

## OPTIMAL DYNAMIC STRATEGIES ON GAUSSIAN RETURNS

NICK FIROOZY AND ADRIANO KOSHIYAMA

Department of Computer Science  
University College London 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. Assuming Gaussian returns and Gaussian dynamic weights or *signals*, (e.g., linear filters of past returns, such as simple moving averages, exponential weighted moving averages, forecasts from ARIMA 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 modelling the interaction of strategy weights with returns, we demonstrate that positive skewness and excess kurtosis are essential components of all positive Sharpe dynamic strategies, which is generally observed empirically; demonstrate that *total least squares* (TLS) or *orthogonal least squares* is more appropriate than OLS for maximizing the Sharpe ratio, while *canonical correlation analysis* (CCA) is similarly appropriate for the multi-asset case; derive standard errors on Sharpe ratios which are tighter than the commonly used standard errors from Lo; and derive standard errors on the skewness and kurtosis of strategies, apparently new results. We demonstrate these results are applicable asymptotically for a wide range of stationary time-series. In two follow-up papers, we use the same basic set-up to introduce an over-fitting penalization on Sharpe ratios which is meant to be a better predictor of out-of-sample performance, contrasting this to other over-fitting techniques such as criteria and CV, and consider utility-optimising non-linear transforms of signals, their properties and standard errors. Possible future extensions of the work to normalised signals, to multi-period returns, and non-linear transforms are discussed together with extensions to multi-asset dynamic strategies.

Algorithmic Trading, Dynamic Strategies, over-fitting, Quantitative Finance, Signal Processing

*MSC Numbers:* 60G10, 62E15, 62P05, 62F99, 91G70, 91G80

*JEL Classifications:* C13, C58, C61, G11, G19

## 1. INTRODUCTION

CTAs (*Commodity Trading Advisors*) or managed-future accounts are a subset of asset managers with over \$341bn of assets under management [Barclay Hedge \(2017\)](#) as of Q2 2017. The predominant strategy which CTAs employ is trend-following. Meanwhile, bank structuring desks have devised a variety of *risk-premia* or *styles* strategies (including momentum, mean-reversion, carry, value, etc) which have been estimated to correspond to between approximately \$150bn [Miller \(2016\)](#) to \$200bn [Allenbridge IS \(2014\)](#) assets under management. Responsible for over 80% of trade volume in equities and a large (but undocumented amount due to the OTC nature) of the FX market, [Credit Suisse \(2017\)](#), high-frequency trading firms (HFTs) and e-trading desks in investment banks are known to make use of many strategies which are effectively short-term

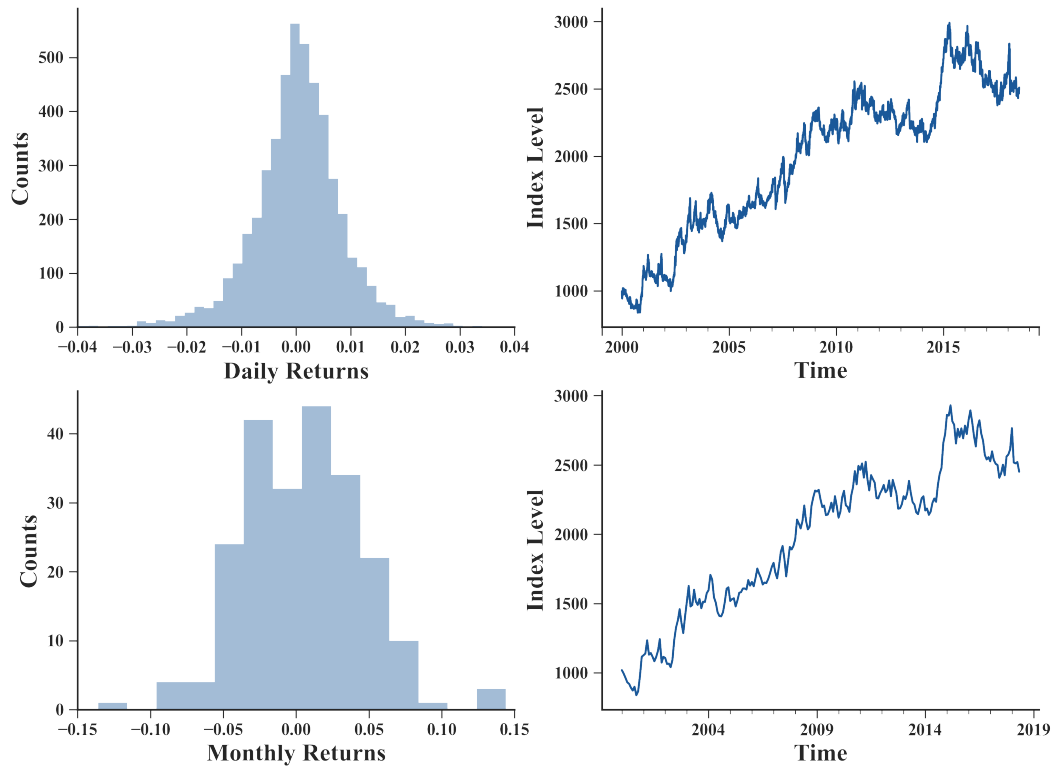


FIGURE 1. SocGen Trend Followers Index: daily returns and monthly returns profiles

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 optimisation, has only been undertaken in relatively limited contexts.

The corresponding statistics for the SG Trend index area in the table below and except for some noise show that skewness and excess kurtosis are largely positive for CTAs.

TABLE 1. Soc Gen Trend Index, Daily and Monthly Statistics

	Daily	Monthly
Ann Avg Return (%)	5.695	5.752
Volatility (%)	13.283	14.088
Sharpe Ratio	0.429	0.408
Skewness	-0.448	0.186
Exc Kurtosis	3.845	0.807

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).<sup>1</sup> 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 like 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 which determines the strategy's behaviour.

Equities, and in particular SPX is known to mean-revert over short horizons (e.g., shorter than 1m, typically on the order of 5-10 days) and trend only over longer horizons (i.e., 3m-18m), and mean-revert again over even longer horizons (i.e., 2y-5y) as has been well-established by the quant equities literature following on the study of Jegadeesh and Titman (1993) and the work of Fama and French (1992). This distinct form of behaviour, with reversals on small-scale, trend on an intermediate and reversion on a long scale, is frequently observed across a large number of asset classes and strategies can be designed to take advantage of the behaviour of asset-prices across each time-scale.

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 for allocating 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 macro-traders (CTAs, and other trend followers) include:

- 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, Damped HW
- Difference between current price and moving average:<sup>2</sup>

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

- Forecasts from ARMA(p, q) models:

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

- Differences between SMAs:

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

<sup>1</sup>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 auto-correlation function

<sup>2</sup>We note that if we replace  $P$  by  $\log(P)$  and  $R_t = \log(P_t) - \log(P_{t-1})$ , this filter amounts to  $X_t = \sum \frac{T-k}{T} R_{t-k}$ , i.e., a *triangular* filter on returns, which bears some similarity to EWMA on returns.

- Differences between EWMA:

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

and variations using volatility or variance weighting such as z-scores (SMAs or EWMA weighted by a simple or weighted standard deviation, see [Harvey et al \(2018\)](#)), and transformations of each of the signals listed above (e.g. allocations depending on sigmoids of moving averages, reverse sigmoids, Winsorised signals, etc.). Other signals commonly used in equity algorithmic trading include economic and corporate releases, and sentiment as derived from unstructured datasets such as news releases.

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

Frequently, exponential smoothers have been the effective *best* models in various economic forecasting competitions (see, e.g., 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 (e.g., inventories, employment, monetary aggregates). EWMA (and the related (S)MA), and HW remain some of the most commonly used filtering methods for CTAs and HFT shops.

In the case of returns which are normal with fixed autocorrelation function (ACF), i.e., those which are covariance stationary, signals created from linear combinations of historic returns are indeed normal random variables which are jointly normal with returns. External datasets (e.g., unstructured data, corporate releases), are less likely to contain normally distributed variables although there is an argument for asymptotic normality. Irrespective, 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 were the first to look at the empirical properties of momentum strategies [Fung and Hsieh \(1997\)](#), noting (without any theoretical foundation) the resemblance of strategy returns to straddle pay-offs.<sup>3</sup> Potters and Bouchaud [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 of [1.3, 1.7] and kurtosis in the range [8.8, 15.3] respectively (see [Hoffman-Kaminski \(2016\)](#)).

<sup>3</sup>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.”

Bruder and Gaussel [Bruder and Gaussel \(2011\)](#) and [Hamdan et al \(2016\)](#) (see Appendix 2 for a superlative use of SDE-based methods for analyzing a wide variety of dynamic strategies) used SDEs to study the power-option like behaviour of pay-offs. Martin and Zou considered general but IID discrete time distributions (see [Martin-Zou \(2012\)](#) and [Martin-Bana \(2012\)](#)) to study the *term-structure* of skewness over various horizons and the effects of certain non-linear transforms on the term structure of return distributions. More recently, Bouchaud et al [Bouchaud et al \(2016\)](#) considered more general discrete-time distributions to study the convexity of pay-offs, and the effective dependence of returns on long-term vs short-term variance. Other studies have focused predominantly on the empirical behaviour of returns, the relationship to macro-financial conditions, the persistence of trend-following returns, and the benefits from their inclusion into broader portfolios.

In the larger portion of the theoretical studies, the assumptions have been minimal in order to consider more general return distributions. Due to their generality, the derived results are somewhat more restrictive. Rather than opting for the most general, we choose more specific distributional assumptions, in the hope that we can obtain broader, possibly more practical results. Aside from this current study, the authors have extended this work further to consider the endemic problem of over-fitting (see [Koshiyama and Firoozye \(2018\)](#)), proposing total least squares with covariance penalties as a means of model-selection, showing their outperformance to standard methods, using OLS with AIC.

