Algorithmic Finance 4 (2015) 53–68
DOI 10.3233/AF-150043
IOS Press



# Multi-scale capability: A better approach to performance measurement for algorithmic trading

Ricky Cooper[a], Michael Ong[b] and Ben Van Vliet[a,*]
[a]*Stuart School of Business, Illinois Institute of Technology, Chicago, IL, USA*
[b]*Stuart School of Business, Michael K. Ong Risk Advisory, Chicago, IL, USA*

**Abstract**. This paper develops a new performance measurement methodology for algorithmic trading. By adapting capability from the quality control literature, we present new criteria for assessing control, expected tail loss and risk-adjusted performance in a single framework. The multi-scale capability measure we present is more descriptive and more appropriate for algorithmic trading than the traditional measure used in finance. It is robust to non-normality and the multiple time horizon decision processes inherent in algorithmic trading. We also argue that an algorithmic trading strategy, indeed any investment strategy, which satisfies the criteria to be multi-scale capable also satisfies any definition of prudence. It will be unlikely to harm the investor or external market participants in the event of its failure, while providing a high likelihood of satisfactory risk-adjusted performance.

Keywords: Risk-adjusted performance measure, term structure of capability, algorithmic trading, prudence

## 1. Introduction

Trading strategies must be able to generate sufficient returns to cover their associated cost with some amount of certainty if they are to warrant an allocation. This cost is often an opportunity cost in the form of the risk-free rate or a benchmark return (e.g. for traditional buy-and-hold strategies) or a hurdle rate (e.g. for hedge funds). For algorithmic and/or high frequency trading strategies, this cost is the expense incurred to research, build and operate these data and technology-intensive systems. Kumiega, et al. (2013) (henceforth KNV) describe algorithmic strategies that do reliably cover their costs as being *capable*[1], and they introduce a new measure of capability which captures this concept.

KNV opens up new opportunities for additional contributions. In this paper, we add to the notion of capability that, if it is going to be a robust performance measurement methodology, it should take into account that returns may be non-normal, and that allocation decisions are often made considering multiple time scales.[2] This is especially true for algorithmic trading strategies, where the algorithm itself operates on one time scale (say, milliseconds), the capital allocation decision to the strategy is made on a second time scale (say, daily), and the funding decisions of investors in the trading firm may be made on a third time scale (say, monthly). Unifying measurement of capability across these disparate time scales in a distributionally-generalized framework requires additional criteria. Thus, the performance methodology

*Corresponding author: Ben Van Vliet, Stuart School of Business, Illinois Institute of Technology, 565W. Adams, Chicago, IL 60661, USA. Tel.: +1 312 906 6513; Emails: bvanvliet@stuart.iit.edu; rcooper3@iit.edu (R. Cooper); mong.prof@gmail.com. (M. Ong).

[1]For high frequency trading firms and/or algorithmic trading, infrastructure costs demand a much higher rate of return than the risk free rate to break even. Thus, minimizing fixed operating costs is key component of performance.

[2]Indeed, this scenario occurs in all investment strategies where the expected return of the position over some holding period does not align with the investors expected holding period. Short term performance may impact longer term allocation decisions.

we propose in this paper is a significant extension of KNV.

Specifically, this paper makes four new contributions to the concept of capability. First, our methodology presents five criteria for answering the question of which trading strategy is "good enough" to justify an allocation. While we also borrow concepts from the quality management literature[3], our extension of KNV capability defines decision criteria which rank on longer term goals under the premise that shorter term behavior is also acceptable. Our framework provides for *both* a determination of an algorithmic trading strategy's ability to deliver returns at an acceptable level of risk over the relevant time scale, *and* a ranking of strategies according to a term structure of capability. In doing so, we combine into a single measure operational control, expected tail loss or drawdown, and risk-adjusted performance.

Second, we avoid any assumption of normality. Where KNV assumes normality by way of sampling and the central limit theorem, our measure explores capability using non-normal return distributions. Similarly, where KNV uses sampling and normality for stability assessment and control monitoring (which is is inherently slow[4]), we examine each outcome and assess its conformance to non-normal expected behavior, building on the methodology of Cooper and Van Vliet (2012) (henceforth CV). In our example implementations of capability, we consider a non-normal case and provide[5] the necessary tests and statistics, including a new algorithm for deriving the distribution of the mean.

Third, we investigate serially correlated behavior. Serial correlation in the time series of returns will cause incorrect control violations. Serial correlation in the time series of mean returns will cause misstatement of capability. We describe methods for resolving both these issues.

Fourth and finally, we describe how the capability methodology presented is manipulation-proof in the sense of Goetzmann et al. (2007), and how it can serve as a proxy for prudence in the operation of automated and/or algorithmic agents. This is especially relevant given the recent scrutiny of some forms of algorithmic trading by regulators and the media.

The remainder of this paper is organized as follows. In Section II, we review the relevant literatures in finance and quality management. In Section III, we elaborate on the methods of control and downside risk assessment presented in CV. Section IV presents our new, generalized capability measure and the term structure of capability. Section V presents three example implementations of algorithmic trading strategy assessment according to our methodology. Section VI discusses the impact of serial correlation on our capability measure. Section VII briefly describes how non-constant costs affect capability. Section VIII argues the methodology developed is manipulation-proof and makes the link with prudence. Section IX concludes.

## 2. Background

Within the finance literature, Markowitz (1952) ushered in Modern Portfolio Theory (MPT) with his analysis of risk-averse investors facing a multivariate normal universe of asset returns. This idea eventually flowed into the widely used measures of risk-adjusted return used in the financial industry. The Sharpe Ratio (Sharpe, 1966, 1994) measures expected return minus the risk-free rate per unit of standard deviation. The Information Ratio, developed in Grinold (1989), replaces the risk free rate with an appropriate benchmark return, and standard deviation of returns with the standard deviation of active returns. Jensen (1968), then later Fama and French (1992), as well as Carhart (1997), replace a specific portfolio benchmark with systematic risk premiums and use the portion of portfolio returns that are correlated with these premiums to measure risk. KNV point out that these measures often do not apply to algorithmic strategies, because they do not take into account the significant research and development and technological infrastructure costs associated with the activity. We note that some attention in the finance literature has been given to manipulation of these traditional performance measures, as in Goetzmann et al. (2007). We briefly address this topic with respect to capability.

[3] As we show, although intended to capture the nuances of industrial processes, the spirit of capability measurement in the literature of quality management is surprisingly close to (and in ways more robust than) popular measures in finance.

[4] All other things being equal, slowness is preferably avoided, especially in high frequency trading where out of control performance could lead to large losses while samples are being collected.

[5] While any distribution—normal or otherwise—will work within our framework without any loss of generality, our reference implementation uses the generalized lambda distribution (GLD), largely because it allows for very general specifications of skewness and kurtosis.

Concurrent with the development of these ideas in finance, the quality management literature in industrial engineering developed its own criteria for prudent operations in manufacturing by way of statistical process control (SPC) and measures of process capability. SPC and its constituent charts (e.g. X-bar and R charts) have been the primary tools for the design and control of industrial processes since Shewhart (1931). Shewhart defines a state of statistical control as "a phenomenon will be said to be controlled when, through the use of past experience, we can predict, at least within limits, how the phenomenon may be expected to vary in the future." A state of statistical control exists when a process generates independent and identically distributed deviations from its mean. Such a process is free from variation due to external, or otherwise assignable, causes. (In engineering, the term stable is used rather than stationary to mean that the expected value and variance of the distribution do not change. We use stable to mean in control, as in KNV). The goal of SPC is to ensure a stable process, so that descriptive statistics have meaning. Additionally, as any process will inevitably change, process monitoring with SPC can detect violations of randomness (often by way of single, extreme observation or improbable sequences of observations (Alwen and Roberts, 1988). The process can then be terminated before additional costs or losses are incurred.

Relating SPC to algorithmic finance, however, is a recent development. Hassan et al. (2010), Bilson et al. (2010) and KNV apply traditional sample-based SPC techniques to control algorithmic trading systems. CV provides a new methodology for real time control of high frequency trading systems, using continuously updated statistics with no reliance on normality.[6] As non-normal returns should be explicitly modeled (for example, see Harvey et al. (2010) and Jondeau and Rockinger (2006)), these contributions connect SPC to finance theory. In this paper, we consolidate these ideas by placing a continuously updated, non-normal statistical control methodology and performance measurement in the literature on algorithmic trading.

