INSTITUTO
UNIVERSITÁRIO
DE LISBOA

ESSAYS ON THE ESTIMATION OF DISTRIBUTION MOMENTS

Pedro Nogueira Serrasqueiro

PhD in Management, specialization in Quantitative Methods

Supervisor:

Doctor José Dias Curto, Associate Professor with Habilitation, ISCTE - Instituto Universitário de Lisboa

September, 2021

Department of Quantitative Methods for Management and Economics

ESSAYS ON THE ESTIMATION OF DISTRIBUTION MOMENTS

Pedro Nogueira Serrasqueiro

PhD in Management, specialization in Quantitative Methods

Jury:

Doctor Diana Aldea Mendes, Associate Professor, ISCTE - Instituto Universitário de Lisboa
Doctor Nuno Marques, Assistant Professor, Universidade Nova de Lisboa
Doctor Andreia Dionísio, Associate Professor, Universidade de Évora
Doctor Pedro Ribeiro, Invited Assistant Professor, ISCTE - Instituto Universitário de Lisboa
Doctor José Dias Curto, Associate Professor with Habilitation, ISCTE-IUL

September, 2021

À família.

Resumo

No conjunto de quatro artigos aqui compilados, estudam-se aspectos particulares dos primeiros quatro momentos das distribuições de probabilidade: média, variância, assimetria e curtose. A estrutura dos primeiros três artigos é semelhante, na medida em que se estabelecem hipóteses e/ou se estudam propriedades teóricas, posteriormente testadas em contexto de simulação e ilustradas por aplicações a situações reais; o quarto artigo surge no contexto da actual pandemia SARS-CoV-2 e é de natureza empírica. O primeiro artigo (Capítulo 2) analisa a relação entre as médias geométrica e harmónica no âmbito da agregação de rácios de natureza financeira. No segundo artigo (Capítulo 3), a respeito da estimação de modelos de regressão linear, propõe-se uma medida de ordenação de regressores em termos da sua importância relativa para a explicação da variância em torno da média da variável dependente. O terceiro artigo (Capítulo 4) aborda as dificuldades conhecidas na estimação dos coeficientes de assimetria e curtose. Recorrendo ao Método dos Momentos Generalizado, propõem-se intervalos de confiança e testes de hipóteses robustos para os coeficientes, atendendo à heteroscedasticidade e à autocorrelação tipicamente presentes nas séries financeiras temporais. No último artigo (Capítulo 5) procurou-se, com recurso à classe de modelos GARCH e à introdução de variáveis *dummy*, investigar e quantificar o impacte da recente pandemia por SARS-CoV-2 na volatilidade (desvio-padrão) dos retornos de um conjunto de acções cotadas e índices bolsistas dos mercados de capitais norte-americanos.

Palavras-chave: Média, Variância, Assimetria, Curtose, Estimação, COVID-19
JEL: C10, C13

Abstract

The set of four articles herein compiled focuses on specific aspects of the first four moments of probability distributions: mean, variance, skewness and kurtosis. The structure of the first three articles is similar, where hypotheses are formulated and/or theoretical properties are studied and later tested with simulation and applied to real life cases. The fourth article is of empirical nature, contextualized by the current SARS-CoV-2 pandemic. The first article (Chapter 2) analyzes the relationship between the geometric and harmonic means with respect to the aggregation of financial ratios. In the second article (Chapter 3), regarding the estimation of linear regression models, a new measure is proposed to rank independent variables according to their relative importance to explain the variability around the mean of the dependent variable. The third article (Chapter 4) approaches known difficulties with the estimation of skewness and kurtosis. Applying the Generalized Method of Moments, confidence intervals and hypothesis tests are derived, taking into account the heteroskedasticity and autocorrelation typically present in financial time series. In the last article (Chapter 5), models of the GARCH family are estimated with the introduction of dummy variables to investigate and quantify the impact of SARS-CoV-2 in the volatility (standard deviation) of returns of a set of US-listed stocks and indices.

Keywords: Mean, Variance, Skewness, Kurtosis, Estimation, COVID-19
JEL: C10, C13

Contents

Glossary

Chapter 1

LLN	Law of Large Numbers
CLT	Central Limit Theorem

Chapter 2

P/E	Price-to-Earnings ratio
ABC	Activity Based Costing
TA	True Average
VWA	Value-weighted Average
EY	Earnings Yield
EPS	Earnings per Share
SE	Squared Error
HA	Harmonic Average
GA	Geometric Average
AA	Arithmetic Average
MED	Median

Chapter 3

VIF	Variance Inflation Factor
EV	Explanatory Variable
DV	Dependent Variable
SSR	Regression Sum of Squares
SST	Total Sum of Squares

Chapter 4

HAC	Heteroskedasticity-Autocorrelation Consistent
GMM	Generalized Method of Moments
MSE	Mean Squared Error
IID	Independent and Identically Distributed
ARCH	Autoregressive Conditional Heteroskedasticity
SE	Standard Error
SGT	Skewed Generalized t-student distribution

Chapter 5

FATANG	Facebook, Amazon, Tesla, Apple, Netflix and Google
GARCH	Generalized Autoregressive Conditional Heteroskedasticity
APARCH	Asymmetric Power Autoregressive Conditional Heteroskedasticity
ADCC EGARCH	Asymmetric Dynamic Conditional Correlation Exponential GARCH
GED	Generalized Error Distribution
MLE	Maximum Likelihood Estimation

CHAPTER 1

Introduction

The set of articles contained in this thesis is motivated by the interest of researchers and practitioners in the application of statistical methods to financial risk management. The classical problem faced by individuals, asset managers and institutions, among others, is how to allocate financial resources to assets (e.g. stocks, bonds, funds and commodities) in such a way as to provide acceptable levels of risk-adjusted returns. In other words, market participants are generally concerned about maximizing returns for increasing levels of risk, as well as monitoring the evolution of portfolio value while maintaining and safeguarding invested capital. Traditionally, this has been achieved by resorting to statistical techniques that make use of the first two moments (mean and variance) of probability distributions. Arguably, the most paradigmatic approach is that of Portfolio Selection by **Markowitz1952** (cited 51.466 times at the time of writing), where the well-known "efficient frontier" was first postulated, and is based on a general rule that investors "do (or should) consider expected return a desirable thing and variance of return an undesirable thing". Put simply, an investor willing to take higher risk should, on average, expect a higher return on investment. The literature on portfolio theory, allocation and risk management has since been greatly expanded, and makes heavy use of the Normal distribution to describe the probabilistic behavior of asset returns.

The choice of the Normal distribution is grounded on two fundamental results of statistical theory, namely the Law of Large Numbers (LLN) and the Central Limit Theorem (CLT). The LLN was first proved by **bernoulli1713**, where he describes in detail the problem of estimating the proportion of white balls in an urn containing an unknown total number of balls (white and black), and concludes that by choosing a sufficiently large number of samples, the proportion of white balls in the urn can be accurately estimated. The proof is finalized with the following remark (p. 38-44):

> *Whence, finally, this one thing seems to follow: that if observations of all events were continued throughout all eternity, and hence the ultimate probability would tend toward perfect certainty, everything in the world would be perceived to happen in fixed ratios and according to a constant law of alternation, so that even in the most accidental and fortuitous occurrences we would be bound to recognize, as it were, a certain necessity and, so to speak, a certain fate.*

Financial markets are a complex web of agents and instruments that express, via prices and their variability, the expectation and will of said agents. Within the framework of financial risk management, at least from a contemporary point of view, while the location (center) of return distributions seems to be well described by a vast collection of theoretical models (informally, we seem to get by "normally" most of the time), the tails of observed (empirical) distributions contain improbable and very consequential events that are hard to predict according to any known tractable law. Markets, institutions and individuals are often caught barehanded by these improbable events, of which the financial crisis of 2007 is a flagrant example. An additional remark by **bernoulli1713** illustrates this point:

> *It certainly remains to be inquired whether after the number of observations has been increased, the probability is increased of attaining the true ratio between the number of cases in which some event can happen and in which it cannot happen, so that this probability finally exceeds any given degree of certainty; or whether the problem has, so to speak, its own asymptote - that is, whether some degree of certainty is given which one can never exceed.*

The CLT, fundamental to probability theory, is at the heart of many statistical considerations of asymptotic convergence to the Normal distribution when general results are not attainable for small samples. While there are several versions of the theorem, the general result states that, under certain conditions, the sum of n independent and identically distributed summands converges in probability to the Normal distribution as n grows to infinity. A widely applied corollary relates the sum to the sample mean, i.e. the sampling distribution of the sample mean is normally distributed, for sufficiently large n, regardless of the original distribution of summands. The history of the CLT is humorously and well encapsulated in the first paragraphs of a paper by **cam1986**, entitled "The Central Limit Theorem around 1935":

> *In the beginning there was de Moivre, Laplace, and many Bernoullis, and they begat limit theorems, and the wise men saw that it was good and they called it by the name of Gauss. Then there were new generations and they said that it had experimental vigor but lacked in rigor. Then came Chebyshev, Liapounov and Markov and they begat a proof and Polyá saw that it was momentous and he said that its name shall be called the Central Limit Theorem [...].*
>
> *The appelation "central" is due to Polyá who used it because of the central role of the theorem in probability theory, not as the modern French do, because it describes the behavior of the center of the distribution as opposed to its tails.*

While both the CLT and the LLN seem, to a certain extent (and especially after being formulated), intuitive, the variable of interest in this context is the sample size n or, rather, the sufficiency of n. How large must a sample be so that statistical inference provides reliable measures of a probability distribution, and what does it depend on? How long does it take to reach the asymptote? An interesting analysis of this problem is provided in **taleb2018**, where a measure of required sample size is derived for different parametrizations of the stable family of distributions.

Financial research is often referred to as being blessed with overwhelming amounts of data and a haven for scientists; in fact, while many data points and sampling frequencies are available at little or no cost, and a trade-off between noise and sample size seems to exist, they may be deemed insufficient for convergence, due to the presence of very large outliers in the data. As will be shown in Chapter 4, outliers are extremely consequential to the estimation of distribution parameters, especially in the case of higher-order moments. In this regard, a paper by **fama1963**, alluding to previous work by **mandelbrot1961** is particularly illustrative:

> *If Mandelbrot's hypothesis is upheld, it will radically revise our thinking concerning both the nature of speculative markets and the proper statistical tools to be used when dealing with speculative prices. Prior to the work of Mandelbrot the usual assumption, which we shall henceforth call the Gaussian hypothesis, was that the distribution of price changes in a speculative series is approximately Gaussian or normal. In the best-known theoretical expositions of the Gaussian hypothesis, arguments are based on the central-limit theorem to support the assumption of normality. If the price changes from transaction to transaction are independent, identically distributed, random variables with finite variance, and if transactions are fairly uniformly spaced through time, the central-limit theorem leads us to believe that price changes across differencing intervals such as a day, a week or a month will be normally distributed since they are simple sums of the changes from transaction to transaction. [...] Mandelbrot contends, however, that past research has overemphasized agreements between empirical distributions of price changes and the normal distribution and has neglected certain departures from normality which are consistently observed.*

In fact, it has now long been recognized that the behavior of financial market returns does not agree with the often assumed normal distribution. Departures from the normal distribution are often illustrated by resorting to higher-order moments such as skewness and kurtosis, and it has become an irrefutable stylized fact that financial market returns have negative asymmetry and excess kurtosis (see, for example, **Bastianin2020**, **WU2019** and **BeraPremaratne2017**).

According to **kimwhite04**, the role of higher moments has become increasingly important in financial literature, mainly because the variance (or standard deviation), the traditional measure of risk, fails to fully capture the true risk of market returns' distributions. Estimation of skewness and kurtosis comes with its own problems, as the standard estimators (standardized third and fourth moments) are essentially based on distances to the sample (arithmetic) mean raised to the third and fourth powers and therefore substantially sensitive to outliers. Additionally, while corrected versions of the standard estimators are known to be unbiased under normality (**JoanesGill1998**), their reliability in small samples decreases drastically under different distributional assumptions. In Chapter 2, alternative location measures, namely the geometric and harmonic means, are investigated with respect to their increased robustness to outliers when aggregating financial ratios such as Price-to-Earnings (P/E). Besides the violation of the distributional assumption (of normality), departures from the assumption that observations are independent over time are also frequent and have been extensively documented. Accordingly, in Chapter 4, after assessing the performance of standard estimators in the presence of outliers, the Generalized Method of Moments is applied to derive the asymptotic distribution of estimators of skewness and kurtosis and confidence intervals and hypothesis tests are constructed, taking into account the heteroskedasticity and autocorrelation typically present in financial time series. In Chapter 5, an empirical application of several GARCH-family models is pursued, in order to model the conditional variance of returns of a set of listed stocks and indices. Specifically, an Asymmetric Power GARCH (APARCH) model with dummy variables is estimated to investigate if the current COVID-19 pandemic has significantly increased the volatility of returns. Using a Likelihood Ratio test, the proposed model is compared to traditional GARCH models with respect to their predictive ability. In Chapter 3, we focus on variance partitioning and adopt a more general approach that goes beyond the realm of financial markets. Namely, the problem of decomposing the coefficient of determination (commonly designated as R^2) and assigning variation shares to independent variables to determine their relative importance (to "explain" variability of the dependent variable around the mean) is addressed. While it has an unique solution when regressors are mutually independent, this problem is no longer trivial in the presence of collinearity. An attempt to overcome this difficulty is provided by employing the Variance Inflation Factor (VIF) as a weighting criterium.

The articles herein compiled roughly follow a common methodological approach: theoretical properties of estimators are evaluated and devised, tested under simulation and then applied to real life cases and series. The order in which they are presented corresponds to the moment-order of the estimators being addressed (mean, variance, skewness and kurtosis). The one exception is the fourth article (Chapter 5), presented last due to its empirical nature when compared to the first three. Overall, this compilation reflects the author's desire to contribute to the existing literature and to consolidate his understanding by attempting to tackle some of the aforementioned estimation issues.

CHAPTER 2

Averaging ratios: geometric vs. harmonic mean

ABSTRACT

Ratios represent a special kind of relation between two magnitudes. Averaging ratios is common practice among academics and Finance practitioners. For example, how does one best aggregate (average) firms' price-to-earnings (P/E) at portfolio level? The arithmetic mean is the natural alternative. However, for financial ratios, it is generally accepted that the much less familiar harmonic mean is more valuable because it solves the upward bias problem encountered while using arithmetic mean and is a more intuitive and logical approach. To the best of our knowledge, there is no statistical evidence to show the superiority of the harmonic mean when computing the average of ratios. By bootstrapping P/E ratios and earnings yields of the companies listed in eight large stock indices, we compare the traditional averages and it is shown that geometric, and not the harmonic mean, as it is commonly accepted, is more suitable to average ratios.

2.1. Introduction

Ratios represent a special kind of relation between two magnitudes, and are of common use in everyday life, as well as in scientific fields (**Brown2011**) such as Mathematics, Engineering, Statistics, Economics, Finance, Medicine, etc. In Finance, market multiples such as market-to-book value, price to cash flow, price to sales, dividend payout ratios (**gilletal2010**; **HusnaSatria2019**; **Sriram2020**; **HusainSunardi2020**) and price-earnings (or P/E) ratio are often used (**agrrawal2010**; **Musallam2018**). Other examples are the estimation and inference about the Sharpe ratio and the average excess return per unit of volatility or total risk. Distribution properties and estimation of the ratio of independent random variables are the subject of several studies, e.g. **NadKotz2005**. For the conversion of odds ratios and hazard ratios to risk ratios see **Weele2020**.

When computing an average, the arithmetic mean[1] is the natural alternative (see **Berishvili2020** for usage of the arithmetic mean to calculate the industry average financial ratios for Georgia). As a consequence, alternative approaches to capture the center of a probability distribution are frequently discarded (**coggeshall1886**). When ratios are at play, the choice of averaging method does matter (as we show), and sometimes the much less familiar harmonic mean provides a more logical approach to average the ratio between two magnitudes (**agrrawal2010**). See also **MartinStanford2007** for the application of the harmonic mean to the ABC inventory classification systems.

We define three different situations when datasets consist of ratios. The first, which we name "All Equal", when either the numerators or denominators are the same (or we can admit the equality) in all the ratios to be averaged. An example is the average number of spectators per football match in the last five seasons. To obtain the exact value (the realized value, represented by "TA", the true average), we divide the sum of all spectators by the number of matches. Alternatively, as the data is available per year, we can compute the ratio per season and then compute the average of ratios. The question is what type average to employ. If it is reasonable to admit that the number of matches is constant (the denominator of ratios per season), the arithmetic average of ratios is exactly the TA. Consider only two seasons, where the number of spectators is 301,000 and 400,400 and the number of matches is 70 per season. The average number of spectators per match is: $TA = (301000 + 400400)/(70 + 70) = 5010$.

The two ratios per season are $R_1 = 301000/70 = 4300$, $R_2 = 400400/70 = 5720$ and the simple arithmetic average of the two ratios is exactly TA: $(4300 + 5720)/2 = 5010$. Thus, the simple arithmetic average of ratios results in the true average if the denominators of the ratios are equal. When the equality is in the numerators, it is the harmonic average

[1]We use "average" and "mean" interchangeably.

that results in the TA[2], as exemplified in the next paragraph and we explained in detail in Section 2.2.

The second situation, which we name "Become Equal", when neither the numerators or denominators are equal, but it is reasonable to conceive a constant value for either. For example, suppose one invests in two stocks with prices of 400 and 200, and earnings per share 20 and 40, respectively. The true average of price per unit of earnings (the value-weighted average (VWA)) is: $VWA = (400 + 200)/(20 + 40) = 10$. The two P/E ratios are $R_1 = 400/20 = 20$, $R_2 = 200/40 = 5$ and the arithmetic and harmonic averages are 12.5 and 8, respectively. None of the ratios equals 10, but we can buy two shares of the second stock (as a portfolio is generally composed by more than one share per stock) and now the investment is the same for both stocks: 400. The true average of a portfolio with $1+2$ shares is $(400+400)/(20+80) = 8$, the value of the harmonic average of R_1 and R_2. According to **agrrawal2010**, this is the main reason to use the harmonic mean, because it has a much more intuitive and logical investment assumption, since most portfolios are neither on an equal number of shares per position basis, nor on an equalized earnings per holding basis. So, it makes more sense to talk about portfolios with equal amount invested in each company. When computing the harmonic or the arithmetic mean of ratios we are assuming equal numerators/denominators. If that is a reasonable assumption, the two averages are still appropriate.

The last situation, which we name "All Different", when the numerators and denominators are different and it is not reasonable to assume that either is constant or it is not possible to make it constant. We face two different cases when averaging such ratios.

Firstly, if all data is available, including the individual numerators and denominators of the ratios, it is possible to compute the true mean by summing all the numerators, all the denominators and obtaining the ratio. Secondly, if available data consists only of observed values of ratios (individual values of numerators and denominators are unknown), one must learn from the previous case (when granular data is available) to decide appropriately. By bootstrapping several samples including real financial data (see Section 2.4), we conclude that, when it is unreasonable to assume a constant numerator or denominator ("All Different" situation), the geometric average, and not the harmonic average, as it is commonly accepted, is more suitable to estimate the true average of ratios.

The purpose of this paper is to clarify academics and professionals about differences on averaging methods when data involves ratios. Our contribution is threefold: firstly,

[2]The true or realized average (TA) is crucial to conclude which of the averages is more appropriate, by comparing each one with TA. This is the reason why the realized P/E portfolio value is included in equation (2.18) – the realized value is needed to infer about the most appropriate measure to average the ratios.

starting from the generalized average, we show in Subsection 2.2.1 that the harmonic mean is equivalent to a weighted arithmetic average, where the weights are inversely proportional to the original values and they are computed in such a way that the contribution of each value to the final average is exactly the same. Thus, the weights compensate the original values to equal the contribution of each value to the final average. We also generalize the formula, having the harmonic and arithmetic averages as particular cases. Secondly, in Subsection 2.2.2 we elaborate on different central tendency measures and their interpretation as the center of a probability distribution. Lastly, in Section 2.3, we present and discuss traditional applications of the harmonic average and emphasize the price-to-earnings ratio (P/E). In Section 2.4, by bootstrapping eight real financial time series, we show that geometric average seems more appropriate to estimate the average of ratios. At the end of the paper present our concluding remarks. We note that our results apply to strictly positive values only.

2.2. The Generalized or Power Average (GA)

The classical methods of averaging data are the three Pythagorean means: the familiar arithmetic mean, the geometric mean (the nth root of the product of data points) and the harmonic mean (the reciprocal of the arithmetic mean of the reciprocals of the data points). The three means are particular cases of the generalized or power mean. For recent applications of the generalized mean see, for example, **gouetal2019**, **Luetal2020**, **priam2020** and **kolahdouzetal2020**.

If k is a non-zero real number, and $x_1, x_2, \ldots, x_n$ are positive real numbers, then the generalized or power mean with exponent k of these positive real numbers is:

$$\bar{X}^k_{GA_1}(x_1, x_2, \ldots, x_n) = \left(\frac{1}{n} \sum_{i=1}^{n} x_i^k \right)^{\frac{1}{k}} \tag{2.1}$$

and the three Pythagorean means are special cases of $\bar{X}^k_{GA_1}$ for particular values of k. For $k = -1$ and $k = 1$ we get the harmonic and the arithmetic mean and when $k \to 0$ we get the geometric mean. In the bootstrap simulation study (see Section 2.4) the constant k is estimated and confidence intervals are computed to conclude about the appropriateness of each averaging method.

The arithmetic mean, or simply the mean or average, is the sum of all x_i divided by the number of observations (n):

$$\bar{X}_A = \frac{1}{n} \sum_{i=1}^{n} x_i = \frac{\sum_{i=1}^{n} x_i}{n}, \quad i = 1, 2, \ldots, n, \tag{2.2}$$

where x_i represents the different values assumed by the X variable. See **Chen1995** for statistical inference about the arithmetic mean in the case of positively skewed distributions.

