Hayette Gatfaoui (France)

Deviation from normality and Sharpe ratio behavior: a brief simulation study

Abstract

Sharpe ratio has been widely used in the portfolio management industry as well as fund industry (Robertson, 2001; Scholz and Wilkens, 2005). Users often forget the main core assumption describing the appropriateness of such a risk-adjusted performance measure, namely asset return normality. This concern is of huge significance insofar as performance indicators drive the asset allocation policy, performance forecasts and cost of capital assessment among others (Farinelli et al., 2008; Lien, 2002; Christensen and Platen, 2007). We employ a brief simulation study to assess the impact of deviations from normality on the performance measures and rankings inferred from Sharpe ratio's estimates. Our analysis allows for assessing the possible bias in both performance measurement and ranking, which results from the existence and the magnitude of skewness and kurtosis patterns in asset returns. This study proposes a method to extract an unbiased performance measure (i.e. unbiased Sharpe ratios) from observed classic Sharpe ratios after accounting for the returns' skewness bias. The resulting unbiased Sharpe ratio outperforms its classic counterpart at the stock picking level.

Keywords: asset return, distributional shocks, kurtosis, performance, Sharpe ratio, skewness.
JEL Classification: C15, C16, G12.

Introduction

There is a long debate about the way to soundly assess the performance of asset returns relative to the corresponding risk borne by investors. Formerly, classic performance measures were introduced so as to assess the investment's reward relative to its risk. Under the general name of risk-adjusted performance measures, such performance indicators usually consider risk as the asset returns' standard deviation or corresponding pairwise covariance (when considering the portfolio's level). Sharpe (1964; 1994) introduced one version of risk-adjusted performance measure, where risk is represented by the returns' standard deviation. Moreover, such framework relies strongly on the well-known mean-variance efficiency principle as well as returns' Gaussian distribution assumption.

As long as returns are Gaussian, any investment's reward is conveniently described by its average return on the investment horizon under consideration, whereas the corresponding risk is conveniently captured by returns' standard deviation (i.e. variance). However, deviation from normality may generate a biased assessment of performance while applying Sharpe ratio. Unfortunately, real world often exhibits non-Gaussian returns (Eling, 2006; Madan and McPhail, 2000; Taleb, 2007) so as to emphasize skewness and kurtosis patterns (Black, 2006; Eling and Schuhmacher, 2007; Harvey et al., 2004; Harvey and Siddique, 2000; Ziemba, 2005). Despite its inappropriateness for non-Gaussian asset returns' distribution, Eling and Schuhmacher (2007) show nevertheless that Sharpe ratio yields the same ranking across hedge funds as other appropriate performance measures. But, the existence of at least one statistically significant outlier return is enough to bias upward or downward the obtained Sharpe ratio (Gatfaoui, 2009)[1]. Indeed, biases relative to the Sharpe ratio have been reported by academic and professional research (Hodges, 1998; Klemkosky, 1973; Spurgin, 2001). Moreover, the general trend of the financial market (i.e. market cycle) drives the assessment quality of Sharpe performance measure (Scholz and Wilkens, 2005; Scholz, 2007; Sortino, 2004).

In the lens of such a landscape, we attempt to characterize the bias in Sharpe ratio, which may arise from deviations from Normality. For this prospect, we run a simulation study, where Gaussian simulated returns are disturbed by shocks resulting from various non-Gaussian distribution functions. The inferred bias is assessed relative to its impact on the nature of the initially normally distributed returns, its impact on performance ranking and, finally, its link with the corresponding skewness and kurtosis of disturbed returns. A fundamental Sharpe ratio (i.e. unbiased Sharpe ratio) is also proposed as a performance measure, which is free of skewness bias.

1. Data simulation and disturbances

We introduce here our simulated data set as well as some corresponding empirical features. We start from a stable sample of Gaussian returns to which we apply more or less destabilizing shocks.

We would like to thank participants at the 6th AFE conference (July 2009, Samos, Greece) among which Prof. Ryszard Doman and participants at the QMF conference (December 2010, Sydney, Australia) for their interesting questions and comments. We also thank anonymous journal referee(s).

[1] This author discusses the significance of asymmetry in asset returns through skewness and kurtosis features. A very brief simulation analysis attempts to quantify the bias of one outlier on Sharpe ratio estimates. The incurred bias depends on the investment horizon under consideration, the data frequency and the magnitude of the outlier return's deviation from the average return level.

1.1. Gaussian data simulation. We consider daily data over a four-year investment horizon, namely 1008 observations per series of data (i.e. four years of working days, where one year corresponds to 252 days). The length of the time horizon is supported by the minimum sample size ensuring consistent representations of specific probability distributions (i.e. statistical significance so as to avoid small sample biases). Over such an investment horizon, we simulate a series of twelve Gaussian returns, whose annual average values range from 3% to 7% whereas related standard deviations range from 10% to 80% (see Table 1). This way, we can consider the significance of volatility across various average "historical" return levels.

Table 1. Simulated Gaussian returns' series

Series name	N3-10	N3-20	N3-50	N3-80	N5-10	N5-20	N5-50	N5-80	N7-10	N7-20	N7-50	N7-80
Mean (%)	3	3	3	3	5	5	5	5	7	7	7	7
Standard deviation (%)	10	20	50	80	10	20	50	80	10	20	50	80

For our performance analysis prospect, we also simulate the risk free rate (i.e. interest rate benchmark) over the same time horizon. We simulate it as a generalized Brownian motion (i.e. diffusion process) exhibiting a starting value of 3.5% on an annual basis, an annual trend coefficient of 3.5% and an annual diffusion coefficient (i.e. standard deviation) of 30% (see Figure 1 and Table 2). Basically, the diffusion coefficient represents the standard deviation of daily return changes.

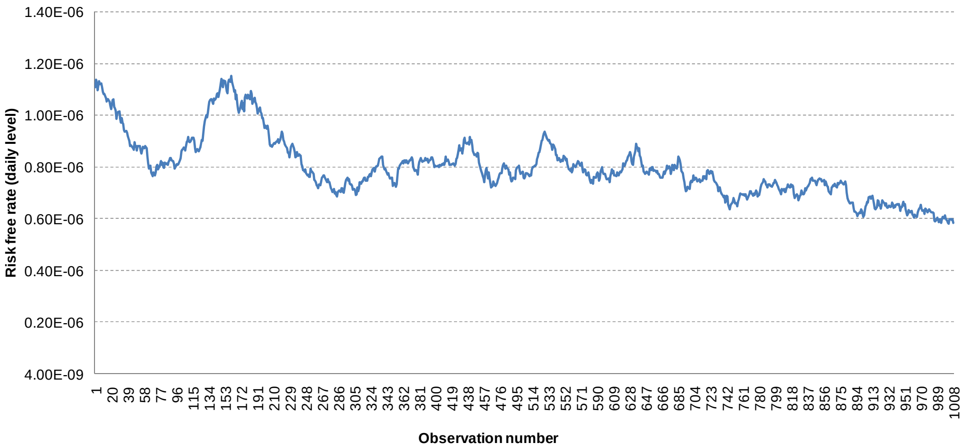

Fig. 1. Daily risk free rate simulation

Table 2. Statistics of the daily simulated risk free rate r

Mean (daily)	7.9915E-05
Standard deviation (daily)	1.1818E-05
Mean (yearly, %)	2.8770E+00
Standard deviation (yearly, %)	2.2422E-02
Median	7.8269E-05
Skewness	8.8902E-01
Excess kurtosis	7.3310E-01
Minimum	5.8304E-05
Maximum	1.1552E-04
Range = Maximum – Minimum	5.7217E-05