[6]HFT is a type of automated trading system that uses pre-defined algorithms to execute trades with positive expected returns on time frames that range anywhere from several micro-seconds to several hours. The need for a rapidly responding control system in this type of environment is apparent. However, even in less rapid response environments, a system that responds quickly to a shift in return distribution could be a money saver over one that is based on traditional SPC techniques.

We note that other authors have done related work with SPC in portfolio construction using time series data. Of particular interest are Schmid and Golosnoy (2007, 2009), Schmid and Severin (1998) and Schmid and Tzotchev (2004), who point out that normality assumptions are materially incorrect. For example, Golosnoy and Schmid (2009) use SPC to test optimal portfolios containing assets with serially correlated returns. However, such portfolios are fully invested at all times and continuously evaluated relative to a benchmark, and serial correlation is often embedded in the benchmark. As we explain in a later section, this kind of non-normality in algorithmic strategies, and other types of serial correlation problems, may be handled in this framework by the way of the exponentially weighted moving average (EWMA) framework of Montgomery and Mastrangelo (1991).

Having a stable return distribution, however, does not necessarily mean that a process is generating outputs that are "good enough." As an example from industrial engineering, a manufacturing process could consistently turn out parts that no customer wants to buy. Of concern in industry is whether or not a supplier is able to consistently produce parts within tolerance limits around a customer's specification (Schneider et al., 1995). What is known as process capability measures the ability of a process to satisfy specifications. A wide body of literature exists on process capability and its measurement in various industries. Spiring et al. (2003) and Yum and Kim (2011) provide extensive bibliographies of the literature. Bothe (2001) provides an application text, describing how (virtually every) process capability index can be used in industry. The most common measures of process capability establish a width of some middle (for two-sided), or left or right (for one-sided) percent of a stable distribution. For normal processes through sampling, which are often used in industry, this width is usually $\mu \pm 3\sigma$. The well-known, one-sided $C_{pl}$ index makes use of this:

$$C_{pl} = \frac{\mu_n - LSL}{3\sigma_n} \tag{1}$$

where $\mu_n$ is the mean of all the samples; $\sigma_n$ is the standard deviation of the sample means; and, $LSL$ is the lower specification limit, the explicit tolerance level that must be satisfied. For example, a random process that has some unacceptable percentage of $\mu$'s less than the $LSL$ is not capable. Although the normal has infinite support, the $-3\sigma$ probability is a proxy for the left tail endpoint. Essentially, events beyond this

should never happen (see Kane (1986) and Wheeler and Chambers (1992).

When applied to algorithmic trading strategies as in KNV, capability in Equation (1) is similar to the well-known financial performance ratios with two notable exceptions. First, in capability assessment, process stability is a pre-requisite, while in the traditional financial ratios, it is not. Second, capability assessment seeks to answer how much and how consistently returns must exceed a risk adjustment threshold to be considered "good enough." On top of KNV, we develop a new methodology, which we call *multi-scale capability*, consisting of five principles, which follow a logical progression. Principles 1 and 2 describe the requirements for stability of the return distribution, and for high frequency trading strategies, real-time operational control, as pre-conditions for capability assessment[7]. Principles 3 and 4 require a complete exploration of left tail events, as required by any definition of prudence. Principle 5 presents our multi-scale capability measure with a sufficiency condition.

## 3. Multiple time scale statistical process control

In order for any performance measure to have meaning, it must be measuring a system that operates with a stable, or in control, return distribution. If a strategy is not in control, then either its reference distribution must be changed until it is in control, or it must be considered incapable. This is the first principle.

1. *The trading strategy must generate a stable distribution of returns, measured at the basic reconciliation level.* By basic reconciliation level, we mean the finest granularity used to assess the trading strategy for capital allocation and trade accounting purposes. Often, this level is daily.

A related but separate principle is that daily return statistics are irrelevant if the (algorithmic) strategy is vulnerable to (intra-reconciliation period or intra-day) catastrophic failure. Such a failure could come from operational breaches, undetected coding errors, sudden information releases to the market, unforeseen interactions with the external trading environment, or a host of other reasons. Given the nature of trading systems, especially those that operate with low latencies, it is essential to continuously monitor the distribution of the real time returns (as well as other critical characteristics). This leads to the second principle of multi-scale capability, which is specific to high frequency trading systems.

2. *High frequency trading systems must operate only when the real-time distribution of returns is in control.* This monitoring function over either trade-by-trade returns or portfolio-wide returns for small time slices[8] is the original intent of the SPC methodology of CV, which we outline next.[9]

Once again, we may consider whether a reasonable basis exists to expect that firms will have these stable distributions. The fact is that higher frequency trading firms spend in the six to eight figures to build low latency infrastructures, and lower frequency, quantitative firms spend similar amounts of money on data and research to build their forecasting systems. It is reasonable to assume that they at least think they have a consistent return generating process (see KNV). If this notion is mistaken, then all participants in the market can improve their strategies with this knowledge. The differentiating feature of CV's statistical control methodology is that it considers every single observation of the process under consideration to continuously assess if the strategy is conforming to the reference distribution[10], as opposed to the traditional SPC methodology of relying on sampling to approximate normality as in KNV[11].

Of course, the reference distribution may be skewed, with long tails—left or right—and be leptokurtotic. CV statistical tests for real time monitoring of such distri-

[7] Certainly, a high frequency trading strategy running at a speed that precludes significant human intervention should have a stable, in control process, or the subject of this research paper is already moot. Indeed, more broadly, any fiduciary is undermined by undetected instability in their investment's return process.

[8] As discussed in CV, the daily return for a high frequency trade is generally either dollar profit or loss divided by maximum dollars invested during the day, or maximum profit or loss divided by allocated capital. The trade-by-trade return is the dollars flowing from any trade that reduces the exposure to an asset divided by the average cost per share or contract of the acquisition of all the shares or contracts in that asset.

[9] Other trading system performance metrics also could be monitored using the techniques described in CV, including number of order requests, trades per unit of time, ratio of winning trades, technological latency, and others.

[10] The reference distribution may be taken from a sample of previous known trading days, paper or probationary trading, from a backtest, or possibly from other strategies that have a desired distribution.

[11] The steps in CV are: define the control variable; define its reference distribution; define the statistical tests to be used; and establish control limits.

butions use the generalized lambda distribution[12, 13, 14] While the literature of quality control is replete with tests of violations of a reference distribution, the most common tests look at how many points fall within or out of specific regions of the reference distribution. Although the regions are customizable, we recommend the following control tests:

- Percentage of returns measured beyond 1% and 99% reference tails;
- Percentage of returns measured beyond 5% and 95% reference tails;
- Percentage of returns in inner and outer 20% of the reference distribution;
- The median (simultaneous upper and lower 50% tails);
- Single event outer 0.01% and 99.99% tail klaxons[15].

Statistical significance can be measured by looking at these tests as repeated Bernoulli trials. For example, the probability of having three 1% tail violations in the last 50 observations is $B(3, 0.01, 50)$, where $B(.)$ is the binomial distribution. These variables may be measured on a rolling $n$-observation basis, or in some cases on a cumulative, running basis. Thus, a strategy could encounter one, or two consecutive, 1% left tail reconciliation days, without triggering a control violation and shutdown[16].

To summarize these first two principles, it is important to monitor the reconciliation distribution on which risk adjusted performance statistics will be built. In a high speed environment, where operational control is essential, real time monitoring can prevent catastrophic losses, unintended market impact, and externalities due to trading systems run amok. Next, Principles 3 and 4 require a complete exploration of left tail events, which are part of the in control return process and, therefore, could occur without an SPC violation (which would trigger a shutdown of the strategy).