In this paper, we consider underlying assets with stationary Gaussian returns and a fixed auto-correlation function (i.e., they are a discrete Gaussian process). While we make no defence of 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 non-linear 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 a EWMA or an ARMA forecast, can be expressed as convolution filters of past returns, i.e., our signal  $X_t$  can be expressed as

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

This is an example of a time-invariant linear filter of a Gaussian process. If we restrict our attention to those filters for which are square summable,  $\sum_1^\infty \phi(k)^2 < \infty$ , 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 distributions 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 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* or IC as given in the *Fundamental Law of Active Management* detailed in [Grinold and Kahn \(1999\)](#). While there is a large body of literature on the IC and its relationship to information ratios, (see for example [Lee \(2000\)](#) for formulas similar to equation (5)), the derivations, resulting formulae and conclusions differ significantly.

We should also mention the work on random matrix theory by Potters and Bouchaud (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 once again different to 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' 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 are some recent efforts to extend to non-Gaussian distributions (see, e.g., Michalowicz et al (2011) for Gaussian-mixture and Kan (2008) for products of quadratic forms and elliptic distributions), and it has 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 is divided into sections on one asset, considered 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 optimisation. 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, more pertinent to the study of dynamic strategies than simple Gaussian measures. In the appendices, we present closed-form solutions to Sharpe ratios in the case of non-zero means. We also discuss extensions to our optimisations in the presence of transaction costs. We touch on the extension to multiple periods as well. As we mentioned, further extensions to over-fitting by the use of covariance penalties (akin to Mallows'  $C_p$  or AIC/BIC) have been presented separately in Koshiyama and Firoozye (2018).

## 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 corresponding auto-correlation function (ACF),  $c(k) = \gamma(k)/\gamma(0)$  at lag  $k$ .

Our main aim is to work with strategies based on linear portfolio weights (or *signals*)  $X_t = \sum_1^\infty a_k R_{t-k}$  for coefficients  $a_k$  generating the corresponding dynamic strategy returns  $S_t = X_t \cdot R_t$  (here, and always, the signal,  $X_t$  is assumed to only have appropriately lagged information). Example strategy weights include exponentially weighted moving averages  $a_k \propto \lambda^k$ , simple moving averages  $a_k = \frac{1}{T} 1_{[1, \dots, T]}$ , forecasts from ARMA models, etc. Most importantly, the portfolio weights  $X$  are normal and *jointly* normal with returns  $R$ . In Appendix B, we show that for a wide set of signals discussed in the Introduction, when applied to Gaussian returns, the signal and returns are jointly Gaussian.

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 intra-day

strategies, this period can be much shorter. The pertinent concern is that the horizon (i.e., one period) is the same horizon over which the rebalancing of strategy weights is done. If weights are rebalanced every five minutes, then the single 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 means of optimising various non-dimensional measures of returns such as the Sharpe ratio via the use of non-linear 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 (e.g., 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.

**2.1. Properties of linear strategies.** Given the joint normality of the signal and the returns, we can explicitly characterise 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' theorem Isserlis (1918), which gives all moments for any multivariate normal random variable in terms of the mean and variance. We also refer to Haldane (1942) who meticulously produces both non-central and central moments for powers and products of Gaussians. While this is a routine application of Isserlis' 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  $X_1, X_2, \dots, X_{2n}$  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  $\rho$  then*

$$\begin{aligned} E[xy] &= \rho \\ E[x^2 y^2] &= 1 + 2\rho^2 \\ E[x^3 y^3] &= 3\rho(3 + 2\rho^2) \\ E[x^4 y^4] &= 3(3 + 24\rho^2 + 8\rho^4) \end{aligned}$$

and thus the central moments of  $xy$  are

$$\begin{aligned} (1) \quad & \mu_1 = \rho \\ (2) \quad & \mu_2 = 1 + \rho^2 \\ (3) \quad & \mu_3 = 2\rho(3 + \rho^2) \\ (4) \quad & \mu_4 = 3(3 + 14\rho^2 + 3\rho^4) \end{aligned}$$

From these one period moments, (and a simple scaling argument giving the dependence on  $\sigma(x)$  and  $\sigma(y)$ ) we can characterise 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  $\rho$ , the Sharpe ratio is given by*

$$(5) \quad \text{SR} = \frac{\rho}{\sqrt{1 + \rho^2}},$$

the skewness is given as

$$(6) \quad \gamma_3 = \frac{2\rho(3 + \rho^2)}{(1 + \rho^2)^{\frac{3}{2}}},$$

and the kurtosis is given by

$$(7) \quad \gamma_4 = \frac{3(3 + 14\rho^2 + 3\rho^4)}{(1 + \rho^2)^2}$$

In the appendix, we extend equations (5) and (6) to the case of non-zero means.

*Proof.* A simple application of Theorem 2.2 give us the following first two moments for our strategy  $S_t = X_t \cdot R_t$ :  $\mu_1 = E[S_t] = E[X \cdot R] = \sigma_X \sigma_R \rho$ . and  $\mu_2 = \text{Var}[S_t] = \sigma_X^2 \sigma_R^2 (\rho^2 + 1)$ . Thus we can derive the following results for the Sharpe ratio,

$$\begin{aligned} \text{Sharpe} &= \frac{\frac{\mu_1}{\mu_2^{1/2}}}{\frac{\sigma_X \sigma_R \rho}{\sigma_X \sigma_R \sqrt{\rho^2 + 1}}} \\ &= \frac{\rho}{\sqrt{\rho^2 + 1}} \end{aligned}$$

Moreover, we can see that the skewness,

$$\begin{aligned} \gamma_3 &= \frac{\frac{\mu_3}{\mu_2^{3/2}}}{\frac{2\rho(3 + \rho^2)}{(1 + \rho^2)^{3/2}}} \\ &= \frac{2\rho(3 + \rho^2)}{(1 + \rho^2)^{3/2}} \end{aligned}$$

Finally, the kurtosis is given by

$$\begin{aligned} \gamma_4 &= \frac{\frac{\mu_4}{\mu_2^2}}{\frac{3(3 + 14\rho^2 + 3\rho^4)}{(1 + \rho^2)^2}} \\ &= \frac{3(3 + 14\rho^2 + 3\rho^4)}{(1 + \rho^2)^2} \end{aligned}$$

□

If we restrict our attention to positive correlations, all three dimensionless statistics are monotonically increasing in  $\rho$ . Consequently, strategies that maximize one of these statistics will maximize the others, although the impact of correlation upon 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 shaded blue histograms correspond to 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.

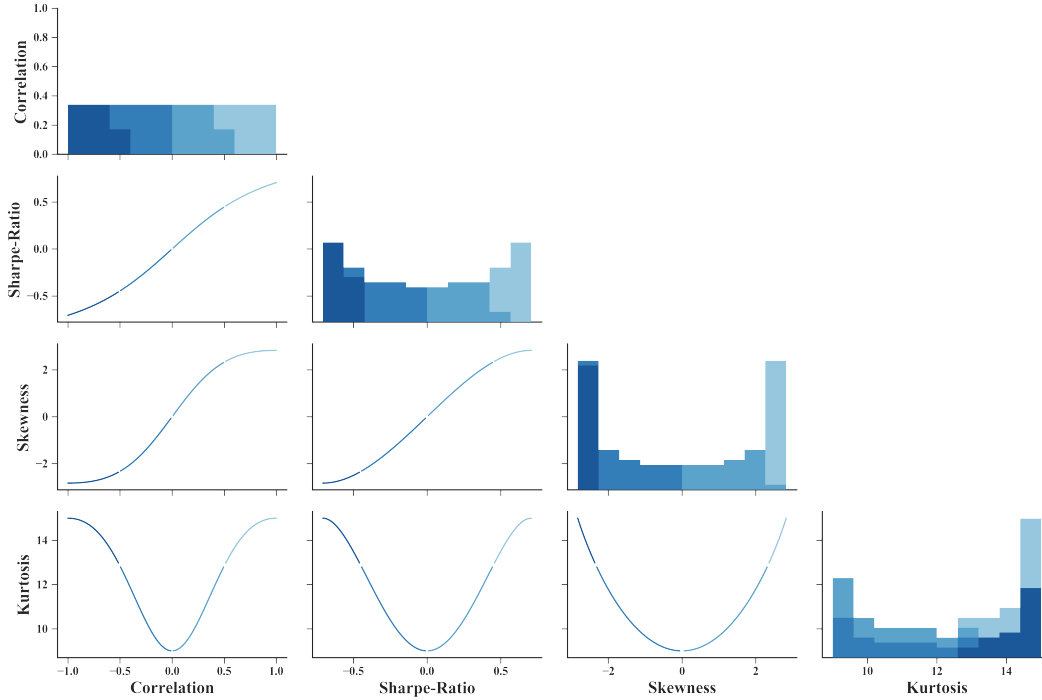


FIGURE 2. **Correlation, Sharpe ratio, Skewness, and Kurtosis pairwise relationship.** A uniform distribution in correlation is bucketed into four ranges  $\{[-1, -0.5], [-0.5, 0], [0, 0.5], [0.5, 1]\}$  as depicted in the bar charts in shades of blue. After transforming the correlation into SR,  $\gamma_3$  and  $\gamma_4$  the frequencies are no longer uniform.

Skewness ranges in  $[-2^{3/2}, 2^{3/2}] \approx [-2.8, 2.8]$ . Unlike the Sharpe ratio, Skewness' dependence on correlation tends to flatten, so to achieve 90% peak skewness, one needs only achieve a 0.60 correlation, while for a 90% peak Sharpe, one needs 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 significantly smaller than the maximal values.

Although we analyse the moments of the strategy  $S_t = X_t R_t$ , the full product density is actually known in closed form (see appendix A, Cui et al (2016) and Nadarajah-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  $\rho \rightarrow 1$ ,

the strategy's density approaches that of a non-central  $\chi^2$ , an effective *best-case* density when considering the design of optimal linear dynamic strategies.

An optimised 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. Annualised Sharpe ratios of between 0.5-1.5 are most common (i.e., correlations of between 3% to 9%) for single asset strategies in this relatively low-frequency regime.

**2.2. Optimisation: Maximal Correlation, Total least squares.** Many algorithmic traders will explain how problematic strategy optimisation is, given the endless concerns of over-fitting, etc. Although these are a concern, the naïve use of strategies which are merely *pulled out of thin air* is equally problematic, where there is no explicit use of optimisation (and, in its place more *eye-balling* strategies or targeting Sharpe ratios rather loosely, effectively a somewhat loose mental optimisation exercise). Practical considerations abound and real-world returns are neither Gaussian nor stationary. We argue irrespectively that using optimisation and a well-specified utility function as a starting point is a means of preventing strategies from being just 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 and Todd (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 the monniker *quantitative* investment strategies.