As a rough guide, we apply a classic performance measure to simulated returns, namely the Sharpe ratio (*SR*). The Sharpe index expresses as follows for any portfolio *P*:

$$SR_P = (\overline{R}_P - \overline{r}) / s_P, \tag{1}$$

where $\overline{R}_P = E[R_P] = \sum_{t=1}^{1008} R_t^P / 1008$ (i.e. arithmetic mean of P portfolio's returns), $\overline{r} = \sum_{t=1}^{1008} r_t / 1008$ (i.e. arithmetic mean of the simulated risk free rate), and $s_P^2 = Var(R_P)$ (i.e. unbiased estimator of the standard deviation or unbiased volatility of portfolio's returns) over our studied investment horizon (i.e. for time $t \in \{1,\ldots,1008\}$). Sharpe ratio measures the portfolio's market risk premium per unit of total risk (as measured by standard deviation). Diversified and undiversified portfolios are often assessed with such a risk-adjusted performance measure. Computing corresponding Sharpe ratios allows for assessing the risk/reward profile of our simulated returns. Namely, checking for the consistency of the reward brought in by the returns relative to the risk borne allows for ranking returns accordingly (i.e. performance ranking inferred from risk-adjusted performance measures). We display in Table 3 the levels of the obtained Sharpe ratios and the inferred ranking (while ordering the estimated ratios from the highest to the lowest).

Table 3. Daily Sharpe ratios and related performance ranking

Statistics	N3-10	N3-20	N3-50	N3-80	N5-10	N5-20	N5-50	N5-80	N7-10	N7-20	N7-50	N7-80
SR	0.0062	0.0031	0.0012	0.0008	0.0188	0.0094	0.0038	0.0024	0.0314	0.0157	0.0063	0.0039
Ranking	6	9	11	12	2	4	8	10	1	3	5	7

The best performing asset is the one whose return follows a Gaussian distribution with a 7% annual mean and a 10% annual standard deviation (i.e. highest average return and lowest standard deviation along with the mean-variance efficiency principle). The worst asset return follows a Gaussian distribution with a 3% annual mean and an 80% annual standard deviation (i.e. lowest average return and highest volatility).

1.2. Simulating disturbances. We introduce now various types of shocks aimed at distorting the previous returns' evolutions. Such distortions are applied both as positive shocks (i.e. returns' increase) and as negative shocks (i.e. returns' decrease) to the studied returns. Namely, we have:

$$\textit{Distorted return at day } t = \textit{Gaussian return at day } t + [\mathrm{c} \times \textit{shock at day } t], \tag{2}$$

$$\textit{Distorted return at day } t = \textit{Gaussian return at day } t - [c \times \textit{shock at day } t], \tag{3}$$

where c is a constant aimed at scaling the applied shock so as to maintain coherent return levels (i.e. neither too small nor too high values). This way, we can study the asymmetric response of returns to positive and negative shocks. The disturbances we consider follow different non-Gaussian probability distributions and are split into two categories. The first category refers to presumed homogeneous shocks so that they imply no structural change in terms of returns' distributional properties. Differently, the second category refers to presumed[1] inhomogeneous shocks so as to deviate returns from their initial Gaussian property (see Table 4). The chosen disturbances are applied all as positive and negative shocks unless mentioned.

Table 4. Simulated non-Gaussian disturbances

Type	Name	Density	Parameter values	c
Homogeneous	Gamma	$f(x)=[a^p/\Gamma(p)]\, x^{p-1} \exp\{-ax\}$, $x > 0$, scale $a > 0$, shape $p > 0$	$a = 0.12$ $p = 0.0055$	1
	Log-normal	$f(x)=x^{-1}\,(2\pi\sigma^2)^{-0.5} \exp\{-[\ln(x)-\mu]^2/(2\sigma^2)\}$, $x > 0$, $\sigma > 0$	$\mu = -7.3$ $\sigma = 0.01$	1
	Student	$f(x)= [(n\pi)^{0.5}\, \beta(1/2,n/2)]^{-1}\,(1+x^2/n)^{-(n+1)/2}$, x real, n positive integer	$n = 80$	1/1000
Inhomogeneous	Log-normal	$f(x)=x^{-1}\,(2\pi\sigma^2)^{-0.5} \exp\{-[\ln(x)-\mu]^2/(2\sigma^2)\}$, $x > 0$, $\sigma > 0$	$\mu = -9$ $\sigma = 2$	1
	Poisson	$f(x)= \lambda^x \exp(-\lambda)/x!$, x positive integer, $\lambda > 0$	λ in [0,1]***	1/250
	Log-logistic*	$f(x)= (\beta/\alpha)\,(x/\alpha)^{\beta-1}\,[1 + (x/\alpha)^{\beta}]^{-2}$, $x \geq 0$, scale $\alpha > 0$, shape $\beta > 0$	$\alpha = 1.1792$ $\beta = 99.7306$	d/100
	Weibull**	$f(x)=(\alpha/\beta)\,[x/\beta]^{\alpha-1} \exp\{-[x/\beta]^{\alpha}\}$, $x > 0$, shape $\alpha > 0$, scale $\beta > 0$	$\alpha = 1$ $\beta = 0.01$	1/120

Notes: * Applied as a negative shock only, and scaled by a random factor d between 0.1 and 1.1 bounds. ** Applied as a positive shock only. *** Random value between 0 and 1 issued from simulations.

In the previous Table, we also need to specify the following beta and gamma functions:

$$\Gamma(u) = \int_0^u s^{u-1} e^{-s}\, ds, \tag{4}$$

$$\beta(1/2, n/2) = \frac{\Gamma(1/2) \times \Gamma(n/2)}{\Gamma\left(\dfrac{n+1}{2}\right)}. \tag{5}$$

All the disturbances under consideration allow for introducing skewness and kurtosis patterns in the Gaussian-based return distributions to some extent. Sometimes the induced skewness and kurtosis effects are not statistically significant (e.g., homogeneous shocks) insofar as they imply no structural change in the Gaussian distribution. But, sometimes skewness and kurtosis impacts generate deviations from normality for the returns under consideration so that they structurally change their corresponding risk/return profiles. The application of positive and negative shocks to a set of Gaussian returns allows for considering the various and asymmetric reactions of different groups of stocks in the market over time (e.g., sector-specific trends, under- and over-reaction to shocks).

[1] We apply the same type of shocks to all Gaussian distributions under consideration. Therefore, due to their mean-variance properties, some returns may be highly distorted by the applied shocks, whereas other returns may not be structurally affected.

2. Simulation results and disturbances' impact

We summarize the impact of the considered disturbances on our Gaussian returns as a function of the induced return skewness and kurtosis. The trade-off between the two higher above-mentioned moments explains the potential structural changes generated in original Gaussian returns (i.e. possible deviation from normality due to statistically significant distortions over time).

2.1. Homogeneous shocks. As a first step, we report the average values of the descriptive statistics characterizing the new disturbed returns (see Table 5).