3. *The expected loss of a one-time tail event must be explicitly modeled*[17] *and deemed acceptable at the relevant time frames.*

For instance, based on common practice in industrial quality control, one might consider a left tail event at say $Q(\alpha)$, the $\alpha = 0.135\%$ percentile[18]. Although one could certainly take the loss at some other percentile (say, the left-tail klaxon level), this value of $\alpha$ is commonly used in quality control. In traditional SPC, $\alpha$ is very small, but in principle, we could consider *any* single event in the left tail as indicative of a potential loss. A financial firm may select a higher level of $\alpha$ depending their level of capital and their ability to absorb short term losses. This notion is meant to capture the same type of information as does expected tail loss or conditional value-at-risk (Artzner et al., 1999). Indeed, one could use expected tail loss as the criterion rather than a quality control-based measure. The important point is that consideration of longer run average risk to return is irrelevant if the potential short run performance is unacceptable.

Another important scenario to consider is the occurrence of unforeseen serial correlation in the reconciliation returns (we deal with persistent correlation later). Every trading strategy can have a bad run, even if it remains in operational control. This could be due to a macroeconomic shift, or a change in the behavior of other market participants. The concern is that a change in the market environment could causes correlated underperformance that is not large enough to trigger a (single) SPC tail event. Nevertheless, such correlated events could be large enough to drain performance over time. This correlation could take time to be recognized and the extent of losses that could be realized prior to an SPC signal must be investigated.

SPC of the reconciliation period returns (as in Principle 1 above) should be able to distinguish between acceptable and unacceptable serial behavior.

[12] Karian and Dudewicz (2011) have shown that the GLD can fit nearly any combination of skewness and kurtosis. The GLD specification also can take on both infinite and finite support. If one wished even more flexibility, a GLD/Generalized Beta Distribution combination estimation method, or Johnson family estimation method could be used with little change to our procedure. All of these estimation methodologies are described thoroughly in Karian and Dudewicz.

[13] Any distribution could be used if one had a prior belief that is would fit better than a GLD distribution.

[14] Pal (2004) contains a good discussion of using the generalized lambda distribution in capability assessment.

[15] A klaxon is a warning signal or alarm bell.

[16] Using the parlance of risk management, the preceding can be thought of as referring to the fact that a longer run average behavior may be fine, but if the Value at Risk (VaR) for a single period, or alternatively, the Expected Tail Loss (ETL), is unacceptable the entire strategy may be unacceptable.

[17] Estimation of the reference distribution could come through Delphi methods (essentially, polling of traders), backtesting, paper trading, simulated trading, probationary trading (i.e. small lots sizes), or past trading of the strategy.

[18] $Q(0.135\%)$ is three standard deviations below the mean for a normal distribution. This is the probability used to assess control in Nelson (1984).

For example, we may consider that three successive 10% left tail events in a row could occur before an SPC event signals non-random serial correlation. Any combination of consecutive events such that the probability $p$ of a single event occurring multiple times $n$, exceeds $\alpha$ (say, $\alpha = 0.135\%$, the typical control limit) may be used for this test. Generally speaking, a trading firm is free to choose the parameters $p$ and $n$ as long as $p^n$ is a probability low enough to trigger a control event. The goal is not to prescribe the best description of possible downside deviations due to unforeseen serially correlated disturbances as much as it is to assert that some consistent standard needs to exist, lest an important component of downside risk is left unexplored. This is the fourth principle.

4. *Possibly correlated losses that could occur prior to the triggering of an SPC control event must be modeled and the accumulated loss deemed acceptable.* As before, this criterion could be stated in terms of expected tail loss if preferred.

Principles 1 through 4 consider partially the tradeoff between the variability and the returns of an algorithimic strategy. However, the conflict goes much deeper. When dealing with money, many people are suspicious of historical return distributions. While all things are estimated with error, if performance in the short run is unacceptably bad, it creates doubt in the estimated distribution of longer run performance. Put another way, by cataloging the extent to which the short run returns could be negative, we build confidence in the longer run projections for a strategy's performance.

## 4. Term structure of capability

To develop the relationship between short run performance and a longer term distribution more fully, we begin with an example of a proprietary trading firm which seeks to remain *on average* cash flow positive. That is, while trading profits from algorithmic strategies will not be positive every reconciliation period (i.e. every day), over the course of an accounting cycle (e.g. bi-weekly payroll, or quarterly performance disclosure) average reconciliation period profits must exceed expenses (see also KNV). Given a stable distribution of returns, the ability of an algorithmic strategy (or any investment strategy) to achieve some specified level of profitability can be examined. This is its capability. Using sample size $n$ and taking several samples, the firm can calculate the mean $\mu_n$ and standard deviation $\sigma_n$ of the sample means $\sigma_n = \sigma/\sqrt{n}$. As in KNV, for the process to be capable using the one-sided $C_{pl}$ index defined in (1), the value must be greater than (for example) one[19]:

$$C_{pl}(n) = \frac{\mu_n - LSL}{3\sigma_n} > 1 \tag{2}$$

Thus, capability says that there is less than a $\mu_n - 3\sigma_n$, or 0.135%, chance that the average daily return of a given sample will be below the *LSL*. (As a reminder, we use 0.135% to follow common practice in industrial quality control, although one could be more or less conservative, or risk averse.) KNV generalizes the *LSL* as $c$, the cost (in percent) allocated to the strategy. The actual nature of this cost depends upon the trading strategy. In high frequency trading, $c$ is the allocated fixed and variable costs to research, build and operate the system[20]. Whatever the case, we can now define *any* strategy as capable over an $n$-period window if equation (2) holds. This is tantamount to saying the strategy is likely to cover its costs (or beat its benchmark) over $n$ consecutive reconciliations with 99.865% certainty.

There are two significant extensions we make to the KNV's $C_{pl}(n)$ measure for finance. First, we note that, in industrial applications, the number of periods $n$ is generally fixed, but with algorithmic trading strategies $n$ can be flexible. One week, one month, or one quarter may all be acceptable time horizons over which to consider capability. Furthermore, as $\sigma_n$ decreases with $n$, the pertinent question is not whether the strategy is capable for a specific $n$, but rather does the *term structure of capability* exceed one within an acceptable time frame. Many investors would be comfortable with a strategy having $C_{pl}(n) > 1$ on an $n > 22$ day (i.e. one month) basis, but certainly not if $C_{pl}(n) > 1$ on an $n > 500$ basis (i.e. two years).

Second, since the reconciliation distribution is not necessarily normal, we generalize the $C_{pl}(n)$ as $GC_{pl}(n)$ where:

[19] Boyles (1991) recommends minimum capability values for one and two-sided specifications in industry. These values are dependent upon whether the process is existing or new, and/or safety critical. In manufacturing, it is common to use 1.33 instead of 1 to allow for a safety margin.

[20] For other lower frequency systems, $c$ could be the risk free rate or the benchmark return. For a pension plan, $c$ could be some hurdle rate. To keep things clear, we assume $c$ is constant. However, if $c$ is not constant, SPC must be performed on the entire numerator as $\mu(t) - c(t)$. The rest of the section proceeds in a straightforward manner. An obvious example would be if $c(t)$ were a benchmark return. In this case, SPC would be on the active return, and all the criteria described would apply to the active return not just the return.

$$GC_{pl}(n) = \frac{\mu_n - c}{\mu_n - Q_n(\alpha)} > 1 \qquad (3)$$

In Equation (3), the distance from the mean to the proxy for the left tail endpoint, which was represented by $3\sigma$ in Equation (2), is replaced by the non-normal equivalent $\mu_n - Q_n(\alpha)$, at some specified level of $\alpha$.

The difficulty in Equation (3) is that the distribution of the $n$ sample size mean $\mu_n$ (and hence the $\alpha$ level) is not known. And, it turns out that $n$ must be quite large before skewness and kurtosis disappear by way of the central limit theorem, something algorithmic traders may not have the patience to wait for. Ideally, we should model $\mu_n$ with a distribution that allows for non-normal values of the third and fourth moments. We solve for the distribution of $\mu_n$ as shown in Appendix 2. In our implementation, we adapt the methodology of Acar et al. (2011) for finding the moments of a generalized lambda distribution[21]. These are the basic steps:

1. Find the non-central moments of the distribution.
2. Apply the Acar et al. methodology to the mean of $n$ independent, identically distributed variables using customized formulas explicitly derived in Appendix 2.
3. Use the four moments obtained in step 2, along with method of moments estimation to obtain a distribution of the sample mean.

