Research Paper

Optimal dynamic strategies on Gaussian returns

Nick Firoozye and Adriano S. Koshiyama

Department of Computer Science, University College London, Gower Street,
London WC1E 6EA, UK; emails: n.firoozye@ucl.ac.uk, a.koshiyama@cs.ucl.ac.uk

ABSTRACT

Dynamic trading strategies, in the spirit of trend-following or mean reversion, represent an only partly understood but lucrative and pervasive area of modern finance. By assuming Gaussian returns and Gaussian dynamic weights or "signals" (eg, linear filters of past returns, such as simple moving averages, exponential weighted moving averages and forecasts from autoregressive integrated moving average models), we are able to derive closed-form expressions for the first four moments of the strategy's returns in terms of correlations between the random signals and unknown future returns. By allowing for randomness in the asset allocation, and by modeling the interaction of strategy weights with returns, we demonstrate that positive skewness and excess kurtosis are essential components of all positive Sharpe dynamic strategies (as is generally observed empirically), and that total least squares or orthogonal least squares are more appropriate than ordinary least squares for maximizing the Sharpe ratio, while canonical correlation analysis is similarly appropriate for the multi-asset case. We derive standard errors on Sharpe ratios that are tighter than the commonly used standard errors from Lo, and derive standard errors on the skewness and kurtosis of strategies that are apparently new results. We demonstrate that these results are applicable asymptotically for a wide range of stationary time series. Possible future extensions of this work to normalized signals, to multi-period returns and

Minor unmarked changes to this paper have been made according to US English, journal style (use of italic for emphasis is not permitted), etc. In addition, marginal queries have been added where major changes have been made or the meaning is unclear. Please check all text carefully throughout.

Text regarding future work has been moved to the conclusions section, as the abstract should describe only the work done in this paper – OK?

to nonlinear transforms, together with extensions to multi-asset dynamic strategies, are discussed.

1 INTRODUCTION

Commodity trading advisors (CTAs) or managed-future accounts are a subset of asset managers with over US\$341 billion of assets under management as of 2017 Q2.[1] The predominant strategy CTAs employ is trend-following. Meanwhile, bank structuring desks have devised a variety of risk premium or styles strategies (including momentum, mean reversion, carry and value) estimated to correspond to between approximately US\$150 billion (Miller 2016) and US\$200 billion (Allenbridge 2014) assets under management. Responsible for over 80% of trade volume in equities and a large but undocumented amount (due to their over-the-counter nature) of the foreign exchange market (Avramovic 2017), high-frequency trading (HFT) firms and e-trading desks in investment banks are known to make use of many strategies that are effectively short-term mean-reversion strategies. In spite of the relatively large industry undergoing recent significant growth, a careful analysis of the statistical properties of strategies, including their optimization, has only been undertaken in relatively limited contexts.

The corresponding statistics for the Société Générale Trend Index area in Figure 1 and Table 1 show that, except for some noise, skewness and excess kurtosis are largely positive for CTAs.

The algorithmic trading strategies we consider are time series strategies (often divided into mean-reverting or reversal strategies, trend-following or momentum strategies) and value strategies (also sometimes known as mean-reversion).[2] Each such time-series-related strategy is a form of signal processing. In more standard signal processing, the major interest is in the de-noised or smoothed signals and their properties. In algorithmic trading, the interest is instead in the relationship between statistics such as the moving average or some other form of smoothed historic returns (unfortunately, usually termed the "signal") and the unknown future returns. We show that, when we consider both to be random variables, it is actually the interaction between these so-called signals and future returns that determines the strategy's behavior.
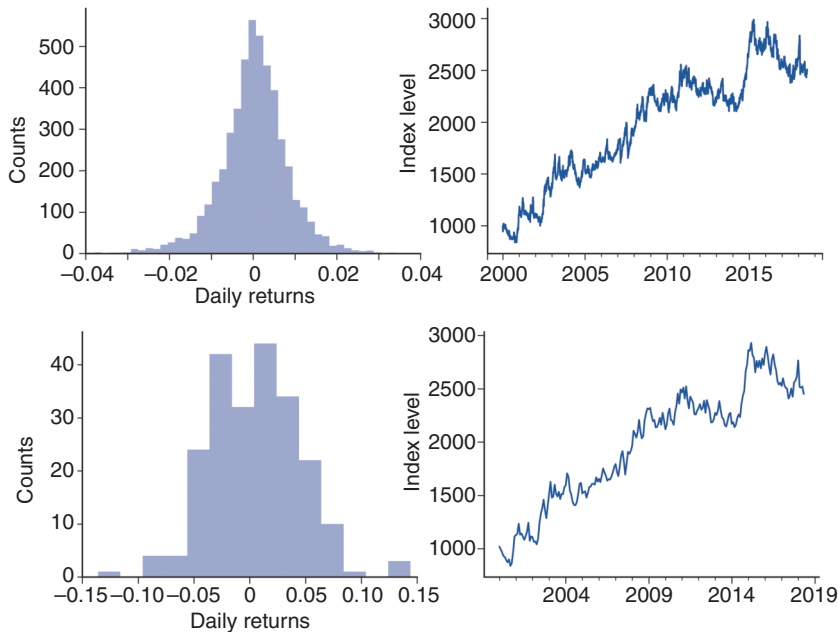
---

[1] See https://www.barclayhedge.com/research/indices/cta/Money_Under_Management.html.

[2] Other common strategies include carry and short-gamma or short-vol. Unlike mean reversion, momentum and value, these do not rely on the specifics of the autocorrelation function.

FIGURE 1 SGI Trend Followers Index: daily returns and monthly returns profiles.

TABLE 1 Société Générale Trend Index daily and monthly statistics.

	Daily	Monthly
Average annual return (%)	5.695	5.752
Volatility (%)	13.283	14.088
Sharpe ratio	0.429	0.408
Skewness	−0.448	0.186
Excess kurtosis	3.845	0.807

Equities, in particular the Standard & Poor's 500, are known to mean-revert over short horizons (eg, less than one month, and typically on the order of five to ten days), trend over longer horizons (ie, three to eighteen months) and mean-revert again over even longer horizons (ie, two to five years), as has been well established by the quant equities literature following on from the study of Jegadeesh and Titman (1993) and the work of Fama and French (1992). This distinct form of behavior, with reversals on a small timescale, trend on an intermediate timescale and reversion on a long timescale, is frequently observed across a large number of asset classes, and

strategies can be designed to take advantage of the behavior of asset prices across each timescale.

Our initial goal is to find a signal, X_t, usually a linear function of historic log (excess) returns $\{R_t\}$, which can be used as a dynamic weight assigned to the underlying asset on a regular basis. We assume log price $P_t = \sum_1^t R_k$. Examples of commonly used signals for macrotraders (CTAs and other trend followers) include the following.

- Simple moving average (SMA):

$$X_t = \frac{1}{T} \sum_1^T R_{t-k}.$$

- Exponentially weighted moving average (EWMA):

$$X_t = c(\lambda) \sum_{k=1}^{\infty} \lambda^k R_{t-k}.$$

- Holt–Winters (HW), or double exponential, smoothing with or without seasonals, and damped HW.

- Difference between current price and moving average:[3]

$$X_t = P_{t-1} - \frac{1}{T} \sum_1^T P_{t-k}.$$

- Forecasts from autoregressive–moving-average ARMA(p, q) models:

$$X_t = \phi_1 R_{t-1} + \cdots + \phi_p R_{t-p} + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}.$$

- Differences between SMAs:

$$X_t = \frac{1}{T_1} \sum_1^{T_1} P_{t-k} - \frac{1}{T_2} \sum_1^{T_2} P_{t-j}.$$

- Differences between EWMAs:

$$X_t = c(\lambda_1) \sum \lambda_1^k R_{t-k} - c(\lambda_2) \sum \lambda_2^k R_{t-k}.$$

---

[3] We note that if we replace P by $\log(P)$ and $R_t = \log(P_t) - \log(P_{t-1})$, this filter amounts to $X_t = \sum (T - k) R_{t-k}/T$, ie, a triangular filter on returns, which bears some similarity to the EWMA on returns.

Variations include using volatility or variance weighting such as z-scores (SMAs or EWMAs weighted by a simple or weighted standard deviation (see Harvey *et al* 2018)), and transformations of each of the signals listed above (eg, allocations depending on sigmoids of moving averages, reverse sigmoids, Winsorized signals). Other signals commonly used in equity algorithmic trading include economic and corporate releases and sentiment as derived from unstructured data sets such as news releases.

The returns from algorithmic trading strategies are well documented (see, for example, Asness *et al* 2013; Baltas and Kosowski 2013; Hurst *et al* 2017; Lempérière *et al* 2014). Although many methods have been used to derive signals by practitioners (see, for example, Bruder *et al* (2011) for a compendium), many of these methods are equally good (or bad) and it makes little practical difference whether we use the ARMA, EWMA or SMA as the starting point for a strategy design (see, for example, Levine and Pedersen 2015). In this paper, we only touch on normalized signals (eg, z-scores) and strategy returns, leaving their discussion for a subsequent study. We meanwhile note that the spirit of this paper's results carries through for the case of normalized signals and strategy returns.

Frequently, exponential smoothers have effectively been the best models in various economic forecasting competitions (see, for example, the results of the first three M-competitions (Makridakis 2000)), showing perhaps that their simplicity bestows a certain robustness, and their original intuition was sound even if the statistical foundation took a significant time to catch up. In fact, EWMA and HW can both be justified as state-space models (see Hyndman *et al* 2008), and this formulation brings with it a host of benefits, from mere intellectual satisfaction to statistical hypothesis tests, change-point tests and a metric for goodness-of-fit. Exponential smoothing with multiplicative or additive seasonals and dampened weighted slopes are used to successfully forecast a significant number of economic time series (eg, inventories, employment, monetary aggregates). EWMA (and the related (S)MA) and HW remain the most commonly used filtering methods for CTAs and HFT shops.

[margin note: Change OK? Or "have been the most effective models"?]

[margin note: "theory"?]

In the case of returns that are normal with fixed autocorrelation function (ie, those that are covariance stationary), signals created from linear combinations of historic returns are indeed normal random variables that are jointly normal with returns. External data sets (eg, unstructured data, corporate releases) are less likely to contain normally distributed variables, although there is an argument for asymptotic normality. Nevertheless, our approach is to assume normality of both returns and signals as a starting point for further analysis.