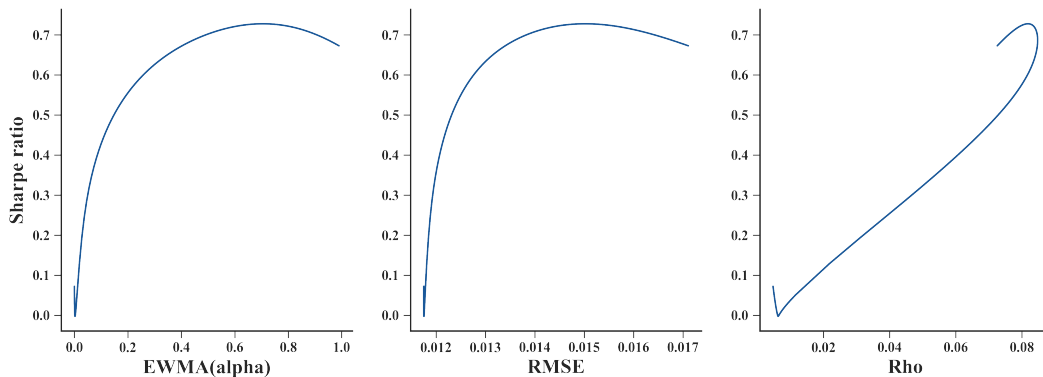


FIGURE 3. **EWMA Strategy Sharpe Ratio vs  $\alpha$ , MSE and correlation** for S&P 500 reversal strategies

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

We can see in figures 3 and 4, a depiction of 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 it is much closer to being linear in the case of the relationship between correlation and Sharpe. For the case of HW (with two parameters), in figure 4 any given MSE can lead to a non-unique Sharpe ratio, sometimes

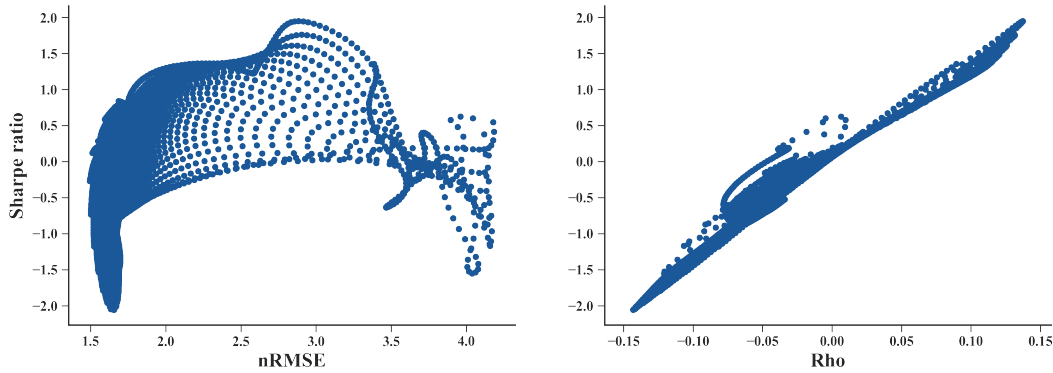


FIGURE 4. **Holt-Winters Strategy Sharpe Ratio vs MSE and correlation** for S&P 500 Reversal Strategies

with a very broad range, leading us to conclude that the optimization is poorly posed. The relationship of correlation to Sharpe is obviously closer to being 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 regression* (TLS) or *orthogonal distance regression*, a form of principal components regression (see, e.g., Golub and Van Loan (1980) and Markovsky and Van Huffell (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 total least squares 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 is a simple form of errors-in-variables (EIV) regression and has been studied since the late 1870s, and is most closely related to principal components analysis. For  $k$  regressors, the TLS fit will produce weights which are orthogonal to the first  $k-1$  principal components.

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

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

but

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

where  $\sigma_{k+1}$  is the smallest singular value for the  $T \times (k+1)$  dimensional matrix  $\tilde{X} = [R, Z]$  (i.e., the concatenation of the features and the returns, see, e.g., Rahman and Yu (1987)<sup>4</sup>). It is well known that, for the case of OLS, the smooth or hat matrix  $\hat{R} = MR$  is given by

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

<sup>4</sup>A more common method for extracting TLS estimates is via a PCA of the concatenation matrix  $\tilde{X}$ , where  $\hat{\beta}^{TLS}$  is chosen to cancel the least significant principal component

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

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

and effectively has a greater number of degrees of freedom than that of OLS, i.e.,

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

with equality only when there is complete collinearity<sup>5</sup>. For this reason, many people see TLS as an *anti-regularisation* method and may result in less-stable response to outliers (see for example, [Zhang \(2017\)](#)). Consequently, there is extensive study of *regularised* TLS, typically using a weighted ridge-regression (or Tikhonov) penalty (see 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 over-fitting 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 which 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 optimising TLS is not appropriate for prediction (see, e.g., [Fuller \(2009\)](#) section 1.6.3). When we seek to maximize the Sharpe ratio of a strategy, the objective should *not be* 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 non-central  $\chi^2$  distribution and the resulting maximal statistics are

$$\begin{aligned} \text{SR}^{max} &= \frac{\sqrt{2}}{2} \approx 0.707 \\ \gamma_3^{max} &= 2\sqrt{2} \approx 2.828 \\ \gamma_4^{max} &= 15.000 \end{aligned}$$

While the estimate for the Sharpe ratio may seem surprisingly low, we comment that these are for a single period, for one single rebalancing. For a daily rebalanced strategy, if we naïvely annualize the Sharpe ratio (by a factor of  $\sqrt{252}$ ), we get a maximal Sharpe of approximately  $\text{SR}_{max} \approx 11.225$ , 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.

We note that our assumption of normality could easily be relaxed by considering non-linear transforms of the signals  $X$  with the end-result that the maximal Sharpe Ratio bounds are relaxed. While this is beyond the scope of the current paper, we note that it is easy to show that simple non-linear strategies, going long one unit if the signal is above a threshold  $k$  and

<sup>5</sup>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 know that the singular values of  $\sigma(L) = \{\lambda_i^2 / (\lambda_i^2 - \sigma_{k+1}^2)\}$  where  $\lambda_i$  are the singular values of  $Z$  (or correspondingly,  $\lambda_i^2$  are the singular values of  $Z'Z$ ), and  $\lambda_1 \geq \dots \geq \lambda_k > 0$  ([Leyang \(2012\)](#)). By the Wilkinson interlacing theorem,  $\lambda_k \geq \sigma_{k+1} \geq 0$  (see [Rahman and Yu \(1987\)](#)). Consequently,

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

with equality iff  $\sigma_{k+1}^2 = 0$  (i.e., when there the  $R^2 = 100\%$  and consequently, OLS and TLS coincide). In other words,  $\text{tr}(M^{TLS}) \geq \text{tr}(M^{OLS})$ .

short one unit if it below  $-k$ , i.e.,  $f_k(X) = 1_{X>k} - 1_{X<-k}$  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 non-zero autocorrelation can lead to violations of Hansen-Jagannathan (or *good deal* bounds).

Noticeable as well from these formulas is that, while Sharpe and skewness may change sign, kurtosis is always bounded below and takes a minimum value of 9 (i.e., an excess kurtosis of 6). Normality of the resulting strategy returns is not a good underlying assumption, since the theoretical value of the Jarque-Bera test would be, at

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

and this is asymptotically  $\chi^2(2)$  (i.e., 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 the strategies on an ongoing basis, and are typically in a quandary over whether the observed change in Sharpe ratio or skewness, when they make changes to their strategies, are in fact statistically significant.

**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 to [Pav \(2016\)](#) for an exhaustive overview of the mechanics of Sharpe ratios, and in particular, Section 1.4, quoting 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, he obtains the large-sample distribution,

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

so a standard error,  $\text{stderr}_{Lo} = \sqrt{(1 + \frac{1}{2} SR^2)/T}$  which he suggests should be approximated using standard error  $\sqrt{(1 + \frac{1}{2} \widehat{SR}^2)/T}$ .

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

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

$$(8) \quad \text{stderr}_{\text{implied}} = \frac{1}{(\hat{\rho}^2 + 1)^{3/2}} \sqrt{\frac{1 - \hat{\rho}^2}{T - 2}}$$

$$(9) \quad \approx (1 - \widehat{\text{SR}}^2) \sqrt{\frac{1 - 2\widehat{\text{SR}}^2}{T - 2}}$$

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

$$(10) \quad \hat{\rho} \sim f_{\rho}(\hat{\rho}) = \frac{(T - 2)(1 - \rho^2)^{(T-1)/2}(1 - \hat{\rho}^2)^{(T-4)/2}}{\pi} \int_0^{\infty} \frac{dw}{(\cosh(w) - \rho\hat{\rho})^{T-1}}$$

The standard errors which approximate those in equation (10) for  $\hat{\rho}$  are

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

(attributed to Sheppard, and used by Pearson, see, e.g., [Hald \(2008\)](#)). Taken together with the results of Theorem 2.3, we apply the delta method to find that the resulting standard errors for our plug-in estimate for the Sharpe ratio,  $\widehat{\text{SR}} = \frac{\hat{\rho}}{\sqrt{\hat{\rho}^2 + 1}}$  is given by

$$\begin{aligned} \text{stderr}_{\text{implied}} &= \frac{\partial \widehat{\text{SR}}}{\partial \hat{\rho}} \cdot \text{stderr}_{\rho} \\ &= \frac{1}{(\hat{\rho}^2 + 1)^{3/2}} \sqrt{\frac{1 - \hat{\rho}^2}{T - 2}}. \end{aligned}$$

which gives us equation (8). If we solve for  $\hat{\rho}$  in terms of  $\widehat{\text{SR}}$ , we are able to derive equation (9).  $\square$

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 are 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 that the tail of the distribution given by Lo is much fatter than ours, in figure (6).