Table 5. Descriptive statistics of disturbed Gaussian returns (daily basis)

Type of shock	Negative shocks				Positive shocks			
	Mean (%)	Std. dev. (%)	Skewness	Excess kurtosis	Mean (%)	Std. dev. (%)	Skewness	Excess kurtosis
Gamma	-0.0332	1.9165	-0.0437	0.2310	0.0606	1.9165	0.0407	0.0464
Log-normal	-0.0339	1.9165	0.0237	0.0363	0.0607	1.9165	0.0237	0.0362
Student	-0.0421	1.9165	0.0298	0.0326	0.0700	1.9165	0.0211	0.0370

On one side, negative disturbances introduce skewness and kurtosis effects. Strikingly, the disturbed returns' averages are highly decreased (between 2 and 7 times their initial levels), whereas related standard deviations remain approximately the same. On the other side, positive disturbances also introduce skewness and kurtosis patterns. Strikingly, the disturbed returns' averages are highly increased (between 2 and 7 times their initial levels), whereas related standard deviations remain generally approximately the same. As a further investigation, we check for structural changes in the general Gaussian behavior of initial return series. For this purpose, we perform a normal distribution test while computing successively the Jarque-Bera (JB), Cramer-Von-Mises (CVM) and Aderson-Darling (AD) test statistics[1]. We display in Table 6 below the results corresponding to a 5% normality test for each statistic under consideration. Disturbed returns exhibit generally a Gaussian behavior though having very different average levels as compared to their original counterparts.

Consequently, the homogenous shocks under consideration do not generate structural changes in the original Gaussian return distribution.

Table 6. Percentage of return series satisfying normality tests

Type of shock	Negative shocks			Positive shocks		
	JB	CVM	AD	JB	CVM	AD
Gamma	75.0000	100.0000	83.3333	83.3333	100.0000	91.6667
Log-normal	91.6667	100.0000	91.6667	91.6667	100.0000	91.6667
Student	91.6667	100.0000	91.6667	91.6667	100.0000	91.6667

2.2. Inhomogeneous shocks. As the previous subsection, we display the average values of the descriptive statistics characterizing the new disturbed returns (see Table 7).

Table7. Descriptive statistics of disturbed Gaussian returns (daily basis)

Type of shock	Negative shocks				Positive shocks			
	Mean (%)	Std. dev. (%)	Skewness	Excess kurtosis	Mean (%)	Std. dev. (%)	Skewness	Excess kurtosis
Log-normal	-0.0505	1.9644	-0.9321	10.4464	0.0791	1.9788	1.3399	19.3100
Poisson	-0.1317	3.3635	-7.4755	132.5174	0.1886	3.6702	9.6854	194.3143
Log-logistic /Weibull	-0.3032	4.4751	-4.4977	234.3234	0.3130	1.9740	0.3152	1.3198

On one side, negative disturbances introduce statistically significant skewness and kurtosis effects. Strikingly, the disturbed returns' averages are extremely decreased (until 63 times their initial levels), whereas related standard deviations are generally impacted to some extent. On the other side, positive disturbances also introduce non-negligible skewness and kurtosis patterns. Strikingly, the disturbed returns' averages are very highly increased (between 2 and 38 times their initial levels), whereas related standard deviations are altered in a non-negligible way. As a further investigation, we check for structural changes in the general Gaussian behavior of initial return series. For this purpose, we perform a normal distribution test while computing successively the Jarque-Bera (JB), Cramer-Von-Mises (CVM) and Anderson-Darling (AD) test statistics. We display in Table 8 below the results corresponding to a 5% normality test for each statistic under consideration. Disturbed returns exhibit generally a non-Gaussian behavior due to the combined effects on average returns, standard deviation, skewness and kurtosis statistics as compared to their original counterparts.

---

[1] We compute the adjusted statistic values, which account for the finite sample property of our return series.

Table 8. Percentage of return series satisfying normality tests

Type of shock	Negative shocks			Positive shocks		
	JB	CVM	AD	JB	CVM	AD
Log-normal	33.3333	75.0000	50.0000	33.3333	66.6667	41.6667
Poisson	33.3333	33.3333	16.6667	8.3333	25.0000	16.6667
Log-logistic /Weibull	8.3333	16.6667	16.6667	50.0000	75.0000	50.0000

Consequently, the inhomogeneous shocks under consideration generate significant structural changes in the original Gaussian return distribution.

3. Impact of disturbances on performance assessment

We assess the impact of the various disturbances under consideration over the performance measure and the implied performance rankings. The significance of skewness and kurtosis as underlined by existing outlier returns, is also emphasized.

3.1. Impact on performance rankings. We focus on the modification of the initial Sharpe ratios implied by the application of specific disturbances to the original Gaussian return series. The Tables 9 and 10 below display the obtained average daily Sharpe ratios as well as the implied new performance ranking, and the number of similar Sharpe-based ranks relative to the initial Gaussian setting.

Table 9. Average Sharpe ratios and rankings implied by homogeneous shocks

	Negative shock		Positive shock	
Shock	Daily SR	NRS*	Daily SR	NRS*
Gamma	-0.0398	1	0.0550	3
Log-normal	-0.0403	1	0.0575	4
Student	-0.0474	2	0.0673	4

Note: * Number of ranking similarities relative to the initial Gaussian reference setting.

As regards homogeneous shocks, negatively disturbed returns exhibit lower Sharpe ratios, whereas positively disturbed returns exhibit higher Sharpe ratios under skewness and kurtosis patterns now. Rankings are structurally changed insofar as only 20.8333% of the obtained performance ranks remain similar to the original Gaussian-based setting.

Table 10. Average Sharpe ratios and rankings implied by inhomogeneous shocks

	Negative shock		Positive shock	
Shock	Daily SR	NRS*	Daily SR	NRS*
Gamma	-0.0497	1	0.0661	4
Log-normal	-0.0904	0	0.1149	2
Log-logistic /Weibull	-0.0761	0	0.2697	2

Note: * Number of ranking similarities relative to the initial Gaussian reference setting.

As regards inhomogeneous shocks, negatively distorted returns exhibit lower Sharpe ratios, whereas positively distorted returns exhibit higher Sharpe ratios with skewness and kurtosis features now. Rankings also highlight a structural change since only 12.5000% of the obtained performance ranks remain similar to the original Gaussian-based setting.

3.2. Significance of higher moments. Given the structural changes implied by the disturbances under consideration, we attempt to quantify the performance assessment and ranking bias in the light of the generated skewness and kurtosis in returns (e.g., outliers' impact). For this prospect, we consider all the available statistics computed previously for our 108 return time series. We build our cross section series so as to consider successively negative and positive homogeneous shocks to returns (i.e. 72 series of homogeneously disturbed returns), and then negative and positive heterogeneous shocks to returns (i.e. 72 series of heterogeneously distorted returns). This way we are able to perform a cross section analysis of return performance across observed statistics.

As a first step, we consider the descriptive statistics describing the obtained mean (MU), standard deviation (SIGMA), skewness (S), excess kurtosis (EK), Sharpe ratio (SR) and ranking (RANK) estimates across our 144 disturbed return series (see Table 11)[1]. Such estimates do exhibit a heterogeneous behavior as represented by the prevailing lag between corresponding mean and median values as well as the related skewness and kurtosis patterns. Moreover, the Jarque-Bera test statistic rejects normality at a 5% level for all the variables under consideration. Specifically, the cross sectional non-Gaussian behavior of the obtained Sharpe ratios underlines the existing bias generated by the non-Gaussian behavior of the disturbed return series (Lo, 2002).

