Risk/Performance Measure Standard Error Vignette

Xin Chen and Douglas R. Martin

August 22, 2016

Abstract

This document introduces the new R package EstimatorStandardError. This package is a result of the Google Summer of Code 2016 (GSoC) project Standard Error of Risk and Performance Measures for Non-Normal and Serially Correlated Asset Returns. In this document, a brief introduction of the theoretical background that the authors consider essential is provided. Then the structure of the package is explained to facilitate further expansion of the package. Last but not least, examples of how to take full advantage of what the package has to offer are shown.

1 Theoretical Background

1.1 Risk/Performance Measures and their Nonparametric Sample Estimators

We define Risk/Performance Measures as a functional $\rho(F)$ of the return distribution function $F(r)$. See the following table for the risk/performance measures we cover in this vignette.

Mean	$\mu(F) = \int_{-\infty}^{\infty} r dF(r)$
Standard Deviation	$\sigma(F) = \sqrt{\int_{-\infty}^{\infty} (r - \mu(F))^2 dF(r)}$
Sharpe Ratio	$SR(F) = \frac{\mu(F)-r_f}{\sigma(F)}$
Value-at-Risk	$VaR(F;\beta) = -q_\beta(F) = -F^{-1}(\beta)$
Expected Shortfall	$ES(F;\beta) = -\frac{1}{\beta}\int_{-\infty}^{q_\beta(F)} r dF(r)$
Sortino Ratio	$SoR(F;MAR) = \frac{\mu(F)-MAR}{SSD(F;MAR)}$
STARR Ratio	$STARR(F;\beta) = \frac{\mu(F)-r_f}{ES(F;\beta)}$

In reality, the true distribution F is unknown so the true functional ρ cannot be computed. There are two general ways of forming estimators of the functional ρ, parametric and non-parametric. For discussion see Martin and Zhang (2015) SSRN reference number. Here we take the non-parametric approach obtained by replacing F by the empirical distribution function F_n that puts mass $\frac{1}{n}$ at each return observation. This results in the following sample estimators of the risk/performance measures $\rho(F)$.

Sample Mean	$\hat{\mu}_n(F_n) = \int_{-\infty}^{\infty} r dF_n(r)$
Sample SD	$\hat{\sigma}_n(F_n) = \sqrt{\int_{-\infty}^{\infty} (r - \hat{\mu}_n)^2 dF_n(r)}$
Sample SR	$\widehat{SR}_n(F_n) = \frac{\hat{\mu} - r_f}{\hat{\sigma}}$
Sample VaR	$\widehat{VaR}_n(F_n) = -\hat{q}_\beta = -F_n^{-1}(\beta)$
Sample ES	$\widehat{ES}_n(F_n; \beta) = -\frac{1}{\beta} \int_{-\infty}^{\hat{q}_\beta} r dF_n(r)$
Sample SoR	$\widehat{SoR}_n(F_n; MAR) = \frac{\hat{\mu}_n - MAR}{Semi\widehat{Deviation}_n(F_n; MAR)}$
Sample STARR	$S\widehat{TARR}_n(F_n) = \frac{\hat{\mu}_n - r_f}{\widehat{ES}_n}$

The sample estimators are themselves random variables therefore have standard errors. The EstimatorStandard Error package is designed to compute the standard error of these estimators for the case of independent and identically distributed (i.i.d.) returns and for for serially correlated returns.

1.2 Influence Functions

Let $T(F)$ be a functional of CDFs and define the perturbation of F as $F_\gamma = (1 - \gamma)F + \gamma\delta_r$, where δ_r is the probability measure that puts mass 1 at point r. The influence function of T is given by (Hampel 1986)

$$
\begin{aligned}
IF(r; T, F) &= \lim_{\gamma \to 0} \frac{T(F_\gamma) - T(F)}{\gamma} \\
&= \frac{d}{d\gamma} T(F_\gamma)\Big|_{\gamma=0}
\end{aligned}
$$

Where δ_r is the probability measure that put mass 1 at point r This means that the influence function is the Gâteaux derivative of the functional T in the direction of δ_r evaluated at F. A basic property of the influence function is that

$$\mathbb{E} IF(R; T, F) = 0$$

Table 1 lists the influence functions of the risk/performance measures covered in this package and Figure 1 shows the plots of the influence functions.