Given the distribution of the mean, we can proceed to the fifth principle of multi-scale capability.

*5. The number of reconciliation periods n for which $GC_{pl}(n)$ is greater than one must be acceptably small.*

The acceptable value of $n$ represents a desired time frame over which the strategy is expected to be profitable. The value of $n$ may differ between firms. If a strategy needs to be capable every month, then $n$ must be less than or equal to 22 days in order to assess $GC_{pl}(n) > 1$. However, a high frequency trading firm may need to capable over every two week time horizon. A hedge fund may need to be capable over every three or six month reporting cycle. The appropriate value of $n$ is up to the firm to decide. The level of $\alpha$ represents the risk tolerance of the trading firm, or how important it is that a given level of profitability must be achieved consistently. A low $\alpha$ means that it is totally unacceptable that $c$ not be exceeded. A high $\alpha$ means that it is more acceptable to miss the expected profitability. Again, different firms may choose different levels of $\alpha$ and shifting $\alpha$ changes the slope of the term structure.

The value of $n$ where the term structure crosses one (i.e. $GC_{pl}(n) = 1$) we call $n^*$, and while this point is important, the other points on the curve also contain information. Each point away from $n^*$ indicates a positive or negative safety margin at that point relative to the both the horizontal and vertical axes. The excess or shortfall of capability (i.e. $y$-axis safety margin) at any given point is $GC_{pl}(n) - 1$. The excess or shortfall of $n$ (i.e. $x$-axis safety margin) at any given point is $n–n^*$. A flatter term structure around a given $n$ generally means a smaller safety margin on the $y$-axis, and larger along the $x$-axis. A steeper term structure means a higher safety margin on the $y$-axis, and smaller along the $x$-axis. The slope of the term structure indicates how much the marginal contribution of $n$ is to capability and the stability of the estimate of capability at the given value of $n$. A steeper term structure indicates greater sensitivity to mis-estimation in the $\lambda$ parameters. If the slope is high, then small perturbations in the parameters will change the estimate of capability more significantly.

A practical implication is that the value of $n^*$ highlights the financing decision facing the firm. Financing arrangements vary across values of $n$, affecting the firm's structure and its chosen financing method. If $c$ is constant for the firm, then $n^*$ determines the financing horizon necessary to ensure capability. If the value of $c$ is flexible for firms with various financing options, then the firm will need to consider a curve that is created by points from a family of term structures each with a different cost.

## 5. Numerical examples

In the previous section, we smoothed over a considerable amount of detail. In this section, we expand on the term structure of capability through three example scenarios. The first is a comparison of our capability measure to the Sharpe ratio for strategies with normally distributed returns. The second is an implementation of our methodology given a highly non-normal return process for a hypothetical high frequency trading strategy. The third considers the case of two trading strategies with crossing term structures of capability.

[21] We use the generalized lambda distribution because it is easy to work with and the algorithm to find the moments of a sample mean is fairly straightforward mathematically. However, one could certainly use another distribution if desired—Pearson, Beta, Johnson, Burr, Edgeworth, Weibull, or any member of the extreme value distribution family. The methodology in Appendix 2 would still apply.

### 5.1. Normally distributed returns and capability

If we consider an individual trading strategy that is in control (according to Principles 1 and 2), and has normally distributed returns, then $GC_{pl}(n)$ is the same as the Sharpe ratio multiplied by the constant $\sqrt{n}$ /3, where $c$ would be the risk free rate[22], and $n$ is the number of periods over which the average is taken as in Equation (4).

$$GC_{pl}(n) = \frac{\mu_n - c}{\mu_n - Q_n(\alpha)} = \frac{\sqrt{n}}{3} \cdot \frac{\mu - c}{\sigma} > 1 \quad (4)$$

The parameter value of three in Equation (4) is in keeping with traditional SPC and process capability assessment in industry. In finance, the value of this parameter is a business decision, based on the importance of not missing the required return target on average over $n$ periods. With three, the chance of not covering costs on average over any $n$ periods is only about one or two in 1000. If three is replaced by two in the denominator, for example, then the chance of not covering costs on average is about five in 100. Additionally, if there are two different trading strategies that both have normally distributed returns, then their two term structures of capability will never cross. The strategy with the higher capability at any $n$ will have the highest capability at every $n$.

While KNV compares (at length) capability to the traditional measures in used finance, we note three characteristics of the term structure of capability for normally distributed returns relative to the Sharpe ratio:

1. The term structure of capability tells the manager whether a strategy is or is not acceptable. It is an absolute standard, without the need for comparison between strategies.
2. The time $n$ to capability (i.e. when the term structure crosses one) is a unique ranking system.
3. Because the risk measure (standard deviation) for a normal distribution is convex, trading strategies that are acceptable in their time to capability individually must also be acceptable as a portfolio[23].

[22]Of course, the appropriate cost is not the risk free rate with technology infrastructure-intensive algorithmic trading strategies.

[23]This may not be true for non-normal distributions, since we cannot guarantee the denominator in equation (4) forms a coherent risk measure. However, it would take very unusual shapes of distributions and correlations between them to make this untrue.

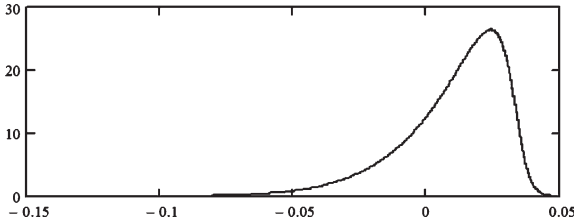


Fig. 1. The density function of the reference distribution.

### 5.2. Non-normal returns and capability

In this section, we consider a more realistic implementation of multi-scale capability. Consider a high frequency trading strategy with a return process that is non-normal, with a small positive expectancy and a long left tail. Many traders would likely recognize this as a strategy that exploits a small statistical regularity (or inefficiency) repeatedly, but on occasion suffers a large loss. Such a strategy would have a probability density of daily returns that looks something like that depicted in Fig. 1, with higher moments that were material in magnitude. The data that generates the density in Fig. 1 has the moments given in Table 1.

As Table 1 shows, the strategy has an expected return of 1%, a 2% standard deviation, and significant left skewness of $-1.25$, and kurtosis of 5. To model this distribution, we use the generalized lambda distribution (GLD) for its ease of use and flexibility for modeling skewness and kurtosis. (It can have infinite or finite support, though most fits to actual data will result in a finite support distribution[24]. If one chooses a different distribution, the methodology we present remains the same.) The generalized lambda distribution which fits these moments is given by $GLD(0.027527, 4.335961, 0.094632, 0.010567)$[25].

Assuming that the strategy is in control with respect to Principles 1 and 2, Principle 3 requires that the one reconciliation period loss be acceptable. This test is easy enough to compute by (A1.1), $Q(0.135\%) = -0.0797$ or $-7.97\%$. Because the distribution is non-normal, this tail event is more extreme

[24]We do not see this as a problem since an accurate estimate of the tails at two, three, or four standard deviations seems more important than having a non-zero probability of say, a 25 standard deviation event.

[25]When using empirical data in practice, there are a variety of estimation techniques available to fit the GLD, as outlined in Appendix 1. As we discuss there, it is important to use a goodness of fit test, especially if the fit is done with method of moments estimation.

Table 1
Moments of the reference distribution of daily returns

| | Central | Non-Central |
|---|---|---|
| Mean | 0.01 | 0.01 |
| Variance | 0.0004 | 0.0005 |
| Skewness | −1.25 | 0.000003 |
| Kurtosis | 5 | 0.00000065 |

than a corresponding $3\sigma$ event under a n assumption of normality (as in KNV) which is $-5\%$. Principle 4 requires examination of (for example) three successive 10% tail events, each with the outcome $Q(10\%) = -1.7372\%$. The accumulated outcome for these three tail events is thus $3 \times Q(10\%)$ equals $-5.2\%$ with probability 0.1%. Alternatively, one could consider, $2 \times Q(1\%)$, or even $5 \times Q(2\%)$, so long as the reconciliation SPC for non-conformance is close to the single event probability.

