# Table of Contents

Executive Summary ................................................................................................................................. 5

Introduction ................................................................................................................................................. 9

1. Issues with Robustness of the Classical Estimator ........................................................................... 13

2. Intuition and Links to Other Estimation Techniques ........................................................................... 19

3. M-estimation of CSV: Bias, Large-sample Theory, and the Algorithm ........................................... 25

4. A Comparison with a Popular Quantile-based Estimator – Asymptotic Variance and Breakdown Points .................................................................................................................. 33

5. Implications for CSV Estimation for Universes of Different Size and for Returns of Different Frequency ........................................................................................................................................ 41

6. Summary ............................................................................................................................................... 51

Appendices ............................................................................................................................................... 53

References ................................................................................................................................................ 59

About Fédération Bancaire Française ...................................................................................................... 61

About EDHEC-Risk Institute .................................................................................................................. 63

About the Authors

Felix Goltz is head of applied research at EDHEC-Risk Institute and research director at ERI Scientific Beta. He does research in empirical finance and asset allocation, with a focus on portfolio construction and indexing strategies. His work has appeared in various international academic and practitioner journals and handbooks. He obtained a PhD in finance from the University of Nice Sophia-Antipolis after studying economics and business administration at the University of Bayreuth and EDHEC Business School.

Lionel Martellini is professor of finance at EDHEC Business School and scientific director of EDHEC-Risk Institute. He has graduate degrees in economics, statistics, and mathematics, as well as a PhD in finance from the University of California at Berkeley. Lionel is a member of the editorial board of the Journal of Portfolio Management and the Journal of Alternative Investments. An expert in quantitative asset management and derivatives valuation, his work has been widely published in academic and practitioner journals and has co-authored textbooks on alternative investment strategies and fixed-income securities.

Stoyan Stoyanov is professor of finance at EDHEC Business School and head of research at EDHEC Risk Institute–Asia. He has ten years of experience in the field of risk and investment management. Prior to joining EDHEC Business School, he worked for over six years as head of quantitative research for FinAnalytica. He has designed and implemented investment and risk management models for financial institutions, co-developed a patented system for portfolio optimisation in the presence of non-normality, and led a team of engineers designing and planning the implementation of advanced models for major financial institutions. His research focuses on probability theory, extreme risk modelling, and optimal portfolio theory. He has published over thirty articles in leading academic and practitioner-oriented scientific journals such as Annals of Operations Research, Journal of Banking and Finance, and the Journal of Portfolio Management, contributed to many professional handbooks and co-authored three books on probability and stochastics, financial risk assessment and portfolio optimisation. He holds a master in science in applied probability and statistics from Sofia University and a PhD in finance from the University of Karlsruhe.
Executive Summary
Volatility is a statistical measure of the dispersion of returns for a given security or market index. It refers to the amount of uncertainty or risk about the size of changes in an underlying security or index value. Higher volatility means that the price of the security can change dramatically in either direction over a given interval of time.

A key distinction exists between systematic and specific volatility: total volatility can be decomposed into systematic volatility, driven by the stock exposure with respect to systematic risk factors, and specific volatility, which is driven by the uncertainty impacting a particular company. The recent financial literature has paid considerable attention to idiosyncratic volatility indicating that idiosyncratic volatility can have forecasting power for future excess returns.

Garcia, Mantilla-Garcia and Martellini (2013) discuss a link between idiosyncratic volatility and cross-sectional volatility (CSV). They show that the cross-sectional variance of stock returns can be regarded as an efficient estimator for the average idiosyncratic variance of stocks within the universe under consideration. Goltz et al. (2011) find this measure of variance to be highly correlated to standard measures of systematic risk when they exist, which further justifies its use in the context of equity volatility measurement. Key advantages of this measure over currently available measures such as sample-dependent historical volatility measures or option-based implied volatility measures are: its observability at any frequency; its model-free nature; and its availability for every region, sector and style of the world equity markets, without the need to resort to any auxiliary option market.

A methodology of CSV estimation has to be based on a statistically robust estimation method because, empirically, cross-sections of returns are polluted by outliers – abnormal returns in the cross-section that deviate significantly from the rest of the sample. The classical method of estimation is sensitive to the outliers in the sample that the resulting noise can hide the dynamics of CSV in the time domain. An estimator is considered robust in a statistical sense if a small proportion of big outliers does not lead to big deviations in the estimated values. The outliers are usually regarded as a contaminant and the goal of any robust approach in statistics is to limit or completely neutralise their influence on the estimated quantity. By construction, the notion of statistical robustness is related to a notion of continuity of the estimator with respect to small deviations from a given reference model.

In this paper, we propose a robust method of estimation which is intuitive and is functionally similar to the weighted-average estimator studied by Garcia, Mantilla-Garcia and Martellini (2013) in the context of one-factor and multi-factor regression models. To meet this objective, we adopt a statistical technique called M-estimation.

Although the interpretation of CSV as the average idiosyncratic volatility of the universe follows from asymptotic arguments, the CSV is computable for any universe. We study the relevance
of robust estimation for universes of different size such as the top 50, 100 and 200 US stocks listed on NYSE and confirm that the properties of robustness become more important for universes of smaller size. Finally, we consider cross-sections of returns of different frequency and conclude that cross-sections of returns of lower frequencies are characterised by fewer outliers. In the case of monthly frequency and the S&P500 universe, the classical estimator combined with mild trimming seems to have a satisfactory performance. The daily and weekly frequencies, however, are characterised by a higher percentage of outliers in the cross-section which justifies the use of a more sophisticated robust estimation technique.
Executive Summary
Introduction

Volatility is a statistical measure of the dispersion of returns for a given security or market index. In other words, volatility refers to the amount of uncertainty or risk about the size of changes in an underlying security or index value. Higher volatility means that the underlying value can potentially be spread out over a larger range of values, signalling a higher riskiness for investors holding the security or index. Indeed, a high volatility means that the price of the security can change dramatically in either direction over a given interval of time. Lower volatility means that the value of a security or index does not fluctuate dramatically, but it changes at a steadier pace over a period of time.

Regardless of the method used in estimating volatility, a key distinction exists between systematic and specific volatility. For any given stock, total volatility can be decomposed into systematic volatility, driven by the stock exposure with respect to systematic risk factors, and specific volatility, which is driven by the uncertainty impacting a particular company. The recent financial literature has paid considerable attention to idiosyncratic volatility. Campbell et al. (2001) and Malkiel and Xu (2002) document that idiosyncratic volatility increased over time, while Brandt et al. (2009) show that this trend completely reversed by 2007, falling below pre-1990 levels and suggest that the increase in idiosyncratic volatility through the 1990s was not a time trend but rather an "episodic phenomenon". Bekaert et al. (2008) confirm that there is no trend both for the United States and other developed countries. A second fact about idiosyncratic volatility is also a source of contention. Goyal and Santa-Clara (2003) put forward that idiosyncratic volatility has forecasting power for future excess returns, while Bali and Cakici (2008) and Wei and Zhang (2005) find that the positive relationship is not robust to the sample chosen.

Garcia, Mantilla-Garcia and Martellini (2013) discuss a link between idiosyncratic volatility and cross-sectional volatility (CSV). They show that the cross-sectional variance of stock returns can be regarded as an efficient estimator for the average idiosyncratic variance of stocks within the universe under consideration. Goltz et al. (2011) find this measure of variance to be highly correlated to standard measures of systematic risk when they exist, which further justifies its use in the context of equity volatility measurement. Key advantages of this measure over currently available measures such as sample-dependent historical volatility measures or option-based implied volatility measures are: its observability at any frequency; its model-free nature; and its availability for every region; sector and style of the world equity markets, without the need to resort to any auxiliary option market.

A methodology of CSV estimation has to be based on a statistically robust estimation method because empirically cross-sections of returns are polluted by outliers – abnormal returns in the cross-section that deviate significantly from the rest of the sample. The classical estimator is sensitive to the outliers in the sample that the resulting noise can hide the dynamics of CSV in the time domain. An estimator is considered robust in a statistical sense.
if a small proportion of big outliers does not lead to big deviations in the estimated values. The outliers are usually regarded as a contaminant and the goal of any robust approach in statistics is to limit or completely neutralise their influence on the estimated quantity. By construction, the notion of statistical robustness is related to a notion of continuity of the estimator with respect to small deviations from a given reference model.

Apart from the continuity aspect, there are other ways to quantify the concept of robustness. One approach is through the so-called breakdown point of the estimator — the smallest proportion of abnormal outliers of arbitrary size that that leads to an explosion of the estimated quantity. An estimator with a high breakdown point is also said to have a high resistance to outliers. The classical estimators of location and scale break down even if one single observation takes an arbitrarily large value and, therefore, have no resistance to outliers. Robust estimators may also have high or low resistance to outliers. For additional information on statistical robustness, we refer the reader to Maronna, Martin, and Yohai (2006) and Huber and Ronchetti (2009).

The objective of the robust approach proposed in this paper is to derive an estimator which is intuitive and functionally similar to the weighted-average estimator studied by Garcia, Mantilla-Garcia and Martellini (2013) in the context of one-factor and multi-factor regression models. To meet this objective, we adopt the M-estimation technique which robustifies the least-squares method and can lead to a weighted-average functional form. The weighting function in the M-estimator is taken from Yohai and Zamar (1997) who prove that this particular form has an optimal balance between bias and efficiency. Having introduced the estimator, we study its properties in the context of a proxy of the S&P 500 universe from 01-Jan-1980 to 31-Jul-2012. We show that the number of outliers detected by the M-estimator sometimes exceeds the breakdown point of the quantile estimator. We report the breakdown point and compare the M-estimator to a popular quantile-based estimator using daily returns from 01-Jan-1980 to 31-Jul-2012.

Although the interpretation of CSV as the average idiosyncratic volatility of the universe follows from asymptotic arguments, the CSV is computable for any universe. We study the relevance of robust estimation for universes of different size such as the top 50, 100, and 200 US stocks listed on NYSE and confirm that the properties of robustness become more important for universes of smaller size. Finally, we consider cross-sections of returns of different frequency and conclude that cross-sections of returns of lower frequencies are characterised by fewer outliers. In the case of monthly frequency and the S&P 500 universe, the classical estimator combined with mild trimming seems to have a satisfactory performance. The daily and weekly frequencies, however, are characterised by a higher percentage of outliers in the cross-section which justifies the use of a more sophisticated robust estimation technique.
Introduction

The paper is structured in the following way. In Section 1, we summarise the connection between cross-sectional dispersion and specific volatility. Section 2 intuitively introduces the M-estimation technique through its connections with the classical method of maximum likelihood and the method of least squares which illustrate the source of robustness. Section 3 discusses the bias and the large sample properties of M-estimators and it also introduces the particular choice of the penalty function and provides the actual calculation algorithm. Section 4 compares the robustness and outlier resistance properties of the estimator discussed in the paper and a popular quantile based estimator of volatility. Finally, Section 5 studies the relevance of robust estimation for universes of different size and returns of different frequency.
1. Issues with Robustness of the Classical Estimator
1. Issues with Robustness of the Classical Estimator

Idiosyncratic variance has recently received substantial attention in the empirical literature. An interesting connection with cross-sectional variance is considered by Garcia, Mantilla-Garcia and Martellini (2013) providing a method of estimation which holds for any frequency, is model-free, and is feasible for any region, sector or style of the equity markets. In this section, we provide an outline of the connection and focus on the problem of estimation of CSV. Although the issue with robustness can be very pronounced in cross-sectional data, the methods discussed in this section can be applicable in a time-series context where variance is estimated from data polluted with atypical observations. Indeed, the field of robust statistics is universal and finds application in cross-sectional and time-series settings and also in univariate and multivariate analysis, see for example Maronna, Martin, and Yohai (2006).