While there is significant need for further study, there have nonetheless been a number of empirical and theoretical results of note in this area. Fung and Hsieh (1997) were the first to look at the empirical properties of momentum strategies, noting (without any theoretical foundation) the resemblance of strategy returns to strad-

dle payoffs.[4] Potters and Bouchaud (2005) studied the significant positive skewness of trend-following returns, showing that for successful strategies the median profitability of trades is negative. The empirical returns of dynamic strategies are far from normal, and common values for skewness and kurtosis for single strategies can have skewness in the range [1.3, 1.7] and kurtosis in the range [8.8, 15.3], respectively (see Hoffman and Kaminski 2016).

Bruder and Gaussel (2011) and Hamdan *et al* (2016) used stochastic differential equations to study the power-option-like behavior of payoffs (see Hamdan *et al* (2016, Appendix 2) for a superlative use of stochastic differential equation-based methods for analyzing a wide variety of dynamic strategies). Martin and Zou considered general but independent and identically distributed (iid) discrete-time distributions (see Martin and Bana 2012; Martin-Zou 2012) to study the term structure of skewness over various horizons and the effects of certain nonlinear transforms on the term structure of return distributions. More recently, Bouchaud *et al* (2016) considered more general discrete-time distributions to study the convexity of payoffs and the effective dependence of returns on long-term versus short-term variance. Other studies have focused predominantly on the empirical behavior of returns, the relationship to macrofinancial conditions, the persistence of trend-following returns and the benefits from their inclusion in broader portfolios.

In the majority of theoretical studies, assumptions have been minimal in order to consider more general return distributions. Due to their generality, the derived results are somewhat restrictive. Rather than opting for the most general assumptions, we choose more specific distributional assumptions in the hope that we can obtain broader, possibly more practical results. We have extended the work in this study further to consider the endemic problem of overfitting (see Koshiyama and Firoozye 2018), proposing total least squares with covariance penalties as a means of model selection, showing they outperform standard methods, using ordinary least squares (OLS) with the Akaike information criterion (AIC).

In this paper, we consider underlying assets with stationary Gaussian returns and a fixed autocorrelation function (ie, they are a discrete Gaussian process). While we make no defense for the realism of using normal returns, we find that normality can be exploited in order to ensure we understand how the returns of linear and nonlinear strategies should work in theory and to further the understanding of the interaction between properties of returns and of the signals as a basis for the development and analysis of dynamic strategies in practice.

Given a purely random mean-zero covariance-stationary discrete-time Gaussian process for returns, the signals listed above, whether an EWMA or an ARMA fore-

Margin notes:

Changes to sentence OK? Is the 'Appendix 2' referred to here in Hamdan *et al* (2016) or is it an online appendix for your present paper?

Changes to sentence OK?

Change OK?

---

[4] Or, as they claimed, the returns of trend-following resemble those of an extremely exotic option (which is not actually traded), daily traded "look-back straddles".

cast, can be expressed as convolution filters of past returns, ie, our signal X_t can be expressed as

$$X_t = \sum_{k \geqslant 1} \phi(k) R_{t-k}.$$

This is an example of a time-invariant linear filter of a Gaussian process. If we restrict our attention to those filters that are square summable, ie, $\sum_1^\infty \phi(k)^2 < \infty$, then it is well known that the resulting filtered series is also Gaussian and jointly Gaussian with R_t.

Our underlying premise is that the important distribution to consider for the analysis of dynamic strategies is a product of Gaussians (rather than a single Gaussian, as would usually apply in asymptotic analysis of asset returns). This product measure can be justified on many levels, and we discuss large sample approximations in the online appendix.

The resulting measure, which determines the success of the strategy, is the correlation between the returns and the signals, a measure which, in the context of measuring an active manager's skill, is known as the "information coefficient", as given in the "Fundamental Law of Active Management" detailed in Grinold and Kahn (1999). While there is a large literature on the information coefficient and its relationship to information ratios (see, for example, Lee (2000) for formulas similar to (2.5)), the derivations, resulting formulas and conclusions differ significantly.

We should also mention the work on random matrix theory by Bouchaud and Potters (2009), which touches on many of the topics we consider in this paper. In particular, their analysis of returns as products of Gaussians or t-distributions is very close to our own. While many of the emphases are different from ours, we believe the general area of random matrix theory to be a fruitful approach to trading strategies.

The primary tool we use to derive results is Isserlis's theorem (Isserlis 1918) (or Wick's theorem, as it is known in the context of particle physics (Wick 1950)). This relates products and powers of multivariate normal random variables to their means and covariances. Wick's theorem has been applied in areas from particle physics to quantum field theory to stock returns, and there have been some recent efforts to extend it to non-Gaussian distributions (see, for example, Michalowicz $et\ al$ (2011) for Gaussian-mixture distributions and Kan (2008) for products of quadratic forms and elliptic distributions). It has also been applied to continuous processes via the central limit theorem (see Parczewski 2014). We have used these theorems in the context of dynamic (algorithmic) trading strategies to find expressions for the first four moments of strategy returns in closed form. While it is not necessarily the aim of all scientific studies of trading strategies to find closed-form expressions, the ease with which we can describe strategy returns makes this direction relatively appealing and allows for a number of future extensions.

The paper considers one asset over a single period. With a normal signal, we will show there is a universal bound on the one-period Sharpe ratio, skewness and kurtosis. We explain the role of total or orthogonal least squares as an alternative to OLS for strategy optimization. We look at the corresponding refinements to measures of Sharpe ratio standard error for these dynamic strategies, improving on the large-sample theory-based standard errors in more common use. We also introduce standard errors on skewness and kurtosis, which are distinct from those for Gaussian returns, and present some basic results about multiple assets and diversification. Finally, we discuss the role of product measures, which are more pertinent than simple Gaussian measures to the study of dynamic strategies. In the online appendixes, we present closed-form solutions to Sharpe ratios in the case of nonzero means. We also discuss extensions to our optimizations in the presence of transaction costs. We touch on the extension to multiple periods as well. As mentioned, further extensions to overfitting by the use of covariance penalties (akin to Mallow's C_p, the AIC or the Bayesian information criterion) are presented separately in Koshiyama and Firoozye (2018).

<div style="text-align: right">Minor changes to this paragraph – OK?</div>

2 SINGLE-PERIOD LINEAR STRATEGIES

We consider the (log) returns of a single asset, $R_t \sim \mathcal{N}(0, \sigma_R^2)$ returns with auto-covariance function at lag k, $\gamma(k) = E[R_t R_{t-k}]$, together with the corresponding autocorrelation function $c(k) = \gamma(k)/\gamma(0)$ at lag k.

Our main aim is to work with strategies based on linear portfolio weights (or signals) $X_t = \Sigma_1^\infty a_k R_{t-k}$ for coefficients a_k generating the corresponding dynamic strategy returns $S_t = X_t R_t$ (here, and always, the signal X_t is assumed to only have appropriately lagged information). Example strategy weights include EWMAs $a_k \propto \lambda^k$, SMAs $a_k = (1/T)\mathbf{1}_{[1,\dots,T]}$ and forecasts from ARMA models. Most importantly, the portfolio weights X are normal and jointly normal with returns R. In Appendix B online, we show that, for the wide set of signals discussed in Section 1, when applied to Gaussian returns, the signal and returns are jointly Gaussian.

<div style="text-align: right">Do you mean the partitioned correlation matrix Σ defined later or should this be a summation sign? Please clarify. Journal style is to use centered dots only for "placeholders" and scalar products. I have assumed that the centered dot denoted simple multiplication of scalars and deleted it here and elsewhere. Please mark any that denote scalar product and should be reinstated.</div>

We restrict our attention to return distributions over a single period. In the case of many momentum strategies, this period can be one day, if not longer. For higher-frequency intraday strategies, this period can be much shorter. The pertinent concern is that the horizon (ie, one period) is the same as that over which the rebalancing of strategy weights is done. If weights are rebalanced every five minutes, then the period should be five minutes. This is a necessary assumption in order to ensure the joint normality of (as yet indeterminate) signals and future returns. Moreover, this assumption will give some context to our results, which imply a maximal Sharpe ratio, maximal skewness and maximal kurtosis for dynamic linear strategies.

We are interested in characterizing the moments of the strategy's unconditional returns, the corresponding standard errors on estimated quantities and the means of optimizing various nondimensional measures of returns, such as the Sharpe ratio, via the use of nonlinear transformations of signals. Our goal is to look at unconditional properties of the strategy. It is important to avoid foresight in strategy design, and this directly impacts the conditional properties of strategies (eg, conditional densities involve conditioning on the currently observed signal to determine properties of the returns, which are just Gaussian). In the context of our study, we are concerned with one-period-ahead returns of the unconditional returns distribution of our strategy, where both the signals and the returns are unobserved, and the resulting distributions (in our case, the product of two normals) are much richer and more realistic; for the interested reader, we have added a more detailed discussion of our framework in Appendix G online.

2.1 Properties of linear strategies

Given the joint normality of the signal and the returns, we can explicitly characterize the one-period strategy returns (see Cui *et al* 2016). To allow for greater extendibility, we prefer to only consider the moments of the resulting distributions. These can be characterized easily using Isserlis's theorem (Isserlis 1918), which gives all moments for any multivariate normal random variable in terms of the mean and variance. We also refer the reader to Haldane (1942), which meticulously produces both noncentral and central moments for powers and products of Gaussians. While this is a routine application of Isserlis's theorem, the algebra can be tedious, so we quote the results.

THEOREM 2.1   (Isserlis 1918)   *If $X \sim \mathcal{N}(0, \Sigma)$, then*

$$E[X_1 X_2 \cdots X_{2n}] = \sum_{i=1}^{2n} \prod_{i \neq j} E[X_i X_j]$$

and

$$E[X_1 X_2 \cdots X_{2n-1}] = 0,$$

where the $\sum \prod$ is over all the $(2n)!/(2^n n!)$ unique partitions of $X_1, X_2, \ldots, X_{2n}$ into pairs $X_i X_j$.

