52. Lead Article.

Glosten, Lawrence R, Jagannathan, Ravi, & Runkle, David E. 1993. On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks. *Journal of Finance*, **48**(5), 1779–1801.

Grinblatt, Mark, & Han, Bing. 2005. Prospect theory, mental accounting, and momentum. *Journal of Financial Economics*, **78**(2), 311 – 339.

Harvey, Campbell R., Liu, Yan, & Zhu, Heqing. 2016. ...and the Cross-Section of Expected Returns. *The Review of Financial Studies*, **29**(1), 5–68.

Hirshleifer, David. 2015. Behavioral Finance. *Annual Review of Financial Economics*, **7**(1), 133–159.

Hubrich, Kirsten, & West, Kenneth D. 2010. Forecast evaluation of small nested model sets. *Journal of Applied Econometrics*, **25**(4), 574–594.

Hurst, Brian, Ooi, Yao Hua, & Pedersen, Lasse Heje. 2017. A Century of Evidence on Trend-Following Investing. *The Journal of Portfolio Management*, **44**(1), 15.

Inoue, Atsushi, & Kilian, Lutz. 2005. In-Sample or Out-of-Sample Tests of Predictability: Which One Should We Use? *Econometric Reviews*, **23**(4), 371–402.

Jagannathan, Ravi, & Wang, Zhenyu. 1996. The Conditional CAPM and the Cross-Section of Expected Returns. *The Journal of Finance*, **51**(1), 3–53.

Jegadeesh, Narasimhan, & Titman, Sheridan. 1993. Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency. *The Journal of Finance*, **48**(1), 65–91.

Jegadeesh, Narasimhan, & Titman, Sheridan. 2002. Cross-Sectional and Time-Series Determinants of Momentum Returns. *The Review of Financial Studies*, **15**(1), 143–157.

Jegadeesh, Narasimhan, & Titman, Sheridan. 2015. Cross-Sectional and Time-Series Determinants of Momentum Returns. *The Review of Financial Studies*, **15**(1), 143–157.

Johnson, Timothy C. 2002. Rational Momentum Effects. *The Journal of Finance*, **57**(2), 585–608.

Jorion, Philippe, & Goetzmann, William N. 1999. Global Stock Markets in the Twentieth Century. *The Journal of Finance*, **54**(3), 953–980.

Karolyi, G., & Kho, Bong-Chan. 2004. Momentum strategies: some bootstrap tests. *Journal of Empirical Finance*, **11**(4), 509–536.

Kim, Abby Y., Tse, Yiuman, & Wald, John K. 2016. Time series momentum and volatility scaling. *Journal of Financial Markets*, **30**, 103–124.

Kosowski, Robert, TImmermann, Allan, Wermers, Russ, & White, Hal. 2006. Can Mutual Fund "Stars"Really Pick Stocks? New Evidence from a Bootstrap Analysis. *The Journal of Finance*, **61**(6), 2551–2595.

Kothari, S.P., Shanken, Jay, & Sloan, Richard G. 1995. Another Look at the Cross-section of Expected Stock Returns. *The Journal of Finance*, **50**(1), 185–224.

LeBaron, Blake. 1999. Technical trading rule profitability and foreign exchange intervention. *Journal of International Economics*, **49**(1), 125 – 143.

Li, David X. 2000. On Default Correlation. *The Journal of Fixed Income*, **9**(4), 43.

Lo, Andrew, & MacKinlay, A Craig. 1990. Data-Snooping Biases in Tests of Financial Asset Pricing Models. *Review of Financial Studies*, **3**(3), 431–67.

MacKenzie, Donald, & Spears, Taylor. 2014. 'The formula that killed Wall Street': The Gaussian copula and modelling practices in investment banking. *Social Studies of Science*, **44**(3), 393–417.

MacKinlay, A. Craig. 1995. Multifactor models do not explain deviations from the CAPM. *Journal of Financial Economics*, **38**(1), 3–28.

Maillard, Sébastien, Roncalli, Thierry, & Teïletche, Jérôme. 2010. The Properties of Equally Weighted Risk Contribution Portfolios. *The Journal of Portfolio Management*, **36**(4), 60.

Maillet, Bertrand, & Michel, Thierry. 2000. Further insights on the puzzle of technical analysis profitability. *The European Journal of Finance*, **6**(2), 196–224.

Mclean, R. David, & Pontiff, Jeffrey. 2016. Does Academic Research Destroy Stock Return Predictability? *The Journal of Finance*, **71**(1), 5–32.

Moskowitz, Tobias J., Ooi, Yao Hua, & Pedersen, Lasse Heje. 2012. Time series momentum. *Journal of Financial Economics*, **104**(2), 228 – 250. Special Issue on Investor Sentiment.

Nelson, Daniel B. 1991. Conditional Heteroskedasticity in Asset Returns: A New Approach. *Econometrica*, **59**(2), 347–370.

Nyström, Kaj, & Skoglund, Jimmy. 2002a. *A Framework for Scenariobased Risk Management*. Preprint. Swedbank.

Nyström, Kaj, & Skoglund, Jimmy. 2002b. *Univariate Extreme Value Theory, GARCH and Measures of Risk*. Preprint. Swedbank.