We further make the following two simplifying assumptions:

1. Homogeneous beta assumption: $\beta_{it} = \beta_t$ for all $i$;
2. Homogeneous residual variance assumption: $E\varepsilon_{it} = 0$ and $\text{cov}(F_t, \varepsilon_{it}) = 0$ for all $i$.

Under these assumptions, Garcia, Mantilla-Garcia and Martellini (2013) show that cross-sectional variance converges towards specific variance in the limit of an increasing large number of constituents:

$$\text{Eq. 1} \quad \text{CSV}^2_{t, w} = \sum_{i=1}^{N_t} w_{it} \left( \bar{r}_{it} - \bar{r}_{w,t} \right)^2 \xrightarrow{N_t \to 00} \sigma^2(t)$$

where $\bar{r}_{t, w}$ is the weighted-return on the portfolio with weights $w_{it}$ at date $t$, CSV$_t$ is the cross-sectional volatility, and $N_t$ is the number of constituents in the universe for a given date $t$.

Under the simplifying assumptions, the estimator in Eq. 1 is asymptotically unbiased and weakly consistent for any non-trivial weighting scheme. Of course, for a finite number of constituents $N_t$, different weighting schemes will generate different proxies for idiosyncratic variance and may introduce a bias. It can be demonstrated that the bias can be easily removed in the equally weighted scheme which also proves to be the estimator with minimal variance within the class of estimators under a strictly positive weighting scheme.

Garcia, Mantilla-Garcia and Martellini (2013) show that relaxing the simplifying assumptions leads to some technical complications which, however, do not significantly change the interpretation. Thus, in the presence of non-homogeneous
1. Issues with Robustness of the Classical Estimator

residuals, cross-sectional variance converges to a weighted average of the residual variances. In the presence of non-homogeneous betas in a one-factor model, cross-sectional variance is an asymptotically biased estimator of the residual variance and the bias is related to the dispersion of the betas. Finally, the authors illustrate that the bias turns out to be empirically insignificant for the US market.

In this paper, we do not discuss the ramifications arising by extending the assumptions behind the model. Rather, we focus on the question of robustness. For this reason, we adopt a simplified representation of stock returns that does not refer to any specific factor model,

Eq. 2

\[ r_{it} = a_t + \varepsilon_{it} \]

where \( a_t \) measures the mean of \( r_t \) and \( \varepsilon_{it} \) are residuals with a mean of zero. In this representation, the mean accumulates the impact of the factors which is restrictive only by assuming that the stock-returns in the universe have equal means which may not be the case in a multi-factor model. However, as argued above, we focus on robustness and will assume that if there was any bias it has been taken care of and the (cross-sectional) sample contains returns with a common mean.

The theory developed by Garcia, Mantilla-Garcia and Martellini (2013) to motivate the estimator in Eq. 1 shows that choosing the weights to be equal, \( w_{it} = 1/N_t \), minimises the variance of the estimator. Using equal weights, and correcting for bias, we obtain

Eq. 3

\[ CSV_{t,EW} = \left( \frac{1}{N_t - 1} \sum_{i=1}^{N_t} (r_{it} - \bar{r}_{t,EW})^2 \right)^{1/2} \]

where \( \bar{r}_{t,EW} \) denotes the average return.

Figure 1. Boxplots of cross-sections of stock returns from the S&P 500 universe in December 2004, which is an example of a calm market, and September 2008 – a month of significant market turbulence. Red crosses denote observations beyond 1.5 times the inter-quartile range, which can be accepted as outliers.
1. Issues with Robustness of the Classical Estimator

The resulting estimator in Eq. 3, however, is not robust in the presence of contamination in the data. In fact, the good large-sample properties of the estimator rely on the assumption that the observations are homogeneous in the sense that they are generated by one and the same stochastic mechanism. In practice, we observe significant outliers in cross-sections of returns which we assume are a result of some sort of contamination. The fact that presence of outliers can severely impact the quality of the estimator in Eq. 1 is well known in statistics; see for example Maronna, Martin, and Yohai (2006). A way to deal with the problem is to invoke the methods of robust statistics.

The presence of outliers in cross-sectional data is illustrated on Figure 1. The figure includes two plots that show boxplots of cross-sections of returns of a proxy of the S&P 500 universe for two months – December 2004 and September 2008. The former month is an example of a calm market and the latter is characterised by significant market turbulence. The red crosses indicate observations that extend beyond 1.5 times the inter-quartile range of the data which corresponds to 99.3% coverage under the assumption of normality. That is, under this assumption, the probability of observing a red cross is below 1%. Under this definition, it is clear that outliers are observed in calm as well as in turbulent markets. In periods of market turbulence, however, the magnitude of the outliers is much bigger.

The word robust has different connotations. For the purposes of this paper, we adopt the following meaning – an estimator is said to be robust if it is continuous with respect to small deviations from the underlying assumptions which are often formalised as a contamination neighborhood of a given reference model. From an intuitive perspective, a small number of abnormal observations should not lead to enormous deviations in parameter estimates. For a formal treatment of robustness, see Huber and Ronchetti (2009) and also Maronna, Martin, and Yohai (2006).

We adopt a technique from robust statistics and use the more general expression in Eq.
1. Issues with Robustness of the Classical Estimator

1. The weights $w_t$ get adapted to the sample so that the outliers likely to be a result of contamination get small weights, or even a weight of zero, while the body of the sample gets weights close to $1/N_t$. In this way, we can regard the new estimator as a modified version of Eq. 3, which has good theoretical properties (minimal bias and variance), where robustness is introduced through decreasing the weights of outliers.

The particular technique we employ is M-estimation. The name is derived from maximum likelihood-type estimation because the functional form of the estimator generalises maximum likelihood. Because the weights adapt to the sample, a key ingredient in M-estimation is the weighting function that determines exactly how the weights decline. We employ a weighting function developed by Yohai and Zamar (1997) which has optimal properties; see Maronna, Martin, and Yohai (2006). The general form of the M-estimator is

\[ Eq. 4 \]

where $r_{it}$ is the residual for the $i$-th stock and the $t$-th period, $b_{i}$ is the estimated coefficient for the $i$-th stock, $\eta$ is a bias correction constant, $w_i = \frac{\psi'(\varepsilon_{it}) + c}{\psi(\varepsilon_{it})}$ and, finally, $\psi$ is the standard normal density, $t^+ = \max(t, 0)$, and $c$ is a constant determining the optimal properties of the estimator. It can be demonstrated that if $c = 0$, then $\psi_i = 1$ and $w_i = 1/N_t$, which results in the equally-weighted expression in Eq. 5.

Although the idea of adjusting the weights in the estimator in Eq. 4 may seem ad-hoc, deep connections exist between M-estimation and other statistical techniques. The next section introduces the approach intuitively and motivates the general form of a weighted average of the estimator of cross-sectional variance. We describe connections with the method of maximum likelihood and also a generalised form of the least squares method which elucidate how the weights are selected and how the robustness properties arise.

Finally, we note that although in principle M-estimation is applicable in a time-series context, its value-added in a direct application for financial data is perhaps limited. The reason is that, in the time domain, volatility tends to cluster, which means that large returns in absolute value are usually followed by returns of similar magnitude. GARCH-type models successfully take this phenomenon into account and empirical literature has demonstrated their value added; see Poon and Granger (2003). As a consequence, the outliers in a given sample are possibly a source of information for the underlying time structure of volatility rather than contaminants distorting statistical inference; that is, ignoring them can reduce the information content relative to the classical estimator which includes all observations on an equal basis. Still, M-estimation can be combined with GARCH-type models to increase the robustness of the model, see Boudt and Croix (2010) for an application in multivariate GARCH estimation.
1. Issues with Robustness of the Classical Estimator
2. Intuition and Links to Other Estimation Techniques
2. Intuition and Links to Other Estimation Techniques

In this section, we introduce the technique of M-estimation intuitively through the connections with the method of maximum likelihood. We describe a way to reformulate the estimator so that it becomes similar to the natural estimator of cross-sectional variance in Eq. 1. We also provide a link to a generalised version of the method of least squares that indicates how the weights in Eq. 1 can arise from a penalty function.

Maximum likelihood estimation and M-estimation

To introduce the intuition behind M-estimation, consider first the maximum likelihood estimator (MLE) of location and scale assuming there is no contamination. ML estimation is widely used in statistics because of its appealing large-sample properties. Without delving into technical details, we only mention that the method relies on the assumption that the sample contains independent observations drawn from a particular distribution. Then, parameters are estimated by maximising the probability of observing the sample, i.e. the estimated values yield the maximum probability of what has already been observed.

If we assume that the observations are described by the model in Eq. 2 and that the standardised residual follows a distribution with a density function \(f(x)\), then the density of the sample is \(f(((x-a_t))/s_t)\), in which \(s_t\) denotes the scale of the residual term in Eq. 2. The MLE of the location and scale is the minimiser of the negative of the log-likelihood,

\[
(a_t, s_t) = \arg\min_{(a_t, s_t)} \left\{ -\sum_{i=1}^{N_t} \ln f \left( \frac{r_{it} - a_t}{s_t} \right) - \ln s_t \right\}
\]

The first-order conditions are:

\[
\sum_{i=1}^{N_t} \left[ f \left( \frac{r_{it} - a_t}{s_t} \right) \frac{r_{it} - a_t}{s_t} \right] = 0
\]

\[
\sum_{i=1}^{N_t} \left[ f \left( \frac{r_{it} - a_t}{s_t} \right) \left( \frac{r_{it} - a_t}{s_t} - 1 \right) \right] = 0
\]

Therefore, the MLE estimates \(\hat{a}_t\) and \(\hat{s}_t\) are numbers such that these two equalities are satisfied.

