

ROBUST STATISTICS AND FINANCIAL DATA

COLIN T. BOWERS[†]

ABSTRACT. This paper is designed to accompany a set of presentation slides that provide a selective introduction to Robust Statistics with a particular focus on financial applications. The topic of robust statistics is introduced, and then methods for measuring fat-tails in the data generating process are discussed. Next, a robust estimator of location, the trimmed mean, is defined, and then some simulations demonstrate its suitability for use with daily financial returns. Finally, robust estimation of a linear model is defined, and then investigated by a series of simulations designed to mimic daily financial returns.

1. BASIC EXAMPLE OF ROBUST STATISTICS

The field of Robust Statistics is concerned with statistical methods for use with non-Gaussian data. In particular, Robust Statistics are especially useful when the data generating process is leptokurtotic (i.e. fat-tails). When data exhibits fat-tails, non-robust (i.e. classical) estimators are typically very sensitive to a handful of outlier observations in the tails of Data Generating Process (DGP). The core idea of Robust Statistics is to ignore, or downplay, the effects of these outlier observations. This usually results in more stable estimates that are a closer reflection of the parameters of the DGP.

For example, Figure 1 contains 20 observations generated from a Students-t distribution with 2 degrees of freedom (a well-known, symmetric, fat-tailed distribution), where the horizontal bars partition 2 observations in each tail. The true mean of the DGP is 0, while the sample mean is 1.38. However, the value of this sample mean is heavily influenced by the tail observations that lie outside the horizontal bars. One robust statistic, known as the *trimmed mean*, suggests removing the tail observations from the sample mean calculation. In this example, if we ignore the observations outside the horizontal bars, then the trimmed mean is 0.68, which is significantly closer to the true mean of the DGP.

Date: 5th August, 2016.

Key words and phrases. Robust, Statistics, Financial Data.

[†]Department of Economics, Macquarie University, North Ryde NSW 2109, Australia. *Email:* colintbowers@gmail.com.

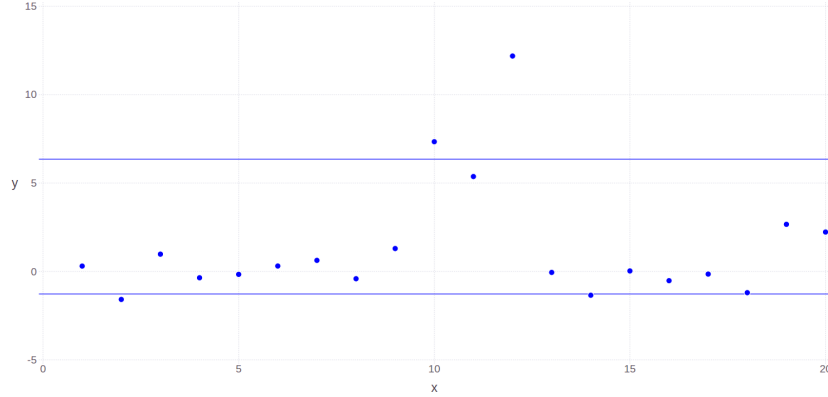


FIGURE 1. Data from Student-t distribution with 2 degrees of freedom

The example in the previous paragraph is, of course, cherry-picked for the purposes of illustration. Figure 2 provides a more complete picture of the behaviour of different location estimators for Students-t distributed data. Figure 2 contains simulated sampling distributions for four different location estimators; the sample mean, the 10% trimmed mean, the 50% trimmed mean, and the sample median.¹ Although all four sampling distributions are centred on the true mean of zero, the sample mean estimates are more likely to be further from zero than estimates from the other three (robust) estimators.

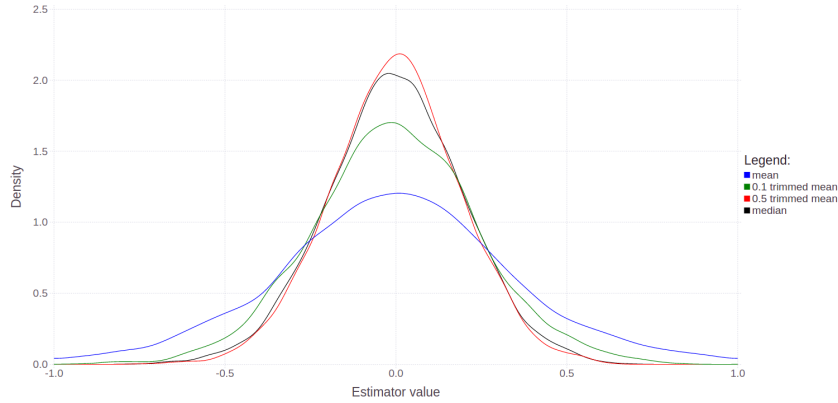


FIGURE 2. Sampling distribution of 4 location estimators

One other important concept can be gleaned from Figure 2. Note that of the three robust estimators, the 50% trimmed mean is, on average, the most accurate. It turns out that there is an optimal² degree of trimming which depends on the fatness of the tails of the