Qi, Min, & Wu, Yangru. 2006. Technical Trading-Rule Profitability, Data Snooping, and Reality Check: Evidence from the Foreign Exchange Market. *Journal of Money, Credit and Banking*, **38**(8), 2135–2158.

Rapach, David E., & Wohar, Mark E. 2006. In-sample vs. out-of-sample tests of stock return predictability in the context of data mining. *Journal of Empirical Finance*, **13**(2), 231 – 247.

Rietz, Thomas A. 1988. The equity risk premium a solution. *Journal of Monetary Economics*, **22**(1), 117–131.

Rossi, Barbara, & Inoue, Atsushi. 2012. Out-of-Sample Forecast Tests Robust to the Choice of Window Size. *Journal of Business & Economic Statistics*, **30**(3), 432–453.

Sagi, Jacob S., & Seasholes, Mark. 2007. Firm-specific attributes and the cross-section of momentum. *Journal of Financial Economics*, **84**(2), 389–434.

Shleifer, Andrei, & Vishny, Robert W. 1997. The Limits of Arbitrage. *The Journal of Finance*, **52**(1), 35–55.

Sullivan, Ryan, Timmermann, Allan, & White, Halbert. 1999. Data-Snooping, Technical Trading Rule Performance, and the Bootstrap. *The Journal of Finance*, **54**(5), 1647–1691.

Sullivan, Ryan, Timmermann, Allan, & White, Halbert. 2001. Dangers of data mining: The case of calendar effects in stock returns. *Journal of Econometrics*, **105**(1), 249 – 286. Forecasting and empirical methods in finance and macroeconomics.

Taylor, Stephen J. 2000. Stock index and price dynamics in the UK and the US: new evidence from a trading rule and statistical analysis. *The European Journal of Finance*, **6**(1), 39–69.

White, Halbert. 2000. A Reality Check for Data Snooping. *Econometrica*, **68**(5), 1097–1126.

Yan, Xuemin (Sterling), & Zheng, Lingling. 2017. Fundamental Analysis and the Cross-Section of Stock Returns: A Data-Mining Approach. *The Review of Financial Studies*, **30**(4), 1382–1423.

A Robustness: An Alternative TSM Strategy

As a robustness test we consider an alternative implementation of Time-Series Momentum. As an estimator for the trend, $\xi_{j,t-1}$, of asset j we use the exponential weighted moving average of past observations,

$$\xi_{j,t-1} = \frac{\sum_{k=1}^{\infty} \lambda^k q_{j,t-k}}{|\sum_{k=1}^{\infty} \lambda^k q_{j,t-k}|} \tag{10}$$

where $\lambda \in (0,1)$ is a parameter that governs the decay. We examine a variety of parameters that are in the neighborhood of those parameters that (ex-post) lead to the highest Sharpe ratios in the in-sample backtests, i.e. $\lambda \in \{0.99, 0.95, 0.90, 0.85, 0.80\}$. A higher (lower) λ gives more (less) weight to more recent return realizations. We calculate the **Long-Short EMA** strategy (LS-EMA) return as:

$$q_{LS-EMA,t} = \frac{1}{|J|} \sum_{j \in J} [\exp(r_{j,t}) - 1] \times \xi_{j,t-1}. \tag{11}$$

where $q_{LS-EMA,vol-adj,t} = \frac{1}{|J|} \sum_{j \in J} \phi_{j,t-1} \times [\exp(r_{j,t}) - 1] \times \xi_{j,t-1}$ is the volatility adjusted variant. If the estimated trend is negative, the exposure becomes short. If the trend is positive the exposure is long. We also implement a **Long-Cash EMA** (LC-EMA). More specifically, the return to this strategy is given by

$$q_{LC-EMA,t} = \frac{1}{|J|} \sum_{j \in J} [\exp(r_{j,t}) - 1] \times \max\{\xi_{j,t-1}, 0\}. \tag{12}$$

where $q_{LC-EMA,vol-adj,t} = \frac{1}{|J|} \sum_{j \in J} \phi_{j,t-1} \times [\exp(r_{j,t}) - 1] \times \max\{\xi_{j,t-1}, 0\}$ is the volatility adjusted variant. Table 6 shows the historical backtest results and 7 shows the Monte-Carlo backtests. Table 6 shows that the peak performance (in terms of the Sharpe ratio) for Momentum strategies on the S&P 500 and the cross-asset class portfolio occurs at $\lambda = 0.85$. The results are similar to those results from Table 3: i) in-sample outperformance does not necessarily translate into out-of-sample outperformance ii) strategies with volatility leveraging exhibit a weaker drawdown behavior than strategies that are volatility unadjusted.

Focusing on the most successful strategies (those with $\lambda = 0.85$), Table 7 again emphasizes that the Long-Short strategies on the S&P 500 are likely to underperform while the Long-Cash strategies fair better. Turning to the strategies on the cross-asset class portfolio, here too, Monte-Carlo highlights the Long-Cash strategies as superior to their Long-Short counterparts both in terms of their drawdown behavior and their long-run Sharpe-ratios. The cross-asset class Long-Cash strategies are also the best performing ones by some distance. Overall, the Monte-Carlo results derived from this alternative Momentum implementation are quantitatively similar to those found in Table 4.