Finally, Principle 5 requires that we find the term structure of capability. To accomplish this, we first find the distribution of the mean using the methods of Appendix 2. Table 2 shows the moments of the distribution of the mean for $n = 5$, and we can see that the distribution of five day average returns retains significant skewness and some kurtosis. This is why measuring capability using non-normal statistics enables greater precision. The moments of the distribution of the sample mean in Table 2 coincide with *GLD*(0.014559, 16.827171, 0.145843, 0.053255). Using this GLD, we can easily find any needed percentile.

To see the entire term structure of capability, we simply vary $n$ and reuse the techniques just discussed. Table 3 presents statistics for values of $n$ from 2 to 79 using an assumed cost per day of $c = 0.001$ (i.e. 10 basis points) and $\alpha = 0.135\%$. The first three columns in Table 3 show the central moments of the $n$-day distribution of the mean, which gets closer to normal as $n$ increases, as expected. But, even at $n = 79$ trading days (about four months) material non-normality remains. The columns in Table 3 labeled $\lambda_1$ through $\lambda_4$ show the

Table 2
Moments of the daily average distribution for a weekly sample

| | Central | Non-Central |
|---|---|---|
| Mean | 0.01 | 0.01 |
| Variance | 0.00008 | 0.000184 |
| Skew | −0.559017 | 0.000003 |
| Kurtosis | 3.4 | 0.000000064 |

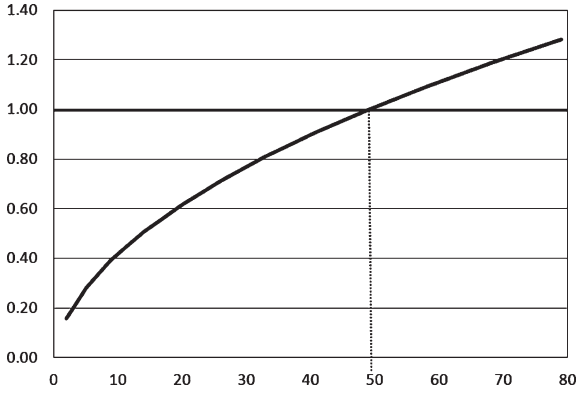


Fig. 2. Term structure of capability via $GC_{pl}(n)$.

GLD parameters that fit the moments in the columns to the left. The right-most column shows the $GC_{pl}(n)$ values that result, and Fig. 2 plots this column, which is the term structure of capability.

As can be seen in Fig. 2 and Table 3, the $GC_{pl}(n)$ is an increasing function on $n$. This should be expected, since the variance of the mean decreases linearly with $n$, regardless of the distribution function (assuming a finite variance exists). In this case, the $GC_{pl}(n)$ crosses one at $n = 50$ days, or ten weeks. So, using the five principles, we can say that this trading strategy is multi-scale capable as long as:

1. The returns remain in control with respect to the reference distribution of reconciliation period returns.
2. The trading strategy runs only when its real time performance metrics are in control.
3. The trading firm is prepared to take a $-7.97\%$ loss over a single reconciliation period with a probability of 0.135%.
4. The trading firm is prepared to take a $-5.2\%$ aggregated loss over three reconciliation periods with a probability of 0.001.
5. The trading firm is prepared to accept that the volatility of returns only turns to their favor on average if considered over a 50 day time horizon.

Clearly, our multi-scale capability method accurately maps the real time and reconciliation period returns, and monitors the stability of both these processes. Further, it answers the question as to whether or not this strategy is "good enough" to overcome its allocated cost within an acceptable time frame.

Table 3
Term structure of capability given non-normality

| $n$ | Variance | Skew | Kurtosis | $\lambda_1$ | $\lambda_2$ | $\lambda_3$ | $\lambda_4$ | $GC_{pl}(n)$ |
|---|---|---|---|---|---|---|---|---|
| 2 | 0.000200 | −0.883883 | 4.000000 | 0.019810 | 8.263772 | 0.121069 | 0.027673 | 0.158391 |
| 5 | 0.000080 | −0.559017 | 3.400000 | 0.014559 | 16.827171 | 0.145843 | 0.053255 | 0.279561 |
| 9 | 0.000044 | −0.416667 | 3.222222 | 0.012740 | 24.959030 | 0.153777 | 0.069407 | 0.394418 |
| 14 | 0.000029 | −0.334077 | 3.142857 | 0.011845 | 32.812585 | 0.156230 | 0.080578 | 0.506403 |
| 20 | 0.000020 | −0.279508 | 3.100000 | 0.011331 | 40.474357 | 0.156558 | 0.088723 | 0.616782 |
| 26 | 0.000015 | −0.245145 | 3.076923 | 0.011042 | 46.992624 | 0.156116 | 0.094147 | 0.711491 |
| 33 | 0.000012 | −0.217597 | 3.060606 | 0.010833 | 53.656968 | 0.155353 | 0.098643 | 0.808985 |
| 41 | 0.000010 | −0.195217 | 3.048780 | 0.010677 | 60.413408 | 0.154446 | 0.102377 | 0.908409 |
| 49 | 0.000008 | −0.178571 | 3.040815 | 0.010571 | 66.506280 | 0.153597 | 0.105194 | 0.998492 |
| 58 | 0.000007 | −0.164133 | 3.034483 | 0.010485 | 72.766128 | 0.152737 | 0.107658 | 1.091401 |
| 68 | 0.000006 | −0.151585 | 3.029407 | 0.010416 | 79.153005 | 0.151895 | 0.109812 | 1.186504 |
| 79 | 0.000005 | −0.140636 | 3.025320 | 0.010359 | 85.636013 | 0.151084 | 0.111695 | 1.283326 |

### 5.3. *Two strategies with crossing term structures of capability*

As previous stated, given two trading strategies with normally distributed returns, the one with the higher $GC_{pl}(n)$ at any $n$ will have the highest capability at every $n$. This is not necessarily true, however, in the presence of non-normal returns. Two term structures of capability could cross, and in this section, we consider this scenario. Consider that the first trading strategy is the non-normal trading strategy from the previous example. The second trading strategy has normally distributed returns with the same mean 0.01 and almost identical variance 0.000484 as the non-normal strategy. Figure 3 adds the term structure of capability for this second, normal strategy (shown as the dashed line) and we can see that at about 18 days, the two term structures cross. For lower values of $n$, the normal strategy is more capable, but for greater values of $n$, the non-normal strategy is more capable. Where the non-normal strategy becomes capable at $n = 50$, the normal strategy becomes capable at $n = 54$.

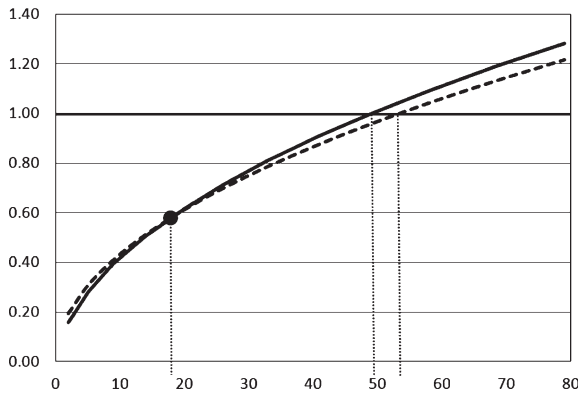


Fig. 3. Crossing term structures of capability.

The reason this crossing occurs is because, in the presence of non-normality, the relationship between the rate of decline in the sample means and increasing values of $n$ is not clear (as it is with two normals). Excess kurtosis in the non-normal strategy implies a greater one time extreme loss, as investigated by Principle 3, and this lowers its capability for low values of $n$. However, the variance of the distribution of mean returns (for the non-normal strategy) consolidates more quickly as $n$ increases than it does for a normal strategy. Thus, while the longer term capability of the non-normal strategy looks better, the short term performance is (potentially) worse. This is what makes investigation of the term structure of capability especially important. Longer term capability of a trading strategy must be understood in light of the possibility of greater short term losses in the presence of non-normality. Our multi-scale capability methodology makes clear the nature of short term performance and capability for differing values of $n$. Investors benefit by looking at both these aspects of performance rather than simply ignoring or assuming away this relationship.