As far as robustness is concerned, MLE is generally non-robust. In fact, the robustness of the estimator is determined by the properties of the density function \(f(x)\). For example, if the normalised residual is assumed to be normally distributed, i.e. \(f(x) = (2\pi)^{-1/2}\exp(-x^2/2)\), then solving the equation above leads to the classical estimators of location and scale,

\[
\hat{a}_t = \frac{1}{N_t} \sum_{i=1}^{N_t} r_{it}
\]

\[
\hat{s}_t = \left( \frac{1}{N_t} \sum_{i=1}^{N_t} (r_{it} - \hat{a}_t)^2 \right)^{1/2}
\]

which are not robust.\(^6\) Assuming, however, that the normalised residual is generated from a Laplace distribution, also known as the double exponential distribution, i.e. \(f(x) = \exp(-|x|)/2\), then the MLE becomes equal to

\[
\hat{a}_t = \text{med}(r_{it})
\]

\[
\hat{s}_t = \frac{1}{N_t} \sum_{i=1}^{N_t} |r_{it} - \hat{a}_t|.
\]

In this case, the estimator of location is the median of the sample and the estimator of scale resembles the mean absolute deviation. Both estimators have better robustness properties than the classical estimators; see Maronna, Martin, and Yohai (2006). The mean absolute

\(^6\) See Maronna, Martin, and Yohai (2006) for examples and other references.
deviation is less sensitive to outliers than the classical estimator of volatility. The median is not only less sensitive to outliers than the sample average, but also exhibits very high outlier resistant properties – half of the sample needs to be polluted with arbitrarily large outliers before the median can become arbitrarily large. In contrast, only one outlier of arbitrary size causes the sample average to become of arbitrarily large.

M-estimators can be introduced by generalising the first-order conditions behind MLE. The general form of the M-estimators for location and scale is,

\[ \sum_{i=1}^{N} \psi \left( \frac{r_{it} - \alpha_t}{\sigma_t} \right) = 0 \]
\[ \sum_{i=1}^{N} \chi \left( \frac{r_{it} - \alpha_t}{\sigma_t} \right) = 0 \]

where the functions \( \psi \) and \( \chi \) should satisfy some fairly general conditions and may not be related to each other. Typically, some symmetry properties are assumed, \( \psi(-x) = -\psi(x) \) and \( \chi(-x) = \chi(x) \). Under general assumptions for these two functions, the M-estimator is consistent and asymptotically normal.\(^7\) For the case of MLE, \( \psi(x) = -d \ln f(x)/dx \) and \( \chi(x) = x \psi(x) \) which demonstrates that MLE is an M-estimator. The particular choice of \( \chi(x) = x \psi(x) \) is common in practice. The function \( \psi(x) \), however, does not need to be related to any probability density function. In fact, \( \psi(x) \) is often defined as the first derivative of a general penalty function, we come back to this question later.

A standard transformation is often employed to express \( \psi \) in a form that allows for a representation through non-normalised weights. Denote \( v(x) = \psi(x)/x \) and rewrite the system as

\[ \sum_{i=1}^{N} v_i (r_{it} - \alpha_t) = 0 \]
\[ \sum_{i=1}^{N} \left[ \frac{v_i (r_{it} - \alpha_t)^2}{\sigma_t^2} - 1 \right] = 0 \]

where \( v_i = \psi(\varepsilon_{it})/\varepsilon_{it} \) in which \( \varepsilon_{it} = (r_{it} - \alpha_t)/\sigma_t \). The numbers \( v_i \) can be interpreted as non-normalised weights\(^8\) that depend on the choice of the function \( \psi(x) \) and the size of the normalised residual. Note that this transformation relies on the choice \( \chi(x) = x \psi(x) \) but there are no limits on the choice of \( \psi(x) \). It is straightforward to derive an estimator for the location parameter from the first equation and an estimator for the scale parameter from the second equation,

\[ \hat{\alpha}_t = \frac{1}{\sum_{i=1}^{N} v_i} \sum_{i=1}^{N} v_i r_{it} \]
\[ \hat{\sigma}_t^2 = \frac{1}{\sum_{i=1}^{N} v_i} \sum_{i=1}^{N} v_i (r_{it} - \hat{\alpha}_t)^2 \]

in which \( v_i \) is defined as above. If we rewrite the estimators in terms of normalised weights, \( w_i = v_i / \sum_{i=1}^{N} v_i \) then we obtain a form very similar to the estimator in Garcia, Mantilla-Garcia and Martellini (2013),

\[ \hat{\alpha}_t = \frac{1}{\sum_{i=1}^{N} w_i} \sum_{i=1}^{N} w_i r_{it} \]
\[ \hat{\sigma}_t^2 = \frac{1}{\sum_{i=1}^{N} w_i} \sum_{i=1}^{N} w_i (r_{it} - \hat{\alpha}_t)^2 \]
2. Intuition and Links to Other Estimation Techniques

where the constant \( C = \frac{N_t}{N_t} \). Finally, we can accept \( s_{t}\) as a biased estimator of the cross-sectional variance.\(^9\)

\[ \text{CSV}_{t,w}^2 = \frac{s_{t}^2}{N_t}. \]

Intuitively, the bias appears because the weights neutralise the effect of large observations. In the absence of contamination, if the sample size increases indefinitely, then \( s_{t}^2 \) converges to a number which is smaller than cross-sectional variance. We come back to the question of bias and the large-sample properties in the next section.

Finally, note that the form of the scale estimator depends on the choice \( \chi(x) = \psi(x)x^{-1} \) for the function \( \chi(x) \) in the general M-estimator. In contrast, \( \hat{a}_t \) has the form of a weighted average for any choice of \( \psi(x) \) in Eq. 7. This observation implies that the function \( \psi \) can be chosen freely without an impact on the functional form of the estimator. It does, however, influence the properties of robustness and can be represented through a penalty function by recourse to regression analysis.

**A link to the least squares technique**

A link exists also between the M-estimation technique and the least-squares regression. In the context of our problem, we can think of Eq. 2 as a regression with no explanatory variables. The least-squares estimation method implies minimising the sum of squares of the normalised residual, which results in \( \psi(x) = x \) and \( \chi(x) = x^2 - 1 \). To preserve consistency with the least-squares approach, the function \( \psi \) is usually chosen so that it behaves like \( \psi(x) = x \) close to the origin and then either flattens out or re-descends and approaches zero. Such choices translate to weights close to \( w_t = 1/N_t \) for normalised residuals close to the origin and weights converging to zero for normalised residuals far out in the tails.

In the robust regression literature, the estimation problem is posed in terms of minimising a general penalty function \( \rho \). In those problems, the scale is often regarded as a nuisance parameter needed only to properly calibrate the sample to the choice of \( \rho \).\(^10\) A version of the minimisation problem, posed for the representation in Eq. 2, takes the form,

\[ \left( \hat{a}_t, \hat{s}_t \right) = \arg\min_{(a_t,s_t)} \sum_{t=1}^{N_t} \rho \left( \frac{r_{it} - a_t}{s_t} \right) s_t \]

The first order conditions for this problem take the form of Eq. 7 with \( \psi(x) = \rho(x) \) and \( \chi(x) = x\psi(x) - \rho(x) \). Repeating the reasoning behind Eq. 9 leads to a CSV estimator with the same general form but with a different normalising constant \( C \).

In a similar vein, the estimator in Eq. 9 can be cast in terms of the general minimisation problem,

\[ \left( \hat{a}_t, \hat{s}_t \right) = \arg\min_{(a_t,s_t)} \left\{ \sum_{t=1}^{N_t} \rho \left( \frac{r_{it} - a_t}{s_t} \right) + \ln s_t \right\} \]

where \( \rho(x) \) is a general penalty function which is linked to \( \psi \) by the same relationship \( \psi(x) = \rho(x) \). The general theory behind M-estimation, however, does not depend on re-formulations involving

---

\(^9\) A scale estimator is not necessarily an estimator of volatility. See the next section for additional details on the bias.

\(^10\) See Section 4 for additional comments. In regression analysis, we are genuinely interested in the slope and the intercept coefficients. As a consequence, the scale of the residual is important only because it controls the penalty assigned to each observation. For this reason, the usual strategy is to adopt an estimator of scale that has a very high breakdown point, e.g. the median of the residual in absolute value.
optimisation problems. For the sake of completeness, note that the optimisation problem behind MLE corresponds to this one when the penalty function is \( p(x) = -\ln(f(x)) \).

**Concluding remarks on MLE and statistical robustness**

In practice, the robustness properties of M-estimators materialise because the function \( \psi(x) \) may level off or re-descend for larger \( x \). This translates into a decaying weight \( \psi(x)/x \) for larger \( x \). It is interesting that these properties may hold for MLE for densities with a power type of tail decay. For example, if \( f(x) = 1/(\pi(1 + x^2)) \) is the Cauchy density, then the corresponding \( \psi(x) = -d \ln(f(x))/dx = 2x/(1 + x^2) \). This function is re-descending as it converges to 0 for large \( x \) and the weighting function equals \( v(x) = \psi(x)/x = 1/(1 + x^2) \).

From an intuitive viewpoint, a density function with fat tails "anticipates" a small proportion of the sample to deviate significantly from the centre of the distribution and appropriately "discounts" the impact of these observations in the estimator by assigning smaller weights to them. In contrast, if the working assumption is the Gaussian distribution, then the outliers are not regarded as tail events because the Gaussian density does not "anticipate" any tail events. Of course, the reasoning behind MLE implies absence of contamination. In the presence of contamination, some observations in the cross-section of returns are either a result of error or are driven by some unrecognised factor or are a result of a different stochastic mechanism. Therefore, assuming a common density for all observations would be a clear misspecification. A better approach is to adopt a methodology that can reject those small in number outliers as non-informative or at least limit their impact.

It is well known that the good large-sample properties of MLE hold only if the data behaves exactly according to the hypothesised reference model and may fail dramatically if the data is in a neighbourhood of the reference model. This is especially true for the property of efficiency – ML estimators may not have minimal asymptotic variance in the presence of contaminants. For example, although the classical estimators of mean and volatility in Eq. 6 are model free, they have good large-sample statistical properties under the assumption of the Gaussian distribution because they arise as ML estimators under this assumption. Both estimators may have a terrible performance in the presence of contamination.