Table 6: Risk Behavior of EMA Momentum Strategy (Historical Data)

	S&P500						Cross-Asset Class Portfolio					
	In-Sample 1985.1-2009.12		**Out-of–Sample** 2009.12-2018.12		**Overall Sample** 1985.1-2018.12		**In-Sample** 1989.2-2009.12		**Out-of–Sample** 2009.12-2018.12		**Overall Sample** 1989.2-2018.12	
	SR	**DD**	**SR**	**DD**	**SR**	**DD**	**SR**	**DD**	**SR**	**DD**	**SR**	**DD**
Buy-and-Hold												
Buy-and-Hold	0.16	-52.56%	0.24	-17.0%	0.18	-52.6%	0.14	-36.7%	0.01	-28.4%	0.10	-36.7%
Buy-and-Hold (vol.-adj.)	0.17	-54.41%	0.26	-22.9%	0.20	-54.4%	0.19	-25.7%	0.01	-27.9%	0.13	-27.9%
LS-EMA												
$\lambda = 0.99$ Momentum	0.14	-56.61%	0.24	-17.0%	0.17	-56.6%	0.17	-28.6%	0.03	-17.9%	0.12	-28.6%
$\lambda = 0.95$ Momentum	0.16	-37.66%	0.09	-38.1%	0.14	-58.8%	0.18	-17.9%	-0.07	-20.5%	0.11	-27.8%
$\lambda = 0.90$ Momentum	0.20	-29.8%	0.00	-38.1%	0.15	-49.5%	0.21	-20.2%	-0.01	-16.4%	0.15	-23.2%
$\lambda = 0.85$ Momentum	0.21	-35.41%	0.05	-29.7%	0.17	-38.6%	0.21	-21.1%	0.03	-15.4%	0.16	-23.1%
$\lambda = 0.80$ Momentum	0.16	-48.14%	0.02	-29.7%	0.13	-48.1%	0.19	-20.1%	0.01	-16.7%	0.14	-23.2%
$\lambda = 0.99$ Mom (vol.-adj.)	0.17	-57.24%	0.26	-22.9%	0.19	-57.2%	0.21	-23.3%	0.02	-17.1%	0.15	-23.3%
$\lambda = 0.95$ Mom (vol.-adj.)	0.18	-36.95%	0.17	-45.0%	0.18	-57.8%	0.26	-12.2%	-0.04	-17.1%	0.17	-21.0%
$\lambda = 0.90$ Mom (vol.-adj.)	0.21	-31.58%	0.09	-45.0%	0.18	-48.0%	0.30	-10.8%	0.02	-12.8%	0.22	-12.8%
$\lambda = 0.85$ Mom (vol.-adj.)	0.23	-36.78%	0.13	-34.0%	0.20	-36.8%	0.30	-11.1%	0.09	-9.5%	0.24	-12.4%
$\lambda = 0.80$ Mom (vol.-adj.)	0.18	-50.88%	0.12	-34.4%	0.17	-50.9%	0.29	-11.0%	0.07	-9.9%	0.22	-13.6%
LC-EMA												
$\lambda = 0.99$ Momentum	0.15	-52.6%	0.24	-17.0%	0.17	-52.6%	0.16	-32.7%	0.02	-22.7%	0.11	-32.7%
$\lambda = 0.95$ Momentum	0.19	-31.5%	0.18	-18.4%	0.19	-32.2%	0.21	-19.3%	-0.02	-21.5%	0.14	-21.5%
$\lambda = 0.90$ Momentum	0.22	-23.7%	0.14	-18.4%	0.20	-23.7%	0.25	-13.5%	0.00	-17.9%	0.17	-17.9%
$\lambda = 0.85$ Momentum	0.23	-26.4%	0.17	-17.3%	0.22	-26.4%	0.26	-9.6%	0.02	-15.6%	0.19	-15.6%
$\lambda = 0.80$ Momentum	0.21	-29.4%	0.16	-17.3%	0.20	-29.4%	0.25	-10.6%	0.01	-15.5%	0.18	-15.5%
$\lambda = 0.99$ Mom (vol.-adj.)	0.17	-54.4%	0.26	-22.9%	0.20	-54.4%	0.21	-24.0%	0.02	-22.4%	0.14	-24.0%
$\lambda = 0.95$ Mom (vol.-adj.)	0.19	-33.3%	0.22	-27.5%	0.20	-47.0%	0.26	-15.8%	-0.01	-18.8%	0.17	-18.8%
$\lambda = 0.90$ Mom (vol.-adj.)	0.21	-29.9%	0.19	-27.5%	0.21	-39.9%	0.30	-10.0%	0.02	-16.1%	0.21	-16.1%
$\lambda = 0.85$ Mom (vol.-adj.)	0.22	-31.8%	0.21	-22.5%	0.22	-31.8%	0.31	-6.7%	0.05	-11.6%	0.23	-11.6%
$\lambda = 0.80$ Mom (vol.-adj.)	0.21	-33.7%	0.20	-24.1%	0.21	-33.7%	0.30	-7.9%	0.04	-11.8%	0.22	-11.8%

Notes: This table summarizes the risk-adjusted performance of the different Momentum strategies in the historical data. There are three periods: in-sample 1985.1-2009.12, out-of-sample 2009.12-2018.12, and overall sample 1985.1-2018.12 for the S&P 500. For the cross-asset class portfolio the in-sample and overall sample periods begin in 1989.2 due to a lack of data availability. 1M means 1 month. The definition of each strategy are given in Section A. 'SR' denotes the the monthly Sharpe ratio calculated under the assumption of a risk free rate of 0%. 'DD' denotes the maximum drawdown during the sample period.