## 6. Serial correlation

The assumption of our application of SPC and capability is that returns are independent and identically distributed. This assumption is sometimes violated in algorithmic trading due to the dynamic nature of

the economic environment. The location of a reference distribution may be conditional upon exogenous market state variables—the time of day, volume, or volatility—which may also introduce serial correlation in the time series of returns. This raises issues that warrant further investigation. First is that serial correlation may induce spurious SPC violations. Second is that serial correlation in mean returns may cause mis-estimation of capability.

Montgomery and Mastrangelo (1991) (henceforth MM) show that for serial correlation that is deemed to be an inherent part of an in control process, it is possible to adjust SPC mechanisms accordingly. (It is important to be sure that serial correlation is not simply due to an uncontrolled state variable, but is (in fact) an inherent part of the strategy's outperformance.) MM demonstrate that using differences in an exponentially weighted moving average (EWMA) of returns will generally result in a stationary time series for SPC. This methodology works well for MA(1), AR(1), and ARMA(1,1) processes. For example, the location of the reference distribution could be calculated using the moving range method if, for example, the process was fit with the AR(1) model $\hat{X}_t = a + bX_{t-1}$. The residuals from this model should be uncorrelated and the control limits in Principal 2 can be computed using the moving range (*MR*) method, where the $MR_i = |x_i - x_{i-1}|$, and the average *MR* value. One could, however, use full ARIMA modeling to accomplish the same goal. We agree with MM, though, that most processes that are in control should have fairly low level serial dependencies.

With respect to risk-adjusted performace, serially correlated returns may cause misstatement of performance measure, including, for example, the Sharpe ratio (Lo (2002)). For normally distributed returns, if positive serial correlation exists in the return series, then the Sharpe ratio will be underestimated because volatility will be overestimated. If negative serial correlation exists, then the Sharpe ratio will be overestimated because volatility will be underestimated. Given the relationship between the Sharpe ratio and our capability measure for normally distributed returns as in equation (4), the term structure of capability will incur a similar bias. $GC_{pl}(n)$ for a given $n$ will be overstated or understated, depending on the sign of the auto-correlation. However, the outcome is the same as in Lo. In the case of normally distributed returns the same correction may be applied directly to our capability measure.

For non-normal returns, however, the required adjustment to $GC_{pl}(n)$ involves a return to Acar et al. (2010). That paper derives a methodology for determing the distribution of the moments of any arbitrary function of a set of random variables. In Appendix 2, we apply their methodology to the simple sample mean $\frac{1}{T}\sum_t X_t$. With some additional effort (and possibly more numerical approximations), this methodology can be applied to any time series formulation. For example, an AR(1) process would require using this methodology on the function, $\frac{1}{(T-1)\sqrt{1-\rho^2}}\sum_t (X_t - \rho X_{t-1})$, where $\rho$ is auto-correlation coefficient. Again, the results are qualitatively similar to Lo, with skewness and causing adjustments as they did in the previous example involving the crossing term structures of capability.

## 7. Non-constant costs

We also note that, relative to the information ratio (IR) of Grinold (1989), if the trading strategy under consideration is being assessed relative to a benchmark (so that the cost function is actually a stochastic variable), then one can simply perform the reference SPC on the active return (AR) distribution and the $GC_{pl}(n)$ ratio can be calculated as in Equation (5):

$$GC_{pl}(n) = \frac{\mu_{AR(n)}}{\mu_{AR(n)} - Q_{AR(n)}(\alpha)} \tag{5}$$

In general, working with everything in terms of relative statistics causes no difficulty. The multi-scale capability standards are simply in terms of a relative reference distribution.

It is also possible to see how Jensen's alpha (1968), and its extension to the Fama and French (1992) or Carhart (1997) alpha measures fit. In the case of capability, risk adjustments to returns become the time varying cost function of capability. The question then becomes: Is the excess return over the risk premium a capable process? The Jensen's alpha and its later derivatives also are enhanced by the capability framework. The cost function can be replaced by the necessary risk adjusted premiums, and capability then becomes an extension of the *t*-statistic of alpha, except that it brings in the ability to accommodate non-normal disturbances and the multi-scale SPC of our methodology.

## 8. Manipulation and prudence

Where other performance measures in finance are subject to manipulation in the sense of Goetzmann et al. (2007), multi-scale capability is manipulation-proof. Goetzmann, et al. focuses on ranking strategies using a single measure that cannot be gamed by levering up risk, or manipulating the aggregation period. Although our multi-scale capability methodology cannot be manipulated, this characteristic comes through SPC monitoring, measuring on multiple time scales, and more accurately mapping return distributions. Trading strategies with unstable return distributions will cause SPC violations. Those strategies that use instruments with non-linear payoffs (i.e. only payoff in certain situations) will cause the time to capability to increase. Among strategies that are capable, the time $n$ (that is necessary to become capable) forms a ranking measure, given the acceptability of short-term performance as previously discussed. These characteristics render multi-scale capability resistant to manipulation.

Also, we believe our methodology defines prudence[26] in algorithmic trading, a concept has evolved since it was first established as court precedent in 1830[27]. Longstreth (1986) points out that the recurring theme in the history of thinking about prudence is that it demands adherence to sound processes that produce strategies with desirable characteristics, including a responsibility to monitor the strategy in light of its purpose, manage risk, and minimize the possibility of large losses. This is traditionally the duty of care to which fiduciaries are obligated. All of these are encompassed within our five principle methodology.

In a more general sense, investors in proprietary trading firms also seek to be prudent with their trading capital. In this sense, prudence means placing money where the expected return is favorable and the chance and magnitude of loss is fully considered. At a complex financial firm, multi-scale capability may be implemented in a different manner at different levels of aggregation. For example, an individual strategy may only need to be capable at a longer $n$ and at a higher level of risk tolerance $\alpha$, where the entire firm may need to be capable at a shorter $n$ and/or a lower level of $\alpha$. Multi-scale capability at the firm level may be acceptable, but capability assessment of individual strategies will uncover where problems may lie.

As we have discussed, SPC is the mechanism for monitoring performance. It should signal when to stop trading and reassess the strategy if the distribution of returns shifts. Equally important, especially in light of recent marketplace debacles caused by high frequency trading, is that the real time control component of multi-scale capability protects the market from operational mishaps or haywire trading systems. This should shield the prudent investor, trader or portfolio manager, as well as external market participants, from possibly devastating consequences.

## 9. Conclusion

Both algorithmic trading and investment strategies bring potential instability to investors. One way to control this risk is through ever more prescriptive rules of responsibility, whether self-imposed or through regulation. However, we believe that both higher frequency trading and lower frequency investment strategies are best controlled by an informative quantitative evaluation process. Therefore, to accomplish this task we have developed a process that merges ideas from both finance and quality management.

The essence of multi-scale capability is that it takes snap-shot measures over various time horizons and unifies them into a coherent inter-temporal decision framework. Given that investors do not want to be hampered by tail events along the path to longer term performance, our methodology brings to the fore the time period over which the investor must wait to know with a high degree of certainty that the strategy will cover its costs. The strategy for which this number is the lowest is the best strategy.

The multi-scale capability methodology protects both the investor and the marketplace with real time monitoring and operational SPC, and verifies the validity of performance with SPC of the input reconciliation data. We conclude then that multi-scale capability is sufficient to assure the prudence of a trading strategy in the legal sense, and that its quantitative nature lends itself to easy outside verification. This is a significant step forward in ensuring market safety and improving

[26] The original statement of the prudent man rule mandated that fiduciaries "observe how men of prudence ... manage their own affairs ... considering the probable income, as well as the probable safety of the capital to be invested" (Massachusetts (1830)).

[27] Over the course of this paper, we have borrowed significantly from the literature of quality control and we note that there is certainly an extensive literature relating quality management with ethical behavior (for an excellent overview of this literature, see Tari (2011)).

the evaluation of all forms of trading and investment strategy.

## References


Acar, E., Rais-Rohani, M., Eamon C., 2010. Reliability estimation using univariate dimension reduction and extended generalized lambda distribution. International Journal of Reliability and Safety 4(2/3), 166-187, reprinted in Handbook of Fitting Statistical Distributions with R. Karian, ZA & EJ Dudewic (eds), New York: CRC Press, 2011.