Haldane's paper quotes a large number of moment-based results for various powers of each normal. We quote the relevant results.

THEOREM 2.2   (Haldane 1942)   *If $x, y \sim \mathcal{N}(0, 1)$ with correlation ρ, then*

$$E[xy] = \rho,$$
$$E[x^2 y^2] = 1 + 2\rho^2,$$
$$E[x^3 y^3] = 3\rho(3 + 2\rho^2),$$
$$E[x^4 y^4] = 3(3 + 24\rho^2 + 8\rho^4),$$

and thus the central moments of xy are

$$\mu_1 = \rho, \tag{2.1}$$
$$\mu_2 = 1 + \rho^2, \tag{2.2}$$
$$\mu_3 = 2\rho(3 + \rho^2), \tag{2.3}$$
$$\mu_4 = 3(3 + 14\rho^2 + 3\rho^4). \tag{2.4}$$

From these one-period moments (and a simple scaling argument giving the dependence on $\sigma(x)$ and $\sigma(y)$) we can characterize Sharpe ratio, skewness, etc, and can also define objective functions in order to determine some sense of optimality for a given strategy.

THEOREM 2.3   (Linear Gaussian)   *For single-asset returns and a one-period strategy, $R_t \sim \mathcal{N}(0, \sigma_R^2)$ and $X_t \sim \mathcal{N}(0, \sigma_X^2)$ jointly normal with correlation ρ, the Sharpe ratio is given by*

$$\text{SR} = \frac{\rho}{\sqrt{1 + \rho^2}}, \tag{2.5}$$

the skewness is given by

$$\gamma_3 = \frac{2\rho(3 + \rho^2)}{(1 + \rho^2)^{3/2}} \tag{2.6}$$

and the kurtosis is given by

$$\gamma_4 = \frac{3(3 + 14\rho^2 + 3\rho^4)}{(1 + \rho^2)^2}. \tag{2.7}$$

In the online appendix, we extend (2.5) and (2.6) to the case of nonzero means.

PROOF   A simple application of Theorem 2.2 gives us the following first two moments for our strategy $S_t = X_t R_t$:

$$\mu_1 = E[S_t] = E[XR] = \sigma_X \sigma_R \rho,$$
$$\mu_2 = \text{var}[S_t] = \sigma_X^2 \sigma_R^2 (\rho^2 + 1).$$

Thus, we can derive the following results for the <u>Sharpe ratio</u>:

Change from "Sharpe" to "SR" as defined above – OK?

$$\text{SR} = \frac{\mu_1}{\mu_2^{1/2}} = \frac{\sigma_X \sigma_R \rho}{\sigma_X \sigma_R \sqrt{\rho^2 + 1}} = \frac{\rho}{\sqrt{\rho^2 + 1}}.$$

Moreover, we can see that the skewness is equal to

$$\gamma_3 = \frac{\mu_3}{\mu_2^{3/2}} = \frac{2\rho(3 + \rho^2)}{(1 + \rho^2)^{3/2}}.$$

Finally, the kurtosis is given by

$$\gamma_4 = \frac{\mu_4}{\mu_2^2} = \frac{3(3 + 14\rho^2 + 3\rho^4)}{(1 + \rho^2)^2}.$$
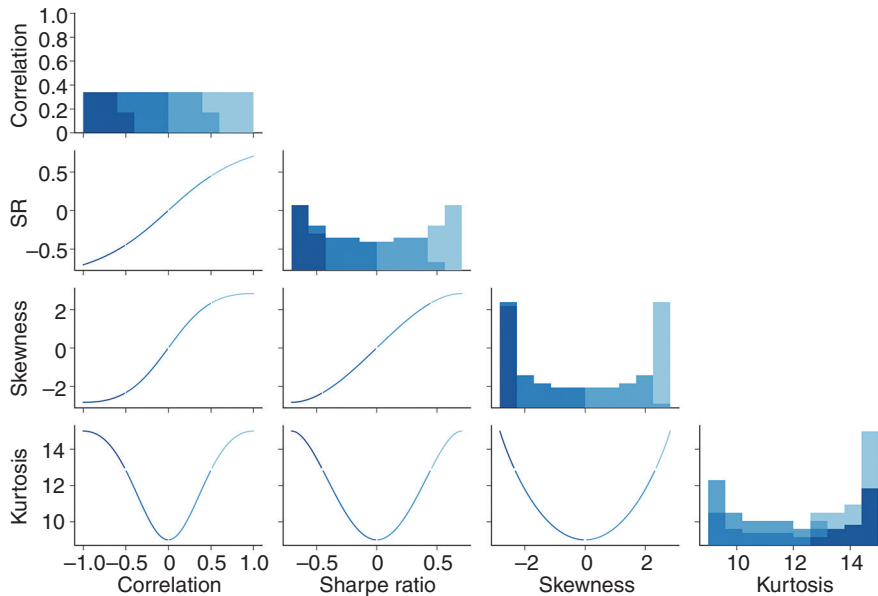
□

If we restrict our attention to positive correlations, all three dimensionless statistics are monotonically increasing in ρ. Consequently, strategies that maximize one of these statistics will maximize the others, although the impact of correlation upon the Sharpe ratio, skewness and kurtosis is different. We illustrate the cross-dependencies in the following charts, depicting the relationships between the variables. In Figure 2, the (blue) shaded histograms correspond to the correlation ranges $\{[-1, -0.5], [-0.5, 0], [0, 0.5], [0.5, 1]\}$. We note that a uniform distribution in correlations maps into a higher likelihood of extreme Sharpe ratios and an even higher likelihood of extreme skewness and kurtosis.

The skewness is in the ranges $[-2^{3/2}, 2^{3/2}] \approx [-2.8, 2.8]$. Unlike the Sharpe ratio, the skewness's dependence on correlation tends to flatten; to achieve 90% peak skewness we need only achieve a 0.60 correlation, while for a 90% peak Sharpe ratio we need a correlation of 0.85. Kurtosis is an even function and varies from a minimal value of 9 to a maximum of 15. In practice, correlations will largely be close to zero, and the resulting skewness and kurtosis will be significantly less than the maximal values.

Although we analyze the moments of the strategy $S_t = X_t R_t$, the full product density is actually known in closed form (see Appendix A online; see also Cui *et al* 2016; Nadarajah and Pogány 2016). It is clear that the distribution of the strategy is leptokurtic even when it is not predictive (when the correlation is exactly zero, the strategy has a kurtosis of 9). In the limit as $\rho \to 1$, the strategy's density approaches that of a noncentral χ^2, an effective "best case" density when considering the design of optimal linear dynamic strategies.

An optimized strategy with sufficient lags (and a means of ensuring parsimony) may be able to capture both mean reversion and trend and result in yet higher correlations. Annualized Sharpe ratios of between 0.5 and -1.5 are most common (ie, correlations of between 3% and 9%) for single-asset strategies in this relatively low-frequency regime.

FIGURE 2 Correlation, Sharpe ratio (SR), skewness and kurtosis pairwise relationships.

A uniform distribution in correlation is bucketed into four ranges $\{[-1, -0.5], [-0.5, 0], [0, 0.5], [0.5, 1]\}$, as depicted by the different shades in the bar charts. After transforming the correlation into SR, γ_3 and γ_4, the frequencies are no longer uniform.

2.2 Optimization: maximal correlation, total least squares

Many algorithmic traders will explain how problematic strategy optimization is, given the endless concerns of overfitting, etc. Although these are a concern, the naive use of strategies that are merely pulled out of thin air is equally problematic, where there is no explicit use of optimization (and, in its place, more eye-balling strategies or targeting Sharpe ratios rather loosely, effectively a somewhat loose optimization mental exercise). Practical considerations abound and real-world returns are neither Gaussian nor stationary. We argue that, regardless, using optimization and a well-specified utility function as a starting point is a means of preventing strategies from being merely untested heuristics. Unlike most discretionary traders' heuristics (or "rules of thumb"), which have their place as a means of dealing with uncertainty (see, for example, Gigerenzer *et al* 1999), heuristic quantitative trading strategies run the risk of being entirely arbitrary, or are subject to a large number of human biases, in marked contrast to "quantitative" investment strategies.

Where optimization is used, the most common optimization method is to minimize the mean-squared error (MSE) of the forecast. Our results show that, rather than

FIGURE 3 EWMA strategy Sharpe ratio versus α, mean square error and correlation for S&P 500 reversal strategies.
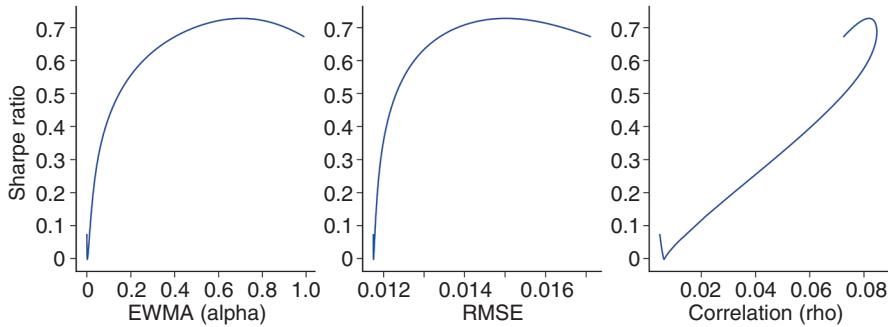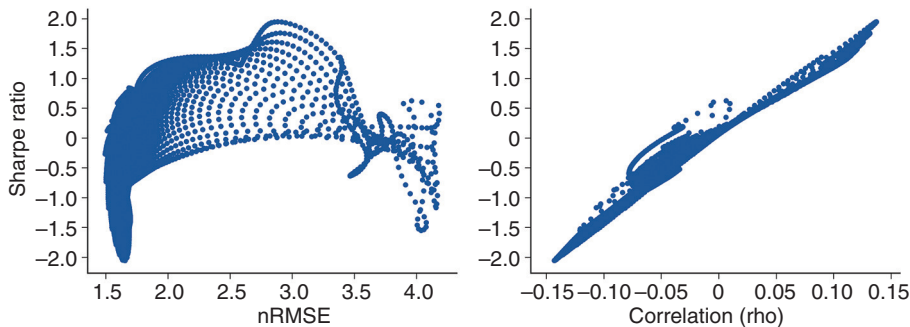

FIGURE 4 Holt–Winters strategy Sharpe ratio versus mean square error and correlation for S&P 500 reversal strategies.

minimizing the $\mathcal{L}^2$ norm between our signal and the forecast returns (or to maximize the likelihood), if the objective is to maximize the Sharpe ratio, we must maximize the correlation.