¹The number of simulation iterations is 5000, and the number of observations per iteration is 50.

²There are various metrics for optimality, but this is an unnecessary complication at this point.

DGP. For the case of Students-t with 2 degrees of freedom, the 10% trimmed mean does not trim enough observations, and the median³ trims too many.

2. TAIL FATNESS AND KURTOSIS

The example in the previous section demonstrates that Robust Statistics are most useful when the DGP exhibits fat tails. The obvious question is: how do we know if the tails of the DGP are fat? The standard measure of tail fatness is *kurtosis*, i.e. the (scaled) fourth centred moment. The Normal distribution has a kurtosis of 3, and fat-tailed, or leptokurtotic, distributions have a kurtosis greater than 3.

Despite its popularity, in my opinion, the kurtosis is not a sensible measure of tail fatness. This is because for distributions with unbounded support (such as the Normal) fatter tails result in undefined higher moments.⁴ For example, for a Students-t distribution with degrees of freedom v , all moments greater than v are undefined. This means that for the Students-t distribution with 2 degrees of freedom, the variance exists, but the skewness and kurtosis are both undefined.

In summary, if one suspects the data is generated by a fat-tailed DGP, it seems odd to choose a measure of tail fatness whose population parameter is more likely to be undefined in the presence of fat tails.

Fortunately, several robust measures of tail fatness have been proposed. In this paper, I use the measure proposed in ?, Section 4. Let F denote a cumulative distribution function with infinite support, and define:

$$(1) \quad I(a, b) = \int_a^b x dF(x).$$

Let $Q(p)$ denote the quantile associated with probability p from F , then Hogg's robust measure of tail fatness can be defined:

$$(2) \quad \kappa_{Hogg} = \frac{I(Q(0.95), \infty) - I(-\infty, Q(0.05))}{I(Q(0.5), \infty) - I(-\infty, Q(0.5))}$$

In words, the numerator is the “average” value in the *far right* of the distribution minus the “average” value in the *far left*, while the denominator is the “average” value in the right minus the “average” value in the left. Roughly speaking, the numerator of this measure gets larger as the tails get fatter, while the denominator gets larger as variance increases, and so can be thought of as a scaling factor for controlling for changes in variance.

³The median, by definition, trims all observations except for 1 or 2, depending on whether the total number of observations is odd or even.

⁴That is, the integrals that define those moments evaluate to ∞ .

It is interesting to compare the properties of the robust estimator of tail fatness with the standard sample kurtosis given a Student-t distribution with 2 degrees of freedom DGP. Figure ?? provides simulated average values for these two estimators for various sample sizes. As can be seen, the sample kurtosis diverges with the sample size (a common phenomenon when the population parameter is undefined). In contrast, the robust estimator of fat tails is stable with respect to the sample size, which is a much more desirable property in an estimator.

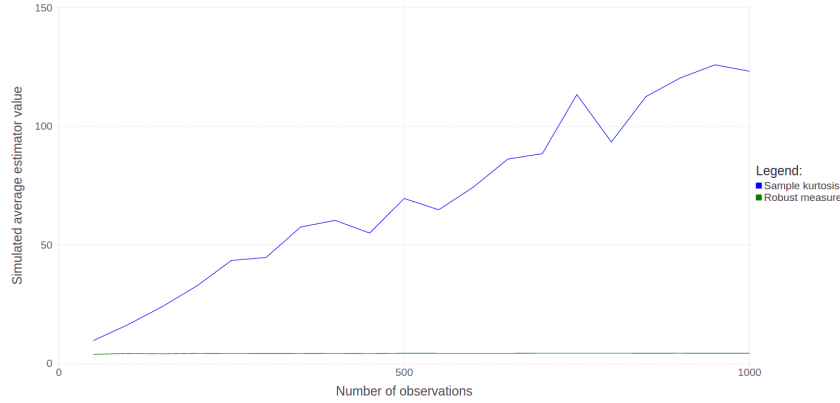


FIGURE 3. Sample kurtosis as a function of sample size when population kurtosis is undefined