Finally, we should point out that in reality outliers may or may not be informative. Clearly, in the context of a contamination model, we assume that they are non-informative. The general philosophy is that a small minority of observations should not exercise a dominant effect on the estimated parameter values. Thus, a contamination model would be consistent if we find empirically a small proportion of abnormally large observations in absolute value. In contrast, if a sample includes a large proportion of abnormal observations, then finding an explanation for their presence would be more appropriate. Clearly, it is impossible to derive a critical value for this proportion because it would also depend on the sample size, 20% outliers in a sample of
2. Intuition and Links to Other Estimation Techniques

10 means 2 extreme observations that could be a result of error while in a sample of 1,000 it implies 200 outliers. See the discussion in Huber and Ronchetti (2009, Chapter 1) for additional details. As far as cross-sectional returns are concerned, the proportion of detected outliers does not exceed 10% for the S&P 500 universe in a period of more than 30 years and in extended periods of time stays below 5%; see Section 4. As a result, we choose an estimation technique that focuses on limiting the impact of outliers rather than building a mixture model to explain them.
3. M-estimation of CSV: Bias, Large-sample Theory, and the Algorithm
3. M-estimation of CSV: Bias, Large-sample Theory, and the Algorithm

Bias and large-sample theory
Garcia, Mantilla-Garcia and Martellini (2013) study the asymptotic properties of the estimator in Eq. 1 with arbitrary weights. In the case of Eq. 4, however, we need more general arguments because the weights are not constants invariant of the sample. In fact, the goal of the estimator is to modify the weights so that every observation gets a weight depending on the extent to which it is an outlier. As a result, the weights in Eq. 4 depend on the particular sample and are, therefore, random quantities from a statistical perspective.

In Section 4, we invoke different limit results which illustrate the properties of the estimator in Eq. 4 as the number of stocks in the cross-section increases indefinitely. We demonstrate that Eq. 4 represents a biased estimator of volatility with nice large-sample properties.

Before going to the particular results, note that location and scale are general concepts. The mean and the volatility of a distribution are just two characteristics describing location and scale. Other characteristics describing scale include the mean absolute deviation, the inter-quantile range and so on. For some distributions, volatility may not be a well-defined concept while other characteristics describing scale can be – there are distributions with infinite volatility while the inter-quantile range, however, is always finite.

As a consequence, any attempt to estimate volatility through a robustified estimator, which by construction deviates from the classical sample estimator, is bound to result in a bias unless the setting is limited to a given parametric family of distributions – a very restrictive hypothesis for the cross-section of returns. The M-estimator for location and scale in Eq. 7 is asymptotically unbiased and consistent estimator of the characteristics $a(F)$ and $s(F)$ solving

\[
\int_{\mathbb{R}} \psi \left( \frac{x - a(F)}{s(F)} \right) dF(x) = 0
\]

\[
\int_{\mathbb{R}} x \left( \frac{x - a(F)}{s(F)} \right) dF(x) = 0,
\]

accepting $\hat{a}$ and $\hat{s}$ as estimators of mean and volatility leads to a bias. This implies that the CSV estimator in Eq. 11 is a biased estimator of cross-sectional volatility. For some distributions, $s(F)$ may happen to coincide with volatility but this is not true in general.

It is impossible to determine the size of the bias in a general setting – as a matter of fact, the bias is arbitrarily large for distributions with infinite volatility. This case, however, is not realistic for financial returns. Because one of the goals in the construction of the robust estimator is to maintain the functional form of a weighted average with the weights deviating from $1/N_t$ only for the extreme outliers, it is clear that the bias of the CSV estimator will increase with tail thickness. This is not necessarily true for $\hat{a}$ – a symmetric fattening of the tails is not expected to increase the bias for odd $\psi$ functions.

A reasonable bias correction strategy is to calibrate the M-estimator of scale for a particular distribution $F$ so that
the $s(F)$ solving the second equation above is indeed the volatility of that distribution. Thus, in the absence of contamination, there is no bias. In contrast to the estimator of volatility, which is not robust, a robust estimator of scale is weakly continuous – “small variations” in the distribution $F$ do not lead to big deviations in scale. Therefore, a calibration to a particular $F$ would only provide a point of reference with the continuity property guaranteeing small deviations in CSV when the distribution varies in a small neighbourhood of $F$. In other words, adding a small amount of contamination to $F$ does not change $s(F)$ much and it stays close to its true value, the volatility of $F$.

There could be different strategies for the calibration. One strategy is to look for consistency with the Gaussian case; that is, if $F$ is the cdf of the Gaussian distribution, then the M-estimator of scale converges in almost sure sense to the volatility of the Gaussian law. This would lead to underestimation of volatility for any distribution with fatter tails than the normal.

Another strategy could be to choose an “average” distribution from a frequently used parametric family. In this way, in cases where the tails are thinner, volatility will be overestimated and in cases where the tails are fatter volatility will be underestimated. It is, however, difficult to choose an “average” model for the cross-section of stock returns.

In any case, a bias correction is not going to modify the behaviour of the scale estimates in the time domain. A calibration to a particular distribution results in,

$$ CSV_{t,w}^2 = K^2 s^2 e $$

where the constant $K$ can be calculated numerically. Note that the particular value of the constant depends also on the choice of $\psi$.

As far as large sample theory is concerned, M-estimators are unbiased and asymptotically normal. The asymptotic variance is determined by the corresponding integrand, e.g. $\psi$. Under general regularity conditions, the M-estimator $\hat{\theta}$ solving

$$ \sum_{i=1}^{n} \psi_0(x_i, \theta) = 0 $$

is a consistent and asymptotically normal estimator of $\theta$. The asymptotic variance is equal to

$$ V(\psi_0, F) = \left( \frac{\int_R \psi_0^2(x, \theta) dF(x)}{\left( \int_R \frac{\partial}{\partial \theta} \psi_0(x, \theta) dF(x) \right)^2} \right)^2 $$

i.e. $n^{1/2}(\hat{\theta} - \theta) \rightarrow N(0, V)$. For additional technical details, see Huber and Ronchetti (2009), Schevlyakov et al. (2008) who focus on re-descending $\psi_0$ functions, and also Maronna, Martin, and Yohai (2006).

The algorithm

The intuition and the properties of M-estimators hold for a general function $\psi(x)$. In an implementation, however, we have to choose a particular $\psi(x)$ or $\rho(x)$, respectively. In the academic literature, many classes of functions have been

---

3. M-estimation of CSV: Bias, Large-sample Theory, and the Algorithm

---

11 - See Huber and Ronchetti (2009) for formal definitions of a neighborhood of in the space of distributions and the notion of continuity.
3. M-estimation of CSV: Bias, Large-sample Theory, and the Algorithm

suggested such as the Huber function, the Talwar function, etc. The choice of $\psi(x)$ for the purposes of CSV estimation is driven by two principles:

- The estimator should resemble the equally weighted estimator discussed by Garcia, Mantilla-Garcia and Martellini (2013) when the normalised cross-sectional returns do not qualify as outliers; that is, $w_t = 1/N_t$ in the notation in Eq. 1. This corresponds to a non-normalised weighting function $\psi(x)/x = 1$, or $\psi(x) = x$ equivalently, for values of the argument close to the origin.

- The impact of big outliers should be neutralised which is achieved if the function $\psi(x)$ not only levels off but re-descends to zero for larger values of the argument. Intuitively, this leads to a non-normalised weighting function that turns into zero for large enough $x$. In terms of the penalty function $\rho(x)$, this condition implies a bounded penalty function. In this way, the impact of severe outliers is almost completely neutralised while the information from medium-sized observations is incorporated.

The two principles are insufficient to uniquely pin down a function $\psi(x)$. For this reason, we resort to a method known as optimal robustness that derives a function $\psi(x)$ based on some optimality conditions. In this case the optimality conditions are balancing between bias and asymptotic variance of the robust estimator in a contamination neighbourhood. Generally, the objective is to find a $\psi(x)$ that leads to an estimator with minimal asymptotic variance subject to a constraint on the maximal bias in a contamination model.\(^\text{12}\)

The problem of optimal balance has been studied in different contexts in the robust statistics literature, we choose the one posed in a regression analysis setting, for further details see Maronna, Martin, and Yohai (2006, Section 5.9.1). We choose a function $\psi(x)$ that belongs to the class of re-descending functions, which has been constructed to provide an optimal balance between bias and efficiency.\(^\text{13}\)

$$\psi(x) = \text{sgn}(x) \left( -\frac{\varphi'(|x|)}{\varphi(|x|)} \right)^+$$

where $\text{sgn}(x)$ denotes the sign of $x$, $\varphi$ is the standard normal density, $t^+ = \max(t, 0)$, and $c$ is a constant determining the optimal properties of the estimator.

The weighting function $\psi(x)/x$ function is defined in Eq. 5 but we do not use it directly in the algorithm. Instead, we use the following polynomial approximation of the corresponding penalty function Eq. 15

$$\rho(x) = \begin{cases} 
\frac{x^2}{2}, & \text{if } |x| \leq 2k \\
f(x), & \text{if } 2k < |x| \leq 3k \\
3.25k^2, & \text{if } |x| > 3k
\end{cases}$$

where

$$f(x) = k^2 \left( 1.792 - 0.972 \left( \frac{x}{k} \right)^2 + 0.432 \left( \frac{x}{k} \right)^6 \
- 0.052 \left( \frac{x}{k} \right)^8 + 0.052 \left( \frac{x}{k} \right)^{10} \right)$$

and $k$ is a parameter. The non-normalised weighting function $\psi(x)/x = \rho'(x)/x$ is equal to 1 if the normalised residual is smaller than $2k$ in absolute value and smoothly converges to zero for values beyond that interval (a plot is provided in Figure 2). The parameter $k$ controls the relative efficiency, we
choose $k = 1.25$, see Scherer and Martin (2005) and the references therein. With this choice of $\rho$, the constant in Eq. 13 calculated numerically with 10 million simulations from a Gaussian distribution equals $K = 1.01462925$ which implies a bias correction of about 1.5%.

Obviously, the non-normalised weights $\nu_i$ depend on the values of location and scale the estimators of which, in turn, depend on $\nu_i$. Thus, we have to rely on an iterative procedure that should converge because of the general theory behind M-estimation. We need to calculate, however, starting values for the location and scale. To this end, we choose simpler robust estimators based on the median, in which $r$ denotes the cross-section of returns. The algorithm follows below.

**Step 0.** Calculate $a_0$ and $s_0$. Set a tolerance of $tol = 10^{-8}$ to be used in calculating the improvement between two consecutive approximations. It is used as an exit condition together with a maximum number of iterations $M = 50$.

**Step 1.** Set $a_{prev} = a_0$ and $s_{prev} = s_0$ and $iter = 1$.

**Step 2.** Calculate $\nu_i = \psi(\hat{\epsilon}_it)/\hat{\epsilon}_it$ in which $\hat{\epsilon}_it = \frac{r_{it} - a_{prev}}{s_{prev}}$.

**Step 3.** Calculate $a_{next}$ and $s_{next}$ from Eq. 9.

**Step 4.** If $iter > 50$ or $\max(|a_{prev} - a_{next}|, |s - s_{next}|) < tol$ stop, multiply $s_{next}$ by $K$ given in Eq. 13 and accept the value as an estimate of $CSV_{r,wt}$. Otherwise, go back to Step 1.