[1] Simulated return series are ordered as in Table 1 for negative shocks (12 negatively disturbed return series) and then for positive shocks (12 positively disturbed return series). This clustering is first applied to each 3 types of homogeneous shock (72 homogeneously disturbed return series) and then to each 3 types of inhomogeneous shock (72 non-homogeneously disturbed return series).

Table 11. Descriptive statistics of cross section statistic estimates

	MU	SIGMA	S	EK	SR	RANK
Mean	0.0002	0.0311	-0.0244	49.3876	0.0238	6.5000
Median	0.0005	0.0315	-0.0049	0.2084	0.0136	6.5000
Std. dev.	0.0024	0.0234	6.1144	129.4086	0.1262	3.4641
Skewness	-0.9183	1.3313	-0.2497	3.3043	1.6524	0.0000
Excess kurtosis	4.3786	3.7988	6.6901	11.4604	5.2099	-1.2168
Jarque-Bera	135.2729	129.1213	270.0378	1050.0920	228.3836	8.8834

As a second step, we run a preliminary analysis while investigating the potential explanatory factors of the bias in Sharpe ratios supported by the following facts. Under the Gaussian setting, Sharpe ratio is a good performance descriptor (i.e. convenient descriptor of the risk/reward profile of asset returns) since the two first distributional moments conform the mean-variance efficiency principle. Under a non-Gaussian setting, the Sharpe ratio becomes, however, a biased indicator and asset performance depends on higher moments such as skewness (e.g., tail existence) and kurtosis (e.g., tail fatness). Performing assets are expected to exhibit an excess risk premium with either a skewness as high as possible, or a skewness as low as possible in magnitude when the risk premium is left-skewed. Moreover, the Sharpe ratio is a function of both the observed historical mean and standard deviation of returns. In the same way, skewness and kurtosis statistics can also be considered as functions of the historical mean and standard deviation. Hence, the common dependence on the mean and standard deviation of returns is so that finally Sharpe ratio can be implicitly linked to skewness and kurtosis to some extent. We check for this pattern while considering the non-parametric correlation coefficients (i.e., Spearman and Kendall coefficients) of cross-sectional Sharpe ratios and related cross-sectional skewness and excess kurtosis (see Table 12).

Table 12. Correlation of cross-sectional Sharpe ratios with other cross-sectional statistics

	Kendall	Spearman
MU	0.7055*	0.8812*
SIGMA	-0.0707	-0.0955
S	0.4005*	0.5848*
EK	-0.0602	-0.0938
Rank	-0.3562*	-0.4214*

Note: * Significant at a 1% two-tailed test level.

The previous table just emphasizes that only skewness is an important factor for explaining the observed bias in cross sectional Sharpe ratios (i.e. obtained from our distorted return series)[1].

4. Estimating unbiased Sharpe ratios

We apply the previous results telling us that the bias describing Sharpe ratios when they are estimated under a non-Gaussian setting comes from the skewness of observed return times series. Then, we try to infer unbiased Sharpe ratios while removing the observed skewness impact from classic Sharpe ratios.

4.1. Methodology. For this purpose, we assume that the cross-sectional estimated Sharpe ratios (SR_i) are biased measures (i.e. distorted observations) of the unbiased Sharpe ratios (USR_i), namely the latent or unobserved component in estimated Sharpe ratios. Therefore, we need to filter out the information provided by observed measures of cross-sectional Sharpe ratios in order to extract the corresponding embedded and unobserved component. Such a component should constitute a pure performance indicator. For this purpose, we adopt the latent factor approach named Kalman filter method. Specifically, we apply a linear state-space representation of the observed cross sectional Sharpe ratios (Hamilton, 1994a,b; Harvey, 1989; Kalman, 1960; Koopman, Shephard and Doornik, 1999).

Before introducing our model, we explain the reasoning underlying the specification we propose. First, we showed that cross-sectional Sharpe ratios are correlated with their corresponding skewness so that the reported performance assessment bias depends on skewness. So, removing the bias from observed Sharpe ratios requires expressing them as a non-linear function of related skewness[2]. Since non-linearity in asset return is often generated by time-varying volatility, we choose to introduce the

[1] The way skewness and kurtosis statistics are computed make them dependent of each other (i.e. they are linked).

[2] It is easy to show the non-linear link between a return's Sharpe ratio and related skewness. One easy way to understand it is to notice from Sharpe ratio's expression that $\mu = \sigma SR + r$. Indeed, we have:

$$\sigma^3 S = g\left(r, \mu, \sigma, E\left[R^3\right]\right) - \sigma^3 SR^3 + 3(\mu - r)\sigma^2 SR^2 - 3\left[\sigma^3 + \sigma(\mu - r)^2\right] SR,$$

where R is the return random variable, r is the average risk free rate of return, $\mu = E[R]$, $\sigma = \mathrm{Var}[R]$ and

$$g\left(\mu, \sigma, E\left[R^3\right]\right) = E\left[(R - r)^3\right] = E\left[R^3\right] - r^3 - 3\left(\sigma^2 + \mu^2\right) + 3\mu r^2.$$

skewness explanatory factor in the volatility describing the error relative to the estimation of Sharpe ratios' equation. Second, Ferson and Harvey (1991) advocate the significance of economic variables in explaining the predictable time-variation of investment returns. For example, the relevant economic variables correspond to the market risk premium and prevailing interest rates for stock-specific and bond-specific investment returns, respectively. Therefore, investment returns exhibit a common component explaining their risk of change (i.e., time-variation) as advocated by Alexander (2005). Since Sharpe ratios are founded on asset risk premiums and given that risk premiums bear a common systematic component (Sharpe, 1964; Fama and French, 1993), observed cross-sectional Sharpe ratios are all linked to each other (to some extent) through such a systematic component. Consequently, cross-sectional Sharpe ratios (SR_i) are serially dependent. We propose to catch such a dependency while stating a first order Markov property for the unbiased Sharpe ratios (USR_i), which are unobserved. Then, we get the following representation:

$$SR_i = USR_i + u_i \,, \tag{6}$$

$$USR_i = a \times USR_{i-1} + v_i \,, \tag{7}$$

where relations (6) and (7) are called dynamic and state equations, respectively, (u_i) and (v_i) are serially independent and uncorrelated Gaussian white noises with a zero mean (i.e. dynamic and state errors). Moreover, the equation errors are assumed to follow a two-dimension Gaussian variable with a diagonal covariance matrix Ω defined by:

$$\Omega_i = \begin{pmatrix} \exp(b)\sqrt{|S_i|}\exp(-S_i^6) & 0 \\ 0 & |\sin(i)|^3 \exp(c) \end{pmatrix},$$

where a, b and c are constant parameters. Our specification is relevant insofar as classic Sharpe ratios under a Gaussian setting represent a special case of our representation. Indeed, a Gaussian world implies $S = 0$ so that we get a null variance for u_i and then $SR_i = USR_i$ (i.e. no dynamic error since the related mean and variance are zero). Moreover, $\sin(i)$ attempts to capture the firm-specific information that explains differences in Sharpe ratios from one firm to another. Indeed, firm-specific performance information such as size and book-to-market indicators, for example, is also important to explain discrepancies in firm performance (Fama and French, 1992; 1993; 1995).