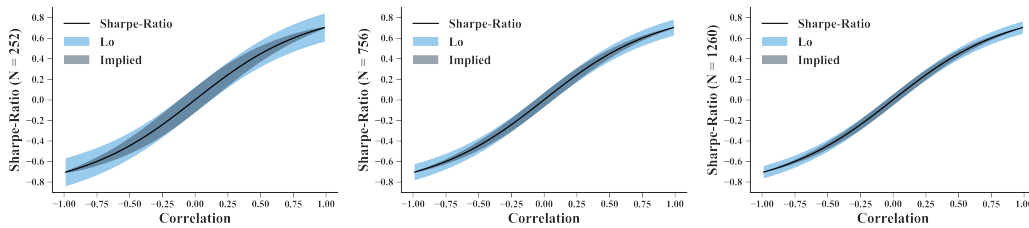


FIGURE 5. **Sharpe ratio and Confidence Interval Comparisons, based on different sample sizes.** We note that the *implied* confidence intervals are within Lo's, although primarily for larger predictive power.

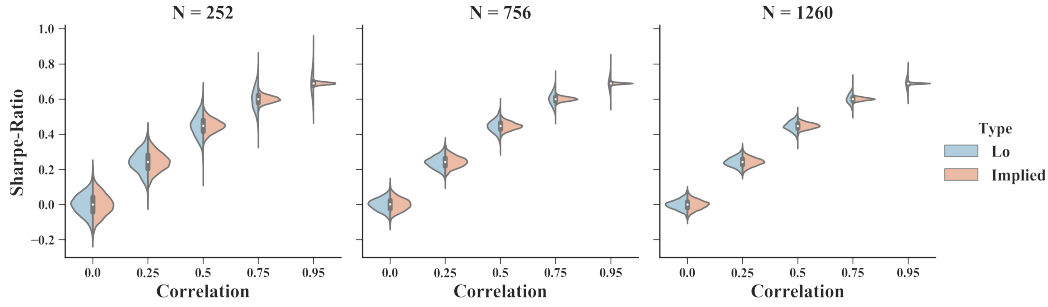


FIGURE 6. **Sharpe ratios full distribution** While the 95<sup>th</sup> percentile shows close agreement between Lo's large-sample standard errors and *implied* standard errors, the distribution of *implied* is far more fat-tailed.

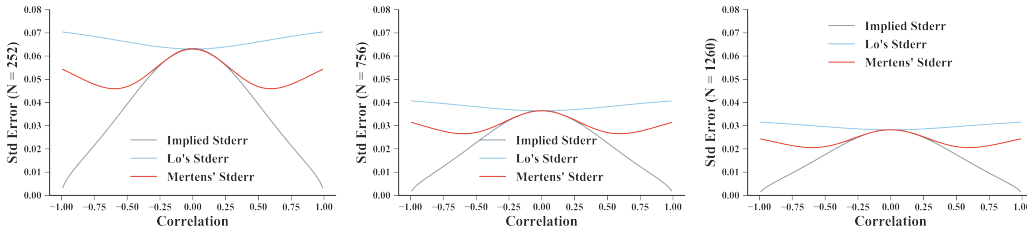


FIGURE 7. **Standard errors based on different sample sizes and formulas.** Ignoring parameter uncertainty, Merten's adjustment to Lo's standard errors improves standard errors to be nearly as tight as *implied*. In practice, parameter uncertainty hurts the performance.

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

$$(11) \quad \text{stderr}_{\text{Mertens}}^2 = \left(1 + \frac{1}{2} \hat{\text{SR}}^2 - \gamma_3 \cdot \hat{\text{SR}} + \frac{\gamma_4 - 3}{4} \cdot \hat{\text{SR}}^2\right).$$

If we use our plug-in estimates for skewness and excess kurtosis (i.e., coming from equations (6 and 7)) into equation (11) we are able to find a modestly tighter estimate of the standard error than Lo. For most smaller amplitude correlations, this estimate 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  $\gamma_3$  and  $\gamma_4$ , Mertens' approximation is always tighter than Lo's; in particular for correlations  $|\rho| < 0.5$ , Mertens' approximation appears almost identical to our own. Irrespective, 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 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, i.e.,

$$\begin{aligned}\gamma_3 &= \frac{\mu_3}{m} u_2^{3/2} \\ \gamma_4 &= \frac{\mu_4}{\mu_2^2}\end{aligned}$$

although it is not hard to consider other definitions of skewness and kurtosis using unbiased estimators of the moments as are 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

$$\begin{aligned}\text{stderr}_{\gamma_3} &= \sqrt{\frac{6(T-2)}{(T+1)(T+3)}} \\ \text{stderr}_{\gamma_4} &= \sqrt{\frac{24T(T-2)(T-3)}{(T+1)^2(T+3)(T+5)}}\end{aligned}$$

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  $\rho$ , and sample size  $T$ , the standard errors are given by<sup>6</sup>*

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

and

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

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

We rely on the delta-method, recognizing that  $\text{stderr}_{\gamma_k} = \partial\gamma_k/\partial\rho \cdot \text{stderr}_\rho$  for  $k = 3, 4$ . Given the following easily calculated derivatives:

$$(12) \quad \frac{\partial\gamma_3}{\partial\rho} = -\frac{6(\rho^2-1)}{(\rho^2+1)^{5/2}}$$

$$(13) \quad \frac{\partial\gamma_4}{\partial\rho} = -\frac{48\rho(\rho^2-1)}{(\rho^2+1)^3}$$

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  $\rho$  in terms of  $\gamma_k$  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 equations [\(12\)](#) 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](#).

---

<sup>6</sup> While  $\rho$  can be expressed in terms of either  $\gamma_3$  or  $\gamma_4$  to eliminate  $\rho$  from these expressions, unlike the case of the standard errors of the Sharpe ratio, the expressions are too complicated to be that useful.



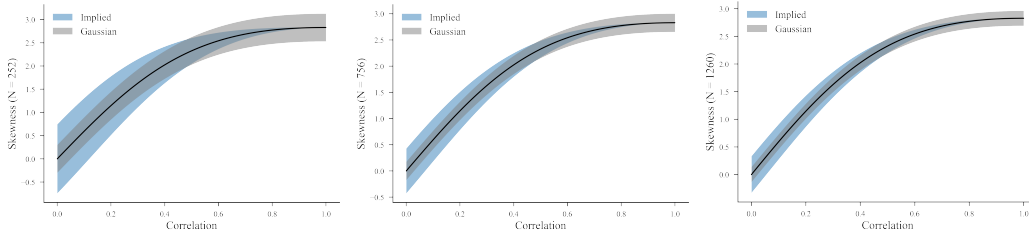


FIGURE 8. **Standard errors for skewness for different sample sizes, implied vs Gaussian** Implied standard errors, especially for skewness are generally larger than those for normal distributions. We argue that the implied standard errors are more appropriate for dynamic strategies.

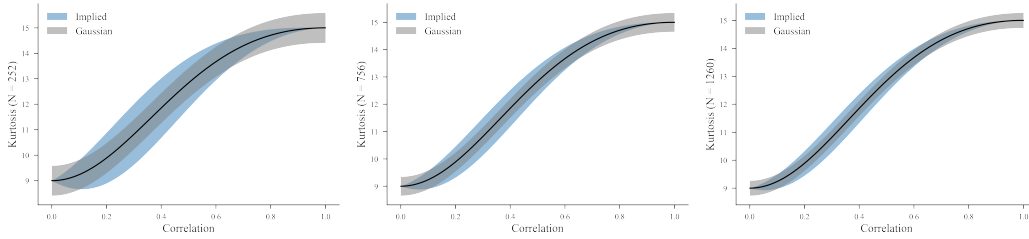


FIGURE 9. **Standard errors for kurtosis for different sample sizes, implied vs Gaussian** Implied kurtosis standard errors are sometimes larger and sometimes tighter than the Gaussian case. We argue that the *implied* standard errors are more appropriate for dynamic strategies.

#### 4. MULTIPLE ASSETS

We 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-Kaminski have noted (Hoffman-Kaminski (2016)) that while single strategies can have skewness ranging from around  $[1.3, 1.7]$  and kurtosis from  $[8.8, 15.3]$ , 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  $\psi$  whose moment generating function is given by Simons (2006):

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

From this we can easily derive four moments:

$$\begin{aligned}\mu_1 &= N\sigma\gamma\rho \\ \mu_2 &= N\sigma^2\gamma^2((N+1)\rho^2+1) \\ \mu_3 &= N(N+2)\sigma^3\gamma^3\rho((N+1)\rho^2+3) \\ \mu_4 &= \sigma^4\gamma^4\left((N+6)(N+4)(N+2)N\rho^4+3(N+2)N(1-\rho^2)^2+\right. \\ &\quad \left.6(N+4)(N+2)N\rho^2(1-\rho^2)\right)\end{aligned}$$

This leads to 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 SR over  $\rho$  leads to  $\frac{\sqrt{N}\sqrt{2}}{2}$ , 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 maximal Sharpe, the corresponding skewness is

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

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 central limit theory should give us asymptotic normality. Effectively, introducing more purely orthogonal assets will increase Sharpe ratios, but decreases 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 *canonical correlation analysis* (CCA),<sup>7</sup>

---

<sup>7</sup>Canonical correlation (from Hotelling (1936), see for example, Rencher and Christiansen (2012)) 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 via a generalized eigenvalue problem

$$\begin{aligned}\Sigma_{RR}^{-1} \Sigma_{RX} \Sigma_{XX}^{-1} \Sigma_{XR} w_k &= r_k^2 w_k \\ \Sigma_{XX}^{-1} \Sigma_{XR} \Sigma_{RR}^{-1} \Sigma_{RX} v_k &= r_k^2 v_k\end{aligned}$$

where  $\Sigma$  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*,  $S_k^{CCA} \equiv (v_k \cdot X)(w_k \cdot R)$  each have returns and variances as characterised by equation (1 and 2) with corresponding correlations  $r_k$  (i.e., with Sharpe ratios given by  $\text{SR}[S_k] = r_k / \sqrt{r_k^2 + 1}$ ) and, due to their independence, 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 asset returns. We assert that this method gives the maximal

resulting in a set of decorrelated strategies (using a and combination of signals to weight a portfolio of assets). The resulting strategies are decorrelated but with unequal returns and variances. Many results of this section would apply after scaling the portfolio returns. The end-result could 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 relatively lower positive skewness, much in line with the discussion of this section. Our simple results here about the final Sharpe ratio and skewness of course depend on independence of the underlying assets and of course the signals themselves, which must only be correlated with their respective asset returns. While this is a not an altogether natural setting, it is suggestive of the gains that can be made in introducing purely orthogonal sources of risk, or perhaps in orthogonalizing (or attempting to) asset returns prior to forming signals, later recombining into a portfolio, and that this may lead to far more desirable properties of portfolios than finding strategies on multiple non-orthogonalized assets.