Table 7: Risk Behavior of EMA Momentum Strategy (Monte-Carlo)

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
Data: S&P 500, 1985.1-2009.12								
Sharpe(LS-EMA $\lambda = 0.85$) > Sharpe(Buy-Hold)	2.9%	3.0%	4.2%	4.5%	6.8%	6.3%	6.6%	6.4%
Sharpe(LS-EMA $\lambda = 0.85$ vol. adj.) > Sharpe(Buy-Hold)	4.0%	4.2%	5.6%	5.8%	8.1%	7.7%	8.0%	7.9%
Sharpe(LC-EMA $\lambda = 0.85$) > Sharpe(Buy-Hold)	13.1%	13.8%	17.2%	17.5%	22.0%	20.8%	21.7%	21.5%
Sharpe(LC-EMA $\lambda = 0.85$ vol. adj.) > Sharpe(Buy-Hold)	13.1%	13.5%	16.3%	16.4%	20.9%	19.7%	20.4%	20.1%
Data: S&P 500, 1985.1-2009.12								
Max DD(LS-EMA $\lambda = 0.85$) < Max DD(Buy-Hold)	85.4%	84.9%	82.8%	82.6%	79.0%	80.0%	79.4%	79.5%
Max DD(LS-EMA $\lambda = 0.85$ vol. adj.) < Max DD(Buy-Hold)	85.2%	84.6%	82.5%	82.3%	79.0%	79.8%	79.2%	79.5%
Max DD(LC-EMA $\lambda = 0.85$) < Max DD(Buy-Hold)	45.7%	45.0%	42.8%	42.6%	39.2%	39.8%	40.0%	39.5%
Max DD(LC-EMA $\lambda = 0.85$ vol. adj.) < Max DD(Buy-Hold)	50.6%	50.1%	47.4%	47.1%	43.3%	44.1%	44.0%	44.1%
Data: Cross-Asset Class Portfolio 1989.2-2009.12								
Sharpe(LS-EMA $\lambda = 0.85$) > Sharpe(Buy-Hold)	23.2%	21.9%	14.8%	17.3%	20.6%	18.3%	32.2%	45.2%
Sharpe(LS-EMA $\lambda = 0.85$ vol. adj.) > Sharpe(Buy-Hold)	27.7%	27.4%	13.6%	17.3%	23.8%	19.9%	38.2%	45.8%
Sharpe(LC-EMA $\lambda = 0.85$) > Sharpe(Buy-Hold)	60.0%	58.7%	46.7%	50.8%	54.5%	52.7%	70.1%	81.2%
Sharpe(LC-EMA $\lambda = 0.85$ vol. adj.) > Sharpe(Buy-Hold)	62.9%	62.4%	43.3%	48.8%	57.4%	53.2%	73.8%	80.0%
Data: Cross-Asset Class Portfolio, 1989.2-2009.12								
Max DD(LS-EMA $\lambda = 0.85$) < Max DD(Buy-Hold)	34.8%	34.9%	41.3%	39.3%	36.7%	37.6%	28.5%	21.3%
Max DD(LS-EMA $\lambda = 0.85$ vol. adj.) < Max DD(Buy-Hold)	27.8%	28.7%	39.5%	34.9%	30.8%	33.3%	21.8%	17.0%
Max DD(LC-EMA $\lambda = 0.85$) < Max DD(Buy-Hold)	5.1%	5.6%	6.7%	6.1%	6.0%	5.7%	3.9%	2.7%
Max DD(LC-EMA $\lambda = 0.85$ vol. adj.) < Max DD(Buy-Hold)	4.0%	3.9%	6.0%	5.7%	4.5%	4.9%	3.1%	2.0%

Notes: Sharpe(LS-EMA $\lambda = 0.85$) > Sharpe(Buy-and-Hold) denotes the fraction of the 10,000 simulated paths (in %) in which the Sharpe Ratio of a volatility unadjusted EMA Long-Short Momentum strategy with $\lambda = 0.85$ is higher than that of a Buy-and-Hold strategy. Similarly, Max DD(LS-EMA $\lambda = 0.85$) < Max DD(Buy-Hold) denotes the fraction of the 10,000 simulated paths (in %) in which the drawdown of a volatility unadjusted Long-Short Momentum strategy with $\lambda = 0.85$ is worse than that of a Buy-and-Hold strategy. The definition of each strategy are given in Section A.