The algorithm is quite stable and converges quickly. For the S&P 500 universe and 20 years of daily data, the average number of iterations is 15.6 and the 95% confidence interval is [9.5, 28].

---

**Figure 2.** The non-normalised weighting function $v(x) = \rho'(x)/x$ in which $\rho(x)$ is given in Eq. 15. $v(x)$ smoothly declines from 1 to 0 symmetrically for positive and negative values of the argument. $v(x) = 1$ leads to the equally weighted CSV estimator.
3. M-estimation of CSV: Bias, Large-sample Theory, and the Algorithm

Figure 3. Boxplots of cross-sections of returns of the S&P 500 universe in four turbulent months together with annualised cross-sectional volatility as estimated by the classical estimator denoted by EW CSV and the robust method denoted by Robust CSV.
Apart from CSV, complementary statistics can also be calculated. One useful statistic is the standard deviation of the estimator based on the asymptotic theory – we can adapt Eq. 14. For the case of scale, \( \psi_0 = \chi \) and the expression becomes,

\[
\begin{align*}
V(\chi, F) &= \frac{s^2 \int_{\mathbb{R}} \chi^2 \left( \frac{x-a}{s} \right) dF(x)}{\left( \int_{\mathbb{R}} \chi' \left( \frac{x-a}{s} \right) \frac{x-a}{s} dF(x) \right)^2}
\end{align*}
\]

where \( \chi(x) = \psi(x)x - 1 \) and \( \chi'(x) = \psi'(x)x + \psi(x) \). The asymptotic variance can be estimated from the sample by

\[
\begin{align*}
V(\chi, F) &= \frac{s^2 \sum_{i=1}^{N_h} \chi^2 \left( \frac{t_n - \bar{t}_n}{s_n} \right)}{\left( \sum_{i=1}^{N_h} \chi' \left( \frac{t_n - \bar{t}_n}{s_n} \right) \frac{t_n - \bar{t}_n}{s_n} \right)^2}
\end{align*}
\]

To obtain the standard deviation of the estimator for a sample of size \( N_h \), we calculate \( \sqrt{V(\chi, F)} / N_h \).

To numerically illustrate the value-added of this approach, we provide boxplots and annualised CSV calculated for different months for the S&P 500 universe including both turbulent and calm periods. The plots in Figure 3 demonstrate that the robust methodology results in smoother time-series of cross-sectional dispersion. As far as the classical estimator is concerned, in some days its behaviour is driven entirely by a few isolated outliers. The calculation of the robust CSV on the plots follows the algorithm outlined in this section.
3. M-estimation of CSV: Bias, Large-sample Theory, and the Algorithm
4. A Comparison with a Popular Quantile-based Estimator – Asymptotic Variance and Breakdown Points
4. A Comparison with a Popular Quantile-based Estimator – Asymptotic Variance and Breakdown Points

The robust regression literature provides many different robust estimators of scale and location. It is a daunting task to compare all of them in the context of CSV estimation for different universes and choose the one that works uniformly well. Even if this was possible, most likely we would end up with a large number of estimators that have similar performance and the choice would again be subjective.

From a practical perspective, it makes sense to choose an estimator that has appealing theoretical properties but which is also intuitive. In the case of Eq. 4, we noted that our goal is to stay close to the equally weighted estimator, because it has attractive interpretation, while robustifying its performance.

Nevertheless, there are other popular estimators with good robust properties having a functional form quite different from the one in Eq. 1. One such popular estimator studied extensively by Pearson and Tukey (1965) is

\[
\hat{s}_q = \frac{\hat{q}_{0.95} - \hat{q}_{0.05}}{3.25}
\]

where \( \hat{q}_p \) is the empirical p-quantile. In this section, we compare the estimator in Eq. 4 to the one in Eq. 17 in terms of their asymptotic variance and an important characteristic of robustness called breakdown point. It is defined as the smallest proportion of abnormally large outliers that could result in an explosion of the estimator given a sample of certain size.

Pearson and Tukey (1965) studied \( \hat{q}_p \) in the context of Pearson’s family of distributions which includes distributions such as Student’s t and the log-normal used frequently in finance. They concluded that \( \hat{q}_p \) is remarkably close to standard deviation of the distributions in Pearson’s family but in contrast to the classical estimator of volatility, it is much more robust.

Apart from the analysis in Pearson and Tukey (1965), quantile-based estimators are sometimes used in the modern robust statistics literature only to remove the effect of the scale through re-normalisation. In some applications, the scale of the sample is often regarded as a nuisance parameter. This is the case of pure location M-estimation problems or robust regression problems – obviously the scale of the sample is not in focus but robust procedures depend on the scale, hence the issue. Since the focus in those problems is on other parameters, the general idea is to remove the effects of the scale of the sample through re-normalisation. As a consequence, the most important property of the scale estimator is to have a high breakdown point; that is, the robustness of the robust regression coefficients, for instance, should not critically depend on the stability of the scale estimator. Popular choices for scale estimators in this case include the median of the absolute deviation and also the inter-quartile range – a version of the estimator in Eq. 17. In the case of CSV, however, the scale is not a nuisance parameter – we have a pure scale estimation problem – which makes the problem more challenging. Nevertheless, the breakdown point
4. A Comparison with a Popular Quantile-based Estimator – Asymptotic Variance and Breakdown Points

remains an important characteristic of outlier resistance although not a goal in itself.

Both asymptotic variance and the breakdown point can depend on the distribution of the sample and to carry out a comparison we need to make an assumption. Although the distributional properties of the cross-section of stock returns have not been extensively studied and it is not clear if Student’s t distribution is a good model, we compare the asymptotic properties of the estimator in Eq. 4 and $q_p$ assuming the sample follows this distribution because the analysis of the attractiveness of $q_p$ was first done in the context of Pearson’s family which includes Student’s t distribution.

After comparing the asymptotic variance, we move on to the breakdown point which we consider again for Student’s t and also for the empirical cross-sectional distribution of S&P 500 in the past 20 years. We focus on S&P 500 universe because long time series are easily available. We expect similar results to hold for other universes.

**Asymptotic variance**

Asymptotic variances of any pair of scale estimators are poor criteria for comparison because of the arbitrary standardisation of the scale estimators, i.e.

$$n^{1/2}(S(F_n) - S(F)) \rightarrow N(0, V_\alpha)$$

the asymptotic variance of the scaled logarithm,

$$n^{1/2}\log\left(\frac{S(F_n)}{S(F)}\right) \rightarrow N\left(0, \frac{V(S)}{S^2(F)}\right).$$

The idea is that $\log(S(F_n))$ transforms any fixed re-scaling of the estimator into a summand which disappears after centering. The corresponding quantities for the two estimators are calculated in the following equations,

$$V(S_q) = \frac{2t(1-2t)}{[2F^{-1}(t)f(F^{-1}(t))]^2} \quad \text{for } t = 0.05$$

in which $f(x)$ is the density of $F$, assuming it exists, and

$$\frac{V(S_M)}{s^2} = \frac{V(\chi, F)}{s^2}$$

where $V(\chi, F)$ is given in Eq. 16 and $s^2$ is obtained by solving Eq. 12. Note that the scaling constant in Eq. 17 does not appear in Eq. 18 which is a consequence of using the log-transform.

The asymptotic variances in Eq. 18 and Eq. 19 depend on the underlying distribution. As mentioned above, we compare the performance of the two estimators for Student’s t distribution for various degrees of freedom. The calculation of Eq. 19 is performed through the Monte Carlo method with 10 million simulations.

The results are illustrated in Figure 4. The asymptotic standard deviation of the M-estimator is between 10% and 20% smaller if the degrees of
freedom parameter is larger than 5. This numerical result indicated better large-sample properties of the M-estimator.

It is interesting that the M-estimator has better large-sample properties for higher values of the degrees of freedom parameter which means for distributions closer to the Gaussian distribution. Although not reported here, we repeated the same calculation for samples of 100 observations drawn again from the Student’s t distribution. We used 2,000 samples of 100 observations to calculate the variances of the corresponding scaled logarithm. The calculation confirmed the results reported in Figure 4 hold also for the small sample case almost without change.

### Breakdown points

The breakdown point of an estimator characterises its resistance to outliers. It is formalised as the smallest proportion of abnormally large outliers that would result in an explosion of the estimator given a sample of size $N_t$. Clearly, the highest possible value for that proportion is 50% because if more than half of the

---

Figure 4. The asymptotic standard deviation of the scaled logarithm of quantile-based and the M-estimator of scale and the percentage difference between them. The M-estimator has about 20% smaller asymptotic standard deviation.
4. A Comparison with a Popular Quantile-based Estimator – Asymptotic Variance and Breakdown Points

sample is contaminated, there is no way to discriminate between "good" and "bad" data points. The sample average and the sample standard deviation both have a breakdown point of $1/N_t$; it is sufficient for only one data point to become arbitrarily large in order for them to explode.

Although the quantile-based estimator is robust, it has relatively low resistance to outliers. Its breakdown point is 5%; i.e. if more than 5% of the sample is corrupted, the estimator may be driven by the contamination and may take an arbitrary value if the contamination is of arbitrary size.

There are different opinions on whether an estimator should have a breakdown point close to the maximum of $\frac{1}{2}$. Huber and Ronchetti (2009) comment that a breakdown point above 25% is desirable for small samples in order to remove the effects of a few aberrant values. Although the breakdown point is not supposed to be an end in itself, it is a useful characteristic. For the purpose of CSV estimation, we are not so much interested in the very small sample case (e.g. 10 observations). Typical samples include between 50 and 500 stocks for single country universes and between 300 and 1,000 for regional universes.

To study the breakdown point, we adopt the following classical contamination model,

$$F(x) = (1 - \varepsilon)F_0(x) + \varepsilon H(x)$$

where $F_0$ is the cdf of the real model and $H$ is the cdf of the contamination. Very often $H$ is assumed to have a point mass at infinity. Clearly, if $\varepsilon > 0.05$, the quantile based estimator may explode. The calculation for the M-estimator is less-straightforward. In fact, re-writing the system in Eq. 12 for the contamination model yields,

$$(1 - \varepsilon) \int_{-\infty}^{\infty} \psi \left( \frac{x - \alpha}{s} \right) dF_0(x) + \varepsilon \int_{-\infty}^{\infty} \psi \left( \frac{x - \alpha}{s} \right) dH(x) = 0$$

and

$$\int_{-\infty}^{\infty} \chi \left( \frac{x - \alpha}{s} \right) dF_0(x) + \varepsilon \int_{-\infty}^{\infty} \chi \left( \frac{x - \alpha}{s} \right) dH(x) = 0.$$ 

where we have suppressed the dependence of $\alpha$ and $\sigma$ on $F$. Assuming that $H$ has a point mass at infinity and noticing that $\psi(\infty)=0$ and $\chi(\infty)=-1$, we obtain

$$\int_{-\infty}^{\infty} \psi \left( \frac{x - \alpha}{s} \right) dF_0(x) = 0$$

and

$$(1 - \varepsilon) \int_{-\infty}^{\infty} \chi \left( \frac{x - \alpha}{s} \right) dF_0(x) = \varepsilon.$$ 

The analysis in Huber and Ronchetti (2009, Section 6.6) indicates that the breakdown point of the scale estimator is lower than the one for location and, as a consequence, we can focus only on the second equation. Indeed, for any finite value of $s$, the first equation can always be solved for $\alpha$ due to the trimming property of $\psi$. Further on, it does not depend on $\varepsilon$.