## 5. GAUSSIAN RETURNS VS PRODUCTS OF GAUSSIANS 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 together 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 (i.e., daily, weekly, monthly), as has been noted by [Potters and Bouchaud \(2005\)](#), [Hoffman-Kaminski \(2016\)](#) and others. If we measure far longer-horizon returns, asymptotic theory should show that favourable qualities like skewness may disappear.

Consequently, even though we make many comparisons to results stemming from either asymptotic theory (e.g., [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 (CLT) holds, e.g., 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 (e.g., annual or longer, non-overlapping) 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

---

Sharpe ratio for the linear combination of signals and returns, although we leave this proof to a subsequent paper.

measures as large-sample limits in appendix A, although asymptotics are beyond the scope of this current study.

## 6. CONCLUSION

Fully systematic dynamic strategies are used by a large portion of the asset management industry as well as by many non-institutional participants. Meanwhile, they are only partly understood. Many funds and strategies (e.g., especially investment bank *smart-beta* or *styles-based* products) involve investment in strategies which are not optimised in any sense. Strategies which are paid via index-swaps have great 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 to build and extend these results.

It is hoped that this paper does 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 (i.e., why there are stable ACFs in the first place). Given that most asset returns are known to have non-trivial autocorrelations, we can establish many results. In particular, we have derived a number of results merely by applying well-known techniques to dynamic strategies, e.g.:

- Strategy returns can be shown to be positively skewed and leptokurtic.
- Sharpe ratios can be characterized, as can skewness and kurtosis.
- The standard errors for Sharpe, skewness and kurtosis can be derived.
- Strategies designed to optimise Sharpe ratios should be based on TLS rather than minimizing prediction error.
- Gains from adding orthogonal assets/risks can be quantified.

Some of these items are empirically well-known, but others are genuinely new. Meanwhile, we have extended our results to the derivation of over-fitting penalties akin to Mallow's  $C_p$  or AIC and can be used to do model selection and 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 (e.g., z-scores) and optimal non-linear functions of z-scores.<sup>8</sup>
- non-linear strategies which are optimised to specific utility functions, possibly incorporating smoothness constraints, especially when relaxing normality.
- local optimality when relaxing stationarity.
- good-deal bounds in the presence of auto-correlated assets with possible non-stationarity or structural breaks.

---

<sup>8</sup>We note that normalized signals applied to normalized returns series can be represented as the product of two Student t-distributions, which is also relatively well-studied [Nadrajah-Kotz \(2003\)](#); [Joarder \(2007\)](#) and the results are qualitatively very similar to those which we have produced in this study. However, the more commonly used strategy of applying normalized signals to returns, with the resulting strategies then vol-scaled, cannot be derived as a trivial application of well-known results

We note that our assumptions were never meant to be completely realistic: stationary returns with fixed ACF 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 over-fitting. 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 put the general study of dynamic strategies onto a more firm theoretical footing.

Some of our findings should be of note 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 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 over-fitting penalties and optimal non-linear strategies, we save for later papers. With a more solid theoretical footing as a sort of *rule-of-thumb* for the development, optimisation, selection and alteration of dynamic strategies, we only hope that there can be room to improve strategy design.

#### ACKNOWLEDGEMENTS

N. Firoozye would like to give his wholehearted love and appreciation to Fauziah, for hanging on, when the paper was always *almost done*. I am hoping the wait is finally over. Adriano Soares Koshiyama would like to acknowledge the funding for its PhD studies provided by the Brazilian Research Council (CNPq) through the Science Without Borders program.

The authors would also like to thank Brian Healy and Marco Avellaneda for the many suggestions and encouragement. Finally, were it not for the product design method as practised by Nomura's QIS team, the authors would never have been inspired to pursue a mathematical approach to this topic.

#### APPENDIX A. FULL DISTRIBUTIONS FOR SINGLE PERIOD

In general, for  $X$  and  $R$  having joint density  $\psi^{X,R}(x, r)$ , and have  $S_t = X_t R_t$  is known to have the product pdf,

$$(14) \quad \psi^S(s) = \int_{-\infty}^{\infty} \psi^{X,R}\left(x, \frac{s}{x}\right) \frac{1}{|x|} dx$$

and, in the special case where  $X \sim \mathcal{N}(0, \sigma_X^2)$  and  $R \sim \mathcal{N}(0, \sigma_R^2)$  jointly normal with correlation  $\rho$  (i.e.,  $\psi$  being a bivariate gaussian), this results in the closed-form expression:

$$(15) \quad p_s = \frac{1}{\pi \sigma_R \sigma_X} \exp\left(\frac{\rho s}{\sigma_R \sigma_X (1 - \rho^2)}\right) K_0\left(\frac{|s|}{\sigma_R \sigma_X (1 - \rho^2)}\right)$$

where  $K_0(\cdot)$  is a modified Bessel function of the 2<sup>nd</sup> kind (Simons (2006), p 51, eq 6.15). The more general density for non-zero means, is given in Cui et al (2016) as an infinite series. In the special cases of independence and of correlated but zero mean, the expressions become much simpler and we choose to focus on the zero-mean case here. The density is unbounded at zero and has fat tails and positive skewness, becoming more pronounced with higher correlation. We can see the distribution for a variety of correlations in figure (10), with the skewness becoming increasingly pronounced for higher  $\rho$ . In the limit as  $\rho \rightarrow 1$  the distribution converges to that of the central  $\chi^2$  distribution with one degree of freedom.

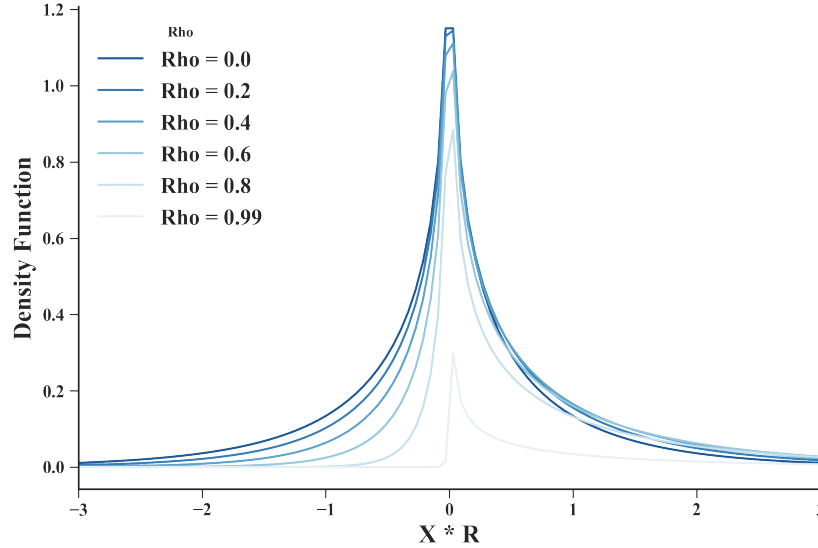


FIGURE 10. **Complete product distributions**, for  $\rho \in \{0, 0.2, 0.4, 0.6, 0.8, 1.0\}$ , normalised to have unit variance, so they can be depicted on one plot. Note the singularity at 0, the increasing asymmetry, nearly truncated left-tails and marginally fatter right-tails with increasing  $\rho$ .

In fact,  $K_0(z) = O(e^{-z}/\sqrt{z})$  for  $z \rightarrow \infty$  and we can see that the tail behaviour of the pdf in equation (15) changes quite significantly from when  $\rho = 0$  and  $K_0(z)$  is the only term to consider, to when  $\rho > 0$ , introducing an asymmetry. The Bessel function is unbounded at  $z = 0$ . Asymptotically, we have the following behaviour:

$$\begin{aligned}
 p_s &= O(e^{-|s|}/\sqrt{|s|}) && \text{for } \rho = 0, |s| \rightarrow \infty \\
 p_s &= O(e^{-s}/\sqrt{|s|}) && \text{for } \rho > 0, s \rightarrow \infty \\
 p_s &= O(e^s/\sqrt{|s|}) && \text{for } \rho > 0, s \rightarrow -\infty \\
 p_s &= O(-\log |s|) && \text{for } |s| \rightarrow 0
 \end{aligned}$$

## APPENDIX B. CONVOLUTION FILTERS AS JOINTLY GAUSSIAN

If we have a purely-random mean-zero covariance-stationary discrete-time Gaussian process  $R_t$ , we note by Wold Decomposition, that all stationary Gaussian processes can be represented as  $MA(\infty)$  in terms of Gaussian innovation process and coefficients in  $l^2$ , with no deterministic component, i.e.,

$$R_t = \sum_{k=0}^{\infty} \phi(k) \epsilon_{t-k}$$

for  $\epsilon \sim \mathcal{N}(0, \sigma^2)$ ,  $\sum_0^{\infty} \phi_k^2 < \infty$  and  $\phi(0) = 1$ .

More specifically, we have

$$\begin{aligned} E[R_t] &= 0 \\ Var(R_t) &= \sigma_R^2 \\ corr(R_t, R_s) &= \gamma(t-s) \end{aligned}$$

(i.e., with ACF  $\gamma$ ), and this would be sufficient to determine  $\phi$  if we so wished.

We are interested in constructing signals:  $X_t$ . A standard signal we will consider is a convolution signal, i.e.,

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

All the signals mentioned in the introduction (e.g., moving average or difference of moving averages or ARMA based forecasts), can be expressed as convolutions with historic returns. A convolution filter is an example of a time-invariant linear filter. If the coefficients  $\phi \in l^2$  then it is well known that the resulting filtered series  $X_t$  are Gaussian<sup>9</sup>. The filtered series  $X_t$  is also jointly Gaussian with  $R_t$ .

$$\begin{aligned} E[X_t] &= 0 \\ Var(X_t) &= \sum_{k,j \geq 1} \phi(k) \phi(j) E[R_{t-k} R_{t-j}] \\ &= \sum_{k,j \geq 1} \phi(k) \phi(j) \gamma(k-j) \sigma_R^2 \end{aligned}$$

(dropping all first order terms because  $E[R_t] = 0$ ) and,

$$\begin{aligned} corr(R_t, X_t) &= \frac{E[\sum_{k \geq 1} \phi(k) R_{t-k}, R_t]}{std(X) \sigma_R} \\ &= \frac{\sum_{k \geq 1} \phi(k) \gamma(k)}{(\sum_{k,j \geq 1} \phi(k) \phi(j) \gamma(k-j))^{1/2}} \end{aligned}$$

cancelling out all  $\sigma_R$  terms.