B Robustness: Monte-Carlo Results for the Overall Sample

As pointed out in the discussion of Section 4, one potential drawback of our analysis is that it makes the implicit assumption that parameter estimates of the times series processes, the empirical distributions and the residual dependency structure remains stable. As a robustness test, Table 8 shows how the Monte-Carlo simulations turn out once the features are estimated from the overall sample not just, the in-sample period. The table confirms that parameter stability is not concern: it shows that the Monte-Carlo computations are consistent with those obtained in-sample in Table 4. In Table 8, the 9M LS-TSM strategies on the S&P 500 are least likely to outperform, while the LC-TSM strategies on the cross-asset class futures portfolio are most likely to outperform. Furthermore, the results are also quantitatively very similar.

Table 8: Risk Behavior of TSM Momentum Strategies (Monte-Carlo)

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
Data: S&P 500, 1985.1-2018.12								
Sharpe(LS-TSM 9M) > Sharpe(Buy-Hold)	0.4%	0.4%	0.4%	0.5%	0.8%	0.7%	0.7%	0.7%
Sharpe(LS-TSM 9M vol. adj.) > Sharpe(Buy-Hold)	0.8%	0.7%	0.8%	0.9%	1.3%	1.2%	1.2%	1.2%
Sharpe(LC-TSM 9M) > Sharpe(Buy-Hold)	6.0%	5.8%	6.4%	7.1%	9.0%	8.7%	8.7%	8.5%
Sharpe(LC-TSM 9M vol. adj.) > Sharpe(Buy-Hold)	5.8%	5.6%	6.2%	7.1%	8.6%	8.3%	8.2%	8.3%
Data: S&P 500, 1985.1-2018.12								
Max DD(LS-TSM 9M) < Max DD(Buy-Hold)	94.0%	94.1%	93.9%	93.1%	91.9%	92.0%	92.0%	92.1%
Max DD(LS-TSM 9M vol. adj.) < Max DD(Buy-Hold)	92.9%	92.9%	92.6%	92.1%	90.6%	90.8%	90.8%	90.8%
Max DD(LC-TSM 9M) < Max DD(Buy-Hold)	51.2%	51.6%	50.9%	49.1%	47.3%	47.4%	47.9%	47.4%
Max DD(LC-TSM 9M vol. adj.) < Max DD(Buy-Hold)	57.2%	57.4%	56.4%	54.8%	52.6%	52.5%	52.9%	52.9%
Data: Cross-Asset Class Portfolio 1989.2-2018.12								
Sharpe(LS-TSM 12M) > Sharpe(Buy-Hold)	18.6%	18.2%	19.1%	19.5%	15.4%	13.9%	40.1%	36.8%
Sharpe(LS-TSM 12M vol. adj.) > Sharpe(Buy-Hold)	20.4%	19.6%	16.4%	16.9%	16.6%	12.8%	43.9%	35.3%
Sharpe(LC-TSM 12M) > Sharpe(Buy-Hold)	54.0%	51.8%	52.7%	54.3%	47.7%	42.8%	79.1%	76.2%
Sharpe(LC-TSM 12M vol. adj.) > Sharpe(Buy-Hold)	55.8%	54.3%	47.9%	48.9%	50.4%	42.3%	81.1%	74.8%
Data: Cross-Asset Class Portfolio 1989.2-2018.12								
Max DD(LS-TSM 12M) < Max DD(Buy-Hold)	38.4%	39.1%	37.6%	36.5%	41.0%	43.8%	21.5%	23.6%
Max DD(LS-TSM 12M vol. adj.) < Max DD(Buy-Hold)	33.8%	35.2%	37.4%	37.6%	37.0%	41.4%	17.9%	22.0%
Max DD(LC-TSM 12M) < Max DD(Buy-Hold)	3.1%	3.8%	3.5%	3.5%	4.0%	4.4%	1.5%	1.5%
Max DD(LC-TSM 12M vol. adj.) < Max DD(Buy-Hold)	3.4%	3.7%	4.1%	3.9%	3.5%	4.8%	1.3%	1.7%

Notes: Sharpe(LS-TSM 9M) > Sharpe(Buy-and-Hold) denotes the fraction of the 10,000 simulated paths (in %) in which the Sharpe Ratio of a volatility unadjusted 9M Long-Short Momentum strategy is higher than that of a Buy-and-Hold strategy. Similarly, Max DD(LS-TSM 9M) < Max DD(Buy-Hold) denotes the fraction of the 10,000 simulated paths (in %) in which the drawdown of a volatility unadjusted 9M Long-Short Momentum strategy is worse than that of a Buy-and-Hold strategy. The definition of each strategy are given in Section 2.5.