Assumptions about the joint multivariate distribution of the state and dynamic errors allow for estimating the model parameters along with the log-likelihood maximization principle. Such a setting requires estimating all the parameters previously introduced with also the two moments, describing the two-dimension Gaussian distribution of u_0 and v_0 (i.e. prior guess, which represents the initial values of the Gaussian white noise vector). Consequently, the optimization process requires estimating 5 parameters.

4.2. Estimation and results. The Kalman filter estimation method yields the following results[1] and parameter estimates (see Table 13 and Figure 2). According to Table 13, all the parameters are significant at a 5% test level. In unreported results, we also tested for the null assumption of each of the parameters and all of the parameters at the same time, but the null hypothesis was rejected in all situations at a 5% Wald test level. Consequently, the model is significant as supported by Figure 2. Figure 2 shows cases, where Sharpe ratios either over- or under-estimate returns' performance, depending on skewness and kurtosis patterns[2]. Moreover, our cross-sectional time series is built so as to account for various reactions of different groups of stocks in the market over time[3]. For example, sector-specific stocks react differently to market shocks across industry areas.

Table 13. Model estimation statistics

	Coefficient	Std. error	Z-statistic	Probability
a	0.7127	0.0673	10.5852	0.0000
b	-2.6716	0.1059	-25.2267	0.0000
c	-4.7734	0.3332	-14.3258	0.0000
	Final state	RMSE*	Z-statistic	Probability
USR	0.0068	0.0618	0.1100	0.9124
Log-likelihood	91.8284	Akaike info criterion		-1.2337
Parameters	3	Schwarz criterion		-1.1719
Accuracy level	1.00E-6	Hannan-Quinn criterion		-1.2086

Note: * Root mean squared error. Convergence achieved after 12 iterations. Initial values: a = 0.7895, b = -2.6430, c = -4.7732.

[1] To avoid any sample bias due to the way we built it (e.g., clusters), we randomly reorganized the sample chronology (i.e. resample process). Then, the model estimation is applied to the reorganized sample.

[2] Overestimation takes place when Sharpe ratios are higher than their unbiased counterparts. The converse remark applies to underestimation cases.

[3] We voluntarily gathered the positively and negatively disturbed returns in order to add clustering effects to the representation. However, the statistical computations are achieved on the reorganized sample. Hence, there is no sample bias. Moreover, the estimation process is more complicated but it also allows for testing for the robustness of the proposed methodology. Indeed, the resample process adds harmonic volatility effects in Sharpe ratio estimates.

Fig. 2. Observed (*SR*) and latent unbiased (*USR*) Sharpe ratios

As a back-test, we tested for both the Gaussian nature of model residuals (i.e. Jarque-Bera, Anderson-Darling and Cramer-Von-Mises goodness of fit tests) and a unit root according to a 5% Phillips-Perron test (i.e. stationary errors). Though, they are non-Gaussian, the dynamic and state errors follow a white noise process as emphasized by relevant distributional and Ljung-Box statistics (see Tables 14 and 15) as well as related correlograms (see Figure 3). Such a non-Gaussian pattern is probably due to the fact that some information is missing. Indeed, Sharpe ratios depend on the financial market climate so that there is still a market-dependent bias to extract (Scholz and Wilkens, 2005; Scholz, 2007; Sortino, 2004). But, such feature is beyond the scope of this paper given that we cannot simulate the market climate randomly. Therefore, the powerful proposed method to infer unbiased Sharpe ratios from observed classic Sharpe ratios in a cross-sectional analysis over a given time horizon requires encompassing relevant economic variables (i.e., additional explanatory factor(s)) so as to remove the market climate bias.

Table 14. Distributional statistics of dynamic and state errors

Statistics	Dynamic errors	State errors
Jarque-Bera	501.6644	33.4089
Cramer-Von-Mises	1.6871	1.2290
Anderson-Darling	8.8817	6.0653
Phillips-Perron	-12.3958	-12.3178

Table 15. Autocorrelations of dynamic and state errors

	Dynamic errors				State errors			
Lag	AC*	PAC**	Q-stat***	Probability	AC*	PAC**	Q-stat***	Probability
1	-0.0560	-0.0560	0.4629	0.4960	-0.0340	-0.0340	0.1719	0.6780
2	0.1770	0.1740	5.0774	0.0790	-0.0280	-0.0290	0.2844	0.8670
3	-0.0080	0.0100	5.0865	0.1660	-0.0020	-0.0040	0.2851	0.9630
4	0.0760	0.0470	5.9536	0.2030	-0.0800	-0.0810	1.2395	0.8720
5	-0.0190	-0.0140	6.0081	0.3050	0.0240	0.0180	1.3258	0.9320
6	-0.1330	-0.1610	8.7102	0.1910	0.0410	0.0380	1.5835	0.9540
7	0.0150	0.0060	8.7454	0.2710	0.0310	0.0340	1.7279	0.9730
8	-0.0170	0.0320	8.7906	0.3600	-0.0720	-0.0740	2.5183	0.9610
9	0.0310	0.0340	8.9415	0.4430	0.0690	0.0710	3.2600	0.9530
10	0.0490	0.0760	9.3202	0.5020	-0.0910	-0.0870	4.5715	0.9180
11	0.0630	0.0540	9.9520	0.5350	-0.0450	-0.0440	4.8846	0.9370
12	-0.0120	-0.0550	9.9738	0.6180	-0.1170	-0.1430	7.0474	0.8540
13	0.1050	0.0840	11.7380	0.5490	0.0860	0.0920	8.2227	0.8290
14	-0.0580	-0.0520	12.2850	0.5830	-0.0060	-0.0260	8.2281	0.8770
15	0.0100	-0.0270	12.3030	0.6560	0.1730	0.1880	13.0900	0.5950
16	0.0450	0.0970	12.6360	0.6990	0.0380	0.0180	13.3220	0.6490
17	-0.1260	-0.1310	15.2750	0.5760	-0.0250	0.0430	13.4260	0.7070
18	0.0170	-0.0140	15.3260	0.6390	0.0900	0.0740	14.7800	0.6770
19	-0.0780	-0.0090	16.3520	0.6340	-0.0190	0.0300	14.8400	0.7330
20	0.0620	0.0160	16.9940	0.6530	0.0410	0.0110	15.1300	0.7690

Notes: * Autocorrelations. ** Partial correlations. *** Ljung-Box Q-statistic.