Consequently,

$$sgn(corr(R_t, X_t)) = sgn(\gamma \cdot \phi)$$

<sup>9</sup>see e.g., Gallagher, R, *Stochastic Processes: Theory for Applications*, 2014, (Cambridge UP: Cambridge), or Gallagher R, *Principles of Digital Communications. MIT Open Coursework*. Section 7.4.2, Theorem 7.4.1.

(i.e., the sign of this infinite inner product matters most for determining usefulness of a given convolution design).

Of the signals mentioned in the introduction, EWMA and SMA in returns, differences of EWMA and SMA in returns, and forecasts from ARMA models are all examples of convolution filters with  $l^2$  coefficients. Most signals constructed in levels (e.g., the difference between a price and its simple moving average), are not, in general, Gaussian, although a difference between a price and one or more EWMA may be Gaussian depending on the data-generating process for the price series (i.e., for MA processes).

Of course, a linear time-invariant filter with  $l^2$  coefficient is just one example of a signal  $X_t$  which is jointly Gaussian with returns  $R_t$ . Similarly, if  $Z_t$  is a set of Gaussian (exogenous) features, then  $X_t = Z_t\beta$  will also be Gaussian and we will assume the  $Z_t$  are jointly Gaussian with  $R_t$ , meaning also  $X_t$  and  $R_t$  will be jointly Gaussian.

#### APPENDIX C. LIMITING BEHAVIOUR FOR CONVOLUTION OF STATIONARY RETURNS

We assert some asymptotic approximation results for dynamic strategies, only outlining their proof. Our claim is that this justifies the use and analysis of product of Gaussian distributions in stationary (or locally stationary) distributions. The proof itself is the direct consequence of much more general work on the limits of quadratic forms by Götze and Tikhonov and by the Wold decomposition theorem.

Letting  $\eta$  be iid random variables with mean zero and unit variance, and letting  $\epsilon$  be iid normal random variables with zero mean and unit variance, we form the quadratic forms:

$$Q_n = \sum_{j,k=1}^n a_{jk}^n \eta_j \eta_k \text{ and } G_n = \sum_{j,k=1}^n a_{jk}^n \epsilon_j \epsilon_k.$$

We write the metric

$$\delta_n(Q_n, G_n) = \sup_x |P\{Q_n \leq x\} - P\{G_n \leq x\}|.$$

We simplify the statement of Theorem 1 from [Götze and Tikhomorov \(1999\)](#):

**Theorem C.1** (Goetze-Tikhomirov). *Let  $\eta$  be IID with*

$$E\eta = 0, E\eta^2 = 1, E|\eta|^3 = \beta_3 < \infty.$$

*Then there is a constant  $C$  such that*

$$\delta_n(Q_n, G_n) \leq C\beta_3^2\Gamma_n$$

where  $\Gamma_n = \max_{1 \leq j \leq n} \sum_{k=1}^n |a_{jk}^n|$ .

Our assertion is a simple application of the results in [Götze and Tikhomorov \(1999\)](#), (see [Götze and Tikhomorov \(2002\)](#) and [Götze et al \(2007\)](#) for further results) which applies to limiting theorems of quadratic forms of random variables.

**Theorem C.2** (Products of Gaussians). *Let  $R_t$  be a covariance stationary process with bounded 3rd moments and mean zero and its Wold decomposition given by  $R_t = \sum_{s=1}^{\infty} b(s)\eta(t-s)$  with  $\eta$  a white-noise process. Let the signal  $X_t$  be a convolution of the lagged returns  $R_t$  with an  $\mathcal{L}^2$  convolution kernel,  $\phi$  and  $X_t = \sum_{s=1}^{\infty} \phi(s)R_{t-s}$ . We let  $R_t^N = \sum_{s=0}^N b(s)\eta(t-s)$  and  $X_t^N = \sum_{s=1}^N \phi(s)R_{t-s}^N$  be truncated sums (only involving the first  $N$  terms),*

$$S_t^N = \frac{1}{\sqrt{N}} X_t^N \cdot \frac{1}{\sqrt{N}} R_t^N$$



be the scaled truncated strategy returns.

Then there is a pair of Gaussians  $\tilde{R}_t^N$  and  $\tilde{X}_t^N$  ( $\tilde{S}_t^N = \tilde{X}_t^N \cdot \tilde{R}_t^N$  be the Gaussian strategy returns) such that

$$\delta_n(S_t^N, \tilde{S}_t^N) \rightarrow 0,$$

or, in other words, that the product of Gaussian approximation can be arbitrarily close to the original strategy.

We note that the product  $S_t = X_t R_t$  is given by the quadratic form:

$$S_t = X_t R_t = \langle A\eta, \eta \rangle$$

where  $A$  is the operator given by

$$A(u, v) = \sum_{s \leq t-1} \phi(t-s)b(s-u)b(t-v)$$

for  $u, v \leq t$ .

$$\begin{aligned} S_t^N &= \frac{1}{\sqrt{N}} X_t^N \cdot \frac{1}{\sqrt{N}} R_t^N \\ &= \langle A^N \eta^N, \eta^N \rangle \end{aligned}$$

where  $A^N$  is an  $n \times n$  matrix

$$A_{u,v}^N = \frac{1}{N} \sum_{s \in [t-n, t-1]} \phi(t-s)b(s-u)b(t-v),$$

for  $u, v$  ranging in  $[t-n, t]$  and  $\eta^N = \{\eta_s\}_{s \in [t, t-N]}$ .

We note that  $A^N$  is lower triangular with no diagonal terms (elements on the diagonal correspond to instantaneously available knowledge, contemporaneous with the observed returns themselves and elements in the upper triangle of the matrix correspond to direct foresight). Moreover, with sufficient conditions on the original series  $R_t$  (i.e., on the Wold coefficients  $b$ ) and on the convolution coefficients  $\phi$ , the  $\Gamma_N = \max_{1 \leq j \leq n} \sum_{k=1}^n |A_{jk}^N|$  can be shown to decay to zero.

A direct application of the theory of quadratic forms would apply when the convolution coefficients are sufficiently well-behaved at infinity.

This is only one of the possible approaches to an asymptotic theory justifying the use of products of Gaussians.<sup>10</sup> While asymptotic approaches are not the main point of this paper, it should be clear that products of Gaussians help to approximate the behaviour of a wide array of dynamic strategies.

#### APPENDIX D. NONZERO MEANS: SHARPE RATIOS AND SKEWNESS

By an abuse of notation, we define  $\text{SR}[R]$  to be  $\mu_R/\sigma_R$  and by an abuse of notation, we define  $\text{SR}[X] = \mu_X/\sigma_X$  (for  $X$  the signal),

**Corollary 1:** If  $R \sim \mathcal{N}(\mu_R, \sigma_R^2)$  and  $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$  then

$$\text{SR}[X \cdot R] = \frac{\text{SR}[R] \cdot \text{SR}[X] + \rho}{(\text{SR}[R]^2 + \text{SR}[X]^2 + 2\rho \text{SR}[R] \cdot \text{SR}[X] + \rho^2 + 1)^{1/2}}$$

<sup>10</sup>Other approaches include assuming infinitesimal Gaussian increments which are observed and “stored” and used in a convolution, then applied as a weight on a strategy which itself is held for a longer time. This effectively results in some product of averages of returns and, obviously, when appropriately scaled can be shown to have a limit of a product of Gaussians.

**Corollary 2:** If  $R \sim \mathcal{N}(\mu_R, \sigma_R^2)$  and  $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$  then

$$\gamma_3[X \cdot R] = \frac{2\rho(\rho^2 + 3 + 3\text{SR}[R]^2 + 3\text{SR}[X]^2)}{(\text{SR}[R]^2 + \text{SR}[X]^2 + 2\rho\text{SR}[R] \cdot \text{SR}[X] + \rho^2 + 1)^{3/2}}$$

We note the one period Sharpe ratio of the strategy may depend on both the interaction between the Sharpe ratios of the Signals (weights) and the Returns, in particular whether they have the same sign or not, together with the sign of the correlation. In fact, the amplitude of the resulting strategy SR may be more dependent on the respective Sharpe ratios rather than  $\rho$  since after all,  $-1 \leq \rho \leq 1$ , while  $\text{SR}[R]$  and  $\text{SR}[X]$  may individually be above 1.

#### APPENDIX E. TRANSACTION COSTS

The sections above consider optimal linear strategies with no transaction costs. If we include transaction costs then the formulas are not nearly as elegant, but the results may still remain tractable.

Maximizing Sharpe ratios are often the result of maximizing a quadratic utility of returns, e.g.,

$$(16) \quad U[XR] = E[XR] - \gamma \text{var}[XR]$$

where  $\gamma$  is a measure of risk-aversion, sometimes called a *Kelly constant*. Extremals of the utility in equation (16) are known to coincide with maximal Sharpe ratios.

We only look at convolution filter strategies, i.e.,  $\phi = (0, \phi_1, \phi_2, \dots, \phi_K)$  which give a corresponding signal as  $X_t = \phi * R_t = \sum_{k=1}^K \phi_k R_{t-k}$ . As we mentioned above, fitting  $\phi$  via TLS instead of OLS is most appropriate in the case of no-transaction costs.