Figures 3 and 4 depict fits of strategies applied to S&P 500 using EWMA and HW filters for a variety of parameters. The relationship between MSE and Sharpe ratio is not monotone in MSE for the EWMA filter, as we see in Figure 3, while the relationship between correlation and Sharpe ratio is much closer to being linear. For the case of HW (with two parameters) in Figure 4, any given MSE can lead to a nonunique Sharpe ratio, sometimes with a very broad range, leading us to conclude that the optimization is poorly posed. The relationship of correlation to Sharpe ratio

Changes to sentence OK?

is obviously closer to linear, with higher correlations almost always leading to higher Sharpe ratios.

In the case of a one-dimensional forecasting problem with (unconstrained) linear signals, optimizing the correlation amounts to using what is known as total least squares (TLS) regression or "orthogonal distance regression", a form of principal components regression (see, for example, Golub and Van Loan 1980; Markovsky and Van Huffel 2007). In the multivariate case, it would be more closely related to canonical correlation analysis (CCA).

Unlike OLS, where the dependent variable is assumed to be measured with error and the independent variables are assumed to be measured without error, in TLS regression, both dependent and independent variables are assumed to be measured with error, and the objective function compensates for this by minimizing the sum squared of orthogonal distances to the fitted hyperplane. This simple form of errors-in-variables regression has been studied since the late 1870s, and it is most closely related to principal component analysis. For k regressors, the TLS fit will produce weights orthogonal to the first $k - 1$ principal components. Changes to sentence OK?

So, if we consider the signal $X = Z\beta$ to be a linear combination of features, with $Z \in \mathbb{R}^k$ a k-dimensional feature space, then we note that

$$\hat{\beta}^{\text{OLS}} = (Z'Z)^{-1}Z'R$$

but

$$\hat{\beta}^{\text{TLS}} = (Z'Z - \sigma_{k+1}^2 I)^{-1}Z'R,$$

where σ_{k+1} is the smallest singular value for the $T \times (k + 1)$-dimensional matrix $\tilde{X} = [R, Z]$ (ie, the concatenation of the features and the returns (see, for example, Rahman and Yu 1987)).[5] It is well known that, for the case of OLS, the smooth or hat matrix $\hat{R} = MR$ is given by

$$M^{\text{OLS}} = Z(Z'Z)^{-1}Z'$$

with $\text{tr}(M^{\text{OLS}}) = k$, the number of features. In contrast,

$$M^{\text{TLS}} = Z(Z'Z - \sigma_{k+1}^2 I)^{-1}Z'$$

and effectively has a greater number of degrees of freedom than that of OLS, ie,

$$\text{tr}(M^{\text{TLS}}) \geqslant \text{tr}(M^{\text{OLS}})$$

with equality only when there is complete collinearity.[6] For this reason, many people see TLS as an antiregularization method that may result in a less stable response to Changes to sentence OK?

---

[5] A more common method for extracting TLS estimates is via a principal component analysis of the concatenation matrix $\tilde{X}$, where $\hat{\beta}^{\text{TLS}}$ is chosen to cancel the least significant principal component.
[6] In this case, it is also known that $\text{tr}(M) = \text{tr}(L)$, where $L = (Z'Z - \sigma_{k+1}^2 I)^{-1}Z'Z$ and we

outliers (see, for example, Zhang 2017, pp. 334–335). Consequently, there has been extensive study of regularized TLS, typically using a weighted ridge-regression (or Tikhonov) penalty (see the discussion in Zhang (2017) for more detail on this large body of research). The stability of TLS in out-of-sample performance is an issue we broach in our study of overfitting penalties (see Koshiyama and Firoozye 2018).

While maximizing correlation rather than minimizing the MSE seems a very minor change in objective function, the formulas differ from those of standard OLS. The end result is a linear fit that takes into account the errors in the underlying conditioning information. We believe that it should be of relatively little consequence when the features are appropriately normalized, as is the case for univariate time series estimation, although some authors have suggested that optimizing TLS is not appropriate for prediction (see, for example, Fuller 1987, Section 1.6.3). When we seek to maximize the Sharpe ratio of a strategy, the objective should be not prediction but rather optimal weight choice.

2.3 Maximal Sharpe ratios, maximal skewness, minimal kurtosis

Surprisingly, there appears to be a maximal Sharpe ratio for linear strategies. In the case of normal signals and normal returns, the maximal Sharpe ratio is that of a noncentral χ^2 distribution, and the resulting maximal statistics are

$$\mathrm{SR}^{\max} = \frac{\sqrt{2}}{2} \approx 0.707,$$
$$\gamma_3^{\max} = 2\sqrt{2} \approx 2.828,$$
$$\gamma_4^{\max} = 15.000.$$

While the <u>estimate</u> for the Sharpe ratio may seem surprisingly low, we comment that this is for a single period, for a single rebalancing. For a daily rebalanced strategy, if we naively annualize the Sharpe ratio (by a factor of $\sqrt{252}$), we get a maximal Sharpe ratio, <u>$\mathrm{SR}^{\max}$</u> ≈ 11.225, a level generally well beyond what is attained in practice. The statistics $\gamma_3^{\max}$ and $\gamma_4^{\max}$ do not scale when annualized, but are still large irrespective of the time horizon.

Singular OK?

Suffix set as superscript to match notation in the display above – OK? Also "approximately" deleted here as this is implied by the symbol – OK?

---

know that the singular values of $\sigma(L) = \{\lambda_i^2/(\lambda_i^2 - \sigma_{k+1}^2)\}$, where λ_i are the singular values of Z (or correspondingly, λ_i^2 are the singular values of $Z'Z$) and $\lambda_1 \geqslant \cdots \geqslant \lambda_k > 0$ (Leyang 2012). By the Wilkinson interlacing theorem, $\lambda_k \geqslant \sigma_{k+1} \geqslant 0$ (see Rahman and Yu 1987). Consequently,

$$\mathrm{tr}(M^{\mathrm{TLS}}) = \sum_i \frac{\lambda_i^2}{\lambda_i^2 - \sigma_{k+1}^2} \geqslant k = \mathrm{tr}(M^{\mathrm{OLS}})$$

with equality if and only if $\sigma_{k+1}^2 = 0$ (ie, when $R^2 = 100\%$ and consequently OLS and TLS coincide). In other words, $\mathrm{tr}(M^{\mathrm{TLS}}) \geqslant \mathrm{tr}(M^{\mathrm{OLS}})$.

We note that our assumption of normality could easily be relaxed by considering nonlinear transforms of the signals X with the end result that the maximal Sharpe ratio bounds are relaxed. While this is beyond the scope of this paper, we note that it is easy to show that simple nonlinear strategies, going long one unit if the signal is above a threshold k and short one unit if it is below $-k$ (ie, $f_k(X) = \mathbf{1}_{X>k} - \mathbf{1}_{X<k}$), can be shown to have arbitrarily large Sharpe ratios, depending on the choice of threshold, k. The probability of initiating such an arbitrarily high Sharpe ratio trade likewise decreases to being negligible. Thus, stationary returns with a small nonzero autocorrelation can lead to violations of Hansen–Jagannathan (or "good deal") bounds.

Changes to sentence OK?

It is also noticeable from these formulas that, while Sharpe ratio and skewness may change sign, kurtosis is always bounded below and takes a minimum value of 9 (ie, an excess kurtosis of 6). The normality of the resulting strategy returns is not a good underlying assumption, since the theoretical value of the Jarque–Bera test would be

$$
\begin{aligned}
\mathrm{JB}(n) &= \frac{n-k+1}{6}\left(\gamma_3^2 + \frac{(\gamma_4-3)^2}{6}\right) \\
&\geq \frac{(n-k+1)}{6}\left(\frac{36}{4}\right) \\
&= 1.5(n-k+1),
\end{aligned}
$$

and this is asymptotically $\chi^2(2)$ (ie, rejection of normality at a 0.99 confidence interval of JB > 9.210). Theoretically, we would need a relatively small sample to be able to reject normality.

3 REFINED STANDARD ERRORS

Given that we have closed-form estimates of a number of relevant statistics for dynamic linear strategies, it makes sense to consider the effects of estimation error upon quantities such as the Sharpe ratio. Many analysts and traders who consider dynamic strategies in practice will consider altering them on an ongoing basis, and when they do make changes to their strategies they are typically in a quandary as to whether the observed change in Sharpe ratio or skewness is in fact statistically significant.

Changes to sentence OK?

3.1 Standard errors for Sharpe ratios

While there are formulas for standard errors for Sharpe ratios of generic assets, these are not specific to Sharpe ratios generated by dynamic trading strategies, and as a consequence there is some possibility of refining them.

We refer the reader to Pav (2016) for an exhaustive overview of the mechanics of Sharpe ratios and, in particular, to Section 1.4 therein, which quotes many of the known results about standard errors. Specifically, we look to Lo (2002) for large-sample estimates of standard errors for Sharpe ratios of generic assets, given the asymptotic normality of returns. For a sample of size N and iid returns, Lo obtains the large-sample distribution

$$\widehat{\text{SR}} \sim \mathcal{N}(\text{SR}, \text{stderr}_{\text{Lo}}^2),$$

giving a standard error

Changes to sentence OK?

$$\text{stderr}_{\text{Lo}} = \sqrt{\left(1 + \frac{\text{SR}^2}{2}\right)\frac{1}{T}},$$

which he suggests should be approximated using the standard error

$$\sqrt{\left(1 + \frac{\widehat{\text{SR}^2}}{2}\right)\frac{1}{T}}.$$

While Lo's estimates may be appropriate for generic assets, for Sharpe ratios derived from dynamic strategies we have a somewhat more refined characterization of the variability of the estimated Sharpe ratios. With correlated Gaussian signals and returns, we derive the following result.