3. ESTABLISHING TAIL FATNESS OF FINANCIAL RETURNS

In this section we examine the tail fatness of daily financial returns. Using daily⁵ data from January 2005 to December 2015, I calculate Hogg’s robust measure of tail fatness for several of the most popular stocks listed on the ASX.⁶ The estimates are depicted in Figure 4, along with two reference lines; one is the population value of Hogg’s estimator for the Normal distribution, and the other is for the Student-t distribution with 2 degrees of freedom. As can be seen, according to Hogg’s measure, all the stocks in the sample exhibit fat tails relative to the Normal distribution. This is typical of financial return data.

It is worth briefly diverging here into a quick discussion of how statisticians typically model fat tails in financial returns. One of the most popular approaches is to model returns as Normal, but with a conditional or stochastic volatility, i.e. $r_t | \sigma_t^2 \sim \mathcal{N}(\mu, \sigma_t^2)$. This framework will generate an unconditionally fat-tailed sequence. The intuition is fairly simple: a small number of returns generated when σ_t^2 is large will, on average, be much further into the tails of the unconditional

⁵Close to close.

⁶AMP, ANZ, BHP, CBA, CCL, CWN, JBH, LLC, MQG, NAB, RIO, SUN, TLS, TOL, WBC, WES, WOW.

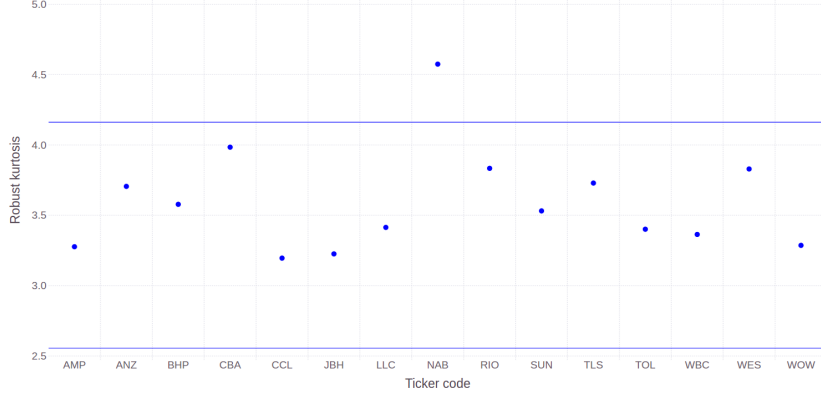


FIGURE 4. Robust kurtosis for daily returns of 15 stocks listed on the ASX. The bottom and top reference lines are the population robust kurtosis for the Normal and the Student-t (2 degrees of freedom) distributions respectively.

distribution than when σ_t^2 is small. To see this in real world data, one needs to only to plot returns from 2006 to 2009, and note how the returns during the market crash of late 2008 and recovery of early 2009 dominate the sample.

Assuming this modelling framework, it follows that the sequence r_t/σ_t should have constant volatility, and thus not exhibit fat-tails relative to the Normal. Unfortunately, we cannot model r_t/σ_t explicitly, since σ_t is unobservable. However, we can substitute in an estimate for σ_t . Figure 5 duplicates Figure 4 except that now I have replaced daily returns as the input with daily returns standardised by a rolling window historical volatility estimate (with a window length of 100). As can be seen, this technique is able to remove some, but not all, of the tail fatness in the return DGP. It is possible that there is a secondary source of tail-fatness in the DGP, but it is also possible that the remaining tail fatness is due to inaccuracies in our proxy for the true volatility process.