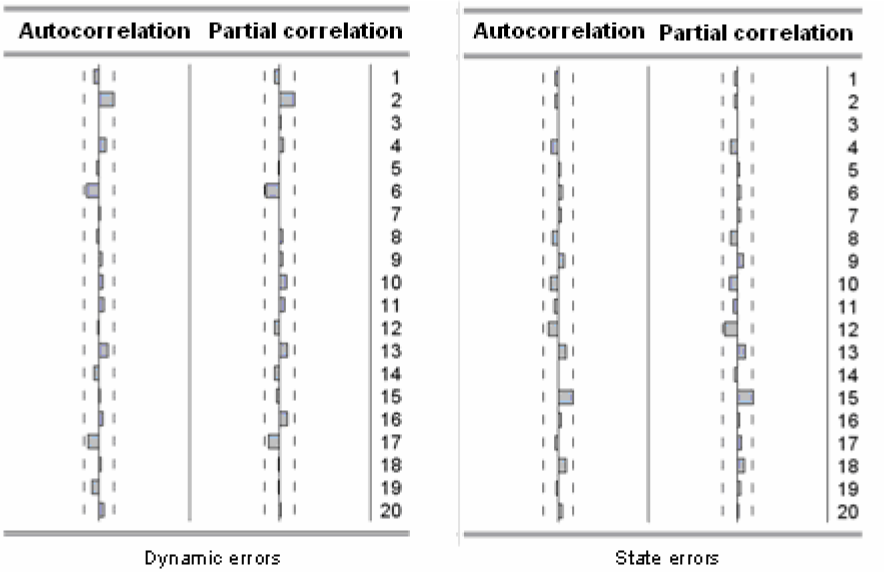

Fig. 3. Correlograms of model residuals

As a rough guide, we also display the variance and volatility of dynamic errors (see Figures 4 and 5)[1], and we report the descriptive statistics relative to the unbiased Sharpe ratios (see Table 16). We notice the asymmetric and heterogeneous nature of such ratios as expected. Moreover, a 5% Jarque-Bera test rejects the normality assumption. Interestingly, the closer to zero skewness magnitude is, the lower the dynamic errors' variance, as expected (see Figure 4). Therefore, the lower dynamic residuals' volatility is the closer observed Sharpe ratios to their unbiased and unobserved counterparts are (i.e. the more unbiased observed Sharpe ratios are since they exhibit a skewness, which is as close as possible to a Gaussian distribution's skewness level).

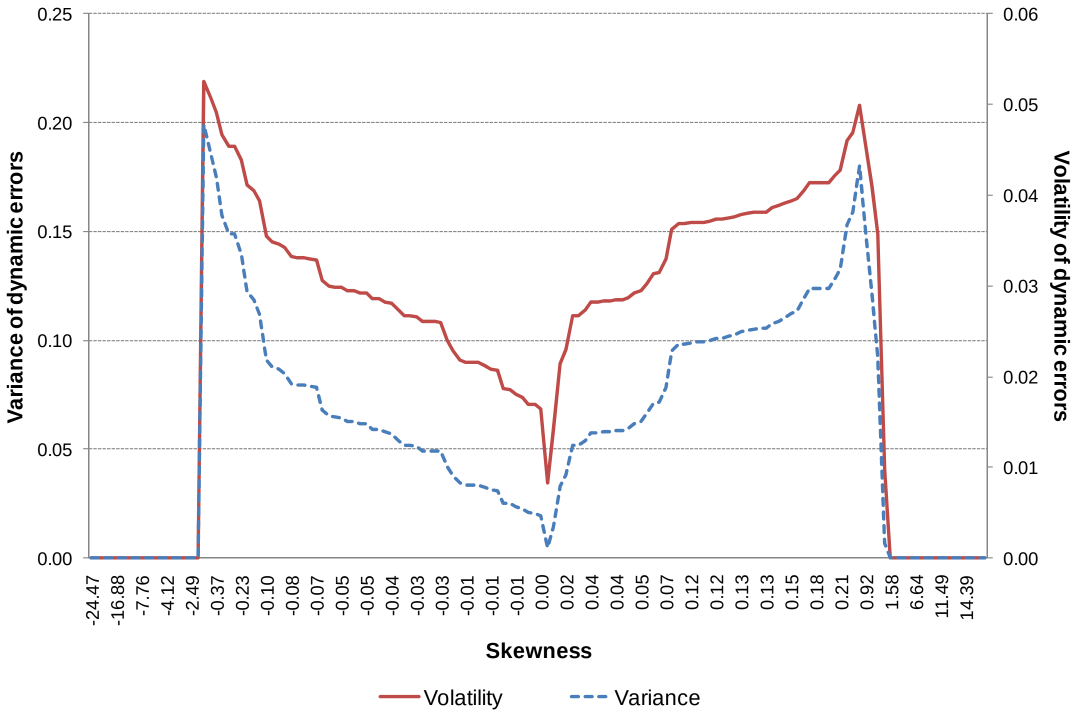

Fig. 4. Ranked variance and volatility of dynamic errors (u_i)

[1] Figure 4 plots variance and volatility as functions of increasing skewness values, whereas Figure 5 plots variance and volatility as functions of increasing observation number. For information prospect, Figure 5 reports also the skewness value corresponding to each observation number.

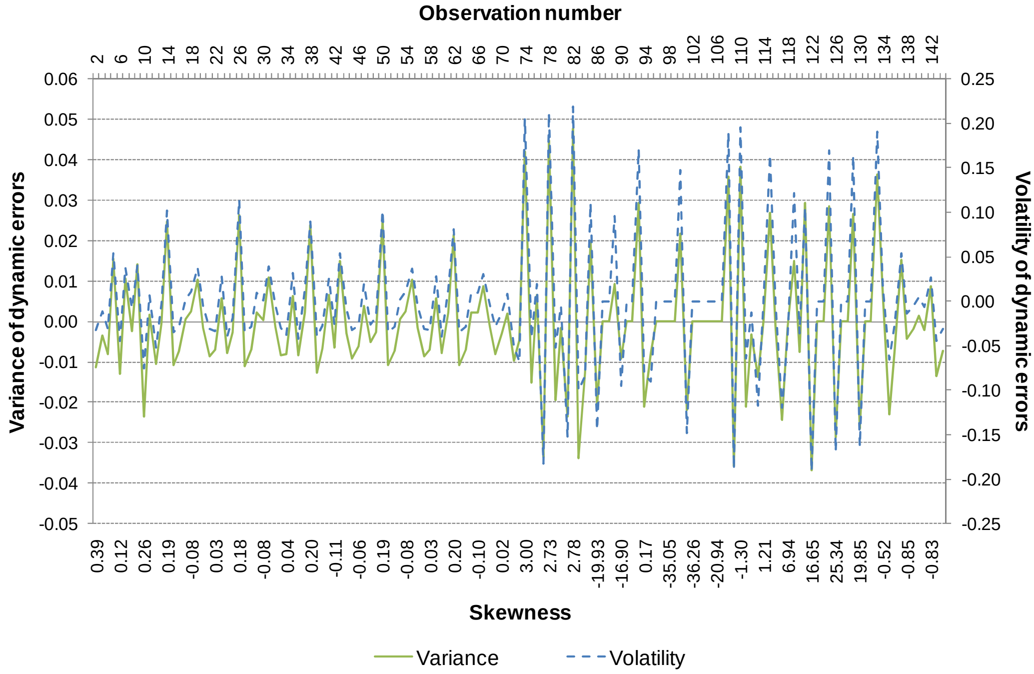

Fig. 5. Variance and volatility of dynamic errors (u_i)

The behavior of the dynamic errors' variance (i.e. volatility, standard deviation) as a function of observed cross section skewness is so that it is non-exploding, and can catch cyclical as well as non- linear patterns. Of course, it is dependent on the nature of the data under consideration. If the proportion of extreme skewness values were high, the shape of the variance would have been more complicated since we look for an average behavior (i.e. general trend over the observed sample). In our present case, 75.6944% of observed cross-sectional skewness values lie between -0.4917 and 1.0237 levels. We implicitly assume that the remaining 24.3056% of observed skewness values exhibit "marginal values" or "extreme levels" (to some extent) that we cannot explain without resorting to some missing relevant explanatory factors. However, we cannot simulate such factors at the moment (e.g., other performance determinants such as market climate, for example). Nonetheless, volatility changes and classic Sharpe ratio changes from one asset return to another exhibit a correlated behavior (see Figure 6). Volatility and Sharpe ratio changes exhibit the same sign (i.e. same trend in changes) in 45.8333% of cases.