COROLLARY 3.1 (Standard errors) *For returns $R_t \sim \mathcal{N}(0, \sigma_R^2)$ and signal $X_t \sim \mathcal{N}(0, \sigma_X^2)$ with correlation ρ and sample size T, the standard errors are given by*

$$\text{stderr}_{\text{implied}} = \frac{1}{(\hat{\rho}^2 + 1)^{3/2}}\sqrt{\frac{1 - \hat{\rho}^2}{T - 2}} \qquad (3.1\text{a})$$

$$\approx (1 - \widehat{\text{SR}^2})\sqrt{\frac{1 - 2\widehat{\text{SR}^2}}{T - 2}} \qquad (3.1\text{b})$$

for $|\widehat{\text{SR}}| < \sqrt{2}/2$.

PROOF   As is well known, for a bivariate Gaussian process of sample size T, the distribution for the sample (Pearson) correlation is given by

$$\hat{\rho} \sim f_\rho(\hat{\rho}) = \frac{(T-2)(1-\rho^2)^{(T-1)/2}(1-\hat{\rho}^2)^{(T-4)/2}}{\pi}\int_0^\infty \frac{\mathrm{d}w}{(\cosh(w) - \rho\hat{\rho})^{T-1}}.$$
$$(3.2)$$

The standard errors that approximate those in (3.2) for $\hat{\rho}$ are

$$\text{stderr}_\rho = \sqrt{\frac{1 - \hat{\rho}^2}{T - 2}}$$

(attributed to Sheppard, and used by Pearson (see, for example, Hald 2008)). Taken together with the results of Theorem 2.3, we apply the delta method to find that the resulting standard error for our plug-in estimate for the Sharpe ratio,

$$\widehat{SR} = \frac{\hat{\rho}}{\sqrt{\hat{\rho}^2 + 1}},$$

is given by

$$\text{stderr}_{\text{implied}} = \frac{\partial \widehat{SR}}{\partial \hat{\rho}} \text{stderr}_\rho = \frac{1}{(\hat{\rho}^2 + 1)^{3/2}} \sqrt{\frac{1 - \hat{\rho}^2}{T - 2}},$$

which gives us (3.1a). If we solve for $\hat{\rho}$ in terms of $\widehat{SR}$, we are able to derive (3.1b).

□

We note that in spite of the fact that Lo's standard errors are very near our estimates for large sample size, the entire sampling distribution from our estimates is much more concentrated than the $\mathcal{N}(0, \text{stderr}_{\text{Lo}}^2)$, potentially leading to tighter confidence intervals at the 99% or higher confidence levels. We can see in Figure 6 that the tail of the distribution given by Lo is much fatter than ours.

Mertens (2002) gives a refinement of Lo's result by including adjustments for skewness and excess kurtosis:

$$\text{stderr}_{\text{Mertens}}^2 = \left(1 + \tfrac{1}{2}\widehat{SR}^2 - \gamma_3 \widehat{SR} + \frac{\gamma_4 - 3}{4} \widehat{SR}^2 \right). \tag{3.3}$$

If we plug our estimates for skewness and excess kurtosis from (2.6) and (2.7) into (3.3), we find a slightly tighter estimate of the standard error than Lo. For most smaller amplitude correlations the estimate given by (3.3) comes very close to our estimate of standard error (see Figure 7), and for small N and low correlations Lo's standard errors are in fact tighter. For large correlations our standard errors are significantly tighter. For large sample sizes there is little difference between them. Using our estimates for γ_3 and γ_4, Mertens's approximation is always tighter than Lo's; in particular, for correlations $|\rho| < 0.5$, Mertens's approximation appears almost identical to our own. Nonetheless, we argue in Section 5 that our standard errors are more appropriate for dynamic strategies if there is any significant difference between the measures.
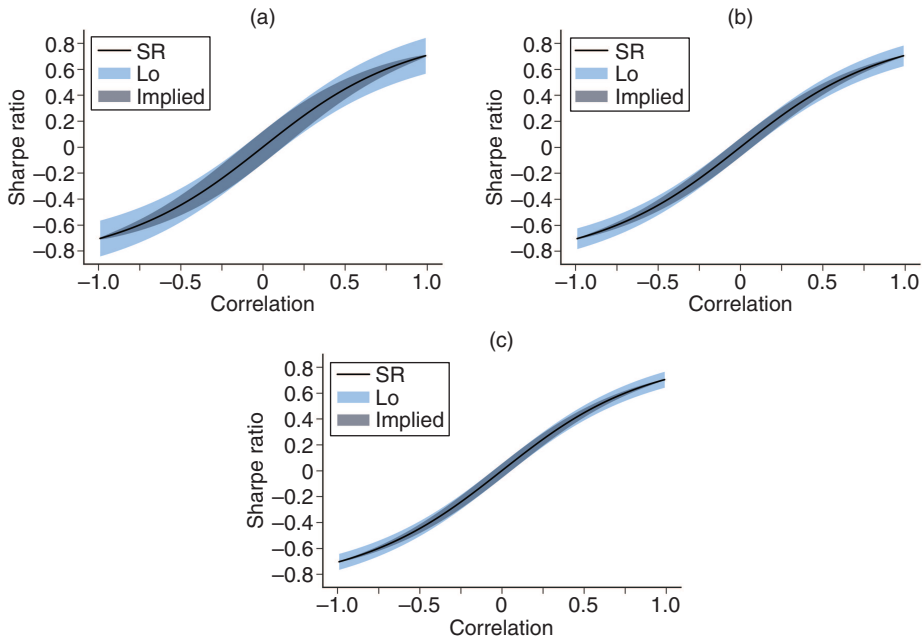
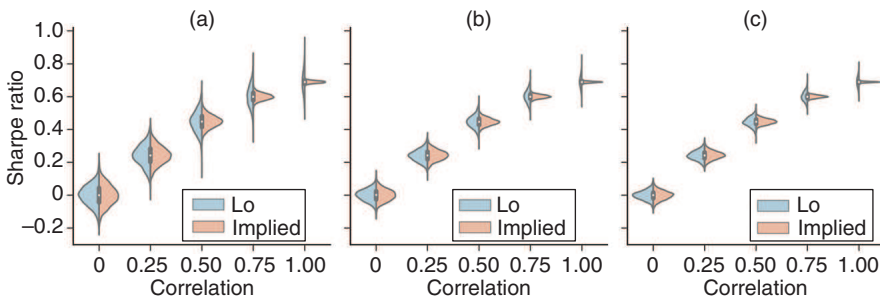## 3.2 Standard errors for higher moments

Using exactly the same procedure, we can easily derive standard errors for both skewness and kurtosis. In terms of classical confidence intervals, we consider Joanes and Gill (1998) and Cramér (1946), which apply to Gaussian (and non-Gaussian)

FIGURE 5 Sharpe ratio and confidence interval comparisons, based on different sample sizes.
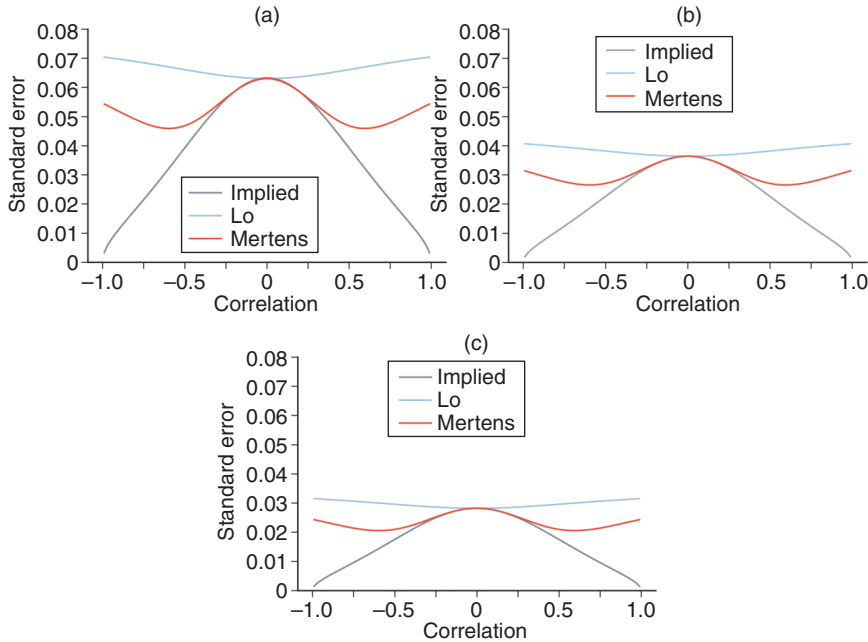

(a) $N = 252$. (b) $N = 756$. (c) $N = 1260$. We note that the implied confidence intervals are within Lo's, although primarily for greater predictive power.

FIGURE 6 Sharpe ratios: full distribution.

(a) $N = 252$. (b) $N = 756$. (c) $N = 1260$. While the 95th percentile shows close agreement between Lo's large-sample standard errors and implied standard errors, the implied distribution is far more fat-tailed.

FIGURE 7 Standard errors based on different sample sizes and formulas.

(a) $N = 252$. (b) $N = 756$. (c) $N = 1260$. Ignoring parameter uncertainty, Merten's adjustment to Lo's standard errors improves standard errors to be nearly as tight as implied. In practice, parameter uncertainty hurts the performance.

distributions, noting that Lo (2002) is a broader result on the large-sample limits of Sharpe ratios. We are concerned with Pearson skewness and kurtosis, ie,

$$\gamma_3 = \frac{\mu_3}{m} u_2^{3/2}, \qquad \gamma_4 = \frac{\mu_4}{\mu_2^2},$$

although it is not hard to consider other definitions of skewness and kurtosis using unbiased estimators of the moments, as given in Joanes and Gill (1998) – in this case originally from Cramér (1946). Given these definitions, under the assumption of normality for the underlying returns (or, correspondingly, using large-sample limits) where the sample size is T, standard errors are given as

$$\text{stderr}_{\gamma_3} = \sqrt{\frac{6(T-2)}{(T+1)(T+3)}}, \qquad \text{stderr}_{\gamma_4} = \sqrt{\frac{24T(T-2)(T-3)}{(T+1)^2(T+3)(T+5)}}.$$

In the case of dynamic strategies, using our assumption of normal signal and normal returns, we are able to derive the following.