Alwan, L.C., Roberts, H.V., 1988. Time series modeling for statistical process control. Journal of Business & Economic Statistics. 6, 87–95.

Artzner, P., Delbaen, F., Heath, J.-M., Eber, D., 1999. Coherent measures of risk. Mathematical Finance. 9, 203–228.

Bilson, J., Kumiega, A., Van Vliet, B., 2010. Trading model uncertainty and statistical process control. Journal of Trading. 5, 39–50.

Bobo, L.J., 1984. Nontraditional investments of fiduciaries: Re-examining the prudent investor rule. Emory Law Journal. 33, 1067–1102.

Bothe, D.R., 2001. *Measuring Process Capability*. Cedarburg, WI: Landmark Publishing, Inc., pp. 1-2.

Boyles, R., 1991. The Taguchi capability index. Journal of Quality Technology. 23, 17–26.

Carhart, M.M., 1997. On persistence in mutual fund performance. Journal of Finance. 52, 57–82.

Cooper, R., Van Vliet, B., 2012. Whole distribution statistical process control for high frequency trading. Journal of Trading. 7, 57–68.

Del Guerico, D., 1996. The distorting effect of the prudent-man laws on institutional equity investments. Journal of Financial Economics. 40, 31–62.

Fama, E.F., French, K.R., 1992. The cross-section of expected stock returns. Journal of Finance. 47, 427–465.

Fleming, A., 1977. Prudent investments: The varying standards of prudence. Real Property, Probate and Trust Journal. 243–255.

Goetzmann, W., Ingersoll, J., Spiegel, M., Welch, I., 2007. Portfolio performance manipulation and manipulation-proof performance measures. Review of Financial Studies. 20, 1503–1546.

Golosnoy, V., Schmid, W., 2009. Statistical process control in asset management. In Applied Quantitative Finance, 2nd ed. W Härdle, N Hautsch, & L Overbeck (eds). Berlin, Germany: Springer Verlag.

Grinold, R., 1989. The fundamental law of active management. Journal of Portfolio Management. 15, 30–37.

Hassan, M.Z., Kumiega, A., Van Vliet, B., 2010. Trading machines: Using SPC to assess performance of automated trading systems. Quality Management Journal. 17, 42–53.

Harvey, C.R., Liechty, J.C., Liechty, M.W., Müller, P., 2010. Portfolio selection with higher moments. Quantitative Finance. 10, 469–485.

Jensen, M.C., 1968. The performance of mutual funds in the period 1945-1964. Journal of Finance. 23, 389–416.

Jondeau, E., Rockinger, M., 2006. Optimal portfolio allocation under higher moments. European Financial Management. 12, 29–55.

Kane, V., 1986. Process capability indices. Journal of Quality Technology. 18, 41–52.

Karian, Z., Dudewicz, E., 2011. Handbook of fitting statistical distributions with R. New York: CRC Press.

King, R., MacGillivray, J., 1999. Theory and methods: A starship estimation method for generalized lambda distributions. Australian & New Zealand Journal of Statistics. 41, 353–374.

Kumiega, A., Neururer, T., Van Vliet, B., 2013. Trading system capability. Quantitative Finance. 14, 383–392.

Lo, A.W., 2002. The statistics of Sharpe ratios. Financial Analysts Journal. 58, 36–52.

Longstreth, B., 1986. Modern Investment Management and the Prudent Man Rule. Oxford University Press.

Markowitz, H., 1952. Portfolio selection. Journal of Finance. 7, 77–91.

Massachusetts, State of, Harvard College v. Armory. 1830. 26 Mass (9 Pick) 446.

Montgomery, D., Mastrangelo, C.M., 1991. Some statistical process control methods for autocorrelated data. Journal of Quality Technology. 23, 179–197.

Nelson, L., 1984. Technical aids. Journal of Quality Technology. 16, 238–239.

Ozturk, A., Dale, R., 1985. Least squares estimation of the parameters of the generalized lambda distribution. Technometrics. 27, 81–84.

Pal, S., 2004. Evaluation of nonnormal process capability indices using generalized lambda distribution. Quality Engineering. 17, 77–85.

Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P., 2007. Numerical Recipes: The Art of Scientific Computing, 3rd ed. Cambridge University Press.

Rahman, S., Xu, H., 2004. A univariated dimension-reduction method for multi-dimensional integration in stochastic mechanics. Probabilistic Engineering Mechanics. 19, 393–408.

Rosenburgh, M., Spieler, A.C., 2009. 21st century pensions: The risk, the hedge and the duty to consider. Journal of International Business and Law. 8, 45.

Schmid, W., Golosnoy, V., 2007. EWMA control charts for monitoring optimal portfolio weights. Sequential Analysis, 26, 195–224.

Schmid, W., Severin, T., 1998. Statistical process control and its application in finance. In Contributions to Economics: Risk Measurement, Econometrics, and Neural Networks, Bol, G., Nakhaeizadeh, G., Vollmer, K.H. (eds). Heidelberg, Germany: Physica-Verlag, pp. 83–104.

Schmid, W., Tzotchev, D., 2004. Statistical surveillance of the parameters of a one-factor Cox-Ingersoll-Ross model. Sequential Analysis. 23, 379–412.

Schneider, H., Pruett, J., Lagrange, C., 1995. Uses of Process Capability Indices in the Supplier Certification Process. Quality Engineering. 8, 225–235.

Sharpe, W.F., 1966. Mutual Fund Performance. Journal of Business. 39, 119–138.

Sharpe, W.F., 1994. The Sharpe Ratio. Journal of Portfolio Management. 21, 49–58.

Shewhart, W.A., 1931. Economic Control of Quality of Manufactured Product. New York: Macmillan.

Spiring, F., Leung, B., Cheng, S., Yeung, A., 2003. A bibliography of process capability papers. Quality and Reliability Engineering International. 19, 445–460.

Tari, J., 2011. Research into quality management and social responsibility. Journal of Business Ethics. 102, 623–638.

Wheeler, D., Chambers, D. 1992. Understanding Statistical Process Control, 2nd edition. Knoxville, TN: SPC Press, Inc.

Yum, B.J., Kim, K.W., 2011. A bibliography of the literature on process capability indices: 2000–2009. Quality and Reliability Engineering International. 27, 251–268.

## Appendix 1

*The Generalized Lambda Distribution (GLD)*

The four parameter GLD is defined by its inverse cumulative distribution, or percentile function:

$$Q(p) = \lambda_1 + \frac{p^{\lambda_3} - (1-p)^{\lambda_4}}{\lambda_2} \tag{A1.1}$$

This leads to the parametric form of its density function:

$$f(x) = \frac{\lambda_2}{\lambda_3 \cdot p^{\lambda_3 - 1} + \lambda_4 (1-p)^{\lambda_4 - 1}} \tag{A1.2}$$

at $x = Q(p)$ as $p$ varies from 0 to 1.

The GLD is capable of taking on either infinite or finite support depending on the signs of $\lambda_3$ and $\lambda_4$. Positive values have finite support and negative have infinite. $\lambda_3$ controls the left tail and $\lambda_4$ the right tail. Most real data will be fit with finite support that extends well beyond the last observation. The normal also is fit best with finite support, though it matches the first four moments exactly and has support plus or minus approximately six standard deviations. A more thorough discussion of these considerations, along with other possible distributions to fit, namely the generalized beta, and the Johnson family, are all discussed in Karian and Dudewicz's exhaustive work.