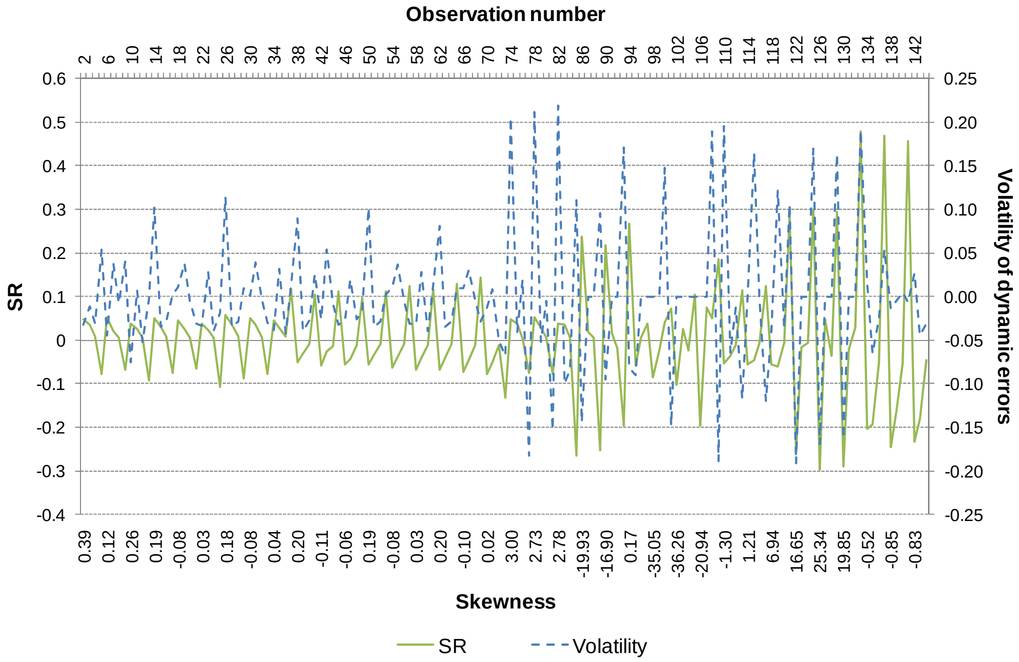

Fig. 6. Changes in Sharpe ratios and dynamic errors' volatility

Table 16. Statistics of the unbiased Sharpe ratios

Mean	0.0065
Standard deviation (%)	0.0546
Median	0.0098
Skewness	-0.5439
Kurtosis	3.7685
Jarque-Bera statistic	10.6441
Probability	0.0049

Finally, we display the obtained unbiased Sharpe ratios (*USR*) for each distorted return and the related performance-based rankings (see Table 17). We also compare the obtained new rankings with the ones related to the biased Sharpe ratios. The unbiased Sharpe ratios yield noticeable structural changes in the rankings based on their biased counterparts.

Table 17. Average unbiased Sharpe ratios and rankings implied by shocks

		Negative shock		Positive shock	
Shock		Daily SR	NRS*	Daily SR	NRS*
Homogeneous	Gamma	0.0198	0	0.00136	0
	Log-normal	0.0026	1	0.02072	1
	Student	-0.0098	0	0.02815	2
Non-homogeneous	Gamma	-0.04968	1	0.06612	4
	Log-normal	-0.09043	0	0.11486	3
	Log-logistic /Weibull	-0.07632	0	0.26974	2

Note: * Number of ranking similarities relative to the biased Sharpe ratios' rankings.

With regard to Tables 6 and 8, homogeneous shocks generally preserve returns' Gaussian property whereas inhomogeneous shocks engender structural changes in returns' distributional behavior. As an extension, Tables 9, 10 and 17 underline the potential structural changes in corresponding performance rankings. Indeed, the ranking stability in between classic Sharpe ratios and unbiased Sharpe ratios depends on the impact on skewness and kurtosis, which results from such shocks, on the one hand. On the other hand, the ranking stability also depends on the engendered moves in average returns' level (i.e. significance of the risk-return profile's alteration). However, a robustness study helps to identify the efficiency of the proposed filtering method and the resulting unbiased Sharpe ratios. Such an analysis answers the following question: Is the *USR* a superior performance measure relative to *SR* while discriminating between winners and poor performing stocks?

Superiority of unbiased Sharpe ratios relative to their biased counterparts.

To answer the previous question, we consider the top and bottom stock return deciles in accordance with classic Sharpe ratios, on one side, and unbiased Sharpe ratios, on the other side. Specifically, we first select the 10% of stocks, which exhibit the highest performance measure across the sample (i.e. a cohort of the top 15 stocks). Then, we select the 10% of stocks, which exhibit the lowest performance measure across the sample (i.e. a cohort of the bottom 15 stocks). We finally analyze the 4 obtained cohorts among which 2 bottom cohorts and 2 top cohorts for classic and unbiased Sharpe ratios, respectively (see Table 18).

Table 18. Robustness of unbiased Sharpe ratios

		Statistics	N+	Mean	Std. dev.	Median	S	EK
Top stocks	*SR*	Average	622	0.0029	0.0101	0.0026	0.3684	1.8364
		Nb > 0	-	15	-	15	12	12
		Nb > Sample mean	-	4	-	-	-	-
	USR	Average	618	0.0034	0.0137	0.0030	0.5842	3.5533
		Nb > 0	-	15	-	15	14	14
		Nb > Sample mean	-	15	-	-	-	-
Bottom stocks	*SR*	Average	458	-0.0024	0.0199	-0.0016	-3.1638	56.1809
		Nb < 0	-	15	-	15	12	4
		Nb > Sample mean	-	10	-	-	-	-
	USR	Average	456	-0.0027	0.0285	-0.0018	-4.5211	86.1777
		Nb < 0	-	14	-	15	13	4
		Nb > Sample mean	-	9	-	-	-	-

Notes: N+ is the number of positive returns. S and EK are the skewness and excess kurtosis statistics, respectively.

With respect to top stocks and compared to *SR*-based selection, more *USR*-based stocks exhibit positive average mean and median returns, on one side. On the other side, the *USR* captures all the stocks whose average returns lie above the mean return of the top 15 stock sample. Moreover, the *USR* captures winners in a better way because returns exhibit higher skewness and excess kurtosis levels on average. Indeed, performers should exhibit positively skewed returns with a positive and high excess kurtosis (i.e. positive fat tail, or equivalently, significant probability of gain). With respect to bottom stocks, the *USR* also captures more performing stocks than *SR* (though, it is less obvious on an average basis). For example, *USR*-based losers exhibit a lower negative skewness with a higher positive excess kurtosis as compared to *SR*-based bottom decile. Indeed, a poor performing stock should exhibit a highly negative skewness with a positive excess kurtosis (i.e. negative fat tail, or equivalently, important risk of loss). As a conclusion, the unbiased Sharpe ratio metric outperforms *SR* metric while identifying and classifying winners and losers.