C Robustness: Alternative Time-Series Models

The AR and MA models by themselves do not assume any sophisticated volatility dynamics. However, stock returns empirically exhibit heteroskedasticity, and exhibit a negative correlation with lagged volatility in what is known as the leverage effect Black (1976). We use the asymmetric EGARCH model as in Nelson (1991) to model these dependencies, as the conditional variance responds asymmetrically to positive and negative shocks. Under an EGARCH(P,Q) process, he conditional variance of asset j has the form

$$\log \sigma_{j,t}^2 = \kappa_j + \sum_{i=1}^{P} \gamma_{j,i} \log \sigma_{j,t-i}^2 + \sum_{k=1}^{Q} \alpha_{j,k} \left(\frac{|\epsilon_{j,t-k}|}{\sigma_{j,t-k}} - E(\frac{|\epsilon_{j,t-k}|}{\sigma_{j,t-k}}) \right) + \sum_{k=1}^{Q} \psi_{j,k} \frac{\epsilon_{j,t-k}}{\sigma_{j,t-k}}, \tag{13}$$

where κ_j is the conditional variance model's constant, $\gamma_{j,i}$ is the GARCH component coefficient, $\alpha_{j,k}$ is the ARCH component coefficient and $\psi_{j,k}$ is the leverage component coefficient. As an alternative asymmetric GARCH model we also employ the GJR model of Glosten *et al.* (1993). Under an GJR(P,Q) process, the conditional variance of asset j has the form

$$\sigma_{j,t}^2 = \kappa_j + \sum_{i=1}^{P} \gamma_{j,i} \sigma_{j,t-i}^2 + \sum_{k=1}^{Q} \alpha_{j,k} \epsilon_{j,t-k}^2 + \sum_{k=1}^{Q} \psi_{j,k} I[\epsilon_{j,t-k} < 0] \epsilon_{j,t-k}^2, \tag{14}$$

where the binary indicator $I[\epsilon_{t-1} < 0] = 1$, and 0 otherwise; κ_j is the conditional variance model's constant, $\gamma_{j,i}$ is the GARCH component coefficient, $\alpha_{j,k}$ is the ARCH component coefficient and $\psi_{j,k}$ is the leverage component coefficient.

Table 9 shows the results for a set of eight alternative time-series models: **Model 9:** AR(3)-EGARCH(1,1); **Model 10:** MA(3)-EGARCH(1,1); **Model 11:** ARMA(3,3); **Model 12:** ARMA(3,3)-EGARCH(1,1); **Model 13:** ARMA(3,3)-GJR(1,1); **Model 14:** ARMA(3,3)-EGARCH(1,3); **Model 15:** AR(12)-EGARCH(1,3); **Model 16:** MA(12)-EGARCH(1,3). The Table shows that the Monte-Carlo computations are consistent with those obtained under Models 1-8 in Table 4. Namely, the 9M LS-TSM strategies on the S&P 500 are least likely to outperform, while the LC-TSM strategies on the cross-asset class futures portfolio are most likely to outperform. The Table shows that the obtained probabilities slightly increase (compare Models 15 and 16 with Models 7 and 8), indicating that heteroskedasticity plays some role in generating Momentum.

Table 9: Risk Behavior of TSM Momentum Strategies (Monte-Carlo)

	Model 9	Mod. 10	Mod. 11	Mod. 12	Mod. 13	Mod. 14	Mod. 15	Mod. 16
Data: S&P 500, 1985.1-2009.12								
Sharpe(LS-TSM 9M) > Sharpe(Buy-Hold)	9.9%	9.8%	3.5%	19.2%	14.5%	8.3%	6.1%	6.5%
Sharpe(LS-TSM 9M vol. adj.) > Sharpe(Buy-Hold)	8.9%	8.9%	4.8%	17.1%	13.4%	7.3%	5.1%	5.4%
Sharpe(LC-TSM 9M) > Sharpe(Buy-Hold)	35.9%	35.8%	16.6%	50.5%	45.5%	33.0%	26.9%	28.2%
Sharpe(LC-TSM 9M vol. adj.) > Sharpe(Buy-Hold)	27.3%	27.1%	15.0%	41.1%	36.4%	24.4%	19.7%	20.8%
Data: S&P 500, 1985.1-2009.12								
Max DD(LS-TSM 9M) < Max DD(Buy-Hold)	72.2%	72.3%	85.7%	60.5%	68.3%	74.0%	79.8%	79.2%
Max DD(LS-TSM 9M vol. adj.) < Max DD(Buy-Hold)	75.6%	75.7%	85.1%	64.6%	67.8%	77.9%	83.0%	82.5%
Max DD(LC-TSM 9M) < Max DD(Buy-Hold)	24.6%	24.6%	39.6%	17.8%	18.6%	25.1%	31.4%	30.5%
Max DD(LC-TSM 9M vol. adj.) < Max DD(Buy-Hold)	32.6%	32.7%	45.2%	24.4%	26.1%	33.9%	40.5%	39.3%
Data: Cross-Asset Class Portfolio 1989.2-2009.12								
Sharpe(LS-TSM 12M) > Sharpe(Buy-Hold)	16.8%	14.1%	11.6%	7.3%	4.5%	11.0%	35.6%	48.3%
Sharpe(LS-TSM 12M vol. adj.) > Sharpe(Buy-Hold)	25.1%	22.6%	17.6%	11.9%	9.3%	19.8%	48.4%	58.5%
Sharpe(LC-TSM 12M) > Sharpe(Buy-Hold)	62.1%	57.6%	43.6%	42.3%	32.4%	49.9%	79.4%	92.1%
Sharpe(LC-TSM 12M vol. adj.) > Sharpe(Buy-Hold)	68.5%	66.4%	52.7%	49.6%	42.4%	63.3%	88.7%	93.5%
Data: Cross-Asset Class Portfolio, 1989.2-2009.12								
Max DD(LS-TSM 12M) < Max DD(Buy-Hold)	40.5%	43.0%	48.0%	53.5%	60.8%	47.1%	29.4%	20.0%
Max DD(LS-TSM 12M vol. adj.) < Max DD(Buy-Hold)	29.9%	30.8%	35.8%	41.2%	43.6%	32.2%	16.6%	12.9%
Max DD(LC-TSM 12M) < Max DD(Buy-Hold)	3.6%	3.7%	5.4%	4.9%	6.0%	4.3%	1.8%	0.9%
Max DD(LC-TSM 12M vol. adj.) < Max DD(Buy-Hold)	2.9%	2.9%	4.1%	4.0%	4.2%	2.9%	1.1%	0.7%