Consider the second equation. Because the function $\psi$ is such that extreme outliers are practically trimmed, the reason for breakdown can only be due to no solution for $s$, which implies convergence to 0 in the iterative procedure. The selection of median values to estimate initial values for $\alpha$ and $\sigma$ implies that for any $\varepsilon < 0.5$, the starting values would always be finite. We rewrite the second equation in the following way, which proves more convenient,
4. A Comparison with a Popular Quantile-based Estimator – Asymptotic Variance and Breakdown Points

Eq. 20
\[ (1 - \varepsilon) \frac{1}{s^2} \int_\mathbb{R} \varphi \left( \frac{x - a}{s} \right) (x - a)^2 dF_0(x) = 1 \]

where \( \varphi(x) = \psi(x)/x \). The non-normalised weighting function is equal to 1 for \( x \) between \(-2k\) and \( 2k \) and, smoothly converges to 0 for \( 2k < |x| < 3k \) and is equal to zero for \( |x| > 3k \). Whether this equation can be solved for \( s \) for a given value of \( \varepsilon \) depends on whether the maximum of the left hand-side can reach 1. Some simple analytic properties of the left hand-side are collected in the proposition below.

Proposition 1. Consider the function
\[ I(s) = \frac{1}{s^2} \int_\mathbb{R} \varphi \left( \frac{x - a}{s} \right) (x - a)^2 dG(x) \]

The proof is provided in Appendix 1. The value of the maximum cannot be calculated explicitly without assuming a particular distribution. It is, however, clear that the maximum contamination in this model equals
\[ \varepsilon^* = \frac{l_{\max} - 1}{l_{\max}} \]

where \( l_{\max} = \max_s I(s) \). It is possible to calculate the value analytically for a class of distributions. Intuitively, the more "regular" the distribution, the higher the breakdown point will be.

An illustration is provided on Figure 5 when \( F_0 \) has a Student’s t distribution and \( a = 0 \). The breakdown point for the Cauchy case is 31%.

The finite sample case follows directly when \( G \) is the sample cdf. Eq. 20 transforms to

Eq. 21
\[ (1 - \varepsilon) \sum_{i=1}^N \varphi \left( \frac{x_i - a}{s} \right) \left( x_i - a \right)^2 = 1 \]

The maximum value of the sum and the corresponding breakdown point can be calculated numerically. Note that this

Figure 5. The breakdown point calculated numerically for Student’s t distribution as \( F_0 \). It quickly approaches 50%. For distributions as bad as the Cauchy (dof = 1), the breakdown point is 31%.
4. A Comparison with a Popular Quantile-based Estimator – Asymptotic Variance and Breakdown Points

calculation would imply that we believe all observations in the sample are generated by the correct model which is not the case. In this way, however, we get an intuition about the breakdown point for the particular (already contaminated) sample.

An empirical calculation based on S&P 500 universe is included in Figure 6. The breakdown point stays above 30%.

Finally, note that the behaviour of the left hand-side of Eq. 20 implies that there are two, one, or no solutions to the equation. The case of no solution is uninteresting. The case of one solution implies we are at the breakdown point of the estimator. The case of two solutions is the typical one. One of the solutions is attained for \( \sigma \) very close to zero implying most of the normalised observations are beyond the +/- 3\( k \) range and \( a \) is close to the median. The second solution is attained for \( \sigma \) closer to the corresponding \( C_a \) value.

Although not explicitly clear, it is possible to demonstrate that the algorithm outlined in Section 4 converges to the local solution. When the starting values are selected as in the algorithm, then the procedure converges to the bigger root unless there is a significant number of zeros in the sample (more than 50%) which would lead to the initial value for the scale being exactly zero. This is, however, extremely unlikely for universes of liquid stocks.

Figure 7 shows a time series of the outliers as a percentage of the sample size for the S&P 500 universe from 31-Jan-1980 to 29-Jun-2012. Values above 5%, which is the breakpoint of the quantile estimator, can occur. We use the weighting function of the M-estimator as a criterion for outlier identification – if an observation gets a weight of zero it is identified as an outlier based on the grounds that the M-estimator ignores these observations altogether.

Figure 6. The breakdown point of the CSV estimator calculated numerically for cross-sections of returns of the S&P 500 universe from 31-Jan-1980 to 29-Jun-2012.
4. A Comparison with a Popular Quantile-based Estimator – Asymptotic Variance and Breakdown Points

Figure 7. The number of observations getting a weight of zero in the M-estimator as a percentage of the sample size. The value of 5% is the theoretical breakdown point of the quantile-based estimator.
5. Implications for CSV Estimation for Universes of Different Size and for Returns of Different Frequency
5. Implications for CSV Estimation for Universes of Different Size and for Returns of Different Frequency

Although the interpretation of CSV as a proxy for the average idiosyncratic volatility of the universe results from an asymptotic argument, the CSV as a quantity can be calculated for any universe. Important market indices such as FTSE 100, DAX 30, EUROSTOXX 50, KOSPI 200 are based on universes much smaller in some cases than the one of S&P 500. It is, therefore, of practical interest to verify if the robustness properties would be more important for relatively smaller or for relatively larger universes. Intuitively, if the sample is small, then only a few outliers can cause serious estimation problems. Thus, we would expect the question of robustness to be more important for relatively smaller universes. Likewise, CSV can be calculated from cross-sectional returns of different frequency. Empirically, it is of interest to verify if the standard estimator becomes more reliable for lower frequencies.

We study the impact of the size of the universe with US stocks by looking at sub-samples of the S&P 500 universe. The results may differ if we change the universe to be from an emerging market rather than a developed one but, if anything, the properties of robustness are expected to be more important for emerging market universes because the cross-sections are expected to contain more ill-behaved returns on a relative basis.

Figure 8 includes four plots showing the number of observations getting a weight of zero in the M-estimator as a percentage of the sample size when the cross-section is based on the top 50, 100, 200 and 500 stocks ranked by market cap. The plot based on the top 500 stocks is identical to the one in Figure 7 but scaled for easier comparison to the other plots. To the extent that the numbers plotted can be regarded as a proxy for the share of outliers in the sample, the results indicate that cross-sections based on 50 returns are more likely to lead to stability problems in the classical and the quantile estimator than a cross-section of 200 or 500 returns. In other words, the share of aberrant observations tends to decrease with sample size which is expected to stabilise estimation.

To illustrate the impact of smaller cross-sections on the behaviour of the estimated CSV, we provide boxplots and estimates.
produced from the classical estimator, the quantile estimator, and the robust M-estimator in Figure 9 for the four periods covered in Figure 3. We can see that for September 2001 there is very good agreement between the three estimators. For the other months, there are days (e.g. after 22 of July 2002) in which the quantile estimator has a behaviour similar to the classical estimator and in those days the
cross-sections usually have more outliers or have more asymmetric outliers.\textsuperscript{18} On other days, the quantile and the M-estimator have similar behaviour.

Plots showing the same four periods but when the cross-section is based on the top 200 and 500 stocks by market cap are included in Figure 15 and Figure 16 in Appendix 2 and support the same conclusions. We can also notice that as the number of stocks in the cross-section increases, the behaviour of the quantile estimator becomes more similar to the robust M-estimator which is confirmed by the decrease in the share of outliers illustrated in Figure 8. Visually, the boxplots seem to have more red crosses identifying outliers, which is expected as the sample size grows. Figure 8 implies that the percentage of these outliers decreases with the sample size.

So far we have demonstrated that when calculating CSV using cross-sections of daily returns, the M-estimator has superior properties to the classical one. The same question can be posed for cross-sections of other frequencies which would typically be weekly or monthly. This adds another dimension to the discussion of the impact of the size of the universe since we can

\textbf{5. Implications for CSV Estimation for Universes of Different Size and for Returns of Different Frequency}

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{figure9.pdf}
\caption{Boxplots of cross-sections of returns of the top 50 US stocks ranked by market cap in the four turbulent months from Figure 3 together with annualised cross-sectional volatility as estimated by the classical estimator denoted by EW CSV, the robust method denoted by Robust CSV, and the quantile estimator denoted by Quantile CSV.}
\end{figure}
5. Implications for CSV Estimation for Universes of Different Size and for Returns of Different Frequency

talk about universes of 50 stocks and daily, weekly or monthly returns. In the remaining part of this section, we demonstrate that the properties of robustness can be important for both weekly and monthly returns. The significance, however, decreases with the frequency and for the 500-stock universe a simple procedure such as trimming a few of the largest and a few of the smallest observations in the sample and then applying the classical estimator seems to work quite well.

A comparison of the classical and the robust estimators for daily, weekly, and monthly frequency for the S&P 500 universe is provided in Figure 10. The three plots show that the deviations between the two estimators tend to decrease with the return frequency although even in the case of monthly returns, significant deviations are possible. This observation indicates that a robust estimator is relevant for lower frequency returns or that, alternatively, some form of data cleaning or trimming should be employed. The improvement is also clear from Figure 11 which shows the share of outliers in the sample using weekly and monthly returns and can be compared to Figure 7 which is based on the same universe, but with daily data.
To test the idea that simple trimming might work, we calculate the CSV using the classical estimator having trimmed the two largest and the two smallest returns in the sample, for more information about trimming see for example Maronna, Martin, and Yohai (2006). The result is provided in Figure 12. In the case of monthly returns for a cross-section of 500 stocks, the simple trimming combined with the classical estimator provides a significant improvement.
5. Implications for CSV Estimation for Universes of Different Size and for Returns of Different Frequency

We conclude this section by repeating the same calculation but this time with the top 50 US stocks ranked by market cap. The goal is to see if indeed for lower frequencies we observe higher proportion of outliers for a smaller universe.
5. Implications for CSV Estimation for Universes of Different Size and for Returns of Different Frequency

Figure 13 provides a comparison similar to Figure 12. Indeed, we notice a higher proportion of outliers in both the weekly and the monthly data for the top 50 universe compared to the top 500 universe. Nevertheless, we observe the same relationship as in the top 500 universe – the proportion of outliers in the cross-sections of daily returns is higher than that of the cross-sections of weekly and monthly returns (compare with Figure 8).