Conclusions

We first simulated normally distributed asset returns to which we applied a series of positive and negative shocks, resulting from various probability distributions. The applied shocks are split into two categories, namely homogeneous and inhomogeneous shocks. The homogeneous distortions keep the Gaussian nature of returns, whereas the inhomogeneous counterparts induce structural return changes through deviations from normality. Then, we studied the behavior of the obtained disturbed returns and related skewness, and kurtosis features. The existence of skewness and kurtosis generate changes in terms of the performance rankings induced by Sharpe ratio estimates.

Under a non-Gaussian setting, Sharpe ratios as functions of the two first distributional moments become insufficient descriptors of the risk/return profile of investments. The modifications incurred by existing distribution tails and related tail fatness requires considering also skewness and kurtosis patterns. Given the statistical expressions of skewness and kurtosis as distributional moments, these two higher moments can be considered as non-linear and non-explicit functions of the two first distributional moments (i.e. mean and standard deviation). Consequently, usual Sharpe ratios (i.e. classic risk-adjusted performance measures) are indirectly functions of skewness and kurtosis (since they also depend on returns' mean and standard deviation).

Given this feature, we estimated the unbiased Sharpe ratios while filtering the estimated Sharpe ratios, which are biased by skewness. For this prospect, we applied a latent factor approach (i.e., Kalman filter to smooth classic Sharpe ratios), which allows for estimating the fundamental component in Sharpe ratios after removing the bias generated by return skewness. The robustness of the estimation process has been emphasized, and the superiority of the unbiased Sharpe ratio for stock picking prospects has been proved. Usually, the rankings inferred from unbiased Sharpe ratios exhibit structural changes. Based on return stylized facts and available market information, it is possible to infer unbiased performance measures as long as the bias is described by empirical features. Such a finding is very important for forecast prospects and asset allocation schemes (e.g., stock picking). Accessing sufficient economic and market information will then help forecasting the future one-period performance since Kalman filter can also play the role of a forecasting tool.

References

1. Alexander, C. (2005). The present and future of financial risk management. *Journal of Financial Econometrics*, 3 (1), pp. 3-25.
2. Black, K.H. (2006). Improving the hedge fund risk exposures by hedging equity market volatility, or how the VIX strategy ate my kurtosis, *Journal of Trading*, 1 (2), pp. 6-15.
3. Christensen, M.M., and E., Platen (2007). Sharpe ratio maximization and expected utility when asset prices have jumps. *International Journal of Theoretical and Applied Finance*, 10 (8), pp. 1339-1364.
4. Eling, M., and F., Schuhmacher (2007). Does the choice of performance measure influence the evaluation of hedge funds? *Journal of Banking and Finance*, 31 (9), pp. 2632-2647.
5. Eling, M. (2006). Autocorrelation, bias and fat tails: are hedge funds really attractive investments? *Derivatives Use, Trading Regulation*, 12 (1), pp. 28-47.
6. Fama, E.F., and K.R., French (1992). The cross-section of expected stock returns, *Journal of Finance*, 47 (2), pp. 427-465.
7. Fama, E.F., and K.R., French (1993). Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics*, 33 (1), pp. 3-56.
8. Fama, E.F., and K.R., French (1995). Size and book-to-market factors in earnings and returns, *Journal of Finance*, 50 (1), pp. 131-155.
9. Farinelli, S., Ferreira, M., Rossello, D., Thoeny, M., and L., Tibiletti (2008). Beyond Sharpe ratio: optimal asset allocation using different performance ratios, *Journal of Banking and Finance*, 32 (10), pp. 2057-2063.
10. Ferson, W.E., and C.R., Harvey (1991). The variation of economic risk premiums, *Journal of Political Economy*, 99 (2), pp. 385-415.
11. Gatfaoui, H. (2009). A correction for classic performance measures. eastern finance association annual meeting (EFA), Washington D.C. (29th April- 2nd May).
12. Harvey, C.R., Liechty, J.C., Liechty, M.W., and P. Müller (2004). Portfolio selection with higher moments, Working Paper, Fuqua School of Business, Duke University.

13. Harvey, C.R., and A., Siddique (2000). Conditional skewness in asset pricing tests, *Journal of Finance*, 55 (3), pp. 1263-1296.
14. Hodges, S.D. (1998). A generalization of the Sharpe ratio and its applications to valuation bounds and risk measures. Working Paper, FORC Preprint 98/88, Financial Options Research Centre, University of Warwick.
15. Hamilton, J.D. (1994a). Time series analysis. Princeton: Princeton University Press.
16. Hamilton, J.D. (1994b). State space models. In Robert F. Engle and Daniel L. McFadden (eds.), *Handbook of Econometrics*, Volume 4, Chapter 50. Amsterdam: Elsevier Science B.V.
17. Harvey, A.C. (1989). Forecasting, structural time series models and the Kalman filter, Cambridge: Cambridge University Press.
18. Kalman, R.E. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering, Transactions of the ASME – Series D*, 82(1), pp. 34-45.
19. Klemkosky, R. (1973). The bias in composite performance measures, *Journal of Financial and Quantitative Analysis*, 8 (3), pp. 505-514.
20. Koopman, S.J., Shephard, N. and J.A., Doornik. (1999). Statistical algorithms for models in state space using SsfPack 2.2, *Econometrics Journal*, 2(1), pp. 107-160.
21. Lien, D. (2002). A note on the relationships between some risk-adjusted performance measures. *Journal of Futures Markets*, 22 (5), pp. 483-495.
22. Lo, A.W. (2002). The statistics of Sharpe ratios, *Financial Analysts Journal*, July/August, pp. 36-52.
23. Madan, D.B., and G.S., McPhail. (2000). Investing in skews, *Journal of Risk Finance*, 2 (1), pp. 10-18.
24. Robertson, K. (2001). The Sharpe ratio, *Panel Publishers 401(k) Advisor - Investment CORNER*, May, 3.
25. Scholz, H., and M., Wilkens. (2005). Interpreting Sharpe ratios: the market climate bias, Forthcoming in *Finance Letters*.
26. Scholz, H., and M., Wilkens. (2005). An investor-specific performance measurement: a justification of Sharpe ratio and Treynor ratio, *International Journal of Finance*, 17 (4), pp. 3671-3691.
27. Scholz, H. (2007). Refinements to the Sharpe ratio: comparing alternatives for bear markets, *Journal of Asset Management*, 7 (5), pp. 347-357.
28. Sharpe, W.F. (1994). The Sharpe ratio, *Journal of Portfolio Management*, 21 (1), pp. 49-58.
29. Sharpe, W.F. (1964). Capital asset prices: a theory of market equilibrium under conditions of risk, *Journal of Finance*, 19 (3), pp. 425-442.
30. Sortino, F. (2004). Historic performance measures distort current risk assessment, *Pensions & Investments*, 32 (3), pp. 20-21.
31. Spurgin, R. (2001). How to game your Sharpe ratio, *Journal of Alternative Investments*, 4 (3), pp. 38-46.
32. Taleb, N. (2007). The black swan: the impact of the highly improbable, Penguin.
33. Ziemba, W.T. (2005). The symmetric downside-risk Sharpe ratio, *Journal of Portfolio Management*, 32 (1), pp. 108-122.