The geometric mean[3] also measures the central tendency of a set of numbers, being defined as the root of order n of the product of values x_i:

$$\bar{X}_G = \sqrt[n]{\prod_{i=1}^{n} x_i} = \sqrt[n]{x_1 \times x_2 \times \ldots \times x_n} = (x_1 \times \ldots \times x_n)^{\frac{1}{n}} \text{ or } \bar{X}_G = \exp\left[\frac{1}{n}\sum_{i=1}^{n} \ln(x_i)\right], \tag{2.3}$$

where the capital letter Π represents a series of products. As **galton1897** suggested in one of the earliest papers on geometric average, the distribution of $\bar{X}_G$ will approach log-normality as n increases, for all parent distributions to which the central limit theorem applies. Thus, the distribution of $\bar{X}_G$ will approach the log-normal form, even though the parent distribution of X may not be log-normal (**alfgrossberg1979**). The geometric mean applies only to numbers of the same sign in order to avoid the nth root of a negative number when n is even, which is not defined in the set of real numbers. In general, only positive numbers are admitted.

We note that the geometric mean of a ratio of two variables equals the ratio of their geometric means, and give an example:

$$G\left(\frac{X_i}{Y_i}\right) = \frac{\overline{X}_G}{\overline{Y}_G} \quad \text{or} \quad \sqrt[n]{\prod_{i=1}^{n} \frac{X_i}{Y_i}} = \frac{\sqrt[n]{\prod_{i=1}^{n} X_i}}{\sqrt[n]{\prod_{i=1}^{n} Y_i}}. \tag{2.4}$$

$$G\left(\frac{X_i}{Y_i}\right) = \sqrt[4]{\frac{1}{4} \times \frac{4}{6} \times \frac{6}{10} \times \frac{8}{14}} = \frac{3.722}{7.614} = 0.489.$$

Therefore, if all data is available, including individual numerators and denominators, the geometric mean of the ratios is the ratio of the geometric means of numerators and denominators, respectively.

The equally weighted harmonic mean is expressed as the reciprocal of the arithmetic mean of the reciprocals. Let $\frac{1}{x_1}, \frac{1}{x_2}, \ldots, \frac{1}{x_n}$ be the reciprocals of the given set of observations. Then:

$$\bar{X}_H = \left(\frac{\sum_{i=1}^{n} \frac{1}{x_i}}{n}\right)^{-1} = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}. \tag{2.5}$$

[3]For a recent application of the geometric mean see, for example, **Tofallis2015**.

If different weights $\delta_1, \delta_2, \ldots, \delta_n$ are assigned to the x_i observations, the weighted harmonic mean is defined by:

$$\bar{X}_H = \frac{\sum_{i=1}^{n} \delta_i}{\sum_{i=1}^{n} \frac{\delta_i}{x_i}}. \quad (2.6)$$

The three means are related with regards to their relative size, as $\bar{X}_H \leq \bar{X}_G \leq \bar{X}_A$.

2.2.1. Contributions to the final average

As

$$\bar{X}_H = \sum_{i=1}^{n} x_i \frac{w_i}{\sum_{i=1}^{n} w_i} = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}, \quad (2.7)$$

where $w_i = \frac{\sum_{i=1}^{n} x_i}{x_i}$, the harmonic mean is equivalent to a weighted arithmetic average, where the weights are inversely proportional to the original values and they are computed in such a way that the contribution of each value (x_i) to the final average is exactly the same[4]:

$$x_1 \frac{w_1}{\sum_{i=1}^{n} w_i} = x_2 \frac{w_2}{\sum_{i=1}^{n} w_i} = \ldots = x_i \underbrace{\frac{w_i}{\sum_{i=1}^{n} w_i}}_{p_i} = \ldots = x_n \frac{w_n}{\sum_{i=1}^{n} w_i} = \frac{1}{\sum_{i=1}^{n} \frac{1}{x_i}}. \quad (2.8)$$

If x_m is the minimum value and its (highest) weight is p_m, the weight (smaller) of the other values x_i is given by $p_i = \frac{x_m}{x_i} p_m$; the weight is inversely proportional to the values, where the constant of proportionality is $l = x_i \times p_i$, which is the constant contribution of each value to the harmonic average. All data points, regardless of magnitude, contribute equally to the harmonic average, as more weight is attributed to the smallest observations in the data set. We note, though, that due to its insensitivity to outliers, the harmonic mean can obscure large values that may be consequential.

Different R packages have built-in functions to calculate the harmonic mean. We have written a simple R function for the purpose of our work (shown below for illustration of a coded version of Equation 2.7):

```
HarmonicMean <- function(x){
weigs1 = sum(x)/x
weigs = weigs1/sum(weigs1)
HarmonicMean <- t(x)%*%weigs
return(HarmonicMean)}
```

Equation 2.7 can be generalised to:

$$\bar{X}_{GA} = \sum_{i=1}^{n} x_i \frac{w_i^l}{\sum_{i=1}^{n} w_i^l}, \quad (2.9)$$

[4]See appendix 2.6 for demonstrations.

where l is a real number. If $l = 1$, $\bar{X}_{GA}$ is the harmonic average; if $l = 0$, $\bar{X}_{GA}$ is the arithmetic average, for $w_i \neq 0$. The geometric average is obtained for a particular value of l in the range $0 < l < 1$, which is not always constant.

In the (simple) arithmetic average the weight associated to each data point is constant:

$$\bar{X}_A = \sum_{i=1}^{n} \frac{1}{n} x_i = \frac{1}{n} x_1 + \frac{1}{n} x_2 + \ldots + \frac{1}{n} x_n. \tag{2.10}$$

Therefore, if x_i differ, the contribution of each ($c_i = \frac{1}{n} x_i$) to the arithmetic average is not constant and higher values will have greater contribution, directly proportional to the x_i; the reason for the arithmetic average to be influenced by outlying observations is that the constant of proportionality is $k = \frac{c_i}{x_i}$ and $c_i = k x_i = \frac{1}{n} x_i$ (**LyuZhangChurch2020**).

The simplest way to reduce the importance of extreme observations is to use, as alternative to the arithmetic, the harmonic or geometric means when the data set includes outlying observations. As mentioned, the geometric average is the root of order n of the product of x_i and the contribution of each value to the geometric average is neither constant nor directly proportional to the respective value; it is an intermediate situation, resulting in an intermediate value between the harmonic and the arithmetic averages.

By taking the natural log of both sides of equation (2.3) we obtain:

$$\ln\left(\bar{X}_G\right) = \frac{1}{n} \sum_{i=1}^{n} \ln(x_i) = \frac{1}{n} \ln(x_1) + \frac{1}{n} \ln(x_2) + \ldots + \frac{1}{n} \ln(x_n). \tag{2.11}$$

By taking the anti-logarithm of both sides:

$$\bar{X}_G = \exp\left\{ \frac{1}{n} \sum_{i=1}^{n} \ln(x_i) \right\} = x_1^{\frac{1}{n}} x_2^{\frac{1}{n}} \ldots x_2^{\frac{1}{n}}. \tag{2.12}$$

Based on Jensen's inequality[5], we conclude that the value of the arithmetic average is at least the value of the geometric average: $\bar{X}_A \geq \bar{X}_G$ (where the equality holds only if all the x_i are equal). The contribution of each data point to the geometric average is not directly proportional to x_i, reducing the impact of outlying observations in the final result.

2.2.2. Different meanings of the center

Choosing the appropriate mean to represent the central tendency of a distribution is an old and yet highly controversial matter (see **coggeshall1886**, p. 84): the mean commonly employed by an economist is not a real quantity, but rather a quantity assumed as the representative number of the set, diverging more or less from it. Its 'fictitious" character renders possible the choice among different values, and thus among different methods of obtaining it. The same conclusion holds for median when the number of observations is

[5]See appendix 2.6.2 for demonstrations.

even.

Consider the median as a representation of the central tendency of a distribution. By ranking observations from lowest to highest, the ranks associated to each observation are given by $r_1, r_2, \ldots, r_n$ where r_1 and r_n represent the minimum and the maximum, respectively. The median $\left(\bar{X}_M\right)$ corresponds to the midpoint in terms of observation count (absolute frequency): $r_M = \frac{r_1 + r_n}{2}$. Also, note that the following property holds for median:

$$\sum_{i=1}^{n} (r_i - r_M) = 0, \tag{2.13}$$

where r_i represents the ranks of ordered observations and r_M is the rank corresponding to the median. The sum of differences between ranks and the rank corresponding to the median equals zero.

The arithmetic average $\left(\bar{X}_A\right)$ defines the center of the distribution in terms of the observations' values:

$$\sum_{i=1}^{n} \left(x_i - \bar{X}_A\right) = 0, \tag{2.14}$$

where x_i represents the value of each observation in the data set.

The geometric average $\left(\bar{X}_G\right)$ defines the center of the distribution in terms of the log percentage deviations **Tornqvist1985** around $\bar{X}_G$ and the following property holds:

$$\sum_{i=1}^{n} \ln\left(\frac{x_i}{\bar{X}_G}\right) = 0, \tag{2.15}$$

or, alternatively, the sum of the differences between the log of x_i and the log of the geometric average is zero:

$$\sum_{i=1}^{n} \left[\ln\left(x_i\right) - \ln\left(\bar{X}_G\right)\right] = 0, \tag{2.16}$$

where $\ln\left(x_i\right) - \ln\left(\bar{X}_G\right)$, when multiplied by 100, is the log percentage deviation between each observation and the geometric average. Thus, the geometric average is the value that balances the negative and positive log percentage deviations.

The harmonic average $\left(\bar{X}_H\right)$ defines the center of the distribution in terms of the absolute deviations weighted by $\frac{w_i}{\sum_{i=1}^{n} w_i}$ (see equation (2.7)):

$$\sum_{i=1}^{n} \left(x_i - \bar{X}_H\right) \frac{w_i}{\sum_{i=1}^{n} w_i} = 0. \tag{2.17}$$

Thus, the harmonic mean defines the center of the distribution such that the weighted deviations on its left compensate the weighted deviations on its right. The weights are inversely proportional to the original values[6].

2.3. Traditional Applications

Consider the classic "average speed" example. Suppose a car is traveling from cities A to B. If the average speed from A to B is 80 miles per hour (mph) and the average speed on the return trip is 48 mph, what is the average speed roundtrip? If you compute the arithmetic average between 80 and 48 the average speed is 64 mph. This is not the correct answer because the travel time over the same distance varies over the two legs. This is the common problem when we average ratios with two independent variables in the numerator and the denominator (distance and time in this example; see **agrrawal2010** for details).

Let D, T and S represent the distance, time and speed. In our example, $S_1 = 80$, $S_2 = 48$, $D_1 = D_2 = D$ and the distance travelled equals speed multiplied by time: $D = S \times T$. Time on each leg of the trip is $T_1 = \frac{D}{80}$ and $T_2 = \frac{D}{48}$ and the average roundtrip speed S_A is the total distance travelled ($2D$) divided by the total elapsed time ($D/80 + D/48$):

$$S_A = \frac{2D}{\frac{D}{80} + \frac{D}{48}} = \frac{2}{\frac{1}{80} + \frac{1}{48}} = 60 \text{ mph},$$

This is the harmonic mean applied to the average speeds over the same distance travelled. The harmonic instead of the arithmetic mean provides the correct answer. Therefore, total travel time is the same as if the driver travelled the whole distance ($2D$) at the average speed of 48 mph.

Now consider two trips where the travel time is the same but the distance is different. As $T_1 = T_2 = T$, then $D_1 = S_1 \times T$, $D_2 = S_2 \times T$ and the average speed is given by:

$$S_A = \frac{D_1 + D_2}{T + T} = \frac{S_1 \times T + S_2 \times T}{2T} = \frac{S_1 + S_2}{2} = 64 \text{ mph},$$

which is the arithmetic average.

Given several sub-trips at different speeds, if each one covers the same distance, the average speed is the harmonic mean of all the sub-trip speeds; if the time is the same, then the average speed is the arithmetic mean of all the sub-trip speeds; choosing the correct averaging method depends on whether one is traveling the same distance or the same time.

In Finance, individual market multiples such as market-to-book, price to cash flow, price to sales and price-earnings (P/E) are aggregated very often at the portfolio level to provide a unified number. Our paper focuses on the price-to-earnings ratio (or P/E ratio),

[6] See appendix 2.6.3 for simple applications of the results of this subsection.

which measures the current share price of a company relative to its per-share earnings.

Consider two stocks A and B with prices \$150 and \$60, and earnings per share (EPS) \$10 and \$20, respectively. The price earnings ratios are: $P/E_A = \frac{150}{10} = 15$ and $P/E_B = \frac{60}{20} = 3$ for stocks A and B, respectively. The realized portfolio P/E (including stocks A and B) is the sum of prices divided by the sum of earnings or the weighted average of the price-to-earnings ratios of stocks A and B, where the weights are proportional to the earnings per share of each stock:

$$P/E_{VWA} = \frac{150+60}{10+20} = 15 \times \frac{10}{10+20} + 3 \times \frac{20}{10+20} = 15 \times 0.6667 + 3 \times 0.3333 = 7,$$

the value-weighted average P/E (assuming a portfolio of one share per company). Thus, the average price per unit of earnings (no matter the stock) is 7 (the sum of the two prices: 210, is divided by the sum of the two earnings: 30). The weighted average is closer to the lower individual ratio $P/E_B = 3$ because the earnings weight of stock B is $20/30 = 0.667$ (or 66.67%), against 33.33% of stock A.

The arithmetic mean of the two P/Es is $P/E_{AA} = \frac{15+3}{2} = 15 \times \frac{1}{2} + 3 \times \frac{1}{2} = 9$. Thus, both P/Es are weighted by 1/2 no matter the earnings of each stock. So the less profitable stock drives the average up, due to its higher P/E. The prices of the two stocks are not diluted by the \$30 of earnings, but each price is divided by the respective earnings. The average price per unit of earnings of stock A and B is 15 and 3, respectively, and the simple arithmetic average is 9: the average price per unit of earnings is 9 and 7 depending on whether simple or price-weighted averaging is applied, respectively.

The harmonic mean $P/E_{HA} = \frac{2}{\frac{1}{15}+\frac{1}{3}} = 5$, yields the lowest value, that is, after computing the arithmetic mean of the inverse of the portfolio P/E ratio (the earnings-to-price ratio, 0.2) we take its inverse, resulting in $1/0.2 = 5$, the value of the harmonic mean. If the average of earnings yield is 0.2 (or 20%) the average P/E is 5.

The arithmetic mean P/E (average price per unit of earnings) equals the P/E of an equal earnings portfolio:

$$P/E_A = \frac{\frac{150}{10} + \frac{60}{20}}{2} = \frac{\frac{300}{20} + \frac{60}{20}}{2} = \frac{300}{40} + \frac{60}{40} = \frac{360}{40} = 9,$$

where two shares of A are needed for every share of B, to equalize the earnings accruing from each holding (the common denominator of the ratios): the arithmetic mean assigns equal weights to the earnings in each ratio.

The harmonic mean P/E (the inverse of the average earnings per unit of price) equals the mean of an equal investment portfolio, a much more straightforward financial interpretation than the arithmetic mean:

$$P/E_H = \frac{300 + 300}{\frac{300}{15} + \frac{300}{3}} = 5,$$

where equal weights are assigned to the price in each ratio.

In this example, the realized portfolio P/E is 7. The arithmetic overstates and the harmonic mean understates its value. This reinforces the need of the realized value in equation (2.18) to infer about the most appropriate measure to average the ratios.

According to **matthews2004** and **agrrawal2010**, applying the harmonic mean is beneficial, when compared to the arithmetic mean, because it solves the upward bias resulting from the arithmetic mean ($P/E_H = 5$ against $P/E_A = 9$) and it has a much more intuitive and logical investment assumption, since most portfolios are neither on an equal number of shares per position basis, nor on an equalized earnings per holding basis. So, it is more reasonable to consider portfolios with equal amounts invested in each company; when this is the case, there is broad consensus on the harmonic average ("All Equal" and "Become Equal" situations).

However, as it is unlikely that all the companies share the same position, how should we average price-to-earnings ratios if the investment in each company is not the same? ("All Different" situation) Furthermore, even if the financial reasoning seems logical, intuitive and appropriate, statistical evidence is unclear, because the harmonic mean also shows a downward bias influenced by the smaller ratios.

In the next section we perform a bootstrapping simulation study to estimate k, the exponent in the generalized (or power) mean to shed light and get evidence about the averaging methods available.

2.4. Bootstrapping P/E ratio and Earnings Yield

The bootstrap will be used to compare the appropriateness of the three averages to estimate the realized portfolio P/E (the value-weighted average P/E). The eight bootstrap samples include the companies listed in the Standard and Poor's 500 (421), NASDAQ 100 (90), Euro Stoxx 600 (467), Japanese NIKKEI 225 (171), UK FTSE 100 (74), Shanghai Composite SSEC (1282), French CAC40 (32) and German DAX30 (23) with non-negative P/E ratio on November 27, 2020. The number of sampled companies is in parenthesis. The data set includes stock price (P), earnings per share (EPS), P/E ratio and earnings yield (EY) for all the sampled companies[7]. The earnings yield (which is the inverse of the

[7]The data source is: https://www.investing.com. Data refers to November 27, 2020.

P/E ratio) refers to the earnings per share for the most recent 12−month period divided by the current market price per share.

Descriptive statistics are shown in Tables 1 (P/E ratios) and 2 (EY). The empirical distribution of P/E ratio is always positive asymmetrical and leptokurtic pointing to the presence of outlying observations. This is confirmed by the large difference between the arithmetic average and the other three central tendency measures. In case of Earnings Yield, the distributions are still positive asymmetrical and leptokurtic (DAX index is the exception), but not so strong as in the P/E case. The estimates for k, the exponent of the generalized or power mean in (2.1) that minimizes the squared error defined in (2.18) – the realized value is shown in column `realrat`, seem far from -1, the value of the harmonic average. As most of k estimates are close to zero, we foresee that the geometric mean will be the most appropriate to average the ratios P/E and EY.

As the two financial ratios are non-negative, the bootstrapping simulation study is conducted by splitting the ratios into two different categories: ratios whose value is greater than one (P/E) and ratios ranging from zero to one (EY). Thus, we exclude the case of negative ratios. The statistics of interest are the value-weighted average (VWA, or realized P/E), the realized EY value, the median, and the arithmetic, geometric and harmonic averages.

The value of the ratio results from the division of x by y: $r = \frac{x}{y}$. In the bootstrap simulation process we repeatedly draw samples from the eight original datasets with replacement 10,000 times to create simulated datasets. This process involves drawing random samples from the eight original samples. Thus, for each bootstrapped sample we get n values (the sample size) for r (P/E and EY ratios), x (Price in P/E and Earnings per share in EY) and y (Earnings per share in P/E and Price in EY): $r_1, r_2, \ldots, r_n$; $x_1, x_2, \ldots, x_n$; and $y_1, y_2, \ldots, y_n$.

A useful general principle to consider when deciding which mean is more appropriate is to replace each observation by the mean and see which one produces the correct result in the context of the question being asked. In this paper "the correct result" is the realized ratio, resulting from the division between the sum of numerators and the sum of denominators. By searching for the value of k that minimizes the Squared Error (SE)

$$SE = \left[\left(\frac{1}{n} \sum_{i}^{n} r_i^k \right)^{\frac{1}{k}} - rr \right]^2, \tag{2.18}$$

which is the difference between the average resulting from (2.1) and the realized ratio given by

$$rr = \frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} y_i}, \tag{2.19}$$

we are looking for the measure that is closer to the "correct result", the portfolio P/E realized value (`rr`). In our simulation, it is possible to compute the realized value of each

ratio at a given point in time because the individual values for x and y are available (see the situation 3 "All Different" with all x and y data available). The objective is to find, for each sample, the average value that is closer to the realized value of each ratio. At the end we get 10,000 estimates for k and we can obtain a 95% confidence interval based on the standard error of the sequence of estimates resulting from the simulations performed before. The simulation routines have been programmed in R and they are available upon request. Based on the limits of the central 95% bootstrapped confidence interval it is possible to infer about the true value for k and conclude about the most suitable measure to average the ratios. The results are shown in Tables 3 and 4, column k. The bootstrap is also used to estimate 95% confidence intervals for each measure (see columns HA, GA, MED and AA).
Rather than conducting a hypothesis test for whether the value of k is significantly different from $k = -1$ (harmonic average), $k = 0$ (geometric average) and $k = 1$ (arithmetic average), we calculate bootstrapped 95% confidence intervals for k. This enables us to assess if an interval estimate for k includes the values -1, 0 and 1. The geometric average shall be the preferred option if the lower confidence interval boundary is greater than -1 and the upper confidence interval boundary is less than unity (including the value 0); the harmonic average beats the remaining alternatives if the lower confidence interval boundary is less than -1 and the upper confidence interval boundary is less than zero (including the value -1); the arithmetic average is preferred when the lower limit is higher than zero and the upper limit is higher than 1 (including the value 1).

In the case of P/E (ratios with value higher than 1), we do not reject the null that $k = -1$ in one (DAX) of eight cases, and we do not reject the null that $k = 1$ in one (CAC40) of eight cases. Thus, the inference rules out the possibility that the harmonic or the arithmetic means can produce mostly the lowest difference in (2.18), questioning their suitability to average the ratios. On the other hand, the null of $k = 0$ is not rejected in all the eight stock indexes (the value 0 is always within the bootstrapped confidence intervals for k). As the geometric average is the particular case of the generalized average when $k \to 0$, our simulation suggests that it is the most appropriate measure to average the P/E ratios. Additional confirmation is provided, based on the bootstrapped confidence intervals for each of the central tendency measures. In all cases, no matter the stock index, the realized P/E value (see column `realrat` of Table 1) is always within the intervals for the geometric average (bold font figures). For the remaining three measures, the bootstrapped confidence intervals contain the realized P/E value in a few cases only.