COROLLARY 3.2   (Higher moment standard errors)   *For returns $R_t \sim \mathcal{N}(0, \sigma_R^2)$ and signal $X_t \sim \mathcal{N}(0, \sigma_X^2)$ with correlation ρ and sample size T, the standard errors are given by*[7]

$$\mathrm{stderr}_{\gamma_3} = -\frac{6(\hat{\rho}^2 - 1)}{(\hat{\rho}^2 + 1)^{5/2}} \sqrt{\frac{1 - \hat{\rho}^2}{T - 2}}$$

and

$$\mathrm{stderr}_{\gamma_4} = -\frac{48\hat{\rho}(\hat{\rho}^2 - 1)}{(\hat{\rho}^2 + 1)^3} \sqrt{\frac{1 - \hat{\rho}^2}{T - 2}}$$

for $|\hat{\rho}| < 1$.

We rely on the delta method, recognizing that $\mathrm{stderr}_{\gamma_k} = (\partial \gamma_k / \partial \rho)\, \mathrm{stderr}_\rho$ for $k = 3, 4$, given the following easily calculated derivatives:

$$\frac{\partial \gamma_3}{\partial \rho} = -\frac{6(\rho^2 - 1)}{(\rho^2 + 1)^{5/2}}, \tag{3.4}$$

$$\frac{\partial \gamma_4}{\partial \rho} = -\frac{48\rho(\rho^2 - 1)}{(\rho^2 + 1)^3}. \tag{3.5}$$

As we can tell from the formulas in Corollary 3.2, the derived standard errors for both skewness and kurtosis collapse to zero when $\rho = 1$.

While we can solve for ρ in terms of γ_k for $k = 3, 4$, the formulas are not easy to present (especially for kurtosis), and we believe that the statement in terms of correlation is easier to use.
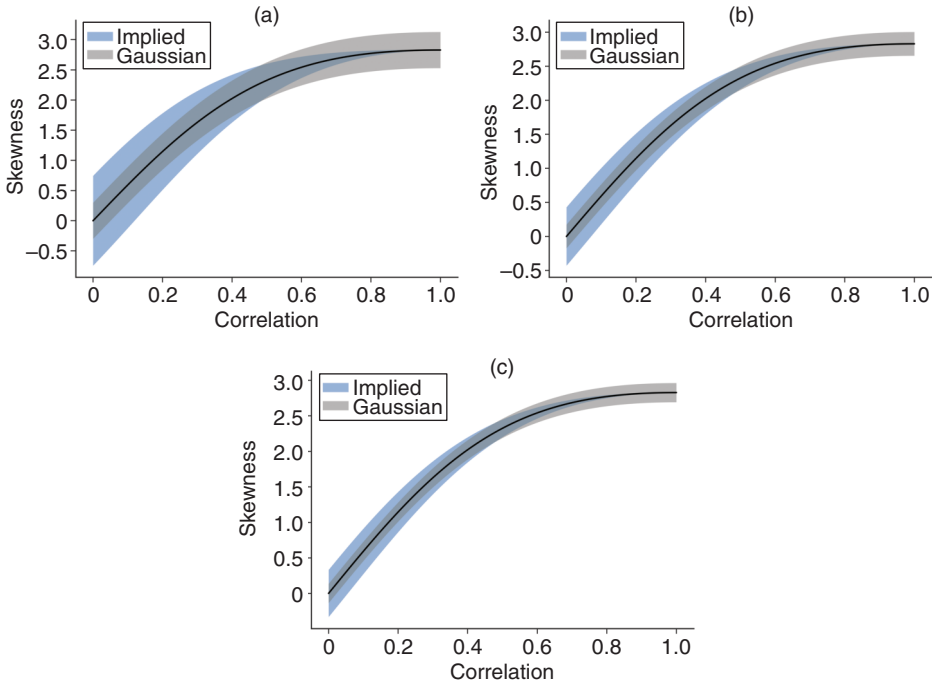
We note that, unlike the argument for using our refined standard errors over those presented in Lo (2002), the rationale for using the skewness and kurtosis standard errors presented in (3.4) is that returns are, for most practical purposes, not close to normal, and the product of two normals is more relevant for dynamic strategies. We elaborate on this in Section 5.

4 MULTIPLE ASSETS

We now consider whether there is a diversification benefit from adding more independent "bets" to our portfolio, and to what extent we can benefit from this. For context we note that portfolios of dynamic strategies can behave very differently from single strategies. For instance, Hoffman and Kaminski (2016) have noted that, while

---

[7] While ρ can be expressed in terms of either γ_3 or γ_4 in order to eliminate it from these expressions, unlike the case of the standard errors of the Sharpe ratio, the expressions are too complicated to be that useful.

FIGURE 8   Standard errors for skewness for different sample sizes, implied versus Gaussian.

(a) $N = 252$. (b) $N = 756$. (c) $N = 1260$. Implied standard errors, especially for skewness, are generally larger than those for normal distributions. We argue that the implied standard errors are more appropriate for dynamic strategies.

single strategies can have skewness ranging around $[1.3, 1.7]$ and kurtosis around $[8.8, 15.3]$, portfolio skewness can be as low as 0.1.
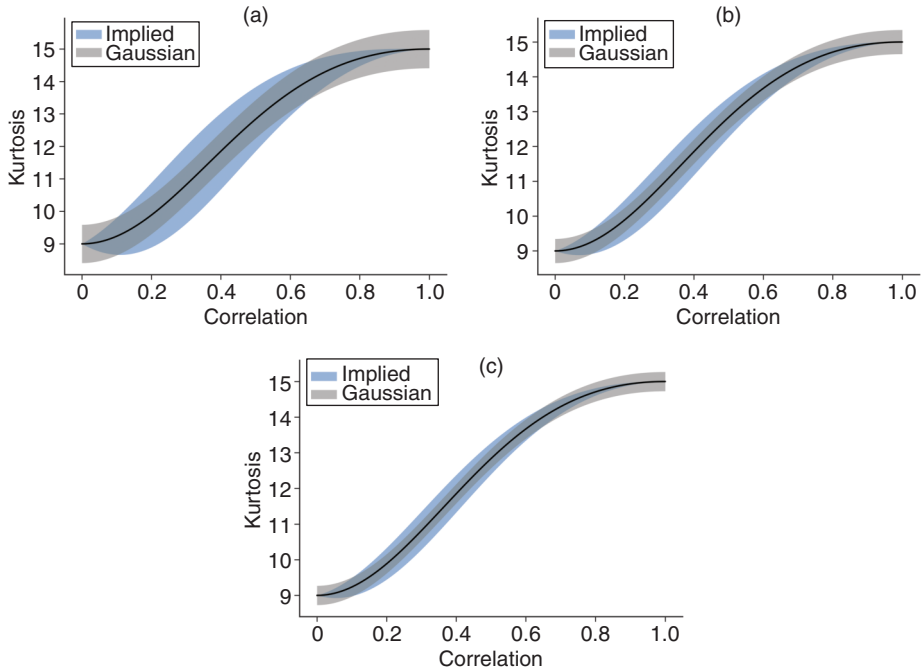
We first consider N independent returns as an N-vector, $R_t \sim \mathcal{N}(0, \sigma^2 I)$, assumed to have the same variance. We devise signals $X_t \sim \mathcal{N}(0, \gamma^2 I)$. The inner product $X_t \cdot R_t$ has a density ψ whose moment generating function is given by (Simon 2006, Chapter 6)

$$M_N(t) = (1 - 2t\sigma\gamma\rho - \sigma^2\gamma^2 t^2(1 - \rho^2))^{-N/2}.$$

From this we can easily derive four moments:

$$\mu_1 = N\sigma\gamma\rho,$$
$$\mu_2 = N\sigma^2\gamma^2((N+1)\rho^2 + 1),$$

FIGURE 9 Standard errors for kurtosis for different sample sizes, implied versus Gaussian.

(a) $N = 252$. (b) $N = 756$. (c) $N = 1260$. Implied kurtosis standard errors are sometimes larger and sometimes tighter than the Gaussian case. We argue that the implied standard errors are more appropriate for dynamic strategies.

$$\mu_3 = N(N + 2)\sigma^3\gamma^3\rho((N + 1)\rho^2 + 3),$$
$$\mu_4 = \sigma^4\gamma^4((N + 6)(N + 4)(N + 2)N\rho^4$$
$$+ 3(N + 2)N(1 - \rho^2)^2$$
$$+ 6(N + f4)(N + 2)N\rho^2(1 - \rho^2)).$$

This leads to the centralized moments

$$\sigma^2 = N(\rho^2 + 1)$$

and

$$\mu_3^c = 2N\rho(\rho^2 + 3).$$

From these we derive the Sharpe ratio,

$$\text{SR} = \frac{\sqrt{N}\,\rho}{\sqrt{\rho^2 + 1}}.$$

Maximizing the Sharpe ratio over ρ leads to $\sqrt{N}\sqrt{2}/2$, clearly showing the benefit of diversification when measuring the Sharpe ratio.

The skewness is

$$\gamma_3 = \frac{1}{\sqrt{N}} \frac{2\rho(\rho^2 + 3)}{(\rho^2 + 1)^{3/2}}$$

and if we consider the maximal Sharpe ratio, the corresponding skewness,

$$\gamma_3^{\max} = \frac{8N}{(2N)^{3/2}} = \frac{2\sqrt{2}}{\sqrt{N}},$$

will show reductions on the order of $1/\sqrt{N}$ in the total number of (orthogonal) assets. This is as expected from large diverse portfolios. In the limit, simple application of the central limit theory should give us asymptotic normality. Effectively, introducing more purely orthogonal assets will increase Sharpe ratios but decrease the (relatively desirable) positive skewness.