Repeating the trimming exercise, however, does not result in any improvement in this case; see Figure 14. Nevertheless, apart from a few data points, the classical estimator does not seem to be drastically different from the M-estimator.

*Figure 13. The number of observations getting a weight of zero in the M-estimator as a percentage of the sample size for cross-sections of weekly and monthly returns based on the top 50 US stocks ranked by market capitalisation.*
5. Implications for CSV Estimation for Universes of Different Size and for Returns of Different Frequency

Figure 14. The annualised CSV calculated by the classical and the robust method for the top 50 US stocks by market capitalisation. The trimmed EW CSV is computed by trimming the two largest and the two smallest returns in the sample and it coincides completely with the EW CSV.
5. Implications for CSV Estimation for Universes of Different Size and for Returns of Different Frequency
6. Summary
As demonstrated by Garcia, Mantilla-Garcia and Martellini (2013), cross-sectional volatility can be accepted as a proxy for idiosyncratic volatility which, together with systematic volatility, can be used as indicator of economic uncertainty. From a practical perspective, an important empirical feature of cross-sections of returns is the presence of significant outliers that pollute volatility estimates. In this paper, we consider in detail the method of simultaneous estimation of location and scale through M-estimation, which deals with the statistical issues resulting from the presence of significant outliers.

We describe the links between M-estimation and other statistical estimation approaches indicating the sources of statistical robustness. We demonstrate that M-estimation provides a natural way to extend the estimator developed by Garcia, Mantilla-Garcia and Martellini (2013) by assigning weights to the observations that depend on the degree to which a given observation in the cross-section is regarded as an outlier. The weights in the method originate from a general penalty function which is selected to be bias efficient. We describe the bias and the large sample properties of the estimator and provide the estimation algorithm. We also compare this estimator to a popular quantile-based estimator which also has robust properties, but is less resistant to outliers.

Finally, we study the behaviour of cross-sections of universes of different sizes and frequencies. The robustness properties of the CSV estimator become more important the smaller the universe is and the higher the frequency is. At the daily and weekly frequencies, the proportion of outliers is substantial and a robust estimator is preferable while at the monthly frequency the classical estimator combined possibly with mild trimming appears to have a satisfactory performance.
Appendices

Appendix 1. Proof of Proposition 1

In this appendix, we re-state and provide a proof to Proposition 1.

Proposition 1. Consider the function

$$I(s) = \frac{1}{s^2} \int_{\mathbb{R}} v\left(\frac{x-a}{s}\right)(x-a)^2 dG(x)$$

where \(a\) is a constant and \(v(x)\) has the properties described above. Assume the random variable \(X\) with a cdf \(G\) is such that \(C_a = E(X-a)^2 < \infty\). The following properties hold:

- \(I(0) = v(a) a^2 (G(a+0) - G(a))\),
- \(I(\infty) = 0\),
- \(I(s)\) is a ratio of two monotonically increasing functions.
- The maximum of \(I(s)\) is attained for \(a < \sigma < \sqrt{C_a}\).

Proof. The behaviour for small values of \(\sigma\) becomes apparent after the substitution \(x = ys + a\),

$$I(s) = \int_{\mathbb{R}} v(y) y^2 dG(y\sigma + a) ightarrow v(a) a^2 (G(a + 0) - G(a))$$

when \(\sigma\) approaches 0 because the integral converges to \(v(a) a^2\) multiplied by the jump size of \(G\) at \(a\). If \(s\) approaches infinity, then from the expression given in the proposition, it is clear that the integral converges to \(C_a\) because \(v(0) = 1\) and, therefore, \(I(s)\) converges to zero because of the explosion of the denominator. To prove the last two bullet point, calculate the derivative of the integral:

$$A = \frac{d}{ds} \int_{\mathbb{R}} v\left(\frac{x-a}{s}\right)(x-a)^2 dG(x) = \int_{\mathbb{R}} v'\left(\frac{x-a}{s}\right)\left(\frac{x-a}{s^3}\right) dG(x)$$

which, after the same substitution, becomes equal to

$$A = s \int_{\mathbb{R}} v'(y) y^3 dG(\sigma y + a)$$

The derivative of \(v\) is non-zero only in the intervals of increase and decrease of \(v\). Therefore, by recourse to the mean-value theorem, we obtain,

$$A = \sigma (v'(\xi_1)(-\xi_1)^3 - v'(\xi_2)\xi_2^3) \geq 0$$

where \(-3k \leq \xi_1 \leq -2k\) and \(-2k \leq \xi_2 \leq 3k\). Since \(v\) is increasing in the first interval and decreasing in the second, it follows that \(A\) is non-negative.

Finally, a ratio of two positive, increasing functions has one maximum. Since \(I(C_a) < 1\), the maximum has to be attained for \(0 < s < \sqrt{C_a}\).
Appendices

Appendix 2. CSV Estimated from Cross-sections of Different Sizes.
In this appendix, we provide boxplots and CSV estimated according to the classical, the quantile, and the M-estimation method. The two figures illustrate that as the size of the cross-section increases and the share of outliers decreases, the quantile estimator starts behaving more like the M-estimator.

Figure 15. Boxplots of cross-sections of returns of the top 200 US stocks ranked by market cap in the four turbulent months from Figure 3 together with annualised cross-sectional volatility as estimated by the classical estimator denoted by EW CSV, the robust method denoted by Robust CSV, and the quantile estimator denoted by Quantile CSV.
Appendices

Figure 16. Boxplots of cross-sections of returns of the top 500 US stocks ranked by market cap in the four turbulent months from Figure 3 together with annualised cross-sectional volatility as estimated by the classical estimator denoted by EW CSV, the robust method denoted by Robust CSV, and the quantile estimator denoted by Quantile CSV.
Appendices
Appendices
References
References

About Fédération Bancaire Française
The French Banking Federation (FBF) is the professional body that represents all banks operating in France. It has 390 members including French and international retail banks, as well as cooperative and mutual banking groups.

The FBF’s mission is to promote the banking and financial industry in France, Europe and around the world, in the interests of all its members, and to voice the sector’s positions, proposals and concerns to government bodies and economic and financial authorities.

It also serves as an intermediary between the banking industry and all those with an interest in banking: politicians, institutions, media, consumers, professional associations, teachers, etc.

The FBF’s missions also involve informing member banks of developments in the industry and regulatory changes, and answering questions about their activities.

The FBF banks in figures:
- 390 banks.
- More than 38,000 bank branches.
- 370,000 employees, thus making it a leading private sector employer in France.
- 23,000 new hires every year.
- Close to 78 million current accounts.*
- 1,945 billion euros in loans. **
- 1,617 billion euros in deposits. ***
- More than 17 billion payment transactions per year.
- More than 58,000 ATMs.
- Has accounted for close to 3% of national GDP over the last ten years.

To fulfil its mandate, the FBF is structured around three departments:
- Banking and Financial Activities and Research
  This department offers a full range of banking expertise, and oversees the FBF’s commissions and committees in areas including retail banking and direct banking, investment banking and capital markets, risk control and capital adequacy requirements, payment systems and instruments, and legal and tax issues. It also negotiates and works with various French and European government and regulatory authorities.
- Information and External Relations
  The role of this department is to anticipate changes in the political, economic and social environment, to promote the role of the banking sector in society and to keep the general public informed. It is responsible for public affairs, including relations with government officials, the media, consumers and young people, among others, and organises the activity of the FBF’s regional committees. It ensures that banks are kept informed about applicable or incoming regulations and about work being conducted by the FBF.
- European and International Affairs
  This department, based in Brussels and in Paris, handles relations with EU institutions and the various bodies representing the banking and financial services industries in Europe. It monitors international issues and relations with global banking associations, and deals with issues that specifically concern foreign banks operating in France.
About EDHEC-Risk Institute
About EDHEC-Risk Institute

The Choice of Asset Allocation and Risk Management
EDHEC-Risk structures all of its research work around asset allocation and risk management. This strategic choice is applied to all of the Institute’s research programmes, whether they involve proposing new methods of strategic allocation, which integrate the alternative class; taking extreme risks into account in portfolio construction; studying the usefulness of derivatives in implementing asset-liability management approaches; or orienting the concept of dynamic “core-satellite” investment management in the framework of absolute return or target-date funds.

Academic Excellence and Industry Relevance
In an attempt to ensure that the research it carries out is truly applicable, EDHEC has implemented a dual validation system for the work of EDHEC-Risk. All research work must be part of a research programme, the relevance and goals of which have been validated from both an academic and a business viewpoint by the Institute’s advisory board. This board is made up of internationally recognised researchers, the Institute’s business partners, and representatives of major international institutional investors. Management of the research programmes respects a rigorous validation process, which guarantees the scientific quality and the operational usefulness of the programmes.

Six research programmes have been conducted by the centre to date:
• Asset allocation and alternative diversification
• Performance and risk reporting
• Indices and benchmarking
• Non-financial risks, regulation and innovations
• Asset allocation and derivative instruments
• ALM and asset allocation solutions

These programmes receive the support of a large number of financial companies. The results of the research programmes are disseminated through the EDHEC-Risk locations in Singapore, which was established at the invitation of the Monetary Authority of Singapore (MAS); the City of London in the United Kingdom; Nice and Paris in France; and New York in the United States.

EDHEC-Risk has developed a close partnership with a small number of sponsors within the framework of research chairs or major research projects:
• Core-Satellite and ETF Investment, in partnership with Amundi ETF
• Regulation and Institutional Investment, in partnership with AXA Investment Managers
• Asset-Liability Management and Institutional Investment Management, in partnership with BNP Paribas Investment Partners
• Risk and Regulation in the European Fund Management Industry, in partnership with CACEIS
• Exploring the Commodity Futures Risk Premium: Implications for Asset Allocation and Regulation, in partnership with CME Group

Founded in 1906, EDHEC is one of the foremost international business schools. Accredited by the three main international academic organisations, EQUIS, AACSB, and Association of MBAs, EDHEC has for a number of years been pursuing a strategy of international excellence that led it to set up EDHEC-Risk Institute in 2001. This institute now boasts a team of 90 permanent professors, engineers and support staff, as well as 48 research associates from the financial industry and affiliate professors.

Founded in 1906, EDHEC is one of the foremost international business schools. Accredited by the three main international academic organisations, EQUIS, AACSB, and Association of MBAs, EDHEC has for a number of years been pursuing a strategy of international excellence that led it to set up EDHEC-Risk Institute in 2001. This institute now boasts a team of 90 permanent professors, engineers and support staff, as well as 48 research associates from the financial industry and affiliate professors.

Founded in 1906, EDHEC is one of the foremost international business schools. Accredited by the three main international academic organisations, EQUIS, AACSB, and Association of MBAs, EDHEC has for a number of years been pursuing a strategy of international excellence that led it to set up EDHEC-Risk Institute in 2001. This institute now boasts a team of 90 permanent professors, engineers and support staff, as well as 48 research associates from the financial industry and affiliate professors.