Notes: Sharpe(LS-TSM 9M) > Sharpe(Buy-and-Hold) denotes the fraction of the 10,000 simulated paths (in %) in which the Sharpe Ratio of a volatility unadjusted 9M Long-Short Momentum strategy is higher than that of a Buy-and-Hold strategy. Similarly, Max DD(LS-TSM 9M) < Max DD(Buy-Hold) denotes the fraction of the 10,000 simulated paths (in %) in which the drawdown of a volatility unadjusted 9M Long-Short Momentum strategy is worse than that of a Buy-and-Hold strategy. The definition of each strategy are given in Section 2.5.

D Robustness: Degrees of Freedom of the Student's t-Copula

A legitimate concern is that our results are sensitive to the Student's t-copula's estimated degrees of freedom (DoF). The financial crisis, for example, has revealed that the use of the Gaussian copula (Student's t-copula with DoF = ∞) as a statistical tool to price and manage the risks of Collateralized Debt Obligations (CDOs) is highly problematic, see e.g. Li (2000) and MacKenzie & Spears (2014). To show that our results for the cross asset class portfolio are robust to varying the copula form, we run the simulations with i) the Student's t-copula estimated DoF$\times$2 ii) the Student's t-copula estimated DoF$\times$0.5. Table 10 shows the results. This table is qualitatively consistent with the results from Table 4 in the sense that the new table also predicts the i) LC-TSM strategy to outperform the LS-TSM strategy according to both risk measures ii) the LC-TSM strategy to outperform Buy-and-Hold in particular under Models 7 and 8. Quantitatively, the results exhibit only marginal deviations from those of Table 4 (within 5 percentage points).

Table 10: Risk Behavior of TSM Momentum Strategies (Monte-Carlo)

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
Data: Cross-Asset Portfolio 1989.2-2009.12, DoF×2								
Sharpe(LS-TSM 12M) > Sharpe(Buy-Hold)	15.8%	13.7%	10.7%	11.6%	17.2%	13.8%	39.7%	52.6%
Sharpe(LS-TSM 12M vol. adj.) > Sharpe(Buy-Hold)	19.4%	17.7%	9.5%	10.8%	22.7%	17.4%	48.3%	52.5%
Sharpe(LC-TSM 12M) > Sharpe(Buy-Hold)	52.2%	48.9%	40.7%	43.2%	55.7%	49.6%	82.1%	89.8%
Sharpe(LC-TSM 12M vol. adj.) > Sharpe(Buy-Hold)	54.6%	52.4%	36.7%	39.4%	61.2%	53.1%	84.9%	87.8%
Data: Cross-Asset Portfolio 1989.2-2009.12, DoF×2								
Max DD(LS-TSM 12M) < Max DD(Buy-Hold)	43.1%	44.3%	50.6%	47.8%	39.6%	43.0%	23.6%	17.2%
Max DD(LS-TSM 12M vol. adj.) < Max DD(Buy-Hold)	35.7%	36.8%	47.5%	45.0%	30.9%	35.6%	17.4%	14.9%
Max DD(LC-TSM 12M) < Max DD(Buy-Hold)	5.1%	5.2%	6.4%	5.7%	4.3%	4.7%	1.5%	1.1%
Max DD(LC-TSM 12M vol. adj.) < Max DD(Buy-Hold)	4.3%	4.3%	7.1%	6.2%	3.4%	4.0%	1.3%	0.9%
Data: Cross-Asset Portfolio 1989.2-2009.12, DoF×0.5								
Sharpe(LS-TSM 12M) > Sharpe(Buy-Hold)	15.3%	13.8%	10.0%	10.9%	17.8%	14.3%	38.9%	52.1%
Sharpe(LS-TSM 12M vol. adj.) > Sharpe(Buy-Hold)	19.3%	17.8%	9.5%	10.9%	22.6%	17.6%	47.7%	51.9%
Sharpe(LC-TSM 12M) > Sharpe(Buy-Hold)	51.6%	48.0%	40.7%	43.3%	56.0%	49.7%	81.8%	90.1%
Sharpe(LC-TSM 12M vol. adj.) > Sharpe(Buy-Hold)	54.6%	51.7%	36.8%	39.7%	61.2%	51.8%	84.9%	88.3%
Data: Cross-Asset Portfolio, 1989.2-2009.12, DoF×0.5								
Max DD(LS-TSM 12M) < Max DD(Buy-Hold)	43.5%	45.8%	50.7%	48.2%	40.1%	43.2%	24.4%	17.3%
Max DD(LS-TSM 12M vol. adj.) < Max DD(Buy-Hold)	35.8%	38.3%	48.1%	44.8%	31.6%	36.1%	17.4%	14.6%
Max DD(LC-TSM 12M) < Max DD(Buy-Hold)	5.0%	5.4%	7.0%	5.7%	4.2%	5.1%	2.0%	1.3%
Max DD(LC-TSM 12M vol. adj.) < Max DD(Buy-Hold)	4.3%	4.8%	6.9%	6.2%	3.4%	4.3%	1.5%	1.3%