Mean	$r - \mu(F)$
Standard Deviation	$(r - \mu(F))^2 - \sigma^2(F)$
Sharpe Ratio	$\frac{r-\mu(F)}{\sigma(F)} - \frac{\mu(F)}{2\sigma^3(F)}((r - \mu(F))^2 - \sigma^2(F))$
Value-at-Risk	$\frac{\mathbb{1}(r \leq -VaR_\alpha(F)) - \alpha}{f(-VaR_\alpha(F))}$
Expected Shortfall	$\frac{-r-VaR_\alpha(F)}{\alpha}\mathbb{1}(r \leq VaR_\alpha(F)) + VaR_\alpha(F) - ES_\alpha(F)$
Sortino Ratio	$-\frac{SoR(F)}{2\sigma_-^2(F)}(r - MAR)^2\mathbb{1}(r \leq MAR) + \frac{r-\mu(F)}{\sigma_-(F)} + \frac{SoR(F)}{2}$
STARR	$\frac{r-\mu(F)}{ES_\alpha(F)} - \frac{STARR(F)}{ES_\alpha(F)}\left[\frac{-r-VaR_\alpha(F)}{\alpha}\mathbb{1}(r \leq VaR_\alpha(F)) + VaR_\alpha(F) - ES_\alpha(F)\right]$

Table 1: Influence Functions

Figure 1 shows the plots of the influence functions.

1.3 Standard Error of Nonparametric Sample Estimators via Influence Functions

Filipova (1962) showed that the difference between the estimator $T(F_n)$ and its asymptotic value $T(F_\theta)$ can be expressed as the following linear combination of influence functions of the returns at each time

$$\sqrt{n}(T(F_n) - T(F_\theta)) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n} IF(r_t; T, F_\theta) + remainder$$

where the remainder goes to zero as $n \to \infty$ in a probabilistic sense For i.i.d. returns the asymptotic variance is given by

$$V(T(F_n)) = \mathbb{E}\{IF^2(R; T, F)\}$$

For finite samples, the non-parametric estimator of the asymptotic variance is obtained by replacing F with the empirical distributionn function F_n:

$$\hat{V}_n(T(F_n)) = \frac{1}{n}\sum_{t=1}^{n} IF^2(r_t; T, F_n)$$

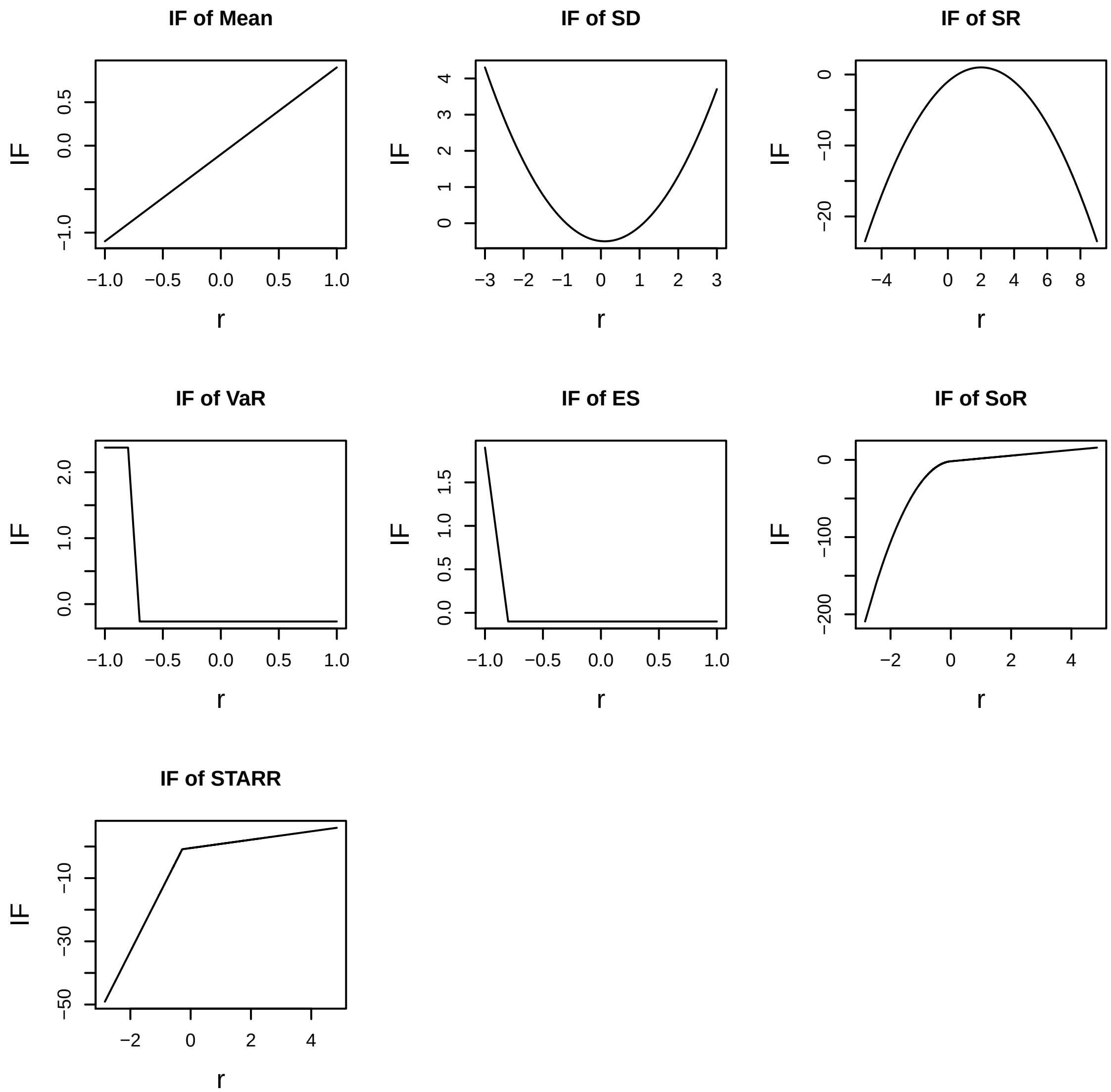

Figure 1: Influence Functions of Risk/Performance Measures

For serially correlated returns of length n, the variance of the sample mean is given by

$$\begin{aligned}\sigma^2(\hat{\mu}) &= \frac{1}{n}\sum_{\ell=-(n-1)}^{n-1}\frac{n-|\ell|}{n}Cov\{IF_1, IF_{1+\ell}\}\\ &= \frac{1}{n}\sum_{\ell=-(n-1)}^{n-1}\frac{n-|\ell|}{n}C(\ell)\end{aligned}$$

As $n \to \infty$, the asymptotic variance is given by

$$\begin{aligned}V(T(F_n)) &= \mathbb{E}\{IF^2(R_1;T,F)\} + 2\sum_{t=2}^{\infty}\mathbb{E}\{IF(R_1;T,F)IF(R_t;T,F)\}\\ &= \mathbb{E}\{IF_1^2(T)\} + 2\sum_{t=2}^{\infty}\mathbb{E}\{IF_1(T)IF_t(T)\}\\ &= \sum_{\ell=-\infty}^{\infty}Cov\{IF_1(T), IF_\ell(T)\}\\ &= \sum_{\ell=-\infty}^{\infty}C(\ell)\end{aligned}$$

The finite sample estimate of this quantity will be discussed in further detail in the following section.

1.4 Computing Standard Errors for Serially Correlated Data using a Frequency Domain Representation and Generalized Linear Model Fitting

The finite sample estimate of the standard error for i.i.d data has already been discussed in the previous section. When the data is serially correlated, one is tempted to simply plug in the sample covariance estimate. Howver, this is not a good idea because it can result in negative estimate for the standard error (Newey and West 1987). In the econometric literature, one kind of solution has been partitularly dominant, which is to use kernel weighted sum of the sample covariances as an approximation of the estimator asymptotic variance, from which one can compute a finite-sample standard error (S.E.) in the usual way by dividing the square root of this variance estimate by the square root of n. This package implements a new and better method of computing standard errors for serially correlated returns based on a frequency domain representation of the variance of the estimator variance with serial correlation. It takes advantage of the following relationship.

- Spectral Density and Autocovariance Functions are a Fourier Transform pair

$$S(f) = \sum_{\ell=-\infty}^{\infty} C(\ell) exp\{-2\pi i f \ell\}$$
$$C(\ell) = \int_{-1/2}^{1/2} S(f) exp\{2\pi i f \ell\} df$$

- Relationship between Spectral Density and Asymptotic Variance

$$S(0) = \sum_{\ell=-\infty}^{\infty} C(\ell)$$
$$= V(T(F_n))$$

This implies that the variance is equal to the spectral density at frequency zero. Therefore, our method consists of roughly three steps.

1. Compute the periodograms as approximations of the spectral density.

2. Fit a polynomial of the periodograms using Generalized Linear Model with Elastic Net regularizatoin.

3. Use the fitted polynomial to estimate the spectral density at frequency zero, which is equal to the variance we are looking for.

Technical details about this method can be found in the Appendix.

2 Structure of the EstimatorStandardError Package

This package implements two types methods of computing the standard error of nonparametric sample estimators. The first type is influence function approach introduced in the previous section. The second type is to use bootstrapping. Wrapper functions are implemented in this package that uses the R boot package function boot() for i.i.d data or tsboot() for serially correlated time series data from the boot package. For details, refer to the documentation of the boot package. The structure of our package is illustrated in Figure 2.

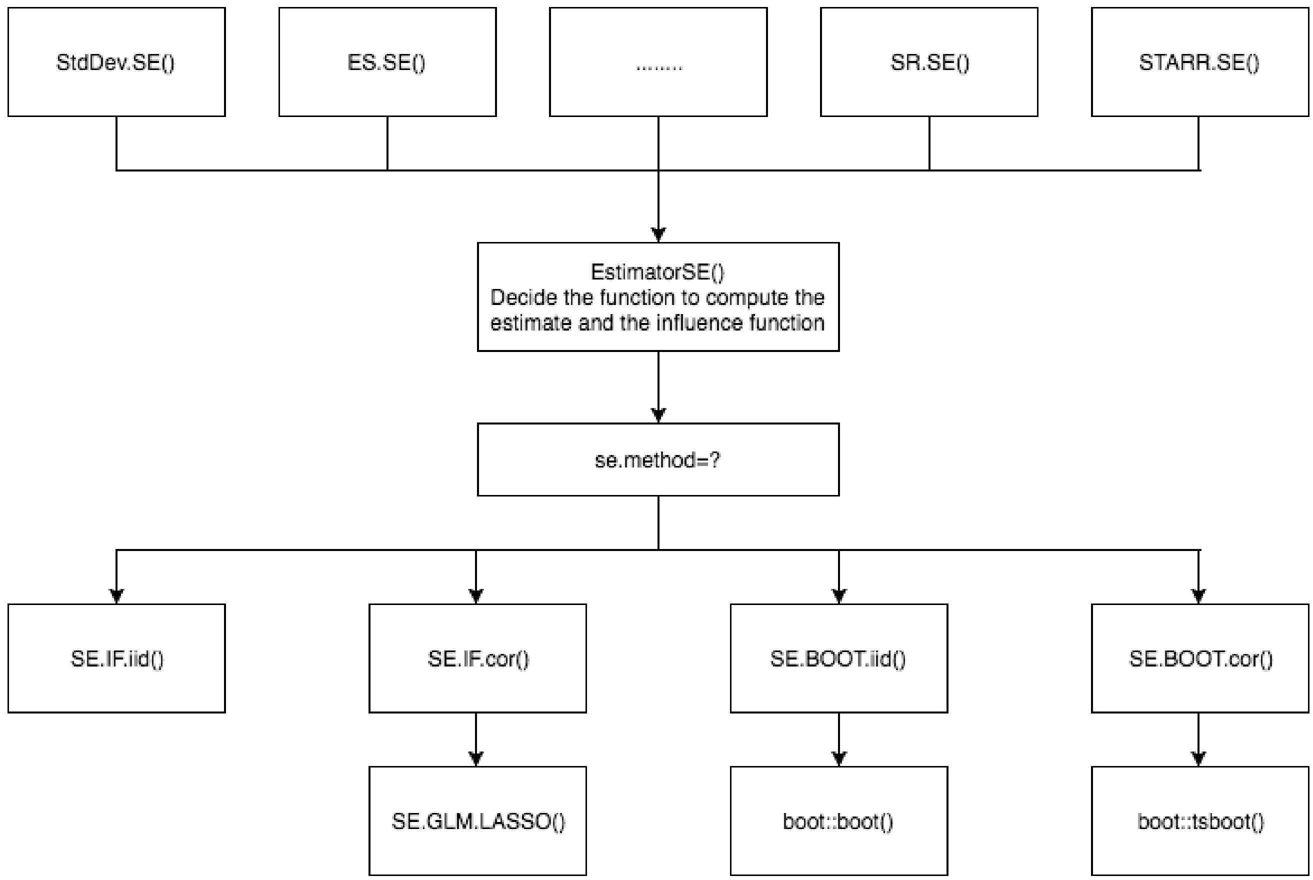

Figure 2: Structure of EstimatorStandardError Package

This structure facilitates future expansion of the package by separating the code associated with specific estimators from the code associated with computing standard errors. To add more estimators, one only need to create three new functions. The first function is the wrapper function xxx.SE(), the second function is the function to compute the estimate and the third fucntion is the influence function xxx.IF(). Then add the new function into the EstimatorSE() function.

3 Using the EstimatorStandardError Package

In this section, we cover how to use the EstimatorStandardError package in practice.

3.1 Installation

3.2 Sample Code

Appendix

We propose the Generalized Linear Model with Elastic Net Regularization (GLM-EN) method for AC Correction

1. Compute the periodograms $\mathbb{I}(n/N)$
2. Build the matrix for indepdent variables

$$R = \begin{bmatrix} 1 & f_1 & \cdots & f_1^d \\ \vdots & \vdots & \vdots & \vdots \\ 1 & f_K & \cdots & f_K^d \end{bmatrix}$$

3. fit GLM-EN model with input matrix R and response vector $\mathbb{I}(n/N)$
4. Predict the value of the GLM-EN model at frequency 0 to get $\hat{S}(0)$

The Generalized Linear Model with Elastic Net Regularization (GLM-EN) model is constructed by adding regularization terms to the classic GLM model

$$\hat{\beta} = \arg\min_{\beta,r} Deviance(\beta, r, y) + \lambda(\alpha||\beta||_1 + (1-\alpha)||\beta||_2)$$

- The ℓ_1 norm encourages sparse solution, which does the job of model selection
- The ℓ_2 norm encourages grouping effect (highly correlated variables have similar regression coefficients) and stabilizes the optimization process