The results confirm also the upward bias of the arithmetic average (where the lower limit of the confidence interval exceeds the realized value) and the downward bias of the harmonic average (where the upper limit of the confidence interval is lower than the realized value). Due to the presence of lower and upper outlying observations, neither the harmonic average nor the arithmetic average seem to be the most suitable measures to average the P/E ratios. Our simulation suggests that, as the geometric average provides intermediate values and is not as influenced by extreme observations. Therefore, it seems more appropriate to estimate the true average of price per unit of earnings, especially when only one share per stock is considered. This result leads us to choose the geometric average as the most appropriate measure to estimate the true average P/E when only one share per stock is considered ("All Different" situation). If the portfolio includes different quantities per stock, but we can admit, as aforementioned, equal investment or equal earnings per stock, the harmonic and arithmetic averages are still the preferred ones ("All Equal" and "Become Equal" situations, respectively). Conclusions for Earnings Yield (ratio between 0 and 1) are generally similar (see Table 4) also pointing to the geometric as a preferable measure to average ratios (proportions) when values range between 0 and 1.

We have adopted a non-dynamic perspective, because, in general, there are no reasons to change the value of the numerator and denominator to keep the ratio constant. Furthermore, the portfolio's realized P/E value is computed for simulation purposes only, as in practice the average is directly computed based on the ratios at a given point in time. A "true value" for the price per unit of earnings is needed (see the Introduction to clarify the importance of the "true average" – TA) for benchmarking. By minimizing the squared error in (2.18), we search for the estimate of k that "makes" the generalized average closer to the "true" value computed before. The resulting average is the harmonic, geometric or the arithmetic if $k = -1$, $k = 0$ or $k = 1$, respectively. A possible issue regarding the portfolio realized P/E value is that it can change, for the same sample, due to a stock split (or reverse stock split), while the companies' P/E ratios remain constant. However, this does not alter our conclusions because, as aforementioned, the realized value is computed for simulation purposes to assess the most suitable measure to average the ratios.For confirmation, we performed additional simulations. The purpose is to test, if, despite the stock slit, the geometric mean is still the most suitable measure to average the ratios. The conclusions seem clear against harmonic and arithmetic averages. The steps of this simulation were as follows:

(1) For each sample, corresponding to a particular stock index[8], we generated a random number between 1 and 10, one per company in the sample.

(2) We divided the price and earnings per share of each company by that random number. As we divide the numerator and denominator by the same constant, the P/E ratio per company remains constant (depicting the effect of a stock split), while the portfolio P/E realized value as well as the estimate for k can change.

[8]We consider the six stock indices with larger sample sizes to avoid the small sample bias.

(3) Based on the changing portfolio P/E realized value, we searched for the value of k that minimizes the squared error in (2.18).

(4) The simulation is repeated 10,000 times (results are shown in Table 5). As the estimate for k can change in the same sample simply due to a stock split, our purpose is not to find an "optimal" k, but to infer about the most appropriate mean to average the ratios, based on the simulated values of k.

Our results show that the estimate for k never reaches the value -1 ("the one of" the harmonic mean) in 4 of the 6 stock indices under analysis. In the case of FTSE100 and NASDAQ100 the minimum value for k is less than minus one, but the number of simulations where the estimate for k could "accommodate" the harmonic mean is very small: 75 (0.75%) and 5 (0.25%), respectively. On the other hand, all the 95% confidence intervals resulting from the simulated 2.5 and 97.5 percentiles (bold figures in the table) include the value 0 for k, but the limits never reach the values of -1 or 1, reinforcing the conclusion that the geometric mean seems more appropriate than the other two central tendency measures to average the P/E ratios. Thus, even if the portfolio realized P/E value and the estimate for k change due to the stock split, the null hypotheses of $k = 0$ is still not rejected considering the confidence interval resulting from the empirical percentiles.

2.5. Conclusions

Ratios represent a special kind of relation between two magnitudes, and averages of ratios are ubiquitous in many fields of application. Our study focuses on which average to use when computing the mean of individual ratios.

In finance, several authors argue that the harmonic mean is preferable to average ratios because the arithmetic mean assigns greater weights to higher ratios and lesser weight to lower ratios in the sample. However, the harmonic mean tends strongly toward the smallest elements, as it tends (compared to the arithmetic mean) to mitigate the impact of large outliers and amplify the impact of the smallest ones. Thus, the arithmetic mean tends to overestimate and the harmonic mean tends do underestimate the realized true value of the mean ratio. On the other hand the geometric mean is the one that provides a better fit to the portfolio realized P/E value. We confirm this by presenting and discussing our results based on real financial time series. Replacing the arithmetic with the geometric mean in the estimation of other widely applied financial ratios is an avenue for further research may prove a valuable alternative to better understand the underlying risk in portfolios of financial assets.

We note two special situations where one of the two means beats the other (and also the geometric average): if equal weight is assigned to the numerator's variable in each ratio, the harmonic mean is more suitable; when the same weight is given to the variable in the denominator, the arithmetic average provides a better result. If neither situation occurs, the geometric average, providing intermediate values, seems to be more appropriate to estimate the true average of P/E, especially in the presence of outlying observations.

2.6. Appendix

2.6.1. Contributions of each data point

See Equation (2.7):

$$\bar{X}_H = \sum_{i=1}^{n} x_i \frac{w_i}{\sum_{i=1}^{n} w_i} = \sum_{i=1}^{n} \left[x_i \frac{\frac{\sum_{i=1}^{n} x_i}{x_i}}{\frac{\sum_{i=1}^{n} x_i}{x_1} + \frac{\sum_{i=1}^{n} x_i}{x_2} + \ldots + \frac{\sum_{i=1}^{n} x_i}{x_n}} \right] = \sum_{i=1}^{n} \left[\frac{x_i \sum_{i=1}^{n} x_i}{x_i \sum_{i=1}^{n} x_i \sum_{i=1}^{n} \frac{1}{x_i}} \right] =,$$

$$= \sum_{i=1}^{n} \left[\frac{1}{\sum_{i=1}^{n} \frac{1}{x_i}} \right] = \frac{1}{\sum_{i=1}^{n} \frac{1}{x_i}} + \frac{1}{\sum_{i=1}^{n} \frac{1}{x_i}} + \ldots + \frac{1}{\sum_{i=1}^{n} \frac{1}{x_i}} = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}},$$

if $x_i \sum_{i=1}^{n} x_i \neq 0$.

See Equation (2.9):

$$x_i \frac{w_i}{\sum_{i=1}^{n} w_i} = x_i \frac{\frac{\sum_{i=1}^{n} x_i}{x_i}}{\frac{\sum_{i=1}^{n} x_i}{x_1} + \frac{\sum_{i=1}^{n} x_i}{x_2} + \ldots + \frac{\sum_{i=1}^{n} x_i}{x_n}} = \frac{x_i \sum_{i=1}^{n} x_i}{x_i \sum_{i=1}^{n} x_i \sum_{i=1}^{n} \frac{1}{x_i}} = \frac{1}{\sum_{i=1}^{n} \frac{1}{x_i}},$$

if $x_i \sum_{i=1}^{n} x_i \neq 0$.

2.6.2. Average inequalities

According to the means definition, see Equations (2.2), (2.3) and (2.5), their logarithms are:

$$\ln\left(\bar{X}\right) = \ln\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right), \quad \ln\left(\bar{X}_G\right) = \frac{1}{n}\sum_{i=1}^{n} \ln\left(X_i\right) \text{ and } \ln\left(\bar{X}_H\right) = -\ln\left(\frac{1}{n}\sum_{i=1}^{n} \frac{1}{X_i}\right).$$

By Jensen's inequality,

$$\ln\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \geq \frac{1}{n}\sum_{i=1}^{n} \ln\left(X_i\right),$$

which can be exponentiated to give the arithmetic mean-geometric mean inequality:

$$\underbrace{\frac{1}{n}\sum_{i=1}^{n} X_i}_{\bar{X}} \geq \underbrace{\left(\prod_{i=1}^{n} X_i\right)^{\frac{1}{n}}}_{\bar{X}_G}, \quad \text{thus } \bar{X} \geq \bar{X}_G.$$

Now comparing the harmonic with the geometric mean (and by Jensen's inequality):

$$-\ln\left(\frac{1}{n}\sum_{i=1}^{n} \frac{1}{X_i}\right) \leq -\frac{1}{n}\sum_{i=1}^{n} \ln\left(\frac{1}{X_i}\right) = \frac{1}{n}\sum_{i=1}^{n} \ln\left(X_i\right),$$

and by exponentiating both sides:

$$\underbrace{\frac{n}{\sum_{i=1}^{n} \frac{1}{X_i}}}_{\bar{X}_H} \leq \underbrace{\left(\prod_{i=1}^{n} X_i\right)^{\frac{1}{n}}}_{\bar{X}_G}, \quad \text{thus } \bar{X}_H \leq \bar{X}_G.$$

2.6.3. Different meanings of the center

Median:

Consider a first data set: $4, 6, 10, 100$ ($n = 4$, even). Thus, the median rank is $r_M = \frac{1+4}{2} = 2.5$, the median is $\bar{X}_M = \frac{6+10}{2} = 8$ and $\sum_{i=1}^{4} (r_i - r_M) = (1-2.5)+(2-2.5)+(3-2.5)+(4-2.5) = 0$.

For a second data set $4, 6, 10, 20, 100$ ($n = 5$, odd), the median rank is $r_M = \frac{1+5}{2} = 3$, the median is $\bar{X}_M = 10$ and $\sum_{i=1}^{5} (r_i - r_M) = (1-3)+(2-3)+(3-3)+(4-3)+(5-3) = 0$. Thus, the median is the center of the distribution in terms of the counting observations: one half of the observations is on the left and one half is on the right of the median, no matter the value of the observations.

Arithmetic average:

Consider again the second data set: $\bar{X}_A = \frac{4+6+10+20+100}{5} = 28$ and $\sum_{i=1}^{5} \left(x_i - \bar{X}_A\right) = (4-28) + (6-28) + (10-28) + (20-28) + (100-28) = 0$.

The arithmetic average is the center of the distribution in terms of the deviations in absolute terms: the arithmetic mean is such that the absolute deviations on its right is compensate by the absolute deviations on its left. So, the center is defined in terms of the absolute deviations (or distances) between each value and the arithmetic average:

$$\underbrace{\underbrace{(4-28)}_{-24}+\underbrace{(6-28)}_{-22}+\underbrace{(20-28)}_{-8}\underbrace{(20-28)}_{-8}}_{-72}+\underbrace{(100-28)}_{+72}.$$

Geometric average:

For the second data set: $\bar{X}_G = \sqrt[5]{4 \times 6 \times 10 \times 20 \times 100} = 13.69$ and

x_i	$\ln(x_i)$	$\ln(x_i) - \ln(\bar{X}_G)$
4	1.386	-123.00L%
6	1.792	-82.45L%
10	2.303	-31.37L%
20	2.996	37.94L%
100	4.605	198.89L%
		Sum = 0

Compounding percentage deviation means:
$4 = 13.69 \times \exp(-123L\%)$, ..., $100 = 13.69 \times \exp(198.89L\%)$, where $L\%$ is the log percentage, see **Tornqvist1985**.

The geometric average is the value that balances the negative and positive log percentage deviations.

Harmonic average:

x_i	$x_i - \bar{X}_H$	$w_i = 1/x_i$	$\frac{w_i}{\sum_{i=1}^n w_i}$	$\left(x_i - \bar{X}_H\right)\left(\frac{w_i}{\sum_{i=1}^n w_i}\right)$
4	-4.671	0.250	0.434	-2.025
6	-2.671	0.167	0.289	-0.772
10	1.329	0.100	0.173	0.231
20	11.329	0.050	0.087	0.982
100	91.329	0.010	0.017	1.584
		0.577	1	0

The harmonic average defines the center of the distribution such that the weighted deviations on its left compensate the weighted deviations on its right. The weights are inversely proportional to the original values.

TABLE 1. Descriptive statistics – Price Earnings Ratio (P/E)

Index	# obs	skew	kurt	realrat	k	HA	GA	MED	AA
S&P500	421	8.0097	102.171	27.89	-0.1037	22.62	28.62	27.62	38.43
DAX	23	4.4375	20.8144	23.85	-0.6752	22.16	30.94	22.95	113.77
NIKKEI	171	12.4934	160.346	23.30	0.0988	16.41	22.21	18.73	82.22
FTSE100	74	4.2071	23.6284	23.96	-0.1333	18.57	24.98	21.69	36.76
CAC40	32	1.2931	4.4211	37.68	0.9478	22.06	29.07	27.63	38.18
NASDAQ100	90	6.5800	49.9882	41.87	0.0395	32.53	41.36	33.48	66.49
SSEC	1268	11.4250	190.5089	33.31	-0.0491	23.36	34.06	31.18	64.10
STOXX600	467	9.0850	126.0639	27.50	0.3191	14.72	26.01	24.04	40.89

skew: skewness coefficient, kurt: kurtosis coefficient; `realrat` is the realized ratio: the sum of prices divided by the sum of earnings per share; k is the estimate for k, the exponent of the generalized or power mean in (2.1) that minimizes the squared error defined in (2.18); the columns HA, GA, MED and AA refer to the harmonic, geometric, median and arithmetic averages, respectively.

Table 2. Descriptive statistics: Earnings Yield – EY

Index	# obs	skew	kurt	realrat	k	HA	GA	MED	AA
S&P500	421	5.68	62.71	0.0359	0.1037	0.0260	0.0349	0.0362	0.0442
DAX	23	0.54	2.48	0.0419	0.6753	0.0088	0.0323	0.0436	0.0451
NIKKEI	171	0.88	3.23	0.0429	-0.0988	0.0122	0.0450	0.0534	0.0609
FTSE100	74	4.94	34.78	0.0417	0.1333	0.0272	0.0400	0.0461	0.0538
CAC40	32	1.23	3.71	0.0265	-0.9478	0.0262	0.0344	0.0362	0.0453
NASDAQ100	90	1.74	8.26	0.0239	-0.0395	0.0150	0.0242	0.0299	0.0307
SSEC	1268	1.84	6.98	0.0300	0.0492	0.0156	0.0294	0.0321	0.0428
STOXX600	467	21.09	451.98	0.0379	-0.3191	0.0245	0.0384	0.0416	0.0679

skew: skewness coefficient, kurt: kurtosis coefficient; `realrat` is the realized ratio: the sum of earnings per share divided by the sum of prices; k is the estimate for k, the exponent of the generalized or power mean in (2.1) that minimizes the squared error defined in (2.18); the columns HA, GA, MED and AA refer to the harmonic, geometric, median and arithmetic averages, respectively.

Table 3. Price Earnings Ratio - P/E (ratios higher than 1)

	k		HA		GA		MED		AA	
					S&P500					
Normal	-0.5751	0.3650	20.78	24.38	**26.67**	**30.50**	**25.65**	**29.41**	33.94	42.82
Basic	-0.5649	0.3657	20.77	24.37	**26.58**	**30.44**	**26.09**	**29.56**	33.64	42.36
Percentile	-0.5731	0.3575	20.87	24.48	**26.79**	**30.65**	**25.68**	**29.15**	33.64	43.22
BCα	-0.6205	0.3235	20.55	24.21	**26.80**	**30.66**	**25.57**	**29.12**	33.64	44.27
					DAX					
Normal	-1.5608	0.2886	**15.88**	**27.64**	**14.94**	**45.28**	**14.97**	**28.77**	-39.21	268.37
Basic	-1.5231	0.3811	**15.05**	**26.73**	**10.47**	**40.87**	**9.55**	**24.12**	-53.30	202.48
Percentile	-1.7316	0.1726	**17.58**	**29.27**	**21.02**	**51.42**	**21.79**	**36.36**	-53.30	280.84
BCα	-1.7141	0.1934	**17.33**	**28.70**	**22.55**	**60.44**	**20.19**	**31.10**	-53.30	439.57
					NIKKEI					
Normal	-0.3326	0.5447	14.71	18.05	**18.93**	**25.41**	16.13	20.58	**-6.68**	**170.82**
Basic	-0.3745	0.5095	14.61	17.93	**18.60**	**25.08**	15.84	20.31	**-16.92**	**136.61**
Percentile	-0.3119	0.5721	14.88	18.20	**19.34**	**25.81**	17.15	21.62	27.82	181.35
BCα	-0.2538	0.6839	14.87	18.19	**19.54**	**26.23**	16.96	21.24	32.87	283.89
					FTSE100					
Normal	-0.6787	0.4151	14.160	22.451	**20.331**	**29.392**	**17.603**	**25.147**	26.045	47.468
Basic	-0.6916	0.3875	14.101	22.401	**19.954**	**29.041**	**16.866**	**24.363**	25.053	46.187
Percentile	-0.6542	0.4249	14.743	23.043	**20.919**	**30.006**	**19.010**	**26.507**	25.053	48.460
BCα	-0.6131	0.4894	13.624	22.044	**20.986**	**30.104**	**18.675**	**25.989**	25.053	52.505
					CAC40					
Normal	-0.0124	1.8694	15.51	27.73	20.95	36.55	**16.49**	**36.28**	**28.08**	**48.12**
Basic	-0.1349	1.7750	14.71	26.77	**20.39**	**35.81**	**10.26**	**33.54**	**27.56**	**47.59**
Percentile	0.1207	2.0306	17.35	29.42	**22.33**	**37.75**	**21.72**	**45.00**	**27.56**	**48.80**
BCα	0.1744	2.1518	16.87	28.53	22.24	37.56	**18.08**	**43.26**	**27.56**	**49.87**
					NASDAQ100					
Normal	-0.5143	0.6067	27.92	36.83	**34.44**	**48.02**	26.43	39.19	**38.97**	**94.02**
Basic	-0.5555	0.5811	27.66	36.58	**33.83**	**47.32**	24.03	36.76	**34.23**	**87.92**
Percentile	-0.5021	0.6346	28.48	37.40	**35.40**	**48.88**	**30.21**	**42.94**	**34.23**	**98.74**
BCα	-0.4329	0.7695	28.15	36.98	**35.86**	**49.73**	**30.21**	**42.94**	**34.23**	**120.23**
					SSEC					
Normal	-0.1911	0.1015	22.22	24.46	**32.25**	**35.84**	29.65	32.82	55.61	72.55
Basic	-0.1914	0.1039	22.21	24.44	**32.18**	**35.78**	29.54	32.95	55.02	71.93
Percentile	-0.2021	0.0931	22.28	24.50	**32.35**	**35.95**	29.41	32.83	55.02	73.17
BCα	-0.1866	0.1079	22.26	24.47	**32.37**	**35.96**	29.37	32.77	55.02	75.01
					STOXX600					
Normal	-0.0898	0.8007	7.28	20.56	**23.92**	**28.06**	21.82	25.87	34.66	47.07
Basic	-0.0662	0.8067	8.92	19.72	**23.86**	**27.98**	21.83	25.66	34.08	46.38
Percentile	-0.1685	0.7044	9.72	20.52	**24.05**	**28.17**	22.42	26.25	34.08	47.71
BCα	-0.0692	0.8230	6.73	19.85	**24.06**	**28.17**	22.34	26.11	34.08	49.66

The lower and upper limits of the (central) 95% bootstrap confidence intervals are computed as follows: Standard (Normal), Percentile (Percentile), Pivotal or Empirical (Basic) and Accelerated bias-corrected (BCα). See, for example, **Manteiga1994**, for bootstrap details. k is the estimate for k, the exponent of the generalized or power mean in (2.1) that minimizes the squared error defined in (2.18). HA: Harmonic Average, GA: Geometric Average, MED: Median and AA: Arithmetic Average. Bold font indicates that the portfolio P/E realized value (see Table 1) is inside the confidence interval for each central tendency measure.