The moments of the GLD are given by:

$$(\text{mean})\mu = E(x) = \lambda_1 + \frac{A}{\lambda_2} \tag{A1.3}$$

$$(\text{variance})\sigma^2 = E(x-\mu)^2 = \frac{B - A}{\lambda_2^2} \tag{A1.4}$$

$$(\text{skewness})\gamma_1 = E(x-\mu)^3/\sigma^3 = \frac{C - 3AB + 2A^3}{\left(B - A^2\right)^{3/2}} \tag{A1.5}$$

$$(\text{kurtosis})\gamma_2 = E(x-\mu)^4/\sigma^4 = \frac{D - 4AC + 6A^2B - 3A^4}{\left(B - A^2\right)^2} \tag{A1.6}$$

Where,

$$A = \frac{1}{1+\lambda_3} - \frac{1}{1+\lambda_4},$$

$$B = \frac{1}{1+2\lambda_3} + \frac{1}{1+2\lambda_4} - 2\beta(1+\lambda_3, 1+\lambda_4),$$

$$C = \frac{1}{1+3\lambda_3} - \frac{1}{1+3\lambda_4} - 3\beta(1+2\lambda_3, 1+\lambda_4) + 3\beta(1+2\lambda_4, 1+\lambda_3),$$

$$D = \frac{1}{1+4\lambda_3} + \frac{1}{1+4\lambda_4} - 4\beta(1+3\lambda_3, 1+\lambda_4) + 6\beta(1+2\lambda_3, 1+2\lambda_4) - 4\beta(1+\lambda_3, 1+3\lambda_4),$$

and,

$$\beta(x, y) = \Gamma(x)\Gamma(y)/\Gamma(x+y) = \exp\left[\ln\Gamma(x) + \ln\Gamma(y) - \ln\Gamma(x)\ln\Gamma(y)\right], \; for x, y > 0. \tag{A1.7}$$

Several methods exist for estimating these four parameters. For our purposes we generally use two. For the term structure work we use method of moments since the output of our algorithm are the four moments of the sample mean distribution. For SPC on real data we tend to prefer the straight-forward least squares method of Ozturk and Dale (1985). This method finds the parameters that fit the GLD to the empirical percentile distribution using a least squares methodology. Many other variations and innovations exist and almost all are discussed thoroughly in Karian and Dudewicz.

Finally, empirical fits of GLD distributions should be checked with a goodness of fit test, or at the very least a visual inspection of the density function should be performed. Some moment combinations track back to multiple GLD specifications. If the wrong fit is found another set of starting values is used for the estimation

procedure (this iterative procedure has been systematized in a process known as starship estimation (King and MacGillivray 1999)). We do not generally find this to be a problem with the method of Ozturk and Dale and do not find it problematic with the method of moments as used in this paper. However, as a precaution we always check.

The test we use for SPC is the modified Kolmorgorov-Smirnov test with Kuiper statistic as outlined in Press et al. (2007). The reader is cautioned to use the third edition of this test as the second contains a minor typographical error that renders the relevant formulas incorrect. For the term structure of capability we check visually that parameters vary smoothly with $n$, and that the resultant distribution looks like the reference distribution, becoming more normal looking as $n$ increases.

## Appendix 2

*Finding the Distribution of the Sample Mean for an Arbitrary Distribution*

To the best of our knowledge, a method for finding the distribution of an $n$-sample mean of independent, identically (but arbitrarily) distributed random variables has not been published previously. So, we present this methodology as a fairly easy solution to the problem.

Finding the distribution of the sample mean is a four step process:

1. Find moments of the original distribution (through GLD estimation)
2. Find the zero moments of the mean of sample size $N$.[28]
3. Convert the zero moments to regular moments, and use method of moments to fit a GLD to the distribution of the sample means.
4. Any confidence level desired can be obtained with $Q(p)$ of this GLD.

To begin, we define the $l$th non-central moment of the distribution of the sample mean as:

$$m_l = E[Y^l(\vec{X})] = E\left[\left(\frac{\sum_{i=1}^{N} Q(p_i)}{N}\right)^l\right] \tag{A2.1}$$

where, $Q(p_{i)}$ is the percentile (i.e. $X$-value) associated with random probability draw $p$. Instead of needing an $N$th order integration to evaluate each of the first four moments of the distribution, the following algorithm is available.

$$m_l = \sum_{i=0}^{l} \binom{l}{i} S_{i,N} \left(-(N-1)N\mu\right)^{l-1} \tag{A2.2}$$

where,

$$\mu = E[Q(p)] \tag{A2.3}$$

$$S_{0,j} = 1 \forall j \in 1, 2, \ldots N \tag{A2.4}$$

The other terms can be found iteratively using:

$$S_{i,1} = E\left[\frac{Q(P) + (N-1)\mu}{N}\right]^i \tag{A2.5}$$

for $i = 1, 2, 3, 4$ and for $j = 2, \ldots, N$

$$S_{i,j} = \sum_{k=1}^{n} \binom{i}{k} S_{i,j-1} E\left[\frac{Q(P)+(N-1)\mu}{N}\right]^{i-k} \tag{A2.6}$$

The progression is to calculate each of $i = 1, 2, 3, 4$ for each $j$, and then to increment $j$. The final $S_{i,N}$'s are then used in the formula. The evaluation of the $S_{i,j}$'s uses Equations (A2.7) through (A2.10). By simple expansions:

$$E\left[\frac{Q(P) + (N-1)\mu}{N}\right]^1 = \mu \tag{A2.7}$$

$$E\left[\frac{Q(P) + (N-1)\mu}{N}\right]^2 = \frac{1}{N^2}\left(E\left[Q(p)^2\right] + \mu^2(N-1)(N+1)\right) \tag{A2.8}$$

$$E\left[\frac{Q(P) + (N-1)\mu}{N}\right]^3 = \frac{1}{N^3}\left(E\left[Q(p)^3\right] + 3E\left[Q(P)^2\right](N-1)\mu + \mu^3(N-1)^2(N+2)\right) \tag{A2.9}$$

[28] We adapt the methodology of Acar et al. (2010), who based their work on the additive decomposition result of Rahman and Xu (2004) for estimating the non-central moments of a general function of random variables.

$$E\left[\frac{Q(P)+(N-1)\mu}{N}\right]^4$$
$$=\frac{1}{N^4}\left(E\left[Q(p)^4\right]+4E\left[Q(P)^3\right](N-1)\mu\right.$$
$$\left.+6E\left[Q(P)^2\right]\mu^2(N-1)^2+\mu^4(N-1)^3(N+3)\right) \tag{A2.10}$$

Furthermore, given the central moments of a GLD[29]—$\mu$, $\sigma^2$, $\alpha_3$, $\alpha_4$—the non-central moments may be defined as:

$$E[Q(p)]=\mu$$
$$E\left[Q(p)^2\right]=\sigma^2+E[Q(p)]^2$$
$$E\left[Q(p)^3\right]=\alpha_3\sigma^3+3E[Q(p)]E[Q(p)^2]-2E[Q(p)]^3$$
$$E\left[Q(p)^4\right]=\alpha_4\sigma^4+4E[Q(p)]E[Q(p)^3]$$
$$-6E[Q(p)]^2E[Q(p)^2]+3E[Q(p)]^4 \tag{A2.11}$$

[29]The formulas for the central moments in terms of the four parameters are given in the Appendix 1.

Table A1
*S* Matrix for $n=5$

| | | | | |
|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 1 |
| 0.01 | 0.02 | 0.03 | 0.04 | 0.05 |
| $1.16\times10^{-4}$ | $4.32\times10^{-4}$ | $9.48\times10^{-4}$ | $1.66\times10^{-3}$ | $2.58\times10^{-3}$ |
| $1.40\times10^{-6}$ | $9.76\times10^{-6}$ | $3.11\times10^{-5}$ | $7.14\times10^{-5}$ | $1.37\times10^{-4}$ |
| $1.77\times10^{-8}$ | $2.28\times10^{-7}$ | $1.05\times10^{-6}$ | $3.14\times10^{-6}$ | $7.39\times10^{-6}$ |

Once the four non-central moments of the *n*-sample mean are known, they can be converted back to central moments. Finally, using method of moments estimation a GLD may be fit and the *n*-sample mean distribution known.

In the text, the reconciliation distribution is given as *GLD*(0.027527, 4.335961, 0.094632, 0.010567). Using equations (A1.3)–(A1.6) the four central moments are determined to be 0.01, 0.0004, −1.25, and 5. Using equation (A2.11) the corresponding non-central moments are 0.01, 0.0005, 0.000003, 0.00000065. With this information we can determine the *S* matrix for an $n=5$ sample mean to be that shown in Table A1. The final column of this *S* matrix is used in (A2.2) to arrive at the sample mean distribution non-central moments in Table 3.