Notes: Sharpe(LS-TSM 9M) > Sharpe(Buy-and-Hold) denotes the fraction of the 10,000 simulated paths (in %) in which the Sharpe Ratio of a volatility unadjusted 9M Long-Short Momentum strategy is higher than that of a Buy-and-Hold strategy. Similarly, Max DD(LS-TSM 9M) < Max DD(Buy-Hold) denotes the fraction of the 10,000 simulated paths (in %) in which the drawdown of a volatility unadjusted 9M Long-Short Momentum strategy is worse than that of a Buy-and-Hold strategy. The definition of each strategy are given in Section 2.5.

UCD CENTRE FOR ECONOMIC RESEARCH – RECENT WORKING PAPERS

WP18/06 John Cullinan, Kevin Denny and Darragh Flannery: 'A Distributional Analysis of Upper Secondary School Performance' April 2018
WP18/07 Ronald B Davies and Rodolphe Desbordes: 'Export Processing Zones and the Composition of Greenfield FDI' April 2018
WP18/08 Costanza Biavaschi, Michał Burzynski, Benjamin Elsner, Joël Machado: 'Taking the Skill Bias out of Global Migration' May 2018
WP18/09 Florian Buhlmann, Benjamin Elsner and Andreas Peichl: 'Tax Refunds and Income Manipulation - Evidence from the EITC' June 2018
WP18/10 Morgan Kelly and Cormac Ó Gráda: 'Gravity and Migration before Railways: Evidence from Parisian Prostitutes and Revolutionaries' June 2018
WP18/11 Kevin Denny: 'Basic Stata Graphics for Economics Students' July 2018
WP18/12 Ronald B Davies and Joseph Francois: 'Irexit: Making the Worst of a Bad Situation' July 2018
WP18/13 Ronald B Davies: 'From China with Love: The Role of FDI from Third Countries on EU Competition and R&D Activities' July 2018
WP18/14 Arnaud Chevalier, Benjamin Elsner, Andreas Lichter and Nico Pestel: 'Immigrant Voters, Taxation and the Size of the Welfare State' August 2018
WP18/15 Michael Spagat and Stijn van Weezel: 'On the Decline of War' August 2018
WP18/16 Stijn van Weezel: 'Apocalypse Now? - Climate Change and War in Africa' August 2018
WP18/17 FM Kiernan: 'The Great Recession and Mental Health: the Effect of Income Loss on the Psychological Health of Young Mothers' October 2018
WP18/18 Ronald B. Davies and Neill Killeen: 'The Effect of Tax Treaties on Market Based Finance: Evidence using Firm-Level Data' October 2018
WP18/19 George Sorg-Langhans, Clemens Struck, and Adnan Velic: Solving Leontief's Paradox with Endogenous Growth Theory
WP18/20 Alan Fernihough, Cormac Ó Gráda: Population and Poverty in Ireland on the Eve of the Great Famine
WP18/21 Cormac Ó Gráda: The Next World and the New World: Relief, Migration, and the Great Irish Famine
WP18/22 Lisa Ryan, Sarah La Monaca, Linda Mastrandrea and Petr Spodniak: 'Harnessing Electricity Retail Tariffs to Support Climate Change Policy' December 2018
WP18/23 Ciarán Mac Domhnaill and Lisa Ryan: 'Towards Renewable Electricity in Europe: An Empirical Analysis of the Determinants of Renewable Electricity Development in the European Union' December 2018
WP19/01 Ellen Ryan and Karl Whelan: 'Quantitative Easing and the Hot Potato Effect: Evidence from Euro Area Banks' January 2019
WP19/02 Kevin Denny: 'Upper Bounds on Risk Aversion under Mean-variance Utility' February 2019
WP19/03 Kanika Kapur: 'Private Health Insurance in Ireland: Trends and Determinants' February 2019
WP19/04 Sandra E Black, Paul J Devereux, Petter Lundborg and Kaveh Majlesi: 'Understanding Intergenerational Mobility: The Role of Nature versus Nurture in Wealth and Other Economic Outcomes and Behaviors' February 2019
WP19/05 Judith M Delaney and Paul J Devereux: 'It's not just for boys! Understanding Gender Differences in STEM' February 2019

UCD Centre for Economic Research Email economics@ucd.ie