4. THE TRIMMED MEAN AND FINANCIAL RETURNS

Having established the presence of fat tails in financial returns, it is of interest to establish whether practitioners should consider replacing standard statistical estimates with robust statistical estimates. Arguably the most famous estimate in statistics is the sample mean:

$$(3) \quad m_{normal} = \frac{1}{T} \sum_{t=1}^T r_t,$$

which is an estimate of the expected value of a distribution, which is equivalent to the centre of a symmetric distribution. A robust analogue to the sample mean is the trimmed mean. Let $r_{[1]}, \dots, r_{[T]}$ denote the

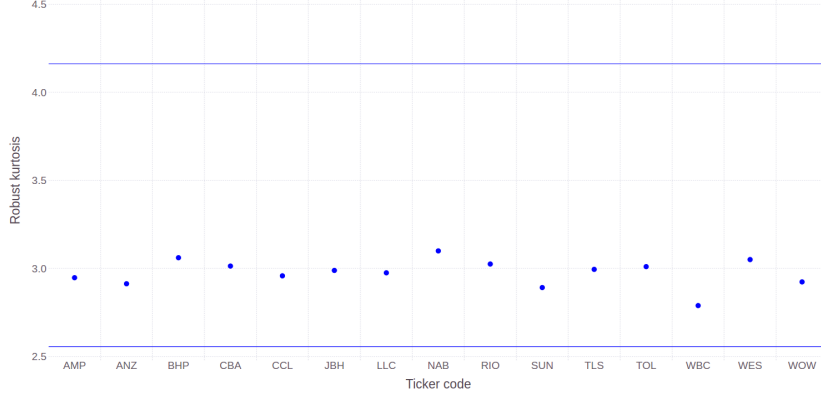


FIGURE 5. Robust kurtosis standardised daily returns of 15 stocks listed on the ASX. The bottom and top reference lines are the population robust kurtosis for the Normal and the Student-t (2 degrees of freedom) distributions respectively.

sorted version of the sequence r_1, \dots, r_T (sometimes called the *order statistics*). Then for some trimming proportion $\alpha \in [0, 1]$, the trimmed mean is:

$$(4) \quad m_{\text{robust}} = \frac{1}{(1 - \alpha)T} \sum_{t=\frac{\alpha}{2}T}^{(1-\frac{\alpha}{2})T} r_{[t]},$$

where for notational convenience I will assume $\alpha/2$ is an integer.

In words, the trimmed mean symmetrically removes the largest and smallest $\alpha/2$ fraction of observations from the dataset, and then computes a sample mean on the remaining data. Note that for $\alpha = 0$, the trimmed mean is equivalent to the sample mean, and for $\alpha = 1$, the trimmed mean is equivalent to the median.

If the underlying DGP is symmetric, then the trimmed mean and sample mean are estimators for the same population statistic. However, if the DGP is fat-tailed, then the trimmed mean is more accurate (on average) than the sample mean as it disregards those observations in the tails of the DGP that add more noise than information. Given the importance of the expected value of financial returns, it is of interest to ask whether the trimmed mean is more accurate when used on financial data than the sample mean. Answering this question is complicated by the fact that the population mean is unobservable. Ideally, what is required is a DGP with roughly the same properties as the DGP for financial returns, but where the population mean is known. Fortunately, we can conjure up such a dataset using a block bootstrap procedure. Full details of the procedure are not provided here, but a classic paper on the subject is ?. Intuitively, we take a sequence of financial returns and centre them, so that their sample mean is 0. We then treat this sequence of returns as if it is the entire population, and

resample with replacement from it using a strategy that ensures any weak dependence structure and heteroskedasticity in the data is preserved. By construction, the resampled sequences will have population mean of zero, although their sample means can (and will) differ from zero due to sampling with replacement.

I resample M sequences, each of length T , and the nature of the resampling scheme ensures that each sequence is independent from the other sequences. I then calculate both the sample mean and the trimmed mean on each sequence, and from the resulting M estimates can construct a kernel density of the sampling distribution of the estimator. Under certain modelling assumptions, this kernel density will converge on the true asymptotic sampling distribution for the estimator as $\{M, T\} \rightarrow \infty$. I performed this procedure on the sequence of NAB returns, and the sequence of CCL returns. I chose these two assets since in Figure 4 they have the largest and smallest robust kurtosis (respectively).

As can be seen in Figure 6, the trimmed mean significantly outperforms the sample mean for NAB data. Perhaps more surprisingly, as Figure 7 demonstrates, the trimmed mean performance is reasonably good relative to the sample mean even for CCL data. This result is surprising, since CCL has relatively stable volatility over the sample period and so its tails are only mildly fat.

