Submission Number: EB-20-00120

Appendix that describe the extreme risk measures and explains the superior predictive ability test.
A Appendix

A.1 Definition of the extreme risk measures

The extreme risk measures, are described as follow:

**Kurtosis** is computed using the Kim and White (2004) robust kurtosis measure

\[
KR = \frac{(K_7 - K_5) + (K_3 - K_1)}{K_6 - K_2} - 1.23
\]

where \( K_i \) is the \( i \)th octile of the asset returns

**Left tail** is defined as the normalised distance between the median and the 10\% percentile of the return distribution

\[
\text{Left tail} = \frac{Q_{50} - Q_{10}}{\sigma}
\]

where \( Q_{50} \) is the median, \( Q_{10} \) is the 10\% percentile and \( \sigma \) is the standard deviation of the asset returns. It proxies for the tail risk for extreme losses.

**Right tail** is defined as the normalised distance between the median and the 90\% percentile of the return distribution

\[
\text{Right tail} = \frac{Q_{90} - Q_{50}}{\sigma}
\]

where \( Q_{50} \) is the median, \( Q_{90} \) is the 90\% percentile and \( \sigma \) is the standard deviation of the asset returns. It proxies for the tail risk for extreme gains.

A.2 Superior Predictive Ability Test

The superior predictive ability (SPA) test proposed by White (2000) and Hansen (2005) is used to compare the performance between competing estimation models. The objective of the SPA test is to compare multiple models with a benchmark model, and determine if the alternative models outperform the benchmark model. Although the term "predictive" is used, the test is not restricted to forecasting models. It can also be used on performance measurement of alternative trading strategies, or the comparison of various regression models (see the examples given in (White 2000)).

In our case, we set the benchmark model to be the McCulloch estimation method. As explained in the section 4, we use the Spearman’s correlation coefficient \( \rho_{j,k} \) calculated between the estimated tail index \( j \) and the extreme risk measures \( k \) to evaluate the performance of the tail index in reflecting the ranking of the extreme risk measure. If the tail index reflects perfectly the ranking of extreme risk measure, then \( \rho_{j,k} \) will be equal to 1.

We proceed to define the loss function of the tail index estimation methods as \( L_j = (1 - |\rho_j|)^2 \), where \( j = \{\text{McCulloch}, \text{GMM}, \text{Hill}\} \).

Following White (2000), we define our null hypothesis that the benchmark estimation method is not inferior to any alternative estimation methods. Defining the vector of relative loss functions as \( d = L_0 - L \), where \( L_0 \) is the loss function of the benchmark McCulloch method, and \( L = [L_{GMM}, L_{Hill}] \) is the loss function of the alternative estimation methods (GMM and the Hill estimator respectively), we formulate the null hypothesis as

\[
H_0 : E(d) \leq 0
\]

To test the benchmark with the set of alternative estimation methods, we should compare the benchmark against the best method with the lowest loss function. Therefore, we can also formulate the null hypothesis with a test statistic defined as

\[
d^* = \max_{j=GMM,Hill} d_j
\]

In addition, following Hansen (2005), we further studentize the test statistic to reduce the possibility of committing a type II error due to the inclusion of poor or irrelevant models in the alternative hypothesis.
Formally, we can rewrite the test statistic for the null hypothesis as:

\[ d^* = \max \left[ \frac{\max_{j=GM,HIll} d_j}{\text{var}(d_j)}, 0 \right] \tag{12} \]

As neither the distribution of Spearman's correlation \( \rho \) nor the vector of relative loss function \( d \) is known to us, it is not possible to deduce a critical value for the test statistic for the null hypothesis (Equation 12). To overcome this problem, we use the bootstrap methodology in section 4.3 to generate an extensive bootstrap sample of \( \rho_{j,b} \) where \( b \) is the bootstrap sample, and use it to simulate the distribution of \( d \). This distribution is subsequently used to compute the significance (p-value) of our test statistic \( d^* \) defined in Equation 12.

More precisely, we obtain the bootstrap distribution of the parameter estimate \( d^* \) under the null hypothesis. Instead of dealing with the null in Equation 10 as a composite, we choose a point in the null space that is least favorable to the alternative, namely \( d^* = 0 \). The null is then imposed by subtracting the estimated parameter \( d^* \) from the parameter estimate obtained on the bootstrapped return series, \( d^*_b \):

\[ d^*_b = \max_{j=GM,HIll} (d^*_j - d^*_j), \quad b = 1, 2, \ldots, B \tag{13} \]

Where \( B \) is the number of bootstraps performed. We count the number of times that a pattern at least as unfavorable as the null is observed in the real data \( d^* \). The number of ‘violations’ is then divided against the number of bootstraps \( B \) to give the p-value for the test and allows us to infer the following:

\[ \hat{p} = 1 - \frac{1}{B} \sum_{b=1}^{B} \{ d^*_b > d^* \} \tag{14} \]

We run the SPA test with 500 bootstraps to compare the performance of the estimated tail indexes in tracking the actual kurtosis measured in the empirical data, with the McCulloch method as the benchmark. We also repeat the same tests on the tail indexes in tracking the left and right tails of the empirical data.

As neither the distribution of Spearman's correlation \( \rho \) nor the vector of relative loss function \( d \) is known to us, it is not possible to deduce a critical value for the test statistic for the null hypothesis (Equation 12). To overcome this problem, we use the bootstrap methodology in section 4.3 to generate an extensive bootstrap sample of \( \rho_{j,b} \) where \( b \) is the bootstrap sample, and use it to simulate the distribution of \( d \). This distribution is subsequently used to compute the significance (p-value) of our test statistic \( d^* \) defined in Equation 12.

More precisely, we obtain the bootstrap distribution of the parameter estimate \( d^* \) under the null hypothesis. Instead of dealing with the null in Equation 10 as a composite, we choose a point in the null space that is least favorable to the alternative, namely \( d^* = 0 \). The null is then imposed by subtracting the estimated parameter \( d^* \) from the parameter estimate obtained on the bootstrapped return series, \( d^*_b \):

\[ d^*_b = \max_{j=GM,HIll} (d^*_j - d^*_j), \quad b = 1, 2, \ldots, B \tag{15} \]

Where \( B \) is the number of bootstraps performed. We count the number of times that a pattern at least as unfavorable as the null is observed in the real data \( d^* \). The number of ‘violations’ is then divided against the number of bootstraps \( B \) to give the p-value for the test and allows us to infer the following:

\[ \hat{p} = 1 - \frac{1}{B} \sum_{b=1}^{B} \{ d^*_b > d^* \} \tag{16} \]

We run the SPA test with 500 bootstraps to compare the performance of the estimated tail indexes in tracking the actual kurtosis measured in the empirical data, with the McCulloch method as the benchmark. We also repeat the same tests on the tail indexes in tracking the left and right tails of the empirical data.