Table 4. Earnings Yield – EY (ratios between 0 and 1)

	k		HA		GA		MED		AA	
	Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper
S&P500										
Normal	-0.3656	0.5722	0.0230	0.0290	**0.0326**	**0.0373**	**0.0338**	**0.0388**	0.0407	0.0477
Basic	-0.3772	0.5583	0.0229	0.0290	**0.0325**	**0.0372**	**0.0333**	**0.0381**	0.0404	0.0475
Percentile	-0.3509	0.5846	0.0231	0.0291	**0.0327**	**0.0374**	**0.0343**	**0.0391**	0.0409	0.0480
BCα	-0.3206	0.6185	0.0225	0.0286	**0.0327**	**0.0374**	**0.0343**	**0.0389**	0.0413	0.0486
DAX										
Normal	-0.2982	1.5754	-0.0217	0.0255	**0.0175**	**0.0456**	**0.0358**	**0.0537**	**0.0338**	**0.0565**
Basic	-0.4261	1.5308	-0.0225	0.0140	**0.0170**	**0.0447**	**0.0376**	**0.0550**	**0.0332**	**0.0560**
Percentile	-0.1803	1.7766	0.0036	0.0400	**0.0200**	**0.0477**	**0.0322**	**0.0495**	**0.0342**	**0.0570**
BCα	-0.1678	1.8013	0.0022	0.0341	**0.0175**	**0.0450**	**0.0275**	**0.0456**	**0.0349**	**0.0577**
NIKKEI										
Normal	-0.5444	0.3232	-0.0102	0.0259	**0.0385**	**0.0513**	0.0483	0.0603	0.0548	0.0671
Basic	-0.5032	0.3725	-0.0114	0.0188	**0.0384**	**0.0510**	0.0485	0.0605	0.0546	0.0669
Percentile	-0.5701	0.3056	0.0056	0.0357	**0.0390**	**0.0517**	0.0463	0.0583	0.0550	0.0673
BCα	-0.7089	0.2510	0.0036	0.0308	**0.0385**	**0.0512**	0.0453	0.0572	0.0552	0.0674
FTSE100										
Normal	-0.4053	0.6764	0.0185	0.0349	**0.0326**	**0.0472**	**0.0393**	**0.0536**	0.0413	0.0663
Basic	-0.3675	0.7078	0.0177	0.0341	**0.0322**	**0.0468**	**0.0399**	**0.0538**	**0.0398**	**0.0643**
Percentile	-0.4412	0.6341	0.0203	0.0367	**0.0332**	**0.0479**	**0.0385**	**0.0523**	0.0434	0.0679
BCα	-0.4985	0.5947	0.0189	0.0348	**0.0330**	**0.0476**	**0.0377**	**0.0523**	0.0454	0.0742
CAC40										
Normal	-1.8664	0.0206	0.0187	0.0328	**0.0250**	**0.0433**	**0.0262**	**0.0473**	0.0334	0.0573
Basic	-1.7713	0.1419	0.0176	0.0319	**0.0240**	**0.0424**	**0.0245**	**0.0502**	0.0329	0.0567
Percentile	-2.0375	-0.1242	0.0205	0.0348	**0.0264**	**0.0448**	**0.0222**	**0.0479**	0.0340	0.0577
BCα	-2.0959	-0.1533	0.0201	0.0339	**0.0265**	**0.0450**	**0.0222**	**0.0470**	0.0351	0.0592
NASDAQ100										
Normal	-0.6128	0.5253	0.0081	0.0206	**0.0202**	**0.0279**	0.0254	0.0350	0.0265	0.0349
Basic	-0.5747	0.5736	0.0077	0.0198	**0.0201**	**0.0278**	0.0266	0.0364	0.0262	0.0347
Percentile	-0.6527	0.4957	0.0103	0.0224	**0.0205**	**0.0282**	**0.0233**	**0.0331**	0.0268	0.0353
BCα	-0.7938	0.4269	0.0088	0.0206	**0.0203**	**0.0279**	0.0230	0.0331	0.0271	0.0356
SSEC										
Normal	-0.1028	0.1941	0.0135	0.0176	**0.0278**	**0.0309**	0.0303	0.0336	0.0408	0.0449
Basic	-0.1013	0.1924	0.0134	0.0176	**0.0278**	**0.0309**	0.0301	0.0337	0.0407	0.0448
Percentile	-0.0941	0.1996	0.0137	0.0178	**0.0278**	**0.0310**	0.0305	0.0341	0.0408	0.0449
BCα	-0.1095	0.1876	0.0134	0.0175	**0.0278**	**0.0310**	0.0305	0.0341	0.0409	0.0450
STOXX600										
Normal	-0.7983	0.0933	0.0206	0.0280	**0.0353**	**0.0415**	0.0384	0.0453	0.0350	0.1003
Basic	-0.7990	0.0632	0.0207	0.0280	**0.0352**	**0.0414**	0.0386	0.0451	**0.0324**	**0.0871**
Percentile	-0.7015	0.1607	0.0209	0.0282	**0.0355**	**0.0416**	0.0381	0.0446	0.0488	0.1035
BCα	-0.8182	0.0721	0.0200	0.0276	**0.0355**	**0.0416**	0.0381	0.0444	0.0503	0.1532

The lower and upper limits of the central 95% bootstrap confidence intervals are computed as follows: Standard (Normal), Percentile (Percentile), Pivotal or Empirical (Basic) and Accelerated bias-corrected (BCα). See, for example, **Manteiga1994**, for bootstrap details. k is the estimate for k, the exponent of the generalized or power mean in (2.1) that minimizes the squared error defined in (2.18). HA: Harmonic Average, GA: Geometric Average, MED: Median and AA: Arithmetic Average. The boldness indicates that the portfolio P/E realized value (see Table 1) is inside the confidence interval for each central tendency measure.

TABLE 5. Price Earnings Ratio - P/E, simulating stock splits

					Percentiles						
Indices	$\# \leq -1$	MIN	0.50%	1.00%	2.50%	5.00%	95.00%	97.50%	99.00%	99.50%	MAX
S&P 500	0	-0.9878	-0.7112	-0.6985	**-0.6152**	-0.5193	0.2671	**0.3377**	0.4054	0.4515	0.6242
NIKKEI 225	0	-0.6376	-0.3174	-0.2763	**-0.2273**	-0.1852	0.3798	**0.4114**	0.4412	0.4570	0.5171
FTSE 100	75	-1.3731	-1.0709	-1.0011	**-0.8526**	-0.7298	0.4006	**0.5005**	0.5949	0.6495	0.8991
NASDAQ 100	5	-1.2136	-0.7867	-0.7189	**-0.5926**	-0.4853	0.4470	**0.5097**	0.5839	0.6247	0.7661
SSEC	0	-0.3006	-0.2457	-0.2266	**-0.2003**	-0.1780	0.0714	**0.0925**	0.1158	0.1335	0.2024
STOXX 600	0	-0.2516	-0.0982	-0.0693	**-0.0185**	0.0260	0.6125	**0.6616**	0.6996	0.7282	0.8328

The estimates for the exponent k in (2.18) are obtained based on 10,000 simulations. $\# \leq -1$ represents the number of simulations where the estimate for k is at most -1, accommodating the harmonic average.

CHAPTER 3

Revisiting relative importance: a VIF-based measure

ABSTRACT

In many multiple regression applications researchers are concerned with relative importance of predictor variables. We revisit the concept and propose a new measure based on part correlations and the Variance Inflation Factor (VIF). The measure assigns shares of variation that combine individual contributions from regressors with a VIF-weighted "common variance" component. The shares add up to the coefficient of determination and rank regressors accordingly. The measure showed acceptable variability for applications when simulated under different conditions. Especially in situations involving many regressors, our measure is advantageous in that it provides an intuitive and additive variable ranking while requiring low computational intensity. Results are discussed with a real life application.

3.1. Introduction

In many linear regression applications, assessing the relative importance of a set of explanatory variables (EV) is one of the key goals of researchers, particularly in sciences that work with observational data such as finance, psychology, weather, etc. Namely, researchers are frequently interested in ranking EV with respect to their individual impact in the dependent variable (DV). When regressors are uncorrelated, the question of relative importance has an unique solution: the standardized coefficients (β_j) and zero-order correlations ($\rho_{Y,X_j}, j = 1, ..., k$) are equal and therefore the single contribution of each regressor can be isolated. However, because observational data are frequently correlated, isolating said contributions is not obvious. Decomposing DV total variation around the mean in the presence of collinearity presents a challenge and an all encompassing solution, if it exists, has not yet been found.

Several methods are available, and many differ on what is meant by "variable importance" (**Bring1994**; **Kruskal1989**; **Achen1982**; **Thomas1999**). Thus, relative importance should be seen as construct, as opposed to a plain, objective quantity to be estimated. Typically, importance metrics focus on partitioning variance based on regression sums of squares and calculate shares of DV variation to quantify the individual contribution of each EV to a model's predictive ability. For the purpose of the present article, we measure predictive ability by the coefficient of determination (R^2). It is interpreted as the explanatory power of the regression, i.e. the percentage of total variation in the DV that can be accounted for by the in-sample variation of EV. We also borrow the following definition of 'relative importance" from **Johnson2004**:

> *The proportionate contribution each predictor makes to R^2, considering both its direct effect (i.e., correlation with the criterion) and its effect when combined with the other variables in the regression equation.*

The nature of the research question should determine the selection of importance metrics. Measures can be broadly classified into three distinct groups: *single analysis*, *multiple analysis* and *variable transformation*. *Single analysis* methods focus on the output of a single regression model and measure relative importance based on estimated correlations and/or regression coefficients (**Hoffman1962**; **Thompson1985**; **Courville2001**). By making use of readily available quantities, single analysis typically requires low computational intensity. These measures are more appropriate to evaluate importance when the entry-order of EV is not relevant or a theoretically meaningful entry order is known (the measure introduced in this paper falls into this category, see **Lmg1980**). *Multiple analysis* methods draw from several regression outputs generated by different permutations of EV (to predict the DV). **Lmg1980** proposed measure LMG (named after the authors) which employs squared part (also called semipartial) correlations to calculate relative importance, by averaging the correlations across all possible combinations of predictors.

Feldman2005 modified LMG by assigning a data-dependent set of average weights (instead of uniform) to orders of regressors, to fulfill the so-called exclusion criterium (i.e. attributing no importance to EV with $\beta = 0$) . Using simulation, **Gromping2006** compared the two measures along a range of covariance matrices and concluded that they provide similar results in most cases, although with substantial differences in variability (the same author provided the package `relaimpo` for R, which calculates several of the available importance metrics). **Budescu1993** introduced the notion of dominance, claiming that if the predictive ability of one EV does not exceed that of another across all permutations of EV, variables cannot be ranked meaningfully and therefore, a dominance analysis should always be conducted (for a recent application see **Stadler2017**). *Variable transformation methods* resemble principal-component analysis in that EV are transformed to be as highly related as possible to the original EV while preserving their individual characteristics. Notably, **Johnson2000** suggested calculating *relative weights*, by regressing the original EV onto their orthogonal counterparts, which are frequently cited and employed in applied research for their convenience and interpretability. For more recent work on transformation methods, see **Lipovetsky2015** and **Garthwaite2019**.

We propose a new measure of relative importance to deal with the known issue of collinearity in single analysis methods. The usage of regression coefficients (in their standardized and unstandardized forms), partial and semipartial correlations and t statistics to evaluate relative importance has been criticized due to its sensitivity to correlation between regressors. We address this issue by combining semipartial correlations with the Variance Inflation Factor (VIF), a quantity that is commonly employed to diagnose collinearity in the estimation of linear regression models, into a measure that is both intuitive and easy to compute.

This paper is organized as follows: in Section 2 we provide the theoretical background for our measure by revisiting familiar collinearity concepts. In Section 3 we test our measure under simulation to evaluate its behavior in different settings. In Section 4 we apply our measure to a real data set and discuss the results. Section 5 contains some concluding remarks, and a snippet of R code to compute the proposed measure is provided in the Appendix.

3.2. Theoretical background

As mentioned in the introduction, allocating variation to regressors when they are linearly uncorrelated has an unique and simple solution. We start by noting that

$$R^2 = \frac{SSR}{SST} = \frac{\sum_{i=1}^{n}\left(\hat{Y}_i - \bar{Y}\right)^2}{\sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2} = 1 - \frac{\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2}{\sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2}, \tag{3.1}$$

where SSR and SST are the Regression and Total Sum of Squares, respectively. By simple manipulation, R^2 can be expressed as a ratio of the sample variances of $\hat{Y}$ and Y:

$$R^2 = \frac{\frac{\sum_{i=1}^{n}\left(\hat{Y}_i - \bar{Y}\right)^2}{n-1}}{\frac{\sum_{i=1}^{n}\left(Y_i - \bar{Y}_i\right)^2}{n-1}} = \frac{S^2_{\hat{Y}}}{S^2_Y}. \tag{3.2}$$

Also, the predicted value of Y_i for k regressors is given by:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \ldots + \hat{\beta}_k X_{ki} \tag{3.3}$$

and therefore the variance of $\hat{Y}$ (the numerator of the last term of Equation 3.2) can be written as:

$$S^2_{\hat{Y}} = \sum_{j=1}^{k} \hat{\beta}_j^2 S^2_{X_j} + \sum_{j=1}^{k}\sum_{l=1}^{k} \hat{\beta}_j \hat{\beta}_l S_{X_j X_l}, j \neq l \tag{3.4}$$

By combining Equations (3.2) and (3.4), we can write:

$$R^2 = \frac{S^2_{\hat{Y}}}{S^2_Y} = \frac{\sum_{j=1}^{k} \hat{\beta}_j^2 S^2_{X_j} + \sum_{j=1}^{k}\sum_{l=1}^{k} \hat{\beta}_j \hat{\beta}_l S_{X_j X_l}}{S^2_Y}, \tag{3.5}$$

where $S^2_{X_j}$ is the sample variance of the explanatory variable X_j and $S_{X_j X_l}$ is the sample covariance between the explanatory variables X_j and X_l. As the second summand in the numerator represents the non-diagonal elements of the covariance matrix, the individual contribution of regressor X_j to R^2 reduces to the squared standardized coefficient $\hat{\beta}_j^2 S^2_{X_j}/S^2_Y$ if all regressors are linearly independent (all non-diagonal elements are zero). If that is not the case, allocating the covariance summand to individual regressors requires a non-trivial choice of weights (i.e. how much covariance should be allocated to each regressor). Additionally, because it can take negative values, the covariance summand may arbitrarily lead to negative shares of R^2, which may be more difficult to interpret.

We propose a heuristic measure based on two familiar concepts: part correlations and the Variance Inflation Factor (VIF), which we approach in subsections 3.2.1 and 3.2.2. Throughout this section, we use an illustrative case of one dependent (DV) and three independent (EV) variables when needed for clarity, without loss of generality.

3.2.1. Part correlations

Let Y be the (standardized) DV, X_1, X_2 and X_3 be three standardized and arbitrarily correlated random variables (EV), and consider the linear estimated model:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} \tag{3.6}$$

Squared part correlations are natural candidates to evaluate relative importance of EV, because they measure the nominal change in R^2 as that particular EV is removed from the model. That is, for X_1:

$$sr_1^2 = R^2_{Y.123} - R^2_{Y.(1)23} \tag{3.7}$$

where (1) means the exclusion of X_1 from the model. It must be noted that sr_i^2 can be interpreted in different ways and can be obtained by (1) regressing X_i on the remaining $k-1$ EV (2) extracting the residuals and (3) squaring the Pearson correlation of those residuals with Y, i.e. if we let $\hat{e}_{X_j}$ be the residuals from regressing X_j on the remaining two variables:

$$\sum sr_i^2 = \underbrace{\frac{Cov^2(Y, \hat{e}_{X_1})}{Var(Y)Var(\hat{e}_{X_1})}}_{sr_1^2} + \underbrace{\frac{Cov^2(Y, \hat{e}_{X_2})}{Var(Y)Var(\hat{e}_{X_2})}}_{sr_2^2} + \underbrace{\frac{Cov^2(Y, \hat{e}_{X_3})}{Var(Y)Var(\hat{e}_{X_3})}}_{sr_3^2} \tag{3.8}$$

Cohen1975 showed that the sum of squared part correlations generally does not exceed R^2 and thus the following holds:

$$\sum sr_i^2 \leq R^2_{Y.123} \tag{3.9}$$

Because each sr_i^2 represents the individual contribution of EV X_i to the model's R^2, when $\sum sr_i^2 < R^2_{Y.123}$, a measure of "shared variance" ($svar$) (i.e. the percentage of DV variance that is commonly explained by the EV) can be obtained as

$$\sum sr_i^2 + svar = R^2_{Y.123} \Leftrightarrow svar = R^2_{Y.123} - \sum sr_i^2 \tag{3.10}$$

i.e. the variance that is not captured individually by the EV is stored in *svar*. We note that *svar* is not covariance as defined in Equation (3.5), because in the presence of collinearity $\hat{\beta}_j^2 S^2_{X_j}/S^2_Y \neq sr_j^2$. Our measure (see Section 3.2.3) is a method of allocating shares of *svar* to each EV and ensuring that R^2 (positive) shares sum to the total R^2.

3.2.2. Variance Inflation Factor (VIF)

The VIF is a commonly employed measure in collinearity diagnostics of linear models. For EV X_j it is given by:

$$VIF_j = \frac{1}{1 - R^2_{X_j}} \tag{3.11}$$

where $R^2_{X_j}$ is the coefficient of determination from regressing X_j on the remaining $(k-1)$ EV. We can see that the typical rule of thumb that excludes variables with $VIF > 10$ implies a threshold of 0.1 for $R^2_{X_j}$, i.e. if more than 90% of X_j variation is "explained" by the remaining EV, then X_j is typically removed and the model re-estimated.

The term *inflation* stems from the fact that the variance of the unstandardized coefficient can be expressed as

$$\hat{Var}(\hat{\beta}_j) = \frac{S^2}{(n-1)S'^2_{X_j}} * \frac{1}{1-R^2_{X_j}} \tag{3.12}$$

where S^2 is the variance of the model's residuals and $s'^2_{X_j}$ is the corrected sample variance of regressor X_j. The second term on the rhs (VIF_j) represents the number of times the presence of X_j "inflates" $v\hat{a}r(\hat{\beta}_j)$, compared to the variance otherwise obtained if X_j was orthogonal to the remaining EV, where $VIF_j = 1$. It is thus a measure of how strongly correlated X_j is with the other predictors.

In the next subsection, we show how to combine squared part correlations with VIF to obtain additive shares of R^2.

3.2.3. An alternative to partition variance

A measure of shared variance can be obtained by calculating $svar$ (Equation (3.10)). The challenge is how to split $svar$ into shares so that $\sum R_j^2 = R^2$. To address the problem, we start by generating a set of weights v:

$$v_j = \frac{VIF_j}{\sum VIF_j}, \tag{3.13}$$

and then compute R_j^2 as

$$R_j^2 = \frac{Cov^2(Y, \hat{e}_{X_j})}{Var(Y)Var(\hat{e}_{X_j})} + v_j * svar = sr_j^2 + v_j * svar, \tag{3.14}$$

where R_j^2 is the nominal share of the model's R^2 allocated to EV X_j. In words, it is the sum of its squared semipartial correlation and its VIF-weighted $svar$. While allocating $svar$ to EV has no unique solution and remains arbitrary, any procedure attempting to do so must be practical and intuitive. The rationale behind using VIF is that the share of commonly explained variance should be higher for EV that are most correlated with their counterparts and vice-versa.

TABLE 1. Desirability criteria for measures of relative importance

Criterium	**Definition**
1. Positivity	All shares (of variance) must be positive
2. Additivity	The sum of shares must be the model's variance
3. Inclusion	Any regressor with $\beta \neq 0$ must receive a positive share
4. Exclusion	Any regressor with $\beta = 0$ must receive a zero share

A set of desirability criteria is commonly accepted in the literature, so as to guide researchers when developing methods for variance decomposition. Table 1 summarizes those criteria. By definition, our measure fulfills all but Criterium 4, as an EV with $\beta = 0$ may still be assigned a share of $svar$ because $VIF_j \geq 1$. Although this criterium is not fulfilled by definition, the typical exclusion of EV with statistically insignificant coefficients avoids the assignment of non-zero shares in most cases. We also note that fulfilment of this criterium has been considered optional when causal interpretations are in mind (see **Gromping2006**), exactly the type of interpretation our measure is meant to address.

We note that, for $svar \geq 0$,

$$\sum R_j^2 = \sum sr_j^2 + v_j * svar \tag{3.15}$$

and that for $svar = 0$ (i.e. orthogonal EV)

$$\sum R_j^2 = \sum sr_j^2 \tag{3.16}$$

where the sums are over all EV. From Equations (3.15) and (3.16) we can see that if EV are statistically independent, each allocated share is both its respective squared part correlation and squared standardized coefficient.

3.2.4. Supression

Suppression can be understood to mean that the dependence between EV is distorting their real relation to the DV, which, as measured by the regression coefficients, may in fact be larger or of opposite sign. Supression can take different forms and may be empirical or induced (for a complete discussion on the different types of supression see **Ludlow2014** and **Tzelgov1991**). For example, when the introduction of a supressor EV X_j increases the variance of Y explained by the EV already in the model (as measured by their squared part correlation) and, in turn, the squared part correlation of X_j also increases, supression is called *cooperative*: the inclusion of a cooperative supressor mutually reinforces the predictive power of the "cooperating" EV. **Cohen1975** showed that while in regular cases (no supresssion) the sum of squared semipartial correlations does not exceed R^2, the bound does not necessarily hold if supression is present, as it may lead to $\sum sr_j^2 > R^2_{Y.123}$ and make *svar* in Equation (3.10) negative. In turn, a negative *svar* may render negative R_j^2 for EV with high VIF_j and low sr_j^2. As negative values of *svar* "may be obtained in situations where some of the variables act as supressors" (see **pedhazur1997**), we follow the approach of **Thompson2006** and recommend that our measure be applied after checking for the presence of supressor variables. This can be achieved by confirming that no predictor simultaneously has a zero (or near-zero) structure coefficient (i.e. the Pearson correlation between predictors and $\hat{Y}$) and a large absolute regression coefficient.

While it would be convenient to discard it as a nuisance to variance partitioning, supression is not common but does occur. For instance, **Pandey2010** reviewed several cases of supression in applied social research, and we note that a careful analysis of the full correlation matrix is indispensable for any method of variance decomposition to deliver accurate conclusions.

3.3. Simulation

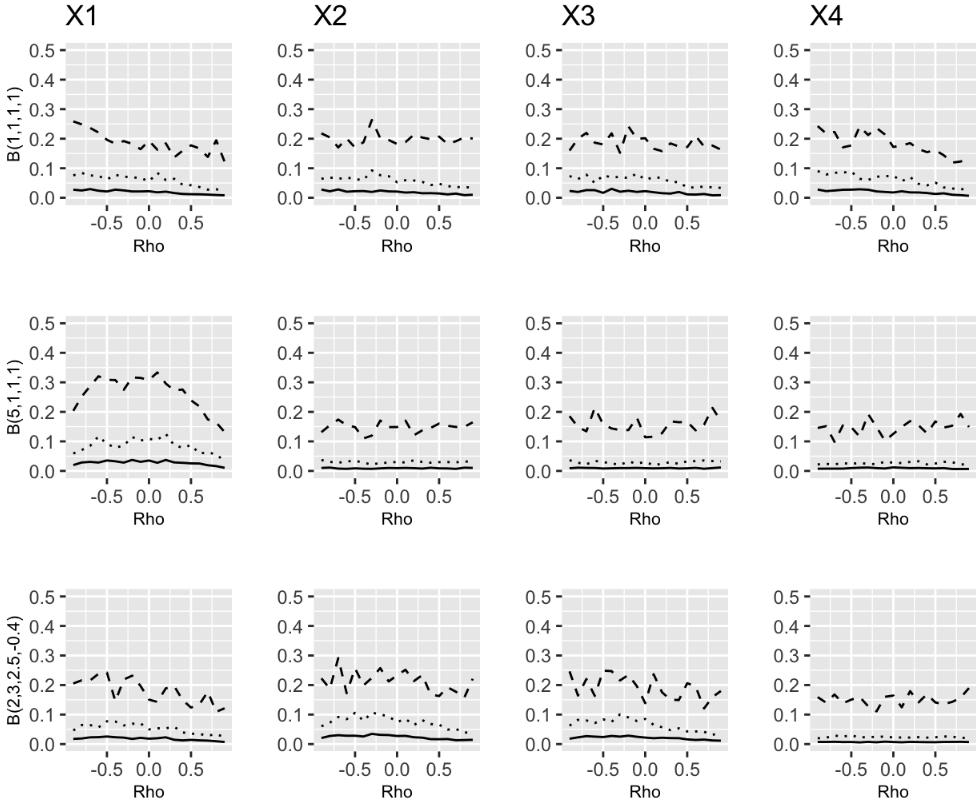
In this section we investigate the behavior of the proposed measure in terms of its variability under different conditions. Table 2 shows the simulation settings. We simulate four multivariate normal regressors with mean 0 and variance 1, i.e. $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + u$, where u is an independent normal term with mean 0. The variance of u is calculated from the correlation matrix in each simulation to control the True R^2, and correlation matrix is simulated as $corr(X_j, X_k) = \rho^{|j-k|}$ to obtain different interactions depending on the choice of ρ. Simulations were repeated 500 times for each ρ.

TABLE 2. Simulation settings

Setting		Observations
Correlation structure of $(X_1, ..., X_4)$	$corr(X_j, X_k) = \rho^{\vert j-k\vert}$	$-0.9 \leq \rho \leq 0.9$ in steps of 0.1
True coefficient vectors	$\beta_1 = (1, 1, 1, 1)^T$	
	$\beta_2 = (5, 1, 1, 1)^T$	
	$\beta_3 = (2, 3, 2.5, -0.4)^T$	
Sample sizes	$n = 10$	sampled from multivariate normal
	$n = 100$	sampled from multivariate normal
	$n = 1000$	sampled from multivariate normal
True R^2	0.25	
	0.75	
	0.90	

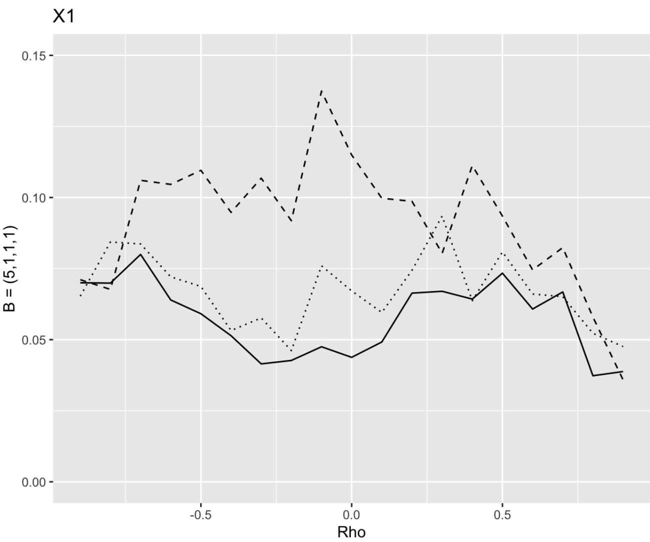
Figure 1 shows the interquartile ranges of R^2_j for a true $R^2 = 0.25$ and across the domain of ρ (lines were drawn for clarity). Overall, the measure provides acceptable levels of variability for applications, with the exception of large absolute β when sample sizes are low (first chart, middle row). We note that the variability tends to decrease slightly with stronger collinearity (because the correlation matrix is designed as $corr(X_j, X_k) = \rho^{|j-k|}$, maximum covariance is obtained for $\rho = 0.9$). Increasing the sample size substantially decreases the variability of our measure. This is to be expected, as part correlations are a function of standardized regression coefficients and their variance decreases with sample size. Accordingly, the lowest variability was found for $n = 1000$. The sample sizes were chosen to differ by orders of magnitude to give insight for different fields of application. Medical and psychological research frequently deal with smaller samples and therefore results should be interpreted with caution, and this holds true for the generality of relative importance measures. Other fields, like finance and economics, typically deal with larger sample sizes which should provide more stable results, depending on the distributional properties of the variables being considered. We note that in its variability, the behavior and size of our measure closely resemble those of LMG (**Lmg1980**) as simulated in **Gromping2006**.

FIGURE 1. Interquartile ranges of R_j^2 for each EV



Each chart shows the interquartile range of 500 simulations of R_j^2 ran for a given covariance matrix as a function of ρ as defined in Table 2. Each row of charts differs in its set of β coefficients as indicated on the left. The simulations were run for $n = 10$ (dashed), $n = 100$ (dotted) and $n = 1000$ (plain).

FIGURE 2. Interquartile ranges of R_1^2 for each level of R^2



Each line represents the interquartile range of the share allocated to EV X_1 according to the level of true R^2 as defined in Table 2: 0.25 (dashed), 0.75 (dotted) and 0.9 (plain). The set of coefficients is as indicated on the left. Sample size was fixed at $n = 100$ (intermediate scenario).

Figure 2 shows the interquartile ranges of R_1^2 (the share of X_1) for different levels of true R^2. Noticeably, the variability decreases with increasing collinearity (ρ) of the same sign as the regression coefficients when sample sizes are lower. This indicates that the measure is more stable when the squared part correlations play a weaker role in determining R^2 (i.e. when individual EV contribution to explanation of variance is relatively lower), therefore leaving a greater amount of shared variance ($svar$) to allocate to each EV. The lowest variability was found for a true $R^2 = 0.90$. Overall, variability levels appear to be acceptable for applications, especially for larger sample sizes.

3.4. Discussion

The application of relative importance measures based on variance decomposition requires an *a priori* analysis of the research problem. For essentially predictive and exploratory problems, the choice should fall on multiple analysis methods that rank importance based on entry-order permutations of EV. For models involving many EV, these methods are computationally intensive. For explanatory problems, provided that variable selection is grounded on a sound theoretical background (i.e. order of EV entry is irrelevant) single analysis methods are computationally cheaper and most provide an intuitive way to decompose the variation of DV.

TABLE 3. Selected regression outputs

Variable	Std. coeff	r_{YX_i}	sr_i	sr_i^2	VIF_i	Weight (v_i)	Share of R^2	Importance rank
Income (X_2)	0.0390	0.9300	0.0104	0.0001	14.0572	0.4666	0.4019	1
Appliances (X_4)	0.3384	0.8861	0.1148	0.0132	8.6947	0.2886	0.2617	2
Air conditioning (X_3)	0.5462	0.9273	0.2935	0.0861	3.4642	0.1150	0.1851	3
Size (X_1)	0.1446	0.8575	0.0726	0.0053	3.9113	0.1298	0.1171	4

In this section, we apply the measure derived in Section 3.2.3 to a practical case, and discuss the results and their limitations. The data set is adapted from **montgomery1981** and has been collected by an utility company during an investigation of factors that influence peak demand Y (kWh) for electricity by residential customers. The explanatory variables are customer's residence size X_1 ($ft^2/1000$), annual family income X_2 (\$/1000), tons of air-conditioning capacity X_3 and the appliance index X_4 (obtained by summing the kilowatt ratings for all major appliances). We assume that the model is well specified and that selected EV are relevant and applied as explanatory of Y. A preliminary inspection of the correlation matrix (Table 4) reveals very strong correlations between EV pointing to low individual contributions to R^2. This is confirmed in Table 3 by the low values of squared part correlations, sr_j^2. Note that the model R^2 is 0.9657 and that Income (X_2) ($sr_j^2 = 0.0001$) seems irrelevant in practice.

Further inspection of Table 3 reveals that all part correlations are positive (sr_i) and never exceed the zero-order correlations (r_{YX_j}) and therefore we can expect $svar > 0$ (in fact, $svar = 0.9657 - 0.1047 = 0.861$, see Equation 3.10). The VIF weights (v_j) sum to 1 and represent the proportion of *svar* allocated to each EV. Particularly noteworthy is weight υ_2 (0.4666) allocated to Income (X_2). Although it is the least relevant EV from an individual perspective, its strong correlation with the remaining EV ($VIF_2 = 14.0572$) make X_2 the most relevant when *svar* is accounted for, with a share $R_2^2 = 0.4019$. This is what our measure is intended to do, and provides an intuitive way to allocate additive (i.e. $\sum R_j^2 = R^2$) shares even when squared partial correlations are low. A practical conclusion from this analysis is that Income X_2 is the most important EV to explain variation in peak demand (Y), not because of its individual contribution but because it is highly correlated with the remaining EV.

TABLE 4. Correlation matrix

	Y	X_1	X_2	X_3	X_4
Y	1				
X_1	0.8575	1			
X_2	0.9300	0.8576	1		
X_3	0.9273	0.7344	0.8270	1	
X_4	0.8861	0.8253	0.9343	0.7191	1

We use the results from this application to a real dataset to show how assigning shares of *svar* based on VIF weights mitigates the issue of multicollinearity affecting measures of relative importance, and may even reverse importance ranks otherwise based solely on squared part correlations when VIF weights differ substantially. Notwithstanding, we note that Income (X_2) has a VIF of 14.0572, implying that it is strongly correlated to the remaining three EV. A typical heuristic employed in research is to exclude from estimation any EV with $VIF \geq 10$. Therefore, we re-estimate the model without X_2 (because $14.0572 > 10$) and analyze the results, shown in Table 5.

TABLE 5. Selected regression outputs (without Income (X_2))

Variable	**Std. coeff**	r_{YX_i}	sr_i	sr_i^2	VIF_i	**Weight** (υ_i)	**Share of** R^2	**Importance rank**
Air conditioning (X_3)	0.5571	0.9273	0.3613	0.1305	2.3774	0.2525	0.3302	1
Appliances (X_4)	0.3622	0.8861	0.1954	0.0382	3.4354	0.3649	0.3268	2
Size (X_1)	0.1494	0.8575	0.0787	0.0061	3.6009	0.3825	0.3087	3

We start by noting that *svar* is positive ($0.9656 - 0.1748 = 0.7908$) and that the re-estimated model assigns $VIF_i < 10$ to all EV, pointing to absence of strong collinearity. Although the relative position (in terms of 'Importance rank') of variables Appliances (X_4) and Size (X_1) is maintained, Air conditioning (X_3) is the most important variable to explain Peak demand (Y) when the redundant EV Income (X_2) is removed. This is because X_3 increases the total R^2 the most when added to the model ($sr_3^2 = 0.1305$), despite being assigned the lowest VIF weight ($\upsilon_3 = 0.2525$). We also note that, although from a strictly numerical perspective it is possible to assign importance ranks ($R_3^2 > R_4^2 > R_1^2$), the assigned shares of total R^2 are similar across all EV, pointing to close to equal importance when collinearity is taken into account. If we were to base the importance rank solely on the squared part correlations, the relative positions of EV would be maintained although their contributions would differ substantially ($sr_3^2 \gg sr_4^2 > sr_1^2$), corroborating the relevance of accounting for collinearity when evaluating relative importance of EV.

3.5. Conclusion

We have discussed the problem of relative importance of regressors and the challenge of partitioning dependent variable variation in the presence of collinearity. As the issue does not have an unique solution, several different methods have been devised with its own pros and cons, and ours has its own limitations, as it must be applied to regression models of an explanatory nature (rather than exploratory or predictive), i.e. inclusion of EV must be solidly justified rather than simply tested for possible satisfactory correlations. In such cases where a predictive analysis is intended, multiple analysis methods such as LMG are preferred. Especially in situations involving many regressors, our measure is advantageous in that it provides an additive variable ranking while requiring low computational power.

A wide set of relative importance metrics is available to researchers, and many differ in their underlying interpretation of variable importance. To draw meaningful conclusions from data, a careful analysis of the correlation matrices must be performed, as no measure should be applied blindly as a general solution . Rather, comparison of results from different methods is advised, as it provides a greater insight on the relationships between variables. Accordingly, our measure is a contribution to the researcher tool kit of relative importance, meant to account for collinearity, be intuitive and easy to compute. For future research, we suggest (1) the investigation of *svar* in terms of its behavior and interpretability as well as its relationships with correlation-type coefficients and (2) the simulation of the established relative importance measures that rely on regression coefficients in the presence of heteroskedasticity.

3.6. Appendix

Code

```
#ASSIGNING R^2 SHARES TO 4 REGRESSORS
#Variables Y and X_j must be loaded from the dataset and the number expanded if needed
#This is a basic code snippet for illustration. NOT RUN.

model <- lm(Y ~ X1 + X2 + X3 + X4) #Estimate model
rsq <- summary(model)$r.squared #Store model R^2

#Calculating squared part correlations for each coefficient (there are faster ways to compute part correlation)
a <- cor(lm(scale(X1) ~ scale(X2) + scale(X3) + scale(X4), data)$residuals, scale(Y), method="pearson")^2
b <- cor(lm(scale(X2) ~ scale(X1) + scale(X3) + scale(X4),data)$residuals, scale(Y), method="pearson")^2
c <- cor(lm(scale(X3) ~ scale(X1) + scale(X2) + scale(X4),data)$residuals, scale(Y), method="pearson")^2
d <- cor(lm(scale(X4) ~ scale(X1) + scale(X2) + scale(X3),data)$residuals, scale(Y), method="pearson")^2

#Summing squared part correlations
sumsemipart <- a + b + c + d

#Computing VIF and VIF weights
vifs <- VIF(model)
vifweights <- vifs/sum(VIF(model))

#Calculating 'svar' (see Equation 10)
svar <- rsq - sumsemipart

# Populate vectors of individual and svar contributions
indiv <- cbind(a,b,c,d)
shared <- vifweights * c(svar)

# Computing R^2 shares (see Equation 14)
totalshare <- indiv+shared #This is a vector with assigned shares to each regressor

# Print results
vifs
rsq
sum(indiv + shared) #Sum of shares = rsq
totalshare #R^2 shares
rowRanks((indiv+shared)) #Ranking shares by size
```

CHAPTER 4

Contributions to the diagnosis of Skewness and Kurtosis

ABSTRACT

This paper is motivated by the widespread interest in higher-order moments for financial risk management, namely skewness and kurtosis. Firstly, we evaluate the behavior of a set of estimators of skewness and kurtosis employing robust central tendency measures under simulation and conclude about their superiority against the standard alternatives available. Secondly, we derive the asymptotic sampling distributions of skewness and kurtosis coefficients in the case of non-i.i.d. random variables applying the Generalized Method of Moments. We add to the existing literature by simulating a conditionally heteroskedastic process and compute heteroskedasticity-autocorrelation consistent (HAC) standard errors for hypothesis testing. The proposed skewness test significantly outperforms the traditional alternatives, while the kurtosis test corroborates known difficulties of accurately estimating kurtosis even for very large samples. In an illustrative example we analyze the daily returns of four major currency pairs in the period 2010-2020.

4.1. Introduction

This paper focuses on the third and fourth moments of financial time series data. The motivation for examining these moments is based on empirical observations about the asymmetry and kurtosis of financial assets' returns. It is a familiar stylized fact that market returns have negative skewness and severe excess kurtosis, i.e. downturns (negative returns) tend to be more severe and less frequent than upturns (positive returns) and the tails of empirical distributions are heavier when compared to the normal distribution. These stylized facts have been supported by a large collection of empirical studies. Some recent papers on this issue include **Bastianin2020**, **WU2019**, **Clark2018**, **BeraPremaratne2017** and **Jaggia**. The departure of financial assets' returns from the normal distribution is further documented in **Bates1996**, **Hwang1999** and **Harvey2000**.

The relevance of skewness and kurtosis goes beyond the realm of empirical finance. See, for example, **hernandez2014**, **AndersonMattson2012** and **Blanca2013** for applications in statistics, engineering and psychology, respectively. **hernandez2014** propose an approach to derive approximations of arbitrary order to estimate high percentiles of sums of positive independent random variables exhibiting heavy tails (excess kurtosis). They conclude that for higher quantiles, and especially for heavier tails, the quality of the estimate improves as more terms are included in the series, up to a certain order. **Bal1988** recognize that many formalizations of the concepts of skewness and kurtosis may arise, as these can be viewed 'vaguely' as a location- and scale-free movement of probability from the shoulders of a distribution into its center and tails.

The role of higher moments has become increasingly important in the financial literature mainly because the traditional measure of risk, variance (or standard deviation), has failed to fully capture the risk of financial assets. According to **kimwhite04** "true risk" may be a multidimensional concept and other measures of distributional shape, such as higher moments, can be used for a better description of risk.

In face of the increasing interest in the skewness and kurtosis of time series data, three papers deserve particular emphasis: **JoanesGill1998** performed a comparison between the measures adopted by well-known statistical computing packages, focusing on bias and mean-squared error for normal samples, and presenting comparisons based on simulation results for non-normal samples. **kimwhite04** provide a survey of robust measures of skewness and kurtosis and carry out extensive Monte Carlo simulations that compare the conventional measures (that is, the standardized third and fourth moments or some variants of these) with selected and robust alternatives. An empirical application of the robust measures to daily S&P500 index returns indicates that the stylized facts related with skewness and kurtosis are also found when the robust measures are used. **BaiSerena2005** discuss the sampling distributions of the coefficients of skewness and kurtosis and propose a joint test of normality for time series observations. They show that when data are serially correlated, consistent estimates of three-dimensional long-run covariance matrices are needed for testing symmetry or kurtosis. Through Monte Carlo simulations they show that the test statistics for symmetry and normality have good finite-sample size and power. However, size distortions render testing for kurtosis almost meaningless except for distributions with thin tails, such as the normal distribution. Skewness and kurtosis are combined to construct a useful test of normality provided that the limiting variance accounts for the serial correlation in the data. All tests are computed and analyzed for 21 macroeconomic time series.

The properties of conventional skewness and kurtosis measures are also analyzed by **Bonato2011** and they are compared with the robust alternatives when fat-tailed distributions which do not possess variance or third or fourth moment are considered. This paper produces two outstanding conclusions. First, for symmetric fat-tailed distribution, skewness is far from being a valid indicator of the presence of asymmetry. Second, Monte Carlo simulation and empirical applications to the series of daily returns on a large cap US stock show why alternative measures are a better tool to describe the skewness and kurtosis of financial returns distribution.

More recently, **BeraPremaratne2017** introduce adjustments on the standard test for skewness to discriminate between symmetric and asymmetric distributions in the presence of excess kurtosis. The main reason for the failure of the standard test is that the expression of its variance is derived under the assumption of no excess kurtosis. They also suggest an adjusted test for kurtosis in the presence of asymmetry. Both tests are applied to simulated and real data. The finite sample properties of the adjusted tests are far superior when compared to those of their unadjusted counterparts.

In most statistical tests involving the sampling coefficients of skewness and kurtosis it is assumed that random variables are independent and identically distributed. Noteworthy exceptions are **BaiSerena2005** and **Lobato2004**, where serial correlation is taken into account. When analyzing financial time series, the dependence of high frequency data (e.g. daily returns) raises the issues of autocorrelation and/or conditional heteroscedasticity, which render the traditional inference methods inappropriate. Other exceptions are **Boutahar2009**, where tests are developed for long-term memory processes.

The contribution of this investigation is as follows: firstly, we examine the behavior of an alternative set of skewness and kurtosis estimators in terms of Mean Squared Error (MSE), and convergence in the absence/presence of a single extreme observation. The estimators are applied to non-normal i.i.d. data. Secondly, due to the characteristics of financial data, we propose heteroskedasticity-autocorrelation consistent (HAC) standard errors for the sample coefficients of skewness and kurtosis, and construct test statistics to detect skewness and kurtosis as financial time series typically exhibit serial correlation and conditional heteroskedasticity. The proposed skewness test significantly outperforms the traditional alternatives, while the kurtosis test corroborates known difficulties of accurately estimating kurtosis even for very large samples. In an illustrative example we analyze the daily returns of four major currency pairs in the period 2010-2020.

The layout of the paper is as follows: in Section 2, we define a set of alternative estimators for skewness and kurtosis based on robust central tendency measures and evaluate their behavior and accuracy using simulation. In Section 3, we derive the asymptotic sampling distribution of the skewness and kurtosis estimators and corresponding statistical tests using the Generalized Method of Moments (GMM). We apply these results in a Monte Carlo setting, compare the proposed estimators to their traditional alternatives and perform an empirical application. Section 4 contains concluding remarks.

4.2. Alternative estimators for non-normal i.i.d. data

Let $\{X_t\}_{t=1}^{T}$ denotes a series with mean μ and variance σ^2. The coefficients of skewness (τ) and kurtosis (κ) are defined as:

$$\tau = \frac{\mu_3}{(\sigma^2)^{3/2}} \quad \text{and} \quad \kappa = \frac{\mu_4}{(\sigma^2)^2}, \tag{4.1}$$

where $\mu_3 = E\left(X - \mu\right)^3$ and $\mu_4 = E\left(X - \mu\right)^4$ are the third and fourth moments of the X_t distribution, respectively. The standard estimators for these parameters are:

$$\hat{\tau} = \frac{\hat{\mu}_3}{(\hat{\sigma}^2)^{3/2}} \quad \text{and} \quad \hat{\kappa} = \frac{\hat{\mu}_4}{(\hat{\sigma}^2)^2}, \tag{4.2}$$

where $\hat{\mu}^3 = \frac{1}{T}\sum_{t=1}^{T}\left(X_t - \hat{\mu}\right)^3$, $\hat{\mu}^4 = \frac{1}{T}\sum_{t=1}^{T}\left(X_t - \hat{\mu}\right)^4$, $\hat{\mu} = \frac{1}{T}\sum_{t=1}^{T} X_t$ and
$\hat{\sigma}^2 = \frac{1}{T}\sum_{t=1}^{T}\left(X_t - \hat{\mu}\right)^2$.

In this section, we compare the performance of estimators $\hat{\tau}$ and $\hat{\kappa}$ (for skewness and kurtosis, respectively) with an alternative set of estimators using simulation of i.i.d. random variables. It has long been established that the estimation of higher order moments may lead to substantial biases, not only because of the moment order (3 and 4 for skewness and kurtosis) but also because traditional estimators rely on the sample mean as the central tendency measure of reference, namely due to the unbiasedness of the standard estimators under normality. This analysis is motivated by the hypothesis that, especially for asymmetric distributions, other central tendency measures, such as the trimmed mean and the median should provide stabler estimates of skewness and kurtosis as they are less sensitive (or not at all) to observations originating from distribution tails.

In addition to $\hat{\tau}$, we define three estimators for skewness ($\hat{\tau}_2$, $\hat{\tau}_3$ and $\hat{\tau}_4$) by replacing the sample mean in $\hat{\tau}$ by the trimmed mean, where the trim-proportion is equal to $1/[2(n-4)^{1/2}]$, the median and an adjusted median as proposed by **DiasCurto2021**, such that:

$$\hat{\tau}_2 = \frac{\frac{1}{T}\sum_{t=1}^{T}(X_t - \hat{\mu}_m)^3}{[\sum_{t=1}^{T}(X_t - \hat{\mu})^2]^{3/2}}, \hat{\tau}_3 = \frac{\frac{1}{T}\sum_{t=1}^{T}(X_t - \hat{\mu}_{med})^3}{[\sum_{t=1}^{T}(X_t - \hat{\mu})^2]^{3/2}}, \hat{\tau}_4 = \frac{\frac{1}{T}\sum_{t=1}^{T}(X_t - \hat{\mu}_{amed})^3}{[\sum_{t=1}^{T}(X_t - \hat{\mu})^2]^{3/2}} \tag{4.3}$$

where $\hat{\mu}_m$ is the trimmed mean, $\hat{\mu}_{med}$ is the median and $\hat{\mu}_{amed}$ is the adjusted median. The adjusted median generates an intermediate value between the median and the arithmetic mean, assigning higher weights to extreme observations, when compared to the median, harmonic and geometric averages, and lower weights when compared to the arithmetic mean. Therefore, outlying observations are still relevant to central tendency estimation, but not as much as in the arithmetic mean. It is calculated as:

$$\hat{\mu}_{amed} = \bar{X}_M + \frac{S_R - S_L}{S_R + S_L}|\bar{X}_A - \bar{X}_M| \tag{4.4}$$

where $S_L = \sum_{x_i<\bar{X}_M}^{n_1}(\bar{X}_M - x_i)$ and $S_R = \sum_{x_i>\bar{X}_M}^{n_2}(x_i - \bar{X}_M)$ are the sum of deviations of x_i to the left and right of the median, respectively, and $\bar{X}_A$ is the arithmetic mean.

Analogously, we define three estimators for kurtosis ($\hat{\kappa}_2$, $\hat{\kappa}_3$ and $\hat{\kappa}_4$), which differ from the estimators in Equation 3 only in their moment order:

$$\hat{\kappa}_2 = \frac{\frac{1}{T}\sum_{t=1}^{T}(X_t - \hat{\mu}_m)^4}{[\sum_{t=1}^{T}(X_t - \hat{\mu})^2]^2}, \hat{\kappa}_3 = \frac{\frac{1}{T}\sum_{t=1}^{T}(X_t - \hat{\mu}_{med})^4}{[\sum_{t=1}^{T}(X_t - \hat{\mu})^2]^2}, \hat{\kappa}_4 = \frac{\frac{1}{T}\sum_{t=1}^{T}(X_t - \hat{\mu}_{amed})^4}{[\sum_{t=1}^{T}(X_t - \hat{\mu})^2]^2} \tag{4.5}$$

In the next subsection, we evaluate the behavior of estimators in Equations 3 and 5 under a simulation setting and compare their accuracy to that of the standard estimators define in Equation 2.

4.2.1. Simulation results

We performed a two-stage simulation study regarding the accuracy of the six alternative estimators of skewness and kurtosis defined above. Namely, we simulate i.i.d. random variables from three asymmetric and three symmetric theoretical distributions [chi-square with 1 degree of freedom: $\chi^2(1)$ (A1), beta: $B(1,10)$ (A2) and gamma: $G(1,6)$ (A3); standard normal: $N(0,1)$ (S1), t-student with 5 degrees of freedom: $t(5)$ (S2), uniform: $U(0,1)$ (S3)]. Since the different estimators differ only in the central tendency measure employed, in the subsequent paragraphs and for ease of interpretation we refer to each estimator by the measure employed therein.

In the first stage, we compute estimates of skewness and kurtosis by applying the alternative estimators as well as their standard counterparts, and calculate Mean Squared Error (MSE) as a loss function to evaluate consistency. The goal is to evaluate our hypothesis that robust central tendency measures yield more accurate estimates, especially for asymmetric distributions and smaller sample sizes. For each sample size, 10000 samples were generated from each distribution and the estimates of skewness and kurtosis compared to the true parameter values. The simulation routines were programmed in R and are available upon request to the authors.

The results are exhibited in Table 1 and confirm our hypothesis. With respect to skewness, for smaller sample sizes (up to $T = 100$), the robust central tendency measures yield lower MSE in the vast majority of cases for asymmetric distributions. For distribution A2 , the arithmetic mean is the preferred measure for sample size $T = 50$ or higher. We also note that while for distributions A1 and A3 either the median or the adjusted median provide the best results for smaller sample sizes, skewness for distribution A2 is more accurately estimated applying the trimmed mean. This is due to the fact that, compared to the chi-square (A1) and gamma (A3), the beta (A2) distribution has a longer left tail. For sample sizes $T \geq 1000$, the arithmetic mean gives lower MSE for all distributions, because increasing sample sizes linearly decrease the weight assigned to each observation, thus reducing the influence of extreme observations. For all symmetric distributions, where negative and positive extreme observations are expected to cancel out, the traditional estimator (i.e. with the arithmetic mean as the central tendency measure) performs best in terms of MSE.

Results of kurtosis estimation further corroborate our hypothesis that higher order moments are more accurately estimated by replacing the arithmetic mean in the standard estimators with more robust central tendency measures. We note that, for $T = 10$, the median results in lower MSE for all distributions (symmetric and asymmetric) except for S3 (uniform). Regarding distribution S2 (t-student), and in contrast to the estimation of skewness, results are best when applying the median up to $T = 50$. For sample sizes $T \geq 1000$, the arithmetic mean achieves lowest MSE for all distributions considered.

In the second stage of our simulation study, we assess the behavior of each estimator by including an extreme outlier in the samples generated from the same distributions. Figures 1-8 contain boxplots describing the distribution of 1000 skewness and kurtosis estimates for increasing sample sizes in the vertical axes. Each figure contains twelve windows corresponding to the simulated distributions and figures are paired for comparison (no outlier and outlier cases). To ensure some consistency in the introduction of the outlier (which we call m), we generate random samples from each distribution and replace the tenth observation (x_{10}) by a multiple of 50 of the first quartile of each sample, i.e., after simulating,

$$x_{10} = m = 50 * F^{-1}(0.25) \tag{4.6}$$

for all samples obtained.

The behavior of all eight estimators is remarkably similar. Figures 1-8 show the box-type convergence of estimators to the true values of skewness and kurtosis in the absence of outliers, although asymmetric distributions exhibit a substantial number of observations above the upper quartile. Noticeably, the performance of all estimators deteriorate as distributions depart from the standard normal. The one exception is the uniform distribution, where estimators converge consistently and similarly to the normal distribution. Introducing the outlier, on the other hand, produces spectacular results. Although convergence occurs when estimating skewness by employing robust central tendency measures, a substantial bias is still present. For example, applying the median ($\hat{\tau}_3$) to estimate the skewness of the gamma distribution results in a median skewness of around 5 (for a true τ of 2) even for sample sizes as high as $T = 500$ (Figure 3). The same type of behavior occurs for the standard normal distribution, where the median skewness is close to -5 (for a true τ of 0) for $T = 5000$. These results point to reliability issues of these estimators in the presence of outliers, which, depending on their relevance for the desired research application, may or may not be removed (i.e. while outlier removal for estimation of a natural quantity such as blood pressure or heart rate may not invalidate results for the generality of individuals, the same cannot be said when considering financial market risk models where extreme (negative) outliers have equally extreme associated payoffs).

Results are even more remarkable in the estimation of kurtosis. For symmetric distributions, the effect is maximized around $T = 1000$, after which the estimators start to converge, but estimates remain unsatisfactory. For example, the trimmed mean estimator yields median kurtosis estimates between 200 and 300 for the normal and uniform distributions, for true parameter values of 3 and 1.8, respectively. The results indicate that it may not be feasible to assign any meaningful interpretation to large values of kurtosis estimators even when apparently sufficient large samples are considered. While it must be true that if an outlier occurs only once (and its magnitude does not depend on sample size), its influence eventually disappears as T grows to infinity, finite sample behavior is poor if extreme observations are present.

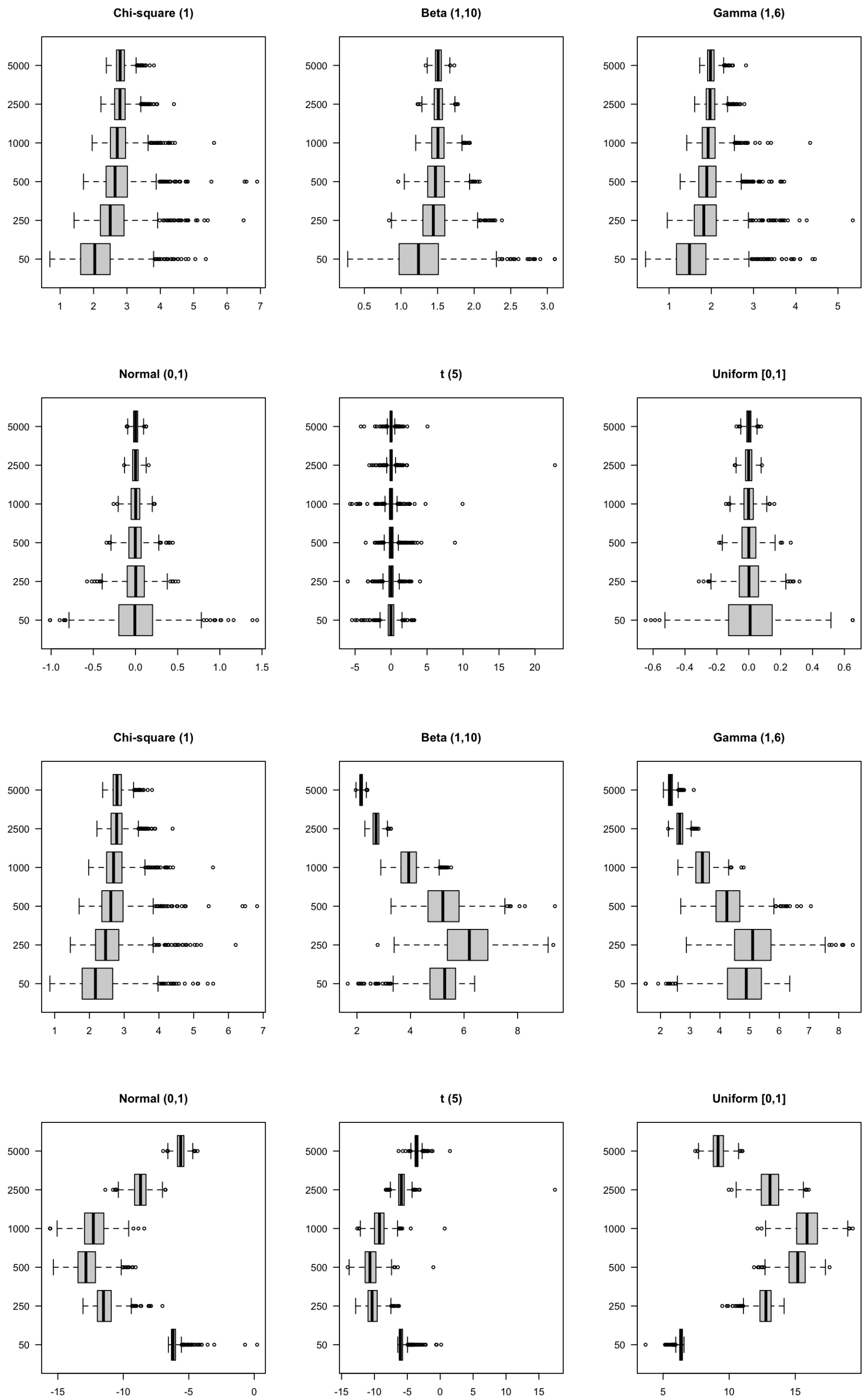
In general, results from our simulation study favor the use of skewness and kurtosis estimators that employ more robust central tendency measures as these provide more accurate estimates in terms of MSE when datasets do not contain extreme observations. Outliers, on the other hand, greatly affect the behavior of all estimators considered as these involve the sample mean and sample standard deviation, known to be more sensitive than e.g. the median to outlying observations. In such cases, the choice of estimators may befall on strictly quantile based measures.

Table 1. Root MSE: Estimators of Skewness and Kurtosis

	Skewness					Kurtosis				
	Mean	Trimmed mean	Median	Adj. Median	Best	Mean	Trimmed mean	Median	Adj. Median	Best
n = 10										
A1	1.813	1.256	1.083	1.508	Median	12.049	10.533	9.848	11.424	Median
A2	0.935	0.829	0.908	1.277	Trim mean	3.550	3.026	2.762	3.185	Median
A3	1.280	1.007	0.996	1.098	Median	6.531	5.688	5.286	6.000	Median
S1	0.493	0.793	0.966	1.014	Mean	1.183	1.140	1.106	1.305	Median
S2	0.661	1.006	1.133	1.172	Mean	6.753	6.408	6.239	6.249	Median
S3	0.402	0.708	0.978	1.068	Mean	0.491	0.712	0.958	1.355	Mean
n = 20										
A1	1.384	1.012	0.711	1.024	Median	10.231	9.039	7.481	9.378	Median
A2	0.699	0.647	0.811	1.280	Trim mean	2.878	2.721	2.541	3.444	Median
A3	0.979	0.797	0.773	0.718	Adj. Median	5.451	4.903	4.182	4.808	Median
S1	0.432	0.577	0.790	0.812	Mean	0.873	0.893	0.887	0.957	Mean
S2	0.764	0.961	1.101	1.120	Mean	5.942	5.802	5.634	5.635	Median
S3	0.308	0.429	0.812	0.850	Mean	0.321	0.379	0.693	0.876	Mean
n = 30										
A1	1.179	0.919	0.705	0.838	Median	9.187	8.274	6.410	8.262	Median
A2	0.594	0.577	0.822	1.281	Trim mean	2.601	2.573	2.615	3.629	Trim mean
A3	0.843	0.730	0.769	0.614	Adj. Median	4.929	4.600	3.903	4.313	Median
S1	0.387	0.481	0.685	0.693	Mean	0.752	0.773	0.777	0.801	Mean
S2	0.770	0.912	1.039	1.044	Mean	5.573	5.501	5.367	5.367	Median
S3	0.253	0.319	0.690	0.714	Mean	0.247	0.266	0.515	0.617	Mean
n = 50										
A1	1.007	0.815	0.829	0.726	Adj, Median	8.270	7.533	6.050	7.397	Median
A2	0.487	0.504	0.832	1.267	Mean	2.345	2.389	2.682	3.711	Mean
A3	0.714	0.651	0.821	0.571	Adj. Median	4.510	4.320	3.967	4.037	Median
S1	0.310	0.379	0.542	0.548	Mean	0.612	0.623	0.624	0.634	Mean
S2	0.804	0.915	1.002	1.003	Mean	5.254	5.242	5.168	5.168	Median
S3	0.196	0.242	0.555	0.565	Mean	0.181	0.189	0.351	0.398	Mean
n = 100										
A1	0.828	0.740	1.012	0.667	Adj, Median	7.459	7.084	6.561	6.866	Median
A2	0.372	0.425	0.859	1.272	Mean	2.004	2.101	2.722	3.779	Mean
A3	0.579	0.576	0.892	0.571	Adj. Median	4.116	4.090	4.247	3.918	Adj. Median
S1	0.236	0.281	0.403	0.403	Mean	0.460	0.466	0.468	0.469	Mean
S2	0.800	0.874	0.924	0.924	Mean	5.374	5.421	5.386	5.383	Mean
S3	0.143	0.170	0.404	0.408	Mean	0.121	0.124	0.204	0.218	Mean
n = 1000										
A1	0.397	0.422	1.212	0.527	Mean	5.073	5.141	7.026	5.257	Mean
A2	0.125	0.172	0.888	1.275	Mean	0.768	0.820	2.350	3.509	Mean
A3	0.249	0.283	0.951	0.542	Mean	2.399	2.458	3.786	2.807	Mean
S1	0.077	0.084	0.130	0.130	Mean	0.152	0.152	0.152	0.152	Mean
S2	0.554	0.569	0.582	0.582	Mean	7.765	7.789	7.793	7.791	Mean
S3	0.045	0.048	0.130	0.130	Mean	0.036	0.036	0.040	0.041	Mean
n = 5000										
A1	0.195	0.220	1.215	0.453	Mean	2.762	2.807	5.902	3.230	Mean
A2	0.058	0.090	0.890	1.272	Mean	0.363	0.395	2.285	3.458	Mean
A3	0.119	0.145	0.951	0.525	Mean	1.237	1.270	3.264	1.975	Mean
S1	0.035	0.036	0.059	0.059	Mean	0.068	0.068	0.068	0.068	Mean
S2	0.411	0.415	0.420	0.420	Mean	10.983	10.990	10.985	10.985	Mean
S3	0.020	0.021	0.059	0.059	Mean	0.016	0.016	0.016	0.016	Mean

Simulation results: root mean squared error values for the standard and alternative estimators of τ (skewness) and κ (kurtosis) for each distribution and sample size n. Column 'Best' designates the estimator yielding the lowest MSE. For ease of interpretation, we refer to the central tendency measure employed in each estimator, i.e. for skewness, 'Mean' corresponds to $\hat{\tau}$, 'Trim mean' to $\hat{\tau}_2$, 'Median' to $\hat{\tau}_3$ and 'Adj. Median' to $\hat{\tau}_4$. The same reasoning applies to estimators of kurtosis.

FIGURE 1. Sampling distributions of $\hat{\tau}$ (box-plots): no outlier (top half) and outlier (lower half) cases



Each set of six windows corresponds to the six distributions ($\chi^2(1)$, $B(1,10)$, $G(1,6)$, $N(0,1)$, $t(5)$ and $U(0,1)$) and each window contains six box-plots for different sample sizes. The numbers on the vertical axes indicate the corresponding sample sizes. Dots beyond the whiskers represent estimates either below the lower or above the upper quartiles. This note applies to all subsequent figures.

FIGURE 2. Sampling distributions of $\hat{\tau}_2$ (box-plots): no outlier (top half) and outlier (lower half) cases

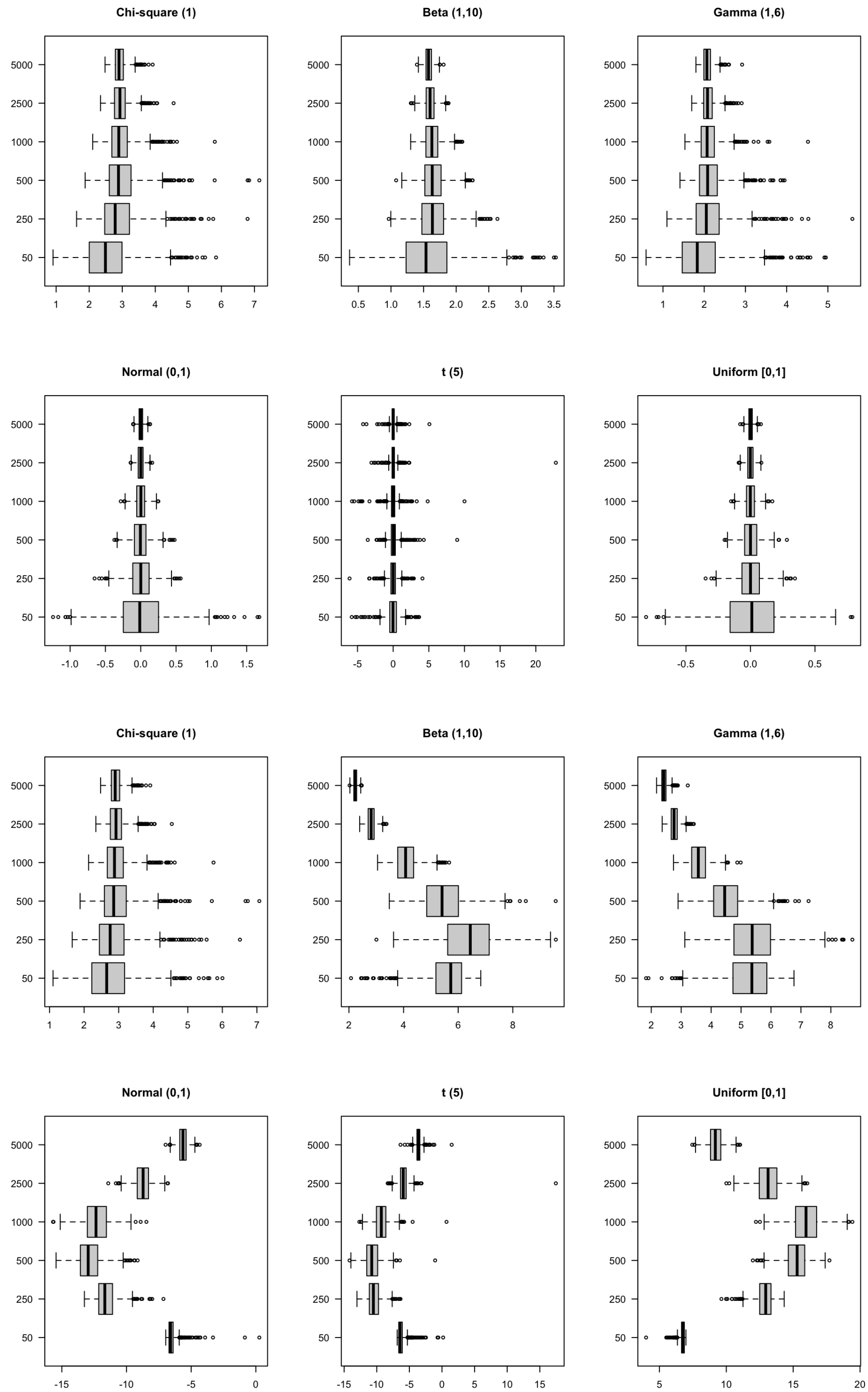


FIGURE 3. Sampling distributions of $\hat{\tau}_3$ (box-plots): no outlier (top half) and outlier (lower half) cases

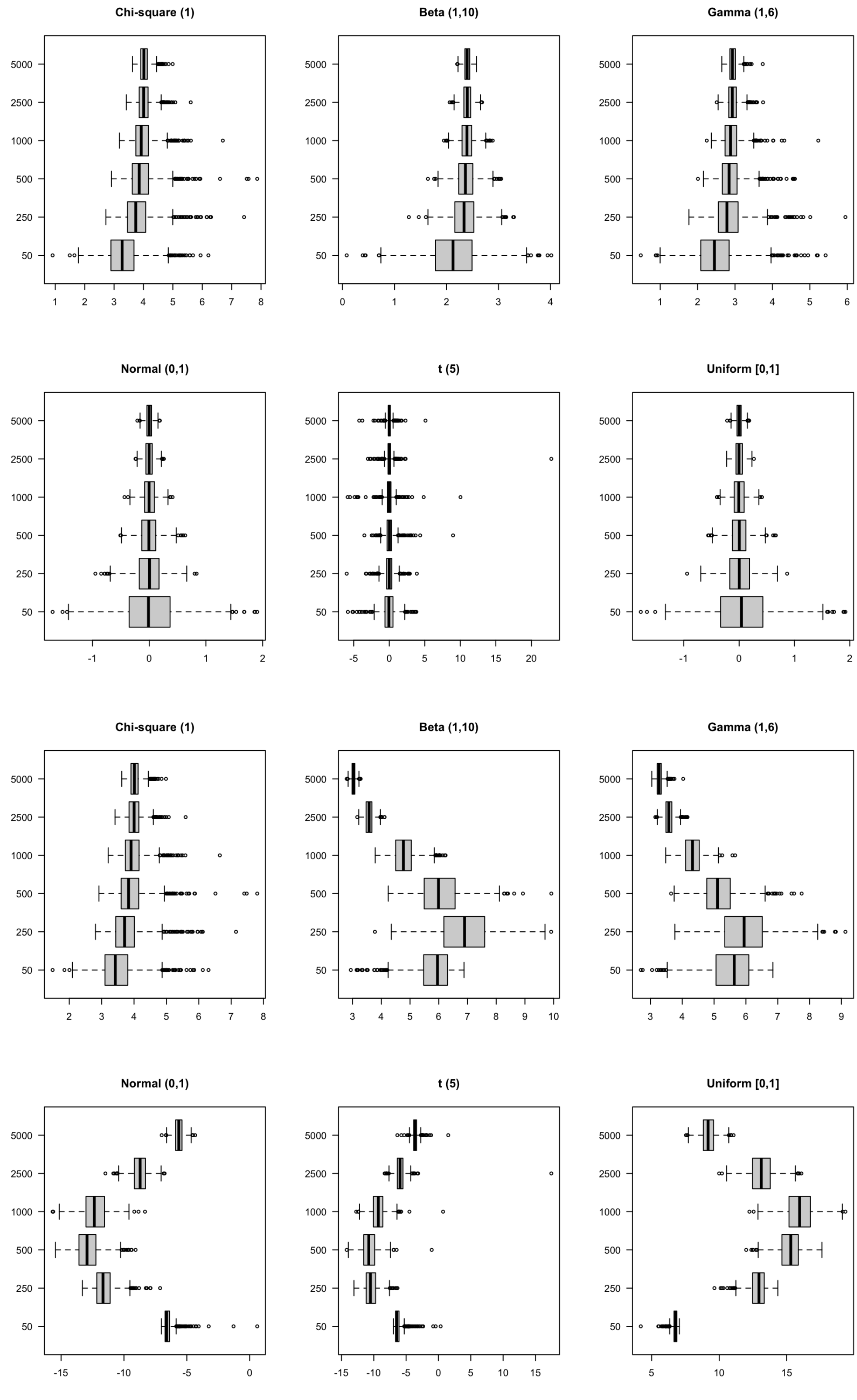


FIGURE 4. Sampling distributions of $\hat{\tau}_4$ (box-plots): no outlier (top half) and outlier (lower half) cases

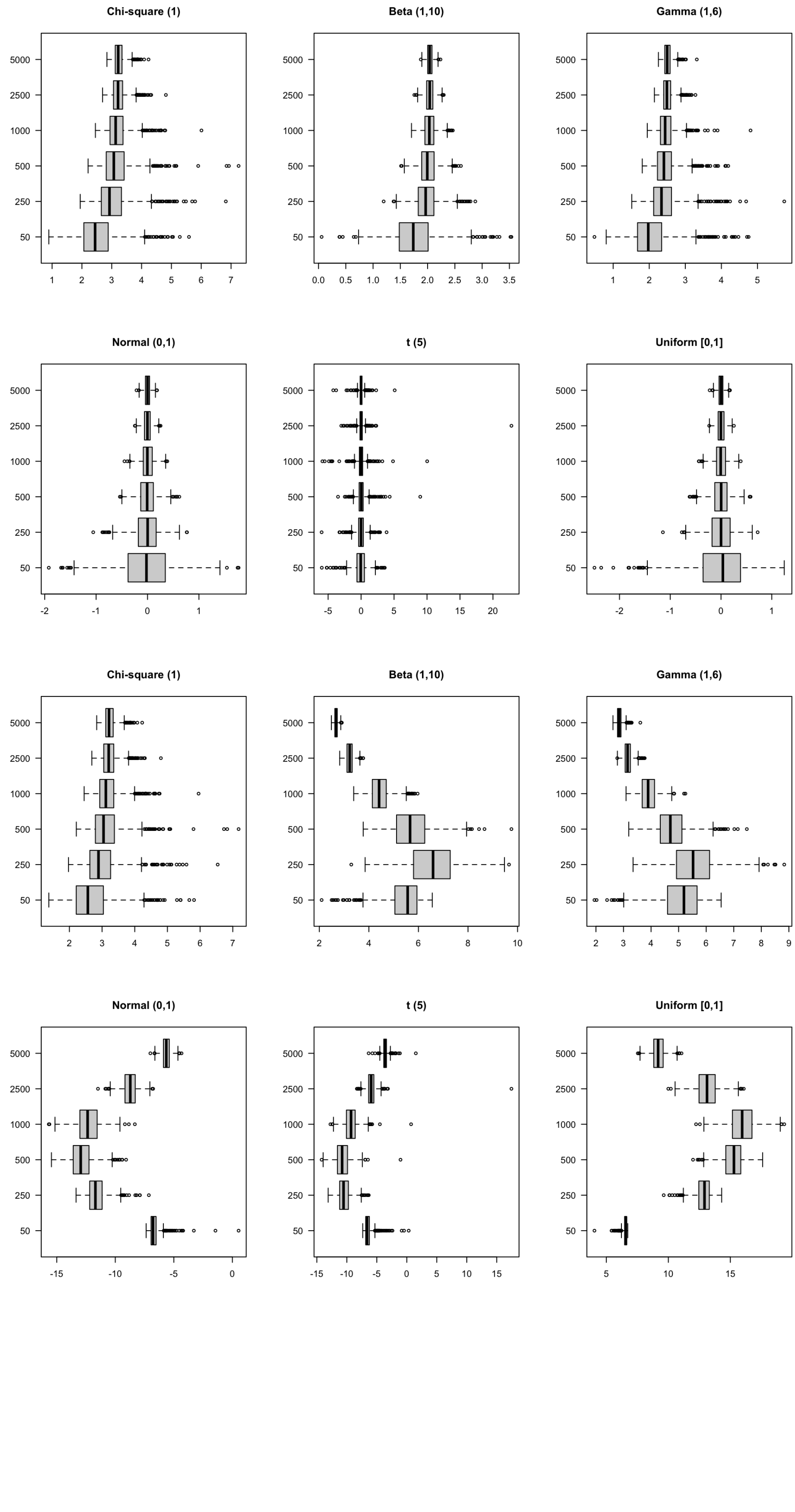


Figure 5. Sampling distributions of $\hat{\kappa}$ (box-plots): no outlier (top half) and outlier (lower half) cases

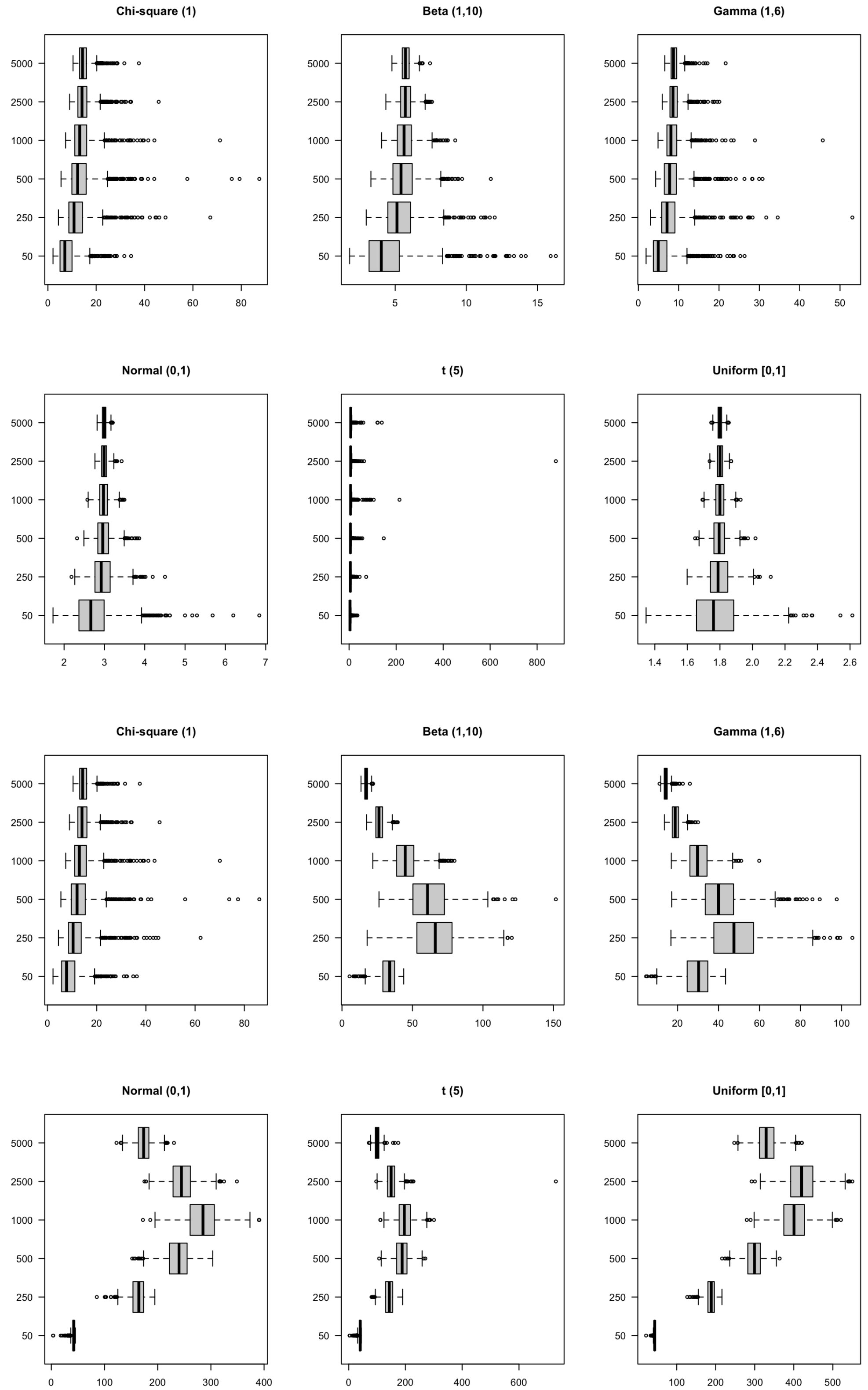


FIGURE 6. Sampling distributions of $\hat{\kappa}_2$ (box-plots): no outlier (top half) and outlier (lower half) cases

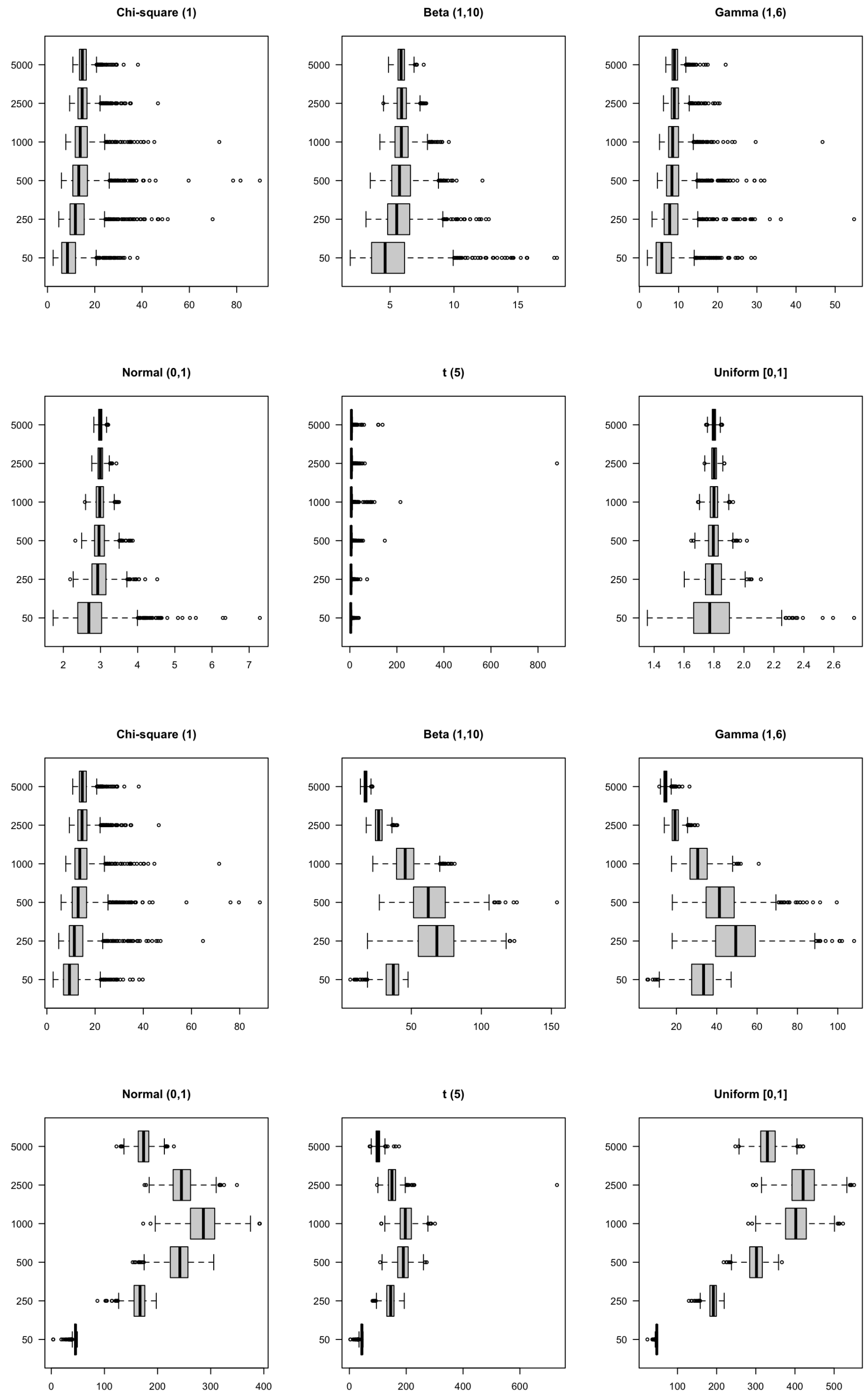


Figure 7. Sampling distributions of $\hat{\kappa}_3$ (box-plots): no outlier (top half) and outlier (lower half) cases

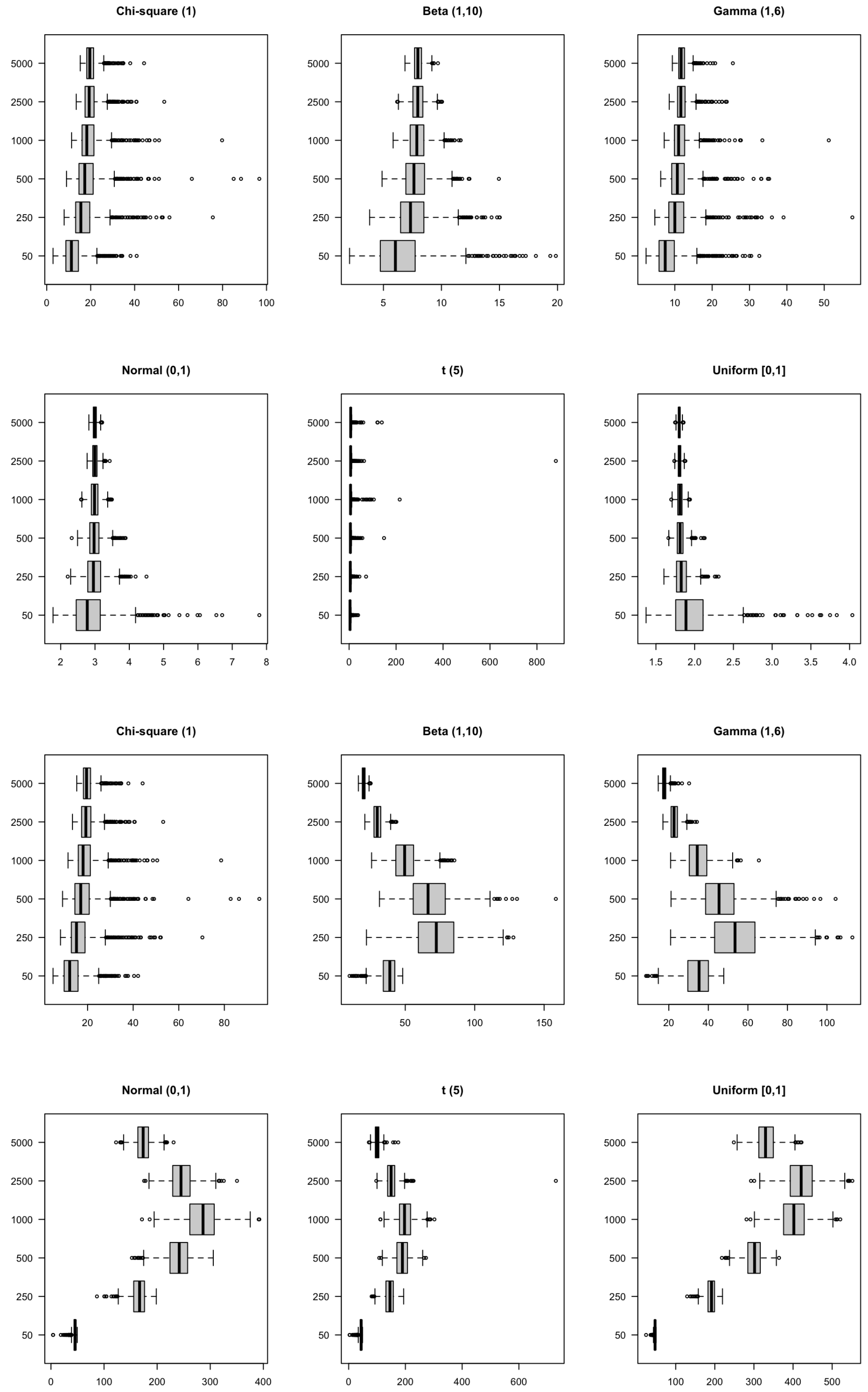
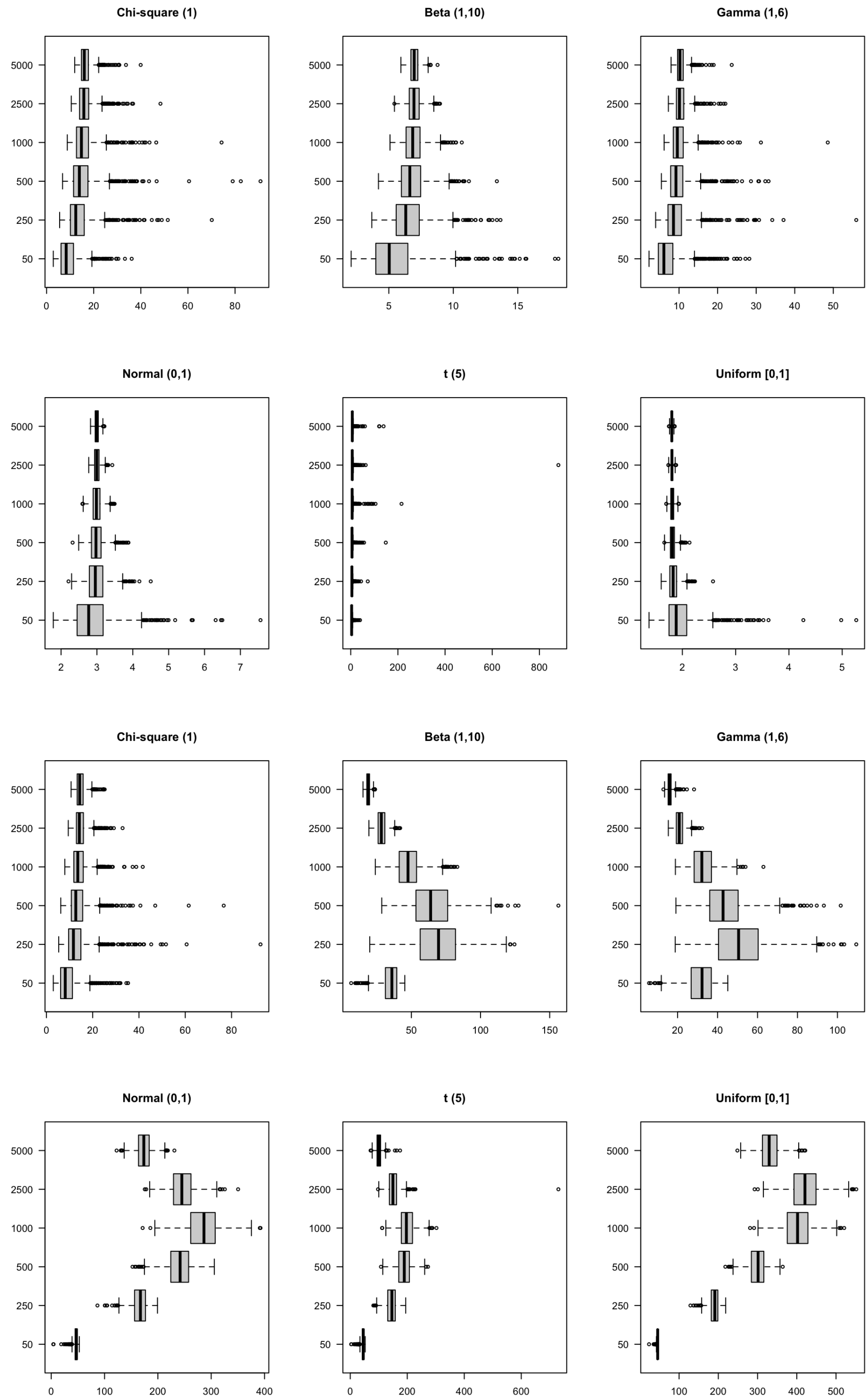


FIGURE 8. Sampling distributions of $\hat{\kappa}_4$ (box-plots): no outlier (top half) and outlier (lower half) cases



4.3. The asymptotic distribution of skewness and kurtosis estimators

In the previous section we evaluated the performance of a set of estimators for skewness and kurtosis for the i.i.d. case. Because in many applications, especially in finance, the i.i.d. assumption is frequently violated, in this section we derive the asymptotic distribution of the standard estimators of skewness and kurtosis and develop confidence intervals using the Generalized Method of Moments (GMM). Additionally, we evaluate the performance of these estimators with respect to the size and power of the corresponding statistical tests by simulating an ARCH(1) process to induce autocorrelation and heteroskedasticity.

If X_t is i.i.d. and normally distributed, then (see, e.g., **kendallstuart69**)

$$\sqrt{T}\hat{\tau} \xrightarrow{d} \sim N(0,6) \quad \text{and} \quad \sqrt{T}\left(\hat{\kappa}-3\right) \xrightarrow{d} N(0,24). \tag{4.7}$$

When the i.i.d. assumption for $(X_1, X_2, \ldots, X_T)$ does not hold, these results are of limited practical value and the asymptotic distribution can be derived by using a "robust" estimator for the coefficients of skewness and kurtosis. If X_t satisfies the assumption of stationarity, then a version of the Central Limit Theorem still applies to most estimators and the corresponding asymptotic distribution can be derived. Following **Lo2002** we apply the GMM to estimate μ, σ^2, μ^3, μ^4 and the results of **Hansen1982** can be used to derive the asymptotic distribution of the coefficients of skewness and kurtosis.

Denote by $\hat{\theta}$ the column vectors $(\hat{\mu} \quad \hat{\sigma}^2 \quad \hat{\mu}^3)'$ and $(\hat{\mu} \quad \hat{\sigma}^2 \quad \hat{\mu}^4)'$ and by θ the corresponding column vector of population values $(\mu \quad \sigma^2 \quad \mu^3)'$ and $(\mu \quad \sigma^2 \quad \mu^4)'$. Hansen shows that:

$$\sqrt{T}\left(\widehat{\theta}-\theta\right) \overset{a}{\sim} N\left(0, V_\theta\right), \quad \text{where} \quad V_\theta \equiv H^{-1}\Sigma\left(H^{-1}\right)', \tag{4.8}$$

$$H \equiv \lim_{T\to\infty} E\left[\frac{1}{T}\sum_{t=1}^{T}\varphi_\theta\left(X_t,\theta\right)\right], \quad \Sigma \equiv \lim_{T\to\infty} E\left[\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\varphi\left(X_t,\theta\right)\varphi\left(X_s,\theta\right)'\right], \tag{4.9}$$

and $\varphi_\theta\left(X_t,\theta\right)$ represents the derivative of $\varphi\left(X_t,\theta\right)$ with respect to θ. Let $\varphi_j\left(X_t,\theta\right)$ denotes the vector function with the following moment conditions:

$$\varphi_\tau\left(X_t,\theta\right) = \begin{bmatrix} X_t-\mu \\ \left(X_t-\mu\right)^2-\sigma^2 \\ \left(X_t-\mu\right)^3-\mu^3 \end{bmatrix} \quad \text{and} \quad \varphi_\kappa\left(X_t,\theta\right) = \begin{bmatrix} X_t-\mu \\ \left(X_t-\mu\right)^2-\sigma^2 \\ \left(X_t-\mu\right)^4-\mu^4 \end{bmatrix}. \tag{4.10}$$

The GMM estimator of θ is given by the solution to:

$$\frac{1}{T}\sum_{i=1}^{T}\varphi_j\left(X_t,\theta\right)=0, \tag{4.11}$$

yielding the standard estimators $\hat{\mu}$, $\hat{\sigma}^2$, $\hat{\mu}^3$ and $\hat{\mu}^4$ defined before.

For the moment conditions in equation (4.10), the matrix H is given by:

$$H_\tau \equiv \lim_{T\to\infty} E\left\{\frac{1}{T}\sum_{t=1}^{T}\begin{bmatrix} -1 & 0 & 0 \\ -2\left(X_t-\mu\right) & -1 & 0 \\ -3\left(X_t-\mu\right)^2 & 0 & -1 \end{bmatrix}\right\} \equiv \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ -3\sigma^2 & 0 & -1 \end{bmatrix}, \tag{4.12}$$

$$H_\kappa \equiv \lim_{T\to\infty} E\left\{\frac{1}{T}\sum_{t=1}^{T}\begin{bmatrix} -1 & 0 & 0 \\ -2\left(X_t-\mu\right) & -1 & 0 \\ -4\left(X_t-\mu\right)^3 & 0 & -1 \end{bmatrix}\right\} \equiv \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ -4\mu^3 & 0 & -1 \end{bmatrix}. \tag{4.13}$$

Therefore, the asymptotic distribution of the coefficients of skewness and kurtosis estimators follows from the delta method:

$$\sqrt{T}\left(\hat{\tau}-\tau\right) \overset{a}{\sim} N\left(0, V_{\theta_\tau}\right), \text{ where } V_{\theta_\tau} = \frac{\partial f\left(\theta_\tau\right)}{\partial \theta_\tau} H_\tau^{-1}\Sigma_\tau \left(H_\tau^{-1}\right)' \frac{\partial f\left(\theta_\tau\right)}{\partial \theta_\tau'}, \tag{4.14}$$

$$\sqrt{T}\left(\hat{\kappa}-\kappa\right) \overset{a}{\sim} N\left(0, V_{\theta_\kappa}\right), \text{ where } V_{\theta_\kappa} = \frac{\partial f\left(\theta_\kappa\right)}{\partial \theta_\kappa} H_\kappa^{-1}\Sigma_\kappa \left(H_\kappa^{-1}\right)' \frac{\partial f\left(\theta_\kappa\right)}{\partial \theta_\kappa'}. \tag{4.15}$$

As $\theta_\tau = \left(\mu \quad \sigma^2 \quad \mu^3\right)'$, $\theta_\kappa = \left(\mu \quad \sigma^2 \quad \mu^4\right)'$, $f\left(\theta_\tau\right) = \tau = \frac{\mu_3}{\left(\sigma^2\right)^{3/2}}$ and $f\left(\theta_\kappa\right) = \kappa = \frac{\mu_4}{\left(\sigma^2\right)^2}$, then

$$\frac{\partial f\left(\theta_\tau\right)}{\partial \theta_\tau} = \begin{bmatrix} -3/\sigma \\ -\frac{3}{2}\mu_3/\sigma^5 \\ 1/\sigma^3 \end{bmatrix} \quad \text{and} \quad \frac{\partial f\left(\theta_\kappa\right)}{\partial \theta_\kappa} = \begin{bmatrix} -4\mu_3/\sigma^4 \\ -2\mu_3/\sigma^6 \\ 1/\sigma^4 \end{bmatrix}. \tag{4.16}$$

In order to estimate the asymptotic variance, an estimator for $\frac{\partial f(\theta_i)}{\partial \theta_i}$ may be obtained by substituting $\hat{\theta}_i$ into equation (4.16) and an heteroskedasticity and autocorrelation[1] consistent (HAC) estimator $\hat{\Sigma}_i$ may be obtained by using the Newey and West's procedure:

$$\hat{\Sigma}_i = \hat{\Omega}_0 + \sum_{j=1}^{m} \omega(j,m)\left(\hat{\Omega}_j + \hat{\Omega}_j'\right), m \ll T, \tag{4.17}$$

$$\hat{\Omega}_j \equiv \frac{1}{T}\sum_{t=j+1}^{T} \varphi\left(X_t, \hat{\theta}\right)\varphi\left(X_t, \hat{\theta}\right)',$$

$$\omega(j,m) = 1 - \frac{j}{m+1},$$

where m is the truncated lag that must satisfy the condition $m/T \to \infty$ as T increases without bound to ensure consistency.

Therefore, for non-i.i.d. random variables, the standard error of the coefficients of skewness and kurtosis estimators can be estimated by $SE\left(\hat{\tau}\right) \overset{a}{=} \sqrt{\frac{V_{\hat{\theta}_\tau}}{T}}$ and $SE\left(\hat{\kappa}\right) \overset{a}{=} \sqrt{\frac{V_{\hat{\theta}_\kappa}}{T}}$. Confidence intervals and the equivalent two-sided t-tests can also be constructed for τ and κ:

$$\hat{\tau} \pm z_{(1-\frac{\alpha}{2})} SE\left(\hat{\tau}\right) \quad \text{and} \quad \hat{\kappa} \pm z_{(1-\frac{\alpha}{2})} SE\left(\hat{\kappa}\right), \tag{4.18}$$

[1]Heteroscedasticity and/or autocorrelation of unknown form are often important specification issues, especially in macroeconomics and financial applications.

where $z_{(1-\frac{\alpha}{2})}$ is the $(1-\frac{\alpha}{2})$ quantile of the standard normal distribution.

4.3.1. Simulation results

We perform a Monte Carlo simulation to compare the performance of the traditional and GMM estimators when attempting to detect the presence of skewness and kurtosis in data sets with varying degrees of persistence. The simulation design is as follows: we generate an ARCH(1) process

$$a_t = e_t \sqrt{\alpha_0 + \rho * a_{t-1}^2} \tag{4.19}$$

where e_t is drawn from five symmetric and five asymmetric distributions, namely from the Skewed Generalized T-Distribution (SGT) family, to assess the size and power of the proposed tests. In choosing to simulate an ARCH(1) process, we add to the results of **BaiSerena2005** by generating a conditionally heteroskedastic series, which we account for by using HAC standard errors. The choice of the SGT family of distributions was made based on the ease of parametrization (SGT distributions assume their location and shape base on five parameters (μ, σ, λ, p, q) where the last two control tail behavior, and the first three control the location, variance and skewness, respectively) and the availability of closed form expressions for population moments. The different parametrizations are provided in the Appendix.

The ARCH(1) series was generated by incorporating the simulated e_t for three different degrees of persistence: $\rho = 0$ (independent), $\rho = 0.5$ and $\rho = 0.8$. It should be noted that estimation of $ARCH(1)$ coefficients does not affect the distribution of tests of skewness and kurtosis, which allows us to apply the traditional and GMM tests to the generated series directly (see **BaiSerena2005** and **White1980**). The asymptotic variance of the GMM estimator is based on the kernel method described in (4.17). The number of simulations for each run was 500, and the significance level set at $\alpha = 0,05$.

4.3.2. Testing skewness

To evaluate the size of the proposed estimator, we obtained samples from five symmetric distributions (S4 is the standard normal distribution, and the remaining four are from the SGT family). We compute the relative frequency of rejection of the true null hypothesis $\hat{\tau} = 0$ for the GMM and standard estimators (Table 2). By standard estimators we mean estimators $\hat{\tau}$ as defined in Section 2 and $\hat{\tau}'$, defined as

$$\hat{\tau}' = \frac{\sqrt{n(n-1)}}{n-2}\hat{\tau} \tag{4.20}$$

We note that both estimators ($\hat{\tau}$ and $\hat{\tau}'$) are unbiased under normality.

Table 2. Size and Power of GMM Test of Skewness

		Sample size											
		50			100			500			1000		
	τ	$\hat{\tau}$	$\hat{\tau}'$	GMM	$\hat{\tau}$	$\hat{\tau}'$	GMM	$\hat{\tau}$	$\hat{\tau}'$	GMM	$\hat{\tau}$	$\hat{\tau}'$	GMM
$\rho = 0$													
Size													
S4	0	0,04	0,04	0,00	0,05	0,05	0,00	0,06	0,06	0,00	0,04	0,04	0,00
S5	0	0,40	0,43	0,14	0,52	0,53	0,20	0,70	0,70	0,13	0,78	0,78	0,10
S6	0	0,44	0,47	0,22	0,58	0,59	0,22	0,77	0,77	0,14	0,83	0,83	0,11
S7	0	0,49	0,51	0,23	0,67	0,68	0,34	0,82	0,83	0,21	0,86	0,86	0,16
S8	0	0,60	0,62	0,33	0,70	0,70	0,39	0,87	0,87	0,25	0,90	0,90	0,20
Power													
A4	-5,46	0,84	0,86	0,43	0,98	0,98	0,86	1,00	1,00	0,95	1,00	1,00	0,96
A5	-3	0,58	0,60	0,26	0,80	0,80	0,57	0,99	0,99	0,86	1,00	1,00	0,89
A6	1,94	0,49	0,51	0,22	0,69	0,71	0,39	0,93	0,93	0,61	0,98	0,98	0,72
A7	4,14	0,64	0,66	0,30	0,86	0,87	0,57	0,99	0,99	0,81	1,00	1,00	0,86
A8	18,8	0,88	0,89	0,53	0,98	0,98	0,84	1,00	1,00	0,91	1,00	1,00	0,92
$\rho = 0.5$													
Size													
S4	0	0,14	0,15	0,00	0,23	0,24	0,00	0,41	0,41	0,00	0,50	0,50	0,00
S5	0	0,43	0,46	0,06	0,54	0,55	0,06	0,82	0,82	0,01	0,84	0,84	0,00
S6	0	0,46	0,48	0,10	0,66	0,67	0,10	0,81	0,81	0,01	0,86	0,86	0,01
S7	0	0,55	0,58	0,17	0,67	0,68	0,15	0,85	0,85	0,04	0,90	0,90	0,01
S8	0	0,57	0,59	0,22	0,75	0,76	0,26	0,88	0,88	0,05	0,94	0,94	0,03
Power													
A4	-5,46	0,78	0,80	0,22	0,93	0,94	0,44	1,00	1,00	0,48	1,00	1,00	0,45
A5	-3	0,56	0,58	0,09	0,80	0,81	0,20	0,97	0,97	0,24	0,98	0,98	0,27
A6	1,94	0,51	0,54	0,08	0,71	0,72	0,15	0,93	0,93	0,12	0,95	0,95	0,10
A7	4,14	0,59	0,62	0,17	0,82	0,83	0,32	0,96	0,96	0,27	0,99	0,99	0,26
A8	18,8	0,88	0,90	0,40	0,97	0,98	0,63	1,00	1,00	0,55	1,00	1,00	0,54
$\rho = 0.8$													
Size													
S4	0	0,33	0,36	0,00	0,44	0,46	0,00	0,76	0,76	0,00	0,81	0,81	0,00
S5	0	0,45	0,46	0,03	0,62	0,63	0,04	0,84	0,84	0,00	0,89	0,89	0,00
S6	0	0,53	0,56	0,08	0,65	0,66	0,03	0,88	0,88	0,00	0,90	0,90	0,00
S7	0	0,51	0,52	0,09	0,68	0,68	0,09	0,89	0,89	0,00	0,93	0,93	0,00
S8	0	0,61	0,63	0,21	0,75	0,76	0,19	0,92	0,92	0,03	0,94	0,94	0,01
Power													
A4	-5,46	0,75	0,78	0,16	0,90	0,91	0,23	0,99	0,99	0,17	0,99	0,99	0,08
A5	-3	0,56	0,60	0,08	0,75	0,75	0,09	0,94	0,94	0,04	0,97	0,97	0,02
A6	1,94	0,52	0,54	0,05	0,68	0,69	0,04	0,92	0,92	0,01	0,97	0,97	0,00
A7	4,14	0,62	0,64	0,09	0,78	0,79	0,15	0,95	0,95	0,07	0,98	0,98	0,05
A8	18,8	0,81	0,83	0,28	0,93	0,94	0,42	0,99	0,99	0,26	1,00	1,00	0,22

The GMM estimator shows good size, generally improving with the level of persistence. As an example, note that for distribution $S4$ and $n = 100$, size decreases from 0.22 to 0.10 and 0.03 for the three increasing levels of persistence. The effect of increasing sample size is noteworthy, especially in the cases where autocorrelation is present: for the sample sizes considered and in relative terms, the same distribution ($S4$) showed a greater decrease of size for $\rho = 0.5$ (0.1 to 0.01) than for $\rho = 0$ (0.22 to 0.11).

The GMM estimator is clearly preferable in terms of size, which validates our application of HAC estimators for standard errors. This is mainly due to the fact that both $\hat{\tau}$ and $\hat{\tau}'$ are unbiased for normal samples but not so otherwise. Additionally, we note that even when sampling from a normal population, the GMM estimator outperforms the traditional estimators across all levels of persistence: the fitness of these estimators to the distribution is outweighed by the presence of conditional heteroskedasticity.

To evaluate test power, we compute the frequency of rejection of a false null hypothesis ($\tau = 0$) based on five asymmetric distributions from the SGT family (three platykurtic and two leptokurtic), considering the true value of skewness of the selected distributions. In spite of performing increasingly better for lower levels of persistence and larger sample sizes, the GMM estimator is outperformed by the standard alternatives, even though its power increases as sample size grows. As an example, note that for distribution $A5$, $\rho = 0.5$ and $n = 500$, GMM showed a power of 0.24, against 0.97 from the competing estimators. As expected, power increases for both alternatives as (absolute) values of skewness increase.

4.3.3. Testing kurtosis

To evaluate the performance of our GMM estimator for kurtosis, we chose five symmetric and five asymmetric distributions with varying degrees of kurtosis. All distributions are from the SGT family, except for S4 (standard normal for size and t-student with 5 degrees of freedom for power). It should be noted that, for the SGT family, the product of the tail behavior parameters must be greater than or equal to the order of the moment being estimated, that is $pq \geq n$ for $E[(X - \mu)^n]$, where p and q control left and right tail behavior, respectively. Hence, parameters were selected to ensure the existence of the fourth moment and provide different degrees of kurtosis (true values of κ are provided in Table 3).

To evaluate size, we test the simulated sample kurtosis against the true alternative (the kurtosis of the simulated ARCH process is calculated according to **Tsay2010**), whereas for power we test the frequency of rejection of $\kappa = 3$, the kurtosis of the normal distribution. We note that in the set of traditional estimators for kurtosis (those commonly applied in software packages), only the Fisher-Pearson standardized coefficient is unbiased under normality:

$$\hat{\kappa}' = \frac{(n-1)}{((n-2)(n-3))}(n+1)\frac{m_4}{m_2^2} \tag{4.21}$$

For this reason, we exclude other standard estimators from the present comparison. Because issues of reliability when estimating kurtosis were anticipated, reported sample sizes were set to much larger values than those for skewness.

Test size is generally poor for both the traditional and the GMM estimators, and their performance is similar. In most cases, the traditional estimator rejected the true null hypothesis in all simulations, even for the i.i.d. case ($\rho = 0$). The one exception is, as expected, when samples are obtained from the normal distribution, where size is low and the GMM estimator is generally outperformed, although sizes converge as the sample size increases. The poor size of both estimators implies that kurtosis cannot be reliably estimated from serially correlated, heteroskedastic, non-normal data, even with sample sizes as large as 5000. The standard estimator $\hat{\kappa}$ is severely biased (generally downwards) and although its values are usually reported by statistical packages its practical value is limited without additional checks of the data distribution.

The GMM estimator exhibits high power. Both tests show high power to reject $\kappa = 3$, and in most cases the null hypothesis was rejected by the traditional estimators in all simulations. This was expected because the standard error substantially decreases with increasing sample size