If we have multiple possibly correlated assets and multiple possibly correlated signals, we assert that an optimal strategy would be to perform CCA,[8] resulting in a set of decorrelated strategies (using a combination of signals to weight a portfolio of assets). The resulting strategies are decorrelated but with unequal returns and variances. Many results in this section will apply after scaling the portfolio

---

[8] Canonical correlation (from Hotelling (1936): see, for example, Rencher and Christiansen (2012, Chapter 11)) is defined by first finding the linear vectors w_1 and v_1 with $|w_1| = |v_1| = 1$, such that $\rho(w_1 \cdot R, v_1 \cdot X)$ is maximized. The resulting correlation is the canonical correlation. The canonical variates are defined by finding subsequent unit vectors w_k and v_k such that $\rho(w_k \cdot R, w_j \cdot R) = \delta_{kj}$, $\rho(v_k \cdot X, v_j \cdot X) = \delta_{kj}$ and $\rho(w_k \cdot R, v_k \cdot X)$ is maximized, leading to $\rho(w_k \cdot R, v_j \cdot X) = r_k \delta_{kj}$. The solution is obtained via a generalized eigenvalue problem:

$$\Sigma_{RR}^{-1} \Sigma_{RX} \Sigma_{XX}^{-1} \Sigma_{XR} w_k = r_k^2 w_k, \qquad \Sigma_{XX}^{-1} \Sigma_{XR} \Sigma_{RR}^{-1} \Sigma_{RX} v_k = r_k^2 v_k,$$

where Σ is the partitioned correlation matrix of (R, X) and the canonical correlates w_k and v_k are the eigenvectors with the same eigenvalues r_k. The corresponding portfolios of canonical strategies $S_k^{\text{CCA}} \equiv (v_k \cdot X)(w_k \cdot R)$ each have returns and variances as characterized by (2.1) and (2.2) with corresponding correlations r_k (ie, with Sharpe ratios given by $\text{SR}[S_k] = r_k/\sqrt{r_k^2 + 1}$). Due to their independence, they can easily be weighted to optimize the portfolio Sharpe ratio. The method of weighting the canonical strategies is, of course, similar to a risk-parity portfolio, due to the independence of the asset returns. We assert that this method gives the maximal Sharpe ratio for the linear combination of signals and returns, although we leave this proof to a subsequent paper.

returns. The end result may easily be optimized using simple mean–variance analysis (reweighting the returns on the independent strategies). We leave the details for another study.

While our optimizer is unlikely to be in use among CTAs, it is still notable that widely diversified CTAs (irrespective of underlying asset correlations) appear to have decent Sharpe ratios but lower positive skewness, much in line with the discussion in this section. Our simple results here about the final Sharpe ratio and skewness depend, of course, not only on the independence of the underlying assets but the signals themselves, which must only be correlated with their respective asset returns. While this is not an altogether natural setting, it is suggestive of the gains that can be made by introducing purely orthogonal sources of risk, or perhaps by orthogonalizing (or attempting to orthogonalize) asset returns prior to forming signals, later recombining them into a portfolio. This may lead to far more desirable properties of portfolios than finding strategies on multiple nonorthogonalized assets.

5 GAUSSIAN RETURNS VERSUS PRODUCTS OF GAUSSIAN RETURNS

While we believe that the assumption of Gaussian returns (and Gaussian signal) is a simplification, we also believe this is far more realistic than the assumption of Gaussian returns for a dynamic strategy. Throughout this paper, we consider Gaussian (log) returns $R \sim \mathcal{N}(0, \sigma_R^2)$ and Gaussian signal $X \sim \mathcal{N}(0, \sigma_X^2)$, which together are jointly Gaussian and form components of the dynamic strategy $S_t = X_t R_t$, whose properties we study.

To be clear, our signal is not considered to have foresight and is fully known as of time t, while the return R_t is from t to $t + \delta t$. All expectations calculated are unconditional, or can be thought of as conditioned on $t_0 < t < t + \delta_t$. Consequently, each element, the signal and the return, will be random variables.

Were we to consider expectations conditional on t, then the resulting strategy returns S_t would be trivially Gaussian. In the unconditional case, the resulting returns are far more interesting and relevant.

CTA returns are known to generally be positively skewed and highly kurtotic over the relevant horizons we are concerned with (ie, daily, weekly, monthly), as has been noted by Potters and Bouchaud (2005), Hoffman and Kaminski (2016) and others. If we measure far longer horizon returns, asymptotic theory should show that favorable qualities such as skewness may disappear.

Consequently, even though we make many comparisons to results either stemming from asymptotic theory (see, for example, Lo 2002) or using exact normality, this comparison does not, in fact, compare like for like. Clearly, Lo (2002) is appropriate for large samples, as is possible under conditions when the central limit theorem

holds, eg, with weak dependence, summing returns over increasingly longer horizons or in the case of a large cross-sectional dimension with increasing numbers of decorrelated assets. For dynamic strategies, asymptotic normality should be expected for large numbers of decorrelated dynamic strategies as well as for long-horizon (eg, annual or longer, nonoverlapping) returns for single dynamic strategies.

Consequently, we believe our standard error results are more appropriate for hypothesis testing on statistics for dynamic strategies. We discuss a strategy for establishing product measures as large-sample limits in Appendix A online, although asymptotics are beyond the scope of the current study.

6 CONCLUSION

Fully systematic dynamic strategies are used by a large portion of the asset management industry as well as by many noninstitutional participants. Meanwhile, they are only partly understood. Many funds and strategies (eg, especially investment bank "smart-beta" or styles-based products) involve investment in strategies that are not optimized in any sense. Strategies that are paid via index swaps have many limits in terms of their adaptability, leading to often highly suboptimal end results. While there have been some very significant results derived in the theoretical properties of these dynamic strategies, there is still much more work left to do. Given that most academic literature in this area considers more general distributions, there has not been a firm foundation on which to build and extend these results.

It is hoped that this paper will form a foundational approach to the study of dynamic strategies and how to optimize them. We make efforts to understand their properties without claiming to understand why they work (ie, why there are stable autocorrelation functions in the first place). Given that most asset returns are known to have nontrivial autocorrelations, we can establish many results. In particular, we derive a number of results merely by applying well-known techniques to dynamic strategies; for example:

- strategy returns are shown to be positively skewed and leptokurtic;

- Sharpe ratios are characterized, as are skewness and kurtosis;

- the standard errors for Sharpe, skewness and kurtosis are derived;

- strategies designed to optimize Sharpe ratios should be based on TLS rather than minimizing prediction error; and

- gains from adding orthogonal assets/risks are quantified.

Some of these items are empirically well known, but others are genuinely new. Meanwhile, we have extended our results to the derivation of overfitting penalties akin to

Mallow's C_p or AIC, and these can be used for model selection and to predict likely out-of-sample Sharpe ratios from in-sample fits (see Koshiyama and Firoozye 2018).

Our study is incomplete. We believe that there is a good deal of interesting work to be done in areas such as

- optimal linear strategies incorporating transaction costs;

- optimal linear strategies relaxing normality;

- normalized linear signals (eg, z-scores) and optimal nonlinear functions of z-scores;[9]

- nonlinear strategies that are optimized to specific utility functions, possibly incorporating smoothness constraints, especially when relaxing normality;

- local optimality when relaxing stationarity; and

- good-deal bounds in the presence of autocorrelated assets with possible nonstationarity or structural breaks.

We note that our assumptions were never meant to be completely realistic: stationary returns with fixed autocorrelation function and Gaussian innovations can only work in theory, not in reality. Many quantitative traders design strategies to overcome the challenges of dealing with real-world data issues and the issues of overfitting. We nonetheless present them as a good starting point for further analysis, hoping to use this work as the basis for further exploration and to place the general study of dynamic strategies onto a more firm theoretical footing.

Some of our findings should be relevant to practitioners. In particular, the use of OLS and other forecast-error-minimizing methods is not necessarily optimal, depending on the problem at hand; total least squares or other correlation-maximizing methods such as CCA may be more efficient. High Sharpe ratios and positive skewness are often quoted as rationales for entering into strategies, and strategies are changed with the rationale of increasing these measures. The relative significance of any of these changes depends on confidence intervals or standard errors, and we have derived these to be specifically suited for dynamic trading strategies. Kurtosis is not studied as often, but as we show, all dynamic strategies should be leptokurtic, and this is an important attribute of these strategies. Other results, such as overfitting penalties and optimal nonlinear strategies, we save for subsequent

Change OK?

Words added – OK?

---

[9] We note that normalized signals applied to normalized returns series can be represented as the product of two Student t distributions, which has also been relatively well studied (Joarder 2007; Nadrajah and Kotz 2004), and the results are qualitatively very similar to those of this study. However, the more commonly used strategy of applying normalized signals to returns, with the resulting strategies then volatility scaled, cannot be derived as a trivial application of well-known results.

papers. With a more solid theoretical footing as a sort of rule-of-thumb for the development, optimization, selection and alteration of dynamic strategies, we only hope that there is room to improve strategy design.

<u>In</u> two follow-up papers, we shall use the same basic set-up to introduce an overfitting penalization on Sharpe ratios, which is meant to be a better predictor of out-of-sample performance, contrasting this to other overfitting techniques such as <u>information criteria and cross-validation</u>. We will also consider utility-optimizing nonlinear transforms of signals, their properties and standard errors.

Text from abstract added here – OK? Citations/reference details needed?

Changes to sentence OK?

DECLARATION OF INTEREST

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

ACKNOWLEDGEMENTS

Minor changes to this paragraph – OK?

REFERENCES

Acar, E. (1990). Expected returns of directional forecasters. In *Advanced Trading Rules*, Acar, E., and Satchell, S. (eds), 2nd edn, pp. 122–151. Butterworth–Heinemann, Oxford (https://doi.org/10.1016/B978-075065516-3.50006-6).

Allenbridge (2014). Quantitative Investment Strategy Survey 2014.

Asness, C. S., Moskowitz, T. J., and Pedersen, L. H. (2013). Value and momentum everywhere. *Journal of Finance* **68**(3), 929–985 (https://doi.org/10.1111/jofi.12021).

Avramovic, A. (2017). We're all high frequency traders now. Report, March 15, Credit Suisse Market Structure, Trading Strategy. URL: http://www.smallake.kr/wp-content/uploads/2017/10/HFT_world.pdf.

Cramér, H. (1946). *Mathematical Methods of Statistics*. Princeton University Press (https://doi.org/10.1515/9781400883868).

Babu, G. J., and Feigelson, E. D. (1992). Analytical and Monte Carlo comparisons of six different linear least squares fits. *Communications in Statistics: Simulation and Computation* **21**(2), 533–549 (https://doi.org/10.1080/03610919208813034).

Details?

Baltas, N., and Kosowski, R. (2013). Momentum strategies in futures markets and trend-following funds. Paper presented at Finance Meeting EUROFIDAI-AFFI, December, Paris (https://doi.org/10.2139/ssrn.1968996).

Bouchaud, J. P., and Potters, M. (2009). Financial applications of random matrix theory: a short review. Preprint (arXiv:0910.1205).

Bouchaud, J. P., Dau, T. L., Deremble, C., Lemperiere, Y., Nguyen, T. T., and Potters, M. (2017). Tail protection for long investors: convexity at work. *The Journal of Investment Strategies* **7**(1), 61–84 (https://doi.org/10.21314/JOIS.2017.093).

Bruder, B., and Gaussel, N. (2011). Risk–return analysis of dynamic investment strategies. Working Paper, Social Science Research Network (https://doi.org/10.2139/ssrn.2465623).

Bruder, B., Dao, T. L., Richard, J. C., and Roncalli, T. (2011). Trend filtering methods for momentum strategies. Working Paper, Social Science Research Network (https://doi.org/10.2139/ssrn.2289097).

Cui, G., Yu, X., Iommelli, S., and Kong, L. (2016). Exact distribution for the product of two correlated Gaussian random variables. *IEEE Signal Processing Letters* **23**(11), 1662–1666 (https://doi.org/10.1109/LSP.2016.2614539).

Fama, E. F., and French, K. R. (1992). The cross-section of expected stock returns. *Journal of Finance* **47**(2), 427–465 (https://doi.org/10.1111/j.1540-6261.1992.tb04398.x).

Fuller, W. A. (1987). *Measurement Error Models*. Wiley Series in Probability and Statistics, Volume 305. Wiley (https://doi.org/10.1002/9780470316665).

Fung, W., and Hsieh, D. A. (1997). Empirical characteristics of dynamic trading strategies: the case of hedge funds. *Review of Financial Studies* **10**(2), 275–302 (https://doi.org/10.1093/rfs/10.2.275).

Gigerenzer, G., Todd, P. M., and the ABC Research Group (1999). *Simple Heuristics That Make Us Smart*. Oxford University Press.

Golub, G. H., and Van Loan, C. F. (1980). An analysis of the total least squares problem. *SIAM Journal on Numerical Analysis* **17**(6), 883–893 (https://doi.org/10.1137/0717073).

Götze, F., and Tikhomirov, A. (1999). Asymptotic distribution of quadratic forms. *Annals of Probability* **27**(2), 1072–1098 (https://doi.org/10.1214/aop/1022677395). Change OK?

Götze, F., and Tikhomirov, A. (2002). Asymptotic distribution of quadratic forms and applications. *Journal of Theoretical Probability* **15**, 423–475 (https://doi.org/10.1023/A:1014867011101).

Götze, F., Tikhomirov, A., and Yurchenko, V. (2007). Asymptotic expansion in the central limit theorem for quadratic forms. *Journal of Mathematical Sciences* **147**, 6891–6911 (https://doi.org/10.1007/s10958-007-0512-8).

Grinold, R. C., and Kahn, R. N. (1999). *Active Portfolio Management*, 2nd edn. McGraw-Hill, New York.

Hald, A. (2008). *A History of Parametric Statistical Inference from Bernoulli to Fisher, 1713–1935*. Sources and Studies in the History of Mathematics and Physical Sciences. Springer (https://doi.org/10.1007/978-0-387-46409-1).

Haldane, J. B. S. (1942). Moments of the distributions of powers and products of normal variates. *Biometrika* **32**(3–4), 226–242 (https://doi.org/10.1093/biomet/32.3-4.226).

Hamdan, R., Pavlowsky, F., Roncalli, T., and Zheng, B. (2016). A primer on alternative risk premia. Working Paper, Social Science Research Network (https://doi.org/10.2139/ssrn.2766850).

Harvey, C. R., Hoyle, E., Korgaonkar, R., Rattray, S., Sargaison, M., and Van Hemert, O. (2018). The impact of volatility targeting. Working Paper, Social Science Research Network (https://doi.org/10.2139/ssrn.3175538).

Hoffman, B., and Kaminski, K. (2016). The taming of the skew. Working Paper, June, Campbell White Paper Series. URL: https://bit.ly/2wihmCp.

Hotelling, H. (1936). Relations between two sets of variates. *Biometrika* **28**(3–4), 321–377 (https://doi.org/10.1093/biomet/28.3-4.321).   Change OK?

Hurst, B., Ooi, Y. H., and Pedersen, L. H. (2017). A century of evidence on trend-following investing. Working Paper, Social Science Research Network (https://doi.org/10.2139/ssrn.2993026).

Hyndman, R., Koehler, A. B., Ord, J. K., and Snyder, R. D. (2008). *Forecasting with Exponential Smoothing: The State Space Approach*. Springer Series in Statistics. Springer (https://doi.org/10.1007/978-3-540-71918-2).

Isserlis, L. (1918). On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables. *Biometrika* **12**(1–2), 134–139 (https://doi.org/10.1093/biomet/12.1-2.134).

Jegadeesh, N., and Titman, S. (1993). Returns to buying winners and selling losers: implications for stock market efficiency. *Journal of Finance* **48**(1), 65–91 (https://doi.org/10.1111/j.1540-6261.1993.tb04702.x).

Joanes, D. N., and Gill, C. N. (1998). Comparing measures of sample skewness and kurtosis. *Journal of the Royal Statistical Society* D **47**, 183–189 (https://doi.org/10.1111/1467-9884.00122).

Joarder, A. (2007). On some characteristics of the bivariate t-distribution. *International Journal of Modern Mathematics* **2**(2), 191–204.

Kan, R. (2008). From moments of sum to moments of product. *Journal of Multivariate Analysis* **99**(3), 542–554 (https://doi.org/10.1016/j.jmva.2007.01.013).

Ko, M. (2006). Functional central limit theorems for multivariate linear processes generated by dependent random vectors. *Communications of the Korean Mathematical Society* **21**(4), 779–786 (https://doi.org/10.4134/CKMS.2006.21.4.779).

Koshiyama, A. S., and Firoozye, N. B. (2019). Avoiding back-testing overfitting by covariance-penalties: an empirical investigation of the ordinary and total least squares cases. (Submitted.)   Details?

Lee, W. (2000). *Theory and Methodology of Tactical Asset Allocation*. Frank J. Fabozzi Series, Volume 65. Wiley.

Lempérière, Y., Deremble, C., Seager, P., Potters, M., and Bouchaud, J. P. (2014). Two centuries of trend following. Preprint (arXiv:1404.3274).

Levine, A., and Pedersen, L. H. (2016). Which trend is your friend? *Financial Analysts Journal* **72**(3), 51–66 (https://doi.org/10.2469/faj.v72.n3.3).

Leyang, W. (2012). Properties of the total least squares estimation. *Geodesy and Geodynamics* **3**(4), 39–46 (https://doi.org/10.3724/SP.J.1246.2012.00039.1).

Lo, A. W. (2002). The statistics of Sharpe ratios. *Financial Analysts Journal* **58**(4), 36–52 (https://doi.org/10.2469/faj.v58.n4.2453).

Magnus, J. R. (1978). The moments of products of quadratic forms in normal variables. *Statistica Neerlandica* **32**(4), 201–210 (https://doi.org/10.1111/j.1467-9574.1978.tb01399.x).

Makridakis, S., and Hibon, M. (2002). The M3-competition: results, conclusions and implications. *International Journal of Forecasting* **16**(4), 451–476 (https://doi.org/10.1016/S0169-2070(00)00057-1).

Markovsky, I., and Van Huffel, S. (2007). Overview of total least-squares methods. *Signal Processing* **87**(10), 2283–2302 (https://doi.org/10.1016/j.sigpro.2007.04.004).

Martin, R., and Bana, A. (2012). Non-linear momentum strategies. *Risk Magazine* **25**(11), 60–65.

Martin, R., and Zou, D. (2012). Momentum trading: 'skews me. *Risk Magazine* **25**(8), 52–57.

Mertens, E. (2002). Comments on variance of the IID estimator in Lo (2002). Research Note. URL: https://bit.ly/2XEmxry.

Michalowicz, J. V., Nichols, J. M., Bucholtz, F., and Olson, C. C. (2011). A general Isserlis theorem for mixed-Gaussian random variables. *Statistics and Probability Letters* **81**(8), 1233–1240 (https://doi.org/10.1016/j.spl.2011.03.022).

Miller, M. (2017). Evaluating bank-sponsored risk-premium strategies. *Pensions & Investments*, April 5. URL: https://bit.ly/34DXDtz.

Nadarajah, S., and Kotz, S. (2004). Multitude of bivariate t distributions. *Statistics* **38**(6), 527–539 (https://doi.org/10.1080/02331880412331319305).

Nadarajah, S., and Pogány, T. K. (2016). On the distribution of the product of correlated normal random variables. *Comptes Rendus Mathématique* **354**(2), 201–204 (https://doi.org/10.1016/j.crma.2015.10.019).

Parczewski, P. (2014). A Wick functional limit theorem. *Probability and Mathematical Statistics* **34**(1), 127–145.

Pav, S. E. (2016). Notes on the Sharpe ratio. Working Paper, March. URL: https://bit.ly/2K8uhuc.

Potters, M., and Bouchaud, J. P. (2005). Trend followers lose more often than they gain. Preprint (arXiv: physics/0508104). URL: https://arxiv.org/abs/physics/0508104.

Rahman, M. D., and Yu, K. B. (1987). Total least squares estimation using linear prediction. *IEEE Transactions on Acoustics, Speech, and Signal Processing* **35**(10), 1440–1454 (https://doi.org/10.1109/TASSP.1987.1165059).

Rencher, A. C., and Christensen, W. F. (2012). *Methods of Multivarite Analysis*, 3rd edn. Wiley (https://doi.org/10.1002/9781118391686).

Simon, M. K. (2006). *Probability Distributions Involving Gaussian Random Variables*. Springer.

Wick, G. C. (1950). The evaluation of the collision matrix. *Physical Review* **80**(2), 268–272 (https://doi.org/10.1103/PhysRev.80.268).

Zhang, X.-D. (2017). *Matrix Analysis and Applications*. Cambridge University Press (https://doi.org/10.1017/9781108277587).