Founded in 1906, EDHEC is one of the foremost international business schools. Accredited by the three main international academic organisations, EQUIS, AACSB, and Association of MBAs, EDHEC has for a number of years been pursuing a strategy of international excellence that led it to set up EDHEC-Risk Institute in 2001. This institute now boasts a team of 90 permanent professors, engineers and support staff, as well as 48 research associates from the financial industry and affiliate professors.
About EDHEC-Risk Institute

- Asset-Liability Management Techniques for Sovereign Wealth Fund Management, *in partnership with Deutsche Bank*
- The Benefits of Volatility Derivatives in Equity Portfolio Management, *in partnership with Eurex*
- Structured Products and Derivative Instruments, *sponsored by the French Banking Federation (FBF)*
- Optimising Bond Portfolios, *in partnership with the French Central Bank (BDF Gestion)*
- Asset Allocation Solutions, *in partnership with Lyxor Asset Management*
- Infrastructure Equity Investment Management and Benchmarking, *in partnership with Meridiam and Campbell Lutyens*
- Investment and Governance Characteristics of Infrastructure Debt Investments, *in partnership with Natixis*
- Advanced Modelling for Alternative Investments, *in partnership with Newedge Prime Brokerage*
- Advanced Investment Solutions for Liability Hedging for Inflation Risk, *in partnership with Ontario Teachers’ Pension Plan*
- The Case for Inflation–Linked Corporate Bonds: Issuers’ and Investors’ Perspectives, *in partnership with Rothschild & Cie*
- Solvency II, *in partnership with Russell Investments*
- Structured Equity Investment Strategies for Long-Term Asian Investors, *in partnership with Société Générale Corporate & Investment Banking*

The philosophy of the Institute is to validate its work by publication in international academic journals, as well as to make it available to the sector through its position papers, published studies, and conferences.

Each year, EDHEC-Risk organises three conferences for professionals in order to present the results of its research, one in London (EDHEC-Risk Days Europe), one in Singapore (EDHEC-Risk Days Asia), and one in New York (EDHEC-Risk Days North America) attracting more than 2,500 professional delegates.

EDHEC also provides professionals with access to its website, www.edhec-risk.com, which is entirely devoted to international asset management research. The website, which has more than 58,000 regular visitors, is aimed at professionals who wish to benefit from EDHEC’s analysis and expertise in the area of applied portfolio management research. Its monthly newsletter is distributed to more than 1.5 million readers.

### EDHEC-Risk Institute: Key Figures, 2011–2012

<table>
<thead>
<tr>
<th>Nbr of permanent staff</th>
<th>90</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nbr of research associates</td>
<td>20</td>
</tr>
<tr>
<td>Nbr of affiliate professors</td>
<td>28</td>
</tr>
<tr>
<td>Nbr of PhD in Finance candidates</td>
<td>72</td>
</tr>
<tr>
<td>Overall budget</td>
<td>€13,000,000</td>
</tr>
<tr>
<td>External financing</td>
<td>€7,550,000</td>
</tr>
<tr>
<td>Nbr of conference delegates</td>
<td>1,860</td>
</tr>
<tr>
<td>Nbr of participants at research seminars and executive education seminars</td>
<td>868</td>
</tr>
</tbody>
</table>
About EDHEC-Risk Institute

The EDHEC-Risk Institute PhD in Finance

The EDHEC-Risk Institute PhD in Finance is designed for professionals who aspire to higher intellectual levels and aim to redefine the investment banking and asset management industries. It is offered in two tracks: a residential track for high-potential graduate students, who hold part-time positions at EDHEC, and an executive track for practitioners who keep their full-time jobs. Drawing its faculty from the world’s best universities, such as Princeton, Wharton, Oxford, Chicago and CalTech, and enjoying the support of the research centre with the greatest impact on the financial industry, the EDHEC-Risk Institute PhD in Finance creates an extraordinary platform for professional development and industry innovation.

Research for Business

The Institute’s activities have also given rise to executive education and research service offshoots. EDHEC-Risk’s executive education programmes help investment professionals to upgrade their skills with advanced risk and asset management training across traditional and alternative classes. In partnership with CFA Institute, it has developed advanced seminars based on its research which are available to CFA charterholders and have been taking place since 2008 in New York, Singapore and London.

As part of its policy of transferring know-how to the industry, in 2013 EDHEC-Risk Institute has also set up ERI Scientific Beta. ERI Scientific Beta is an original initiative which aims to favour the adoption of the latest advances in smart beta design and implementation by the whole investment industry. Its academic origin provides the foundation for its strategy: offer, in the best economic conditions possible, the smart beta solutions that are most proven scientifically with full transparency in both the methods and the associated risks.
EDHEC-Risk Institute Publications
(2010–2013)

2013
• Martellini, L., and V. Milhau. Analyzing and decomposing the sources of added-value of corporate bonds within institutional investors' portfolios (August).
• Deguest, R., L. Martellini, and A. Meucci. Risk parity and beyond - From asset allocation to risk allocation decisions (June).
• Blanc-Brude, F. and O.R.H. Ismail. Who is afraid of construction risk? (March)
• Lixia, L., L. Martellini, and S. Stoyanov. The relevance of country- and sector-specific model-free volatility indicators (March).
• Deguest, R., L. Martellini, and V. Milhau.. The benefits of sovereign, municipal and corporate inflation-linked bonds in long-term investment decisions (February).
• Deguest, R., L. Martellini, and V. Milhau. Hedging versus insurance: Long-horizon investing with short-term constraints (February).
• Padmanaban, N., M. Mukai, L . Tang, and V. Le Sourd. Assessing the quality of asian stock market indices (February).
• Cocquemas, F. Towards better consideration of pension liabilities in european union countries (January).
• Blanc-Brude, F. Towards efficient benchmarks for infrastructure equity investments (January).

2012
• Arias, L., P. Foulquier and A. Le Maistre. Les impacts de Solvabilité II sur la gestion obligataire (December).
• Arias, L., P. Foulquier and A. Le Maistre. The Impact of Solvency II on Bond Management (December).
• Amenc, N., and F. Ducoulombier. Proposals for better management of non-financial risks within the european fund management industry (December).
• Cocquemas, F. Improving Risk Management in DC and Hybrid Pension Plans (November).

- La gestion indicielle dans l'immobilier et l'indice EDHEC IEIF Immobilier d'Entreprise France (September).
- Real estate indexing and the EDHEC IEIF commercial property (France) index (September).
- Goltz, F., S. Stoyanov. The risks of volatility ETNs: A recent incident and underlying issues (September).
- Amenc, N., F. Goltz, V. Milhau, and M. Mukai. Reactions to the EDHEC study "Optimal design of corporate market debt programmes in the presence of interest-rate and inflation risks" (May).
- Schoeffler, P. Optimal market estimates of French office property performance (March).
- Le Sourd, V. Performance of socially responsible investment funds against an efficient SRI Index: The impact of benchmark choice when evaluating active managers – an update (March).
- Sender, S. Shifting towards hybrid pension systems: A European perspective (March).
- Blanc-Brude, F. Pension fund investment in social infrastructure (February).

- Schoeffler P. Les estimateurs de marché optimaux de la performance de l’immobilier de bureaux en France (January).

2011

- Deguest, R., Martellini, L., and V. Milhau. Life-cycle investing in private wealth management (October).
- Le Sourd, V. Performance of socially responsible investment funds against an Efficient SRI Index: The Impact of Benchmark Choice when Evaluating Active Managers (September).
- Charbit, E., Giraud J. R., F. Goltz, and L. Tang Capturing the market, value, or momentum premium with downside Risk Control: Dynamic Allocation strategies with exchange-traded funds (July).
- Scherer, B. An integrated approach to sovereign wealth risk management (June).
- Sender, S. The elephant in the room: Accounting and sponsor risks in corporate pension plans (March).
- Martellini, L., and V. Milhau. Optimal design of corporate market debt programmes in the presence of interest-rate and inflation risks (February).

2010

• Amenc, N., and S. Sender. The European fund management industry needs a better grasp of non-financial risks (December).


• Hitaj, A., L. Martellini, and G. Zambruno. Optimal hedge fund allocation with improved estimates for coskewness and cokurtosis parameters (October).

• Amenc, N., F. Goltz, L. Martellini, and V. Milhau. New frontiers in benchmarking and liability-driven investing (September).

• Martellini, L., and V. Milhau. From deterministic to stochastic life-cycle investing: Implications for the design of improved forms of target date funds (September).

• Martellini, L., and V. Milhau. Capital structure choices, pension fund allocation decisions and the rational pricing of liability streams (July).

• Sender, S. EDHEC survey of the asset and liability management practices of European pension funds (June).


• Amenc, N., and S. Sender. Are hedge-fund UCITS the cure-all? (March).

• Amenc, N., F. Goltz, and A. Grigoriu. Risk control through dynamic core-satellite portfolios of ETFs: Applications to absolute return funds and tactical asset allocation (January).


2012

• Till, H. Who sank the boat? (June).

2011

• Amenc, N., and S. Sender. Response to ESMA consultation paper to implementing measures for the AIFMD (September).
• Uppal, R. A Short note on the Tobin Tax: The costs and benefits of a tax on financial transactions (July).
• Till, H. A review of the G20 meeting on agriculture: Addressing price volatility in the food markets (July).

2010

• Amenc, N., and V. Le Sourd. The performance of socially responsible investment and sustainable development in France: An update after the financial crisis (September).
• Amenc, N., A. Chéron, S. Gregoir, and L. Martellini. Il faut préserver le Fonds de Réserve pour les Retraites (July).
• Amenc, N., P. Schoefler, and P. Lasserre. Organisation optimale de la liquidité des fonds d’investissement (March).
• Lioui, A. Spillover effects of counter-cyclical market regulation: Evidence from the 2008 ban on short sales (March).
Notes
For more information, please contact:
Carolyn Essid on +33 493 187 824
or by e-mail to: carolyn.essid@edhec-risk.com

EDHEC-Risk Institute
393 promenade des Anglais
BP 3116 - 06202 Nice Cedex 3
France
Tel: +33 (0)4 93 18 78 24

EDHEC Risk Institute—Europe
10 Fleet Place, Ludgate
London EC4M 7RB
United Kingdom
Tel: +44 207 871 6740

EDHEC Risk Institute—Asia
1 George Street
#07-02
Singapore 049145
Tel: +65 6438 0030

EDHEC Risk Institute—North America
One Rockefeller Plaza
10th & 11th Floors
New York City - NY 10020 USA
Tel: +1 212 500 6476

EDHEC Risk Institute—France
16-18 rue du 4 septembre
75002 Paris
France
Tel: +33 (0)1 53 32 76 30

www.edhec-risk.com