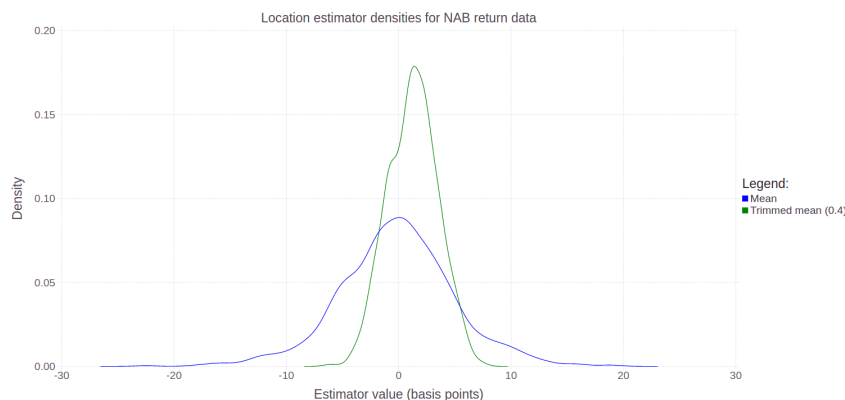


FIGURE 6. Bootstrapped sampling distributions of the sample mean and trimmed mean when applied to NAB daily return data.

Note that the sample period in question here is from 2005 to 2015. Of course, this period includes some major financial events, and correspondingly large shifts in volatility. If one chose a shorter horizon, say, 100 days, perhaps the sample mean would have comparatively better performance, since it is feasible that volatility might be relatively stable over such a short horizon. To investigate this possibility, I isolated the 100 sequential days in NAB returns that correspond to the

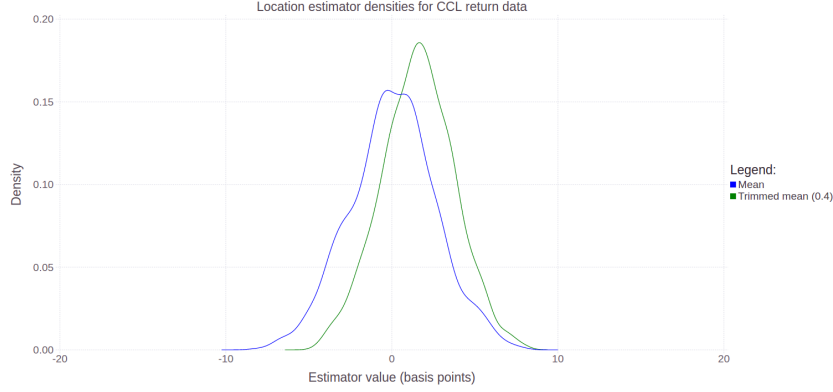


FIGURE 7. Bootstrapped sampling distributions of the sample mean and trimmed mean when applied to CCL daily return data.

largest value of the robust kurtosis, and compare it to the 100 sequential days that correspond to the smallest robust kurtosis. I perform the same bootstrapping exercise on each of these datasets, and the results are in Figure 8 and Figure 9. As expected, the trimmed mean massively outperforms over the interval with the fattest tails. Indeed, for this interval, the sampling distribution of the sample mean is trimodal! Surprisingly, the sample mean has only slightly better performance for the interval with the thinnest tails.

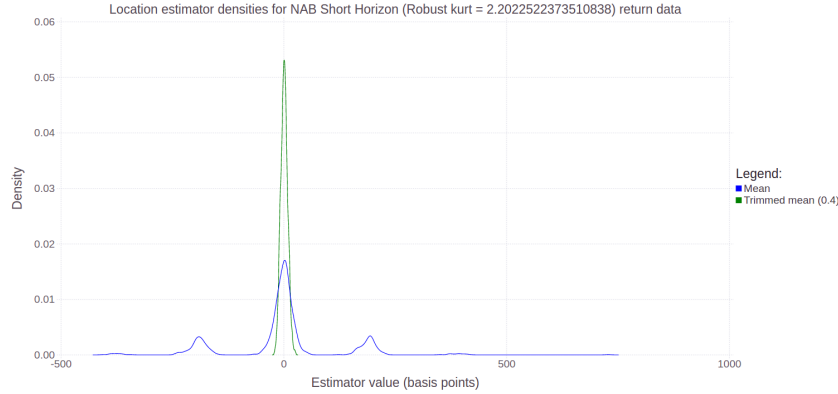


FIGURE 8. Bootstrapped sampling distributions of the sample mean and trimmed mean when applied to a short horizon of NAB daily return data with fat tails.

Taken together, the above analysis suggests that a researcher should always at least consider a robust estimate when estimating the first moment of financial returns using historical price data for that asset.

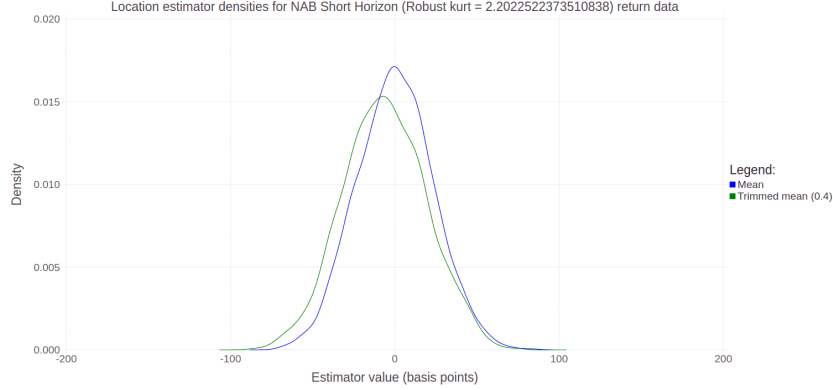


FIGURE 9. Bootstrapped sampling distributions of the sample mean and trimmed mean when applied to a short horizon of NAB daily return data with thin tails.

5. ROBUST ESTIMATION OF A LINEAR MODEL FOR FINANCIAL RETURNS

In this section I examine whether robust estimation methods might be useful for a researcher building a linear model for financial returns. Specifically, I assume the following univariate structure:

$$(5) \quad r_t = \alpha + \beta s_t + e_t,$$

where r_t is a daily return, s_t is a predictive signal, e_t denotes the residual, and $\{\alpha, \beta\}$ are the parameters of the model.

Given a vector of observable returns \mathbf{r} , and a vector of observable signals \mathbf{s} , the parameters in (5) can be estimated by solving the following optimisation problem:

$$(6) \quad \min_{\alpha, \beta} L(\mathbf{e}),$$

where $\mathbf{e} = \mathbf{r} - \alpha - \beta \mathbf{s}$, and L denotes a function for measuring loss. The most popular choice for the loss function is $L(\mathbf{e}) = \|\mathbf{e}\|_2$, i.e. the Euclidean norm, which corresponds to the Ordinary Least Squares (OLS) estimator. It is well understood that if $e_t \sim \mathcal{N}(0, \sigma^2)$, then given a few other regularity conditions, OLS yields the best, linear, unbiased estimator.

However, consider the case where $e_t \sim \mathcal{N}(0, \sigma_t^2)$, that is, the residuals exhibit a pattern of heteroskedasticity. In this case, the Euclidean norm loss function can place too much emphasis on a small number of observations that correspond to a period of high volatility. It is sometimes better to choose a loss function that places less emphasis on observations that are large in magnitude, such as $L(\mathbf{e}) = \|\mathbf{e}\|_1$, which corresponds to the Least Absolute Deviations (LAD) estimator.⁷

⁷ $\|\cdot\|_1$ is sometimes called the taxicab norm.

It is important to note that, unlike Section 4, here it is the statistical properties of the residual, e_t , that are important, rather than the statistical properties of the return, r_t . However, when predicting financial returns, the R^2 associated with a model such as (5) is typically very small, typically much less than 0.05, and so it is likely that r_t and e_t will have similar statistical properties, such as heteroskedasticity. Thus robust estimators such as the LAD estimator may be of interest.

To investigate the effect of LAD versus OLS, I need to simulate (5) in such a way that r_t exhibits similar statistical properties to real-world financial data. I do this as follows. First, I construct a sequence of variances $\hat{\sigma}_t^2$, using a rolling historical variance estimator on NAB daily returns. I then simulate the residual term from (5) using the assumption $e_t \sim \mathcal{N}(0, \hat{\sigma}_t^2)$. I simulate the signal using $s_t \sim \mathcal{N}(0, 1)$, set $\alpha = 0$, and choose β such that the sample R^2 given (5) is 0.05. Using these inputs and (5), I then simulate r_t . Note that the rolling historical variance estimator, $\hat{\sigma}_t^2$ can be thought of as a smoothing function on squared NAB returns where a longer window length implies more smoothing, and thus more mild dynamics for $\hat{\sigma}_t^2$. I deliberately use a window length of 17 days, as this corresponds to simulated returns r_t that have a robust kurtosis roughly the same as that for a Student-t distribution with 2 degrees of freedom (the upper line in Figure 4).

I simulate M datasets, and for each dataset, I estimate (5) using both the OLS and LAD methodologies. I then construct the kernel density estimates of the sampling distributions of parameter estimates from (5). As can be seen in Figure 10 and Figure 11, the LAD estimator is significantly more accurate than the OLS estimator for both the constant term and the slope coefficient in (5). It immediately follows that predictions made using the LAD estimators will, in this framework, be more accurate than their OLS counterparts, on average.

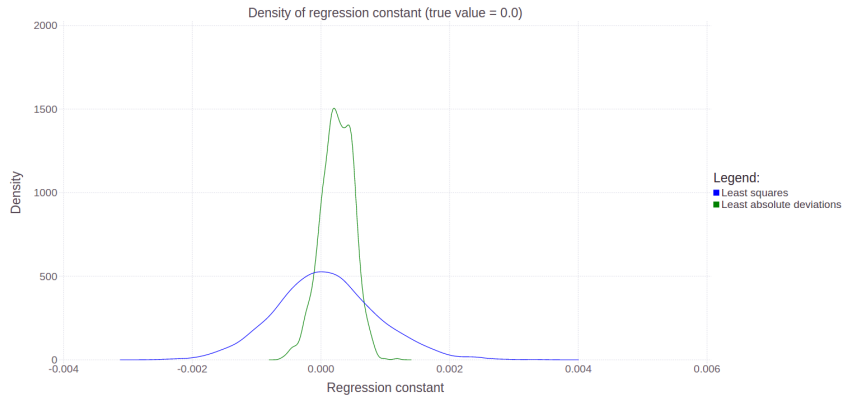


FIGURE 10. Regression constant sampling distribution.

An interesting question for active portfolio managers is whether this increased estimation accuracy translates into better returns in an active



FIGURE 11. Regression coefficient sampling distribution.

management framework. To answer this question, I conduct a very simple trading simulation. Specifically, I assume an economy with 20 assets, plus a cash asset that earns no interest. The returns for all 20 assets are simulated using the process described earlier in this section,⁸ and the returns on the assets are uncorrelated. The portfolio manager is assumed to have access to the signal s_t for all 20 assets, and to know the true functional form of (5), but not the true values for α and β . The portfolio manager estimates α and β using either OLS or LAD on data from the first half of the sample. Then, for the second half of the sample, in every period t , the portfolio manager will go long any asset (with equal weighting) where the forecast return $\hat{r}_t = \hat{\alpha} + \hat{\beta}s_t$ is greater than zero. There are no transaction costs or volume constraints, and starting wealth is 1 million dollars.

The above simulation is repeated M times, and the kernel density estimate of the distribution of terminal wealth given OLS versus LAD are provided in Figure 12. As can be seen, significant material gains are realised by the portfolio manager who uses LAD instead of OLS.

6. CONCLUSION

In this paper I have introduced the topic of robust statistics, discussed ways of measuring tail fatness, and then investigated the efficacy of several popular robust estimators when used with financial data. The general conclusion suggested in this paper is that the heteroskedasticity present in financial returns is sufficient to justify, at the very least, the consideration of robust methodologies when working with financial data.

⁸Although I significantly reduce the R^2 to keep the terminal wealth numbers realistic.

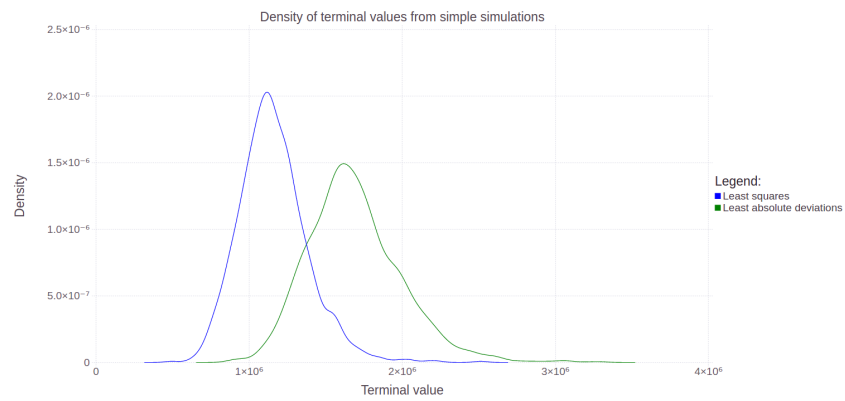


FIGURE 12. Portfolio terminal value.