If we include transaction costs proportional to a constant  $\nu$ , rather than to maximize a quadratic utility in (16), we can add the extra term<sup>11</sup>, e.g.,

$$U[XR] = E[XR] - \gamma \text{var}[XR] - \nu E[|\Delta X|]$$

Given that

$$\Delta X = \phi_1 R_{t-1} + \sum_{k=1}^K \Delta \phi_k R_{t-k} - \phi_K R_{t-K} \equiv \Delta \phi * R$$

where  $\Delta \phi = (0, \phi_1, \Delta \phi_1, \Delta \phi_2, \dots, \Delta \phi_K, -\phi_K)$ . The r.v. is normal,  $\Delta X \sim \mathcal{N}(0, \sigma_{\Delta X}^2)$  and, using the properties of *folded Gaussian variables*, we can characterise

$$E[|\Delta X|] = \sqrt{\frac{2}{\pi}} \sigma_{\Delta X}$$

The entire utility then can be written as

$$U[XR] = \rho \sigma_X \sigma_R - \gamma \sigma_X^2 \sigma_R^2 (1 + \rho^2) - \nu \sqrt{\frac{2}{\pi}} \sigma_{\Delta X}$$

Optimising this utility will be very much like a standard least-squares problem except the term  $\sigma_{\Delta X}$  is a form of regularization.

<sup>11</sup>Alternatively, a term such as  $E[|\Delta X| \cdot P]$  where  $P = P_0 + \sum R_t$  could be added. Again, with work we could equally well characterize this expectation, using properties of distributions derived from Gaussians and some application of Isserlis' theorem

In fact if we let  $C$  being the ACF (Toeplitz) matrix of  $(R_t, \dots, R_{t-k})$ , i.e.,

$$C = \begin{bmatrix} 1 & c(1) & c(2) & \dots & c(k-1) \\ c(1) & 1 & c(1) & \dots & c(k-2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c(k-1) & c(k-2) & c(k-3) & \dots & 1 \end{bmatrix}$$

and  $X = \phi * R$  with  $\phi = (\phi_1, \phi_2, \dots, \phi_k)$  and let  $1_0 = (1, 0, \dots, 0)$  then  $\sigma_X = \sigma_R \sqrt{\phi' \cdot C \cdot \phi}$ ,  $\rho = \phi' \cdot C \cdot 1_0$  and  $\sigma_{\Delta X} = \sigma_R \sqrt{(\Delta\phi)' \cdot C \cdot (\Delta\phi)}$ , effectively penalizing changes in the  $\phi_k$ .

The resulting optimisation problem thus becomes

$$U[XR] = \sigma_R^2 \sqrt{\phi' \cdot C \cdot \phi} \phi' \cdot C \cdot 1_0 - \gamma \sigma_R^4 (\phi' \cdot C \cdot \phi) (1 + (\phi' \cdot 1_0)^2) - \nu \sqrt{\frac{2}{\pi}} \sigma_R \sqrt{(\delta\phi)' \cdot C \cdot (\Delta\phi)}$$

This final regularization term should ensure that the filter weights  $\phi_k$  do not vary too much between themselves (i.e., it is a sort of smoothness constraint analogous to those in a *Lasso* or *Ridge-regression*, but with a slightly different functional form). Unlike the case of an  $\mathcal{L}^2$  penalty as in Ridge regression or an  $\mathcal{L}^1$  penalty as in Lasso, this term though is neither linear nor quadratic.

We do not consider properties of the solutions of optimal trading strategies with transaction costs in this paper.

#### APPENDIX F. MULTIPERIOD RETURNS

Given the ease of analysis of Gaussian returns, it is straightforward to calculate moments of the strategy returns to any horizon. While we do not explore further implications, we produce relevant formulas in this section for future elaboration.

For long-horizon trades we note the following (Magnus (1978))

**Theorem**[Magnus] Let  $A$  be a symmetric matrix and  $R \sim \mathcal{N}(0, V)$  with  $V$  positive definite. Define  $p = R'AR$ . then the expectation, variance, skewness and kurtosis of  $p$  are:

$$\begin{aligned} \mu &= \text{tr}(AV) \\ \sigma^2 &= 2 \text{tr}(AV)^2 \\ \gamma_3 &= 2\sqrt{2} \frac{\text{tr}(AV)^3}{(\text{tr}(AV)^2)^{3/2}} \\ \gamma_4 &= 12 \frac{\text{tr}(AV)^4}{(\text{tr}(AV)^2)^2} \end{aligned}$$

which would allow us to calculate Sharpe ratios, skewness and kurtosis to any horizon. Continuous analogues are feasible using functional central limit theory for Wick products (see Parczewski (2014)).

Given this and the various moment conditions for our Gaussian returns:

$$\begin{aligned} E[R_t] &= 0 \\ E[R_t^2] &= \sigma^2 \\ E[R_t^3] &= 0 \\ E[R_t R_s] &= C(t-s)\sigma^2 \end{aligned}$$

where  $C(0) = 1$  and  $C(-k) = C(k)$ , we can combine for characterising strategy moments.

If the ACF matrix  $\tilde{C}$  is known with certainty, of course, then the linear filter which maximizes the correlation of signal to returns is merely given by finding the eigenvector corresponding to the smallest eigenvalue, i.e.,

$$v_{min} = \underset{v}{\operatorname{argmin}} \frac{v' \tilde{C} v}{|v|^2}$$

and normalizing the first coefficient to be one, i.e.,  $a_k = -v(k+1)/v(1)$ .

For longer horizons, we use the formulas given by Magnus, or equally compute the term-structure by hand:

$$\mu_1(T) = E \sum X_t R_t = \sigma_X \sigma_R \rho T$$

and

$$\begin{aligned} E(\sum X_t R_t)^2 &= \sum \sum E[X_t R_t X_s R_s] \\ &= \sum \sum (E[X_t R_t] E[X_s R_s] + E[X_t X_s] E[R_t R_s] + E[X_t R_s] E[X_s R_t]) \\ &= \sigma_X^2 \sigma_R^2 (T^2 \rho^2 + \sum \sum (C(i-j) D(i-j) + \rho(i-j) \rho(j-i))) \end{aligned}$$

where  $C(k)$  is the ACF for  $R$  and  $D(k)$  is the ACF for signal  $X$ , and  $\rho(k) = E[X_t R_{t-k}]$  and  $\rho(k) = \rho(-k)$  and  $\rho(0) = \rho$  is the contemporaneous correlation.

Consequently,

$$\operatorname{var}(\sum X_t R_t) = \sigma_X^2 \sigma_R^2 (2 \sum_{k=1}^{T-1} (T-k) (C(k) D(k) + \rho(k) \rho(-k)) + T(1 + \rho^2))$$

and consequently, the Sharpe ratio to any horizon is given by

$$SR(T) = \frac{\rho T^{1/2}}{1 + \rho^2 + 2 \sum_{k=1}^{T-1} \frac{T-k}{T} (C(k) D(k) + \rho(k) \rho(-k))}$$

giving us the term-structure of Sharpe ratios by horizon.

## APPENDIX G. SET-UP DETAILS

If we have a purely-random mean-zero covariance-stationary discrete-time Gaussian process  $R_t$ , we note by Wold Decomposition, that all stationary Gaussian processes can be represented as MA( $\infty$ ) in terms of Gaussian innovation process and coefficients in  $l^2$ , with no deterministic component.

Specifically, we have

$$\begin{aligned} E[R_t] &= 0 \\ \operatorname{Var}(R_t) &= \sigma_R^2 \\ \operatorname{corr}(R_t, R_s) &= \gamma(t-s) \end{aligned}$$

(i.e., with ACF  $\gamma$ )

Then we are interested in constructing signals:  $X_t$ . A standard signal we will consider is a convolution signal, i.e.,

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

This is an example of a time-invariant linear filter. If the coefficients  $\phi \in l^2$  then it is well known that the resulting filtered series  $X_t$  are Gaussian<sup>1</sup>. The filtered series  $X_t$  is also jointly Gaussian with  $R_t$ .

We note that if the  $\phi(k)$  can be derived as the coefficients of an ARMA model forecast, or they can be from a simple EWMA, as we have mentioned in the paper.

$$\begin{aligned} E[X_t] &= 0 \\ \text{Var}(X_t) &= \sum_{k,j \geq 1} \phi(k)\phi(j)E[R_{t-k}R_{t-j}] \\ &= \sum_{k,j \geq 1} \phi(k)\phi(j)\gamma(k-j)\sigma_R^2 \end{aligned}$$

(dropping all first order terms because  $E[R_t] = 0$ ) and,

$$\begin{aligned} \text{corr}(R_t, X_t) &= \frac{E[\sum \phi(k)R_{t-k}, R_t]}{\text{std}(X)\sigma_R} \\ &= \frac{\sum \phi(k)\gamma(k)}{(\sum_{k,j \geq 1} \phi(k)\phi(j)\gamma(k-j))^{1/2}} \end{aligned}$$

cancelling out all  $\sigma_R$  terms.

Consequently,

$$\text{sgn}(\text{corr}(R_t, X_t)) = \text{sgn}(\gamma \cdot \phi)$$

(i.e., the sign of this infinite inner product matters most for determining usefulness of a given convolution design). A linear time-invariant filter with  $l^2$  coefficient is just one example of a signal  $X_t$  which is jointly Gaussian with returns  $R_t$ . Similarly, if  $Z_t$  is a set of Gaussian (exogenous) features, then  $X_t = Z_t\beta$  will also be Gaussian and we will assume the  $Z_t$  are jointly Gaussian with  $R_t$ , meaning also  $X_t$  and  $R_t$  will be jointly Gaussian.

#### REFERENCES

- Acar, E. Expected returns of directional forecasters. in *Advanced Trading Strategies*, ed. E Acar and S Satchell, 1990, (Butterworth-Heinemann: Oxford), 122-151.
- Allendbridge IS, Quantitative Investment Strategy Survey, 2014.
- Asness, C.S., Moskowitz, T.J. and Pedersen, L.H. Value and momentum everywhere. *The J. Finance*, 2013 **68**(3), 929-985.
- Avramovic, A., We're All High Frequency Traders Now. *Credit Suisse Market Structure, Trading Strategy*. 15 March 2017, <https://edge.credit-suisse.com/edge/Public/Bulletin/Servefile.aspx?FileID=28410&m=-1290757752> (Accessed 21 Aug 2017)
- Cramér, H. *Mathematical Methods of Statistics*. 1946 (Princeton University Press: Princeton).
- Babu, G.J., and Feigelson, E.D. Analytical and Monte Carlo comparisons of six different linear least squares fits. *Communications in Statistics-Simulation and Computation* **21**(2), 1992, 533-549.
- Baltas, N. and Kosowski, R. Momentum Strategies in Futures Markets and Trend-following Funds (January 5, 2013). Presented at Finance Meeting EUROFIDAI-AFFI Paper, Paris.

<sup>1</sup>see e.g., Gallagher, R, Stochastic Processes: Theory for Applications, Cambridge UP, 2014, or Gallagher R, MIT Open Coursework, Principles of Digital Communications, Section 7.4.2, Theorem 7.4.1.

- December 2012. Available at SSRN: <https://ssrn.com/abstract=1968996> (Accessed 21 August 2017).
- Barclay Hedge: CTA's Asset Under Management. Available online at [https://www.barclayhedge.com/research/indices/cta/Money\\_Under\\_Management.html](https://www.barclayhedge.com/research/indices/cta/Money_Under_Management.html), (accessed 23 Sep 2017).
- Bouchaud, J.P., and Potters, M. Financial applications of random matrix theory: a short review. Available online at arXiv <https://arxiv.org/abs/0910.1205> (2009), (accessed 2 Jan 2016).
- Bouchaud, J.P., Dau, T.L., Deremble, C., Lemprière, Y., Nguyen, T.T., Potters M. Tail Protection for Long Investors: Convexity at Work. *J Inv Strat*, Dec 2017, **7**(1), 61-84.
- Bruder, B. and Dao, T.L., Richard, J.C., and Roncalli, T., Trend Filtering Methods for Momentum Strategies. December 1, 2011. Available online at SSRN: <https://ssrn.com/abstract=2289097>, (accessed 2 July 2014).
- Bruder, B and Gaussel, N. Risk-return analysis of Dynamic investment strategies. Lyxor White paper, Issue 7 (Jun 2011), available online at SSRN: [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2465623](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2465623) (accessed 2 July 2014).
- Cui, G, Yu, X, Iommelli, S, and Kong, L. Exact distribution for the product of two correlated Gaussian random variables. *IEEE Sig Proc Letters*, Nov 2016, **23** (11), 1662-1666.
- Fama, E.F., and French, K.R., The Cross-Section of Expected Stock Returns, *J Finance*, Jun 1992, **47**(2), 427-465.
- Fuller, Wayne A. *Measurement error models*, Vol. 305, 2009, (John Wiley & Sons: New York).
- Fung and Hsieh, Empirical Characteristics of Dynamic Trading Strategies: The case of Hedge Funds, *Rev Fin Stud*, 1997, **10**(2), 275-302.
- Gigerenzer, G., Todd, P.M., and the ABC Research Group. *Simple heuristics that make us smart* 1999, (Oxford University Press: Oxford).
- Götze, F., and Tikhomirov, A. Asymptotic Distribution of Quadratic Forms, *Annals of Prob*, 1999, **50**, 1072-98.
- Götze, F and Tikhomirov, A., Asymptotic Distribution of Quadratic Forms and Applications, *J Theoretical Prob*, 2002, **15**, 423-75.
- Götze, F and Tikhomirov, A., and Yurchenko, V. Asymptotic Expansion in the Central Limit Theorem for Quadratic Forms, *J of Mathematical Sciences*, 2007, **147**(4), 6891-6911.
- Golub, G.H., and Van Loan, C.F. An analysis of the total least squares problem. *SIAM Journal on Numerical Analysis*, 1980, **17**(6), 883-893.
- Grinold, R.C. and Kahn, R.N., *Active Portfolio Management*, 2nd Edition, 1999, (McGraw-Hill: New York).
- Haldane, J.B.S., Moments of the Distributions of Powers and Products of Normal Variates. *Biometrika*, Apr 1942, **32** (3,4), 226-242.
- Hald, A.A *History of Parametric Statistical Inference from Bernoulli to Fisher*, 24 Aug 2008, (Springer Science & Business Media: Berlin), 1713-1935.
- Hamdan, R., Pavlowsky, F., Roncalli, T., and Zheng, B. A Primer on Alternative Risk Premia. April 2016. Available online at SSRN: <https://ssrn.com/abstract=2766850> (accessed 1 Dec 2016).
- Harvey, C.R. and Hoyle, E. and Korgaonkar, R. and Rattray, S. and Sargaison, M. and Van Hemert, O. The Impact of Volatility Targeting. May 21, 2018. Available online at SSRN: <https://ssrn.com/abstract=3175538>. (accessed 2 July 2018).

- Hoffman, B. and Kaminski, K. The taming of the Skew. Campbell White Paper Series, June 2016. Available online at [https://www.valuewalk.com/wp-content/uploads/2016/06/The\\_Taming\\_of\\_the\\_Skew\\_\\_\\_Campbell\\_\\_Company.pdf](https://www.valuewalk.com/wp-content/uploads/2016/06/The_Taming_of_the_Skew___Campbell__Company.pdf), (accessed 8 Feb 2018).
- , Hotelling, H. Relations between two sets of variants. 1936. *Biometrika*, **28**, 321-377.
- Hurst, B., Ooi, Y.H. and Pedersen, L.H. A century of evidence on trend-following investing. (June 27, 2017). Available Online at SSRN: <https://ssrn.com/abstract=2993026>. (Accessed 2 July 2017).
- Hyndman, R., Koehler, A.B., Ord, J.K., and Snyder, R.D. *Forecasting with exponential smoothing: the state space approach*, 2008, (Springer Science & Business Media: Berlin).
- Isserlis, L. On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables. *Biometrika* 1918, **12**(1-2), 134-139.
- Jegadeesh, N. and Titman, S. Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency, *J Finance*, Mar 1993, **48**(1), 65-91.
- Joanes, D.N. and Gill, C. N., Comparing measures of sample skewness and kurtosis. *The Statistician*, 1998, **47**, 183-189.
- Joarder, A., On some characteristics of the bivariate t-distribution. *Intl J Modern Math*, **2**(2), 191-204
- Kan, R. From moments of sum to moments of product. *J Multivariate Anal*, 2008, **99**(3) 542-554.
- COVARIANCE-PENALTIES: AN EMPIRICAL INVESTIGATION OF THE ORDINARY AND TOTAL LEAST SQUARES CASES Item[
- Koshiyama, A.S., and Firoozye, N. B., Avoiding Back-testing Overfitting by Covariance-Penalties: An Empirical Investigation of the Ordinary and Total Least Squares Cases, submitted 2019.
- Lee, W. *Theory and methodology of tactical asset allocation*, 2000, Vol. 65. (John Wiley & Sons: New York).
- Lempérière, Y., Deremble, C., Seager, P., Potters, M., & Bouchaud, J. P. Two centuries of trend following. 2014. Available online at arXiv <https://arxiv.org/abs/1404.3274>, (accessed 17 Jun 2016).
- Levine, A. and Pedersen, L.H., Which Trend Is Your Friend? (May 7, 2015). Financial Analysts Journal, vol. 72, no. 3 (May/June 2016). Available at SSRN: <https://ssrn.com/abstract=2603731>. (accessed 17 Jun 2015).
- Leyang, W. Properties of the Total Least Squares estimation. *Geodesy and Geodynamics*, 2012, **3**(4), 39-46.
- Lo, A.W. The statistics of Sharpe ratios. *Fin Analysts J*, 2002, **58**(4), 36-52.
- Magnus, J.R. The moments of products of quadratic forms in normal variables. *Stat Neerlandica*, 1978, **32**(4), 201-210.
- Makridakis, S., & Hibon, M. The M3-Competition: results, conclusions and implications. *Int J Forecasting*, 2002, **16**(4), 451-476. Chicago
- Markovsky, I. and Van Huffel, S. Overview of total least-squares methods. *Signal Proc*, 2007, **87**(10), 2283-2302.
- Martin, R and Bana, A. Non-linear Momentum Strategies. *Risk Magazine*, Nov 2012, 60-65.
- Martin, R and Zou, D. Momentum Trading: 'Skews me. *Risk Magazine*, Aug 2012, 52-57.
- Mertens, E. Comments on variance of the IID estimator in Lo. Research Note, 2002, Available online at <http://www.elmarmertens.com/research/discussion/#T0C-Correct-variance-for-estimated-Sharpe-Ratios> (Accessed on 2 July 2017).

- Michalowicz, J. V., Nichols, J. M., Bucholtz, F., & Olson, C. C. A general Isserlis theorem for mixed-Gaussian random variables. *Stat & Prob Letters*, 2011, **81**(8), 1233-1240.
- Miller, M. Evaluating bank-sponsored risk-premium strategies, *Pensions & Investments*, APRIL 5, 2017. Available online at <http://www.pionline.com/article/20170405/ONLINE/170409975/evaluating-bank-sponsored-risk-premium-strategies>. (Accessed 8 May 2017).
- Nadarajah, S. and Kotz, S. Multitude of bivariate t-distributions. *J Theo Appl Stats*, 2004 **38**, 527-39.
- Nadarajah, S., and Pogány, T.K. On the distribution of the product of correlated normal random variables. *Comptes Rendus Math*, 2016, **354**(2), 201-204.
- Parczewski, Peter. A Wick functional limit theorem. *Probab. Math. Stat.*, 2014 **34**(1), 127-145.
- Pav, Steven E., Notes on Sharpe ratios. March 2016, Available online at <https://cran.r-project.org/web/packages/SharpeR/vignettes/SharpeRatio.pdf>. (Accessed 27 Jan 2017).
- Potters, M., and Bouchaud, J.P. Trend followers lose more often than they gain. 2005. Available online at arXiv <https://arxiv.org/abs/physics/0508104>. (Accessed 2 Feb 2017).
- Anisur Rahman, M. and Yu, K.B. Total Least Squares Estimation Using Linear Prediction. *IEEE Trans Acoust, Speech and Sig Proc*, Oct 1987, **35**(10), 1440-54.
- Rencher, A.C., and Christensen, W.F., *Methods of Multivariate Analysis*, 3rd ed., 2012, (Wiley: New York), ch 11.
- Simons, M. K, *Probability Distributions Involving Gaussian Random Variables*, 2006, (Springer Verlag: Berlin), 2006, ch 6.
- Wick, G.C., The evaluation of the collision matrix. *Phys Rev*, 1950, **80**(2), 268-272.
- Zhang, Xian-Da, *Matrix Analysis and Applications*, 2017, (Cambridge UP: Cambridge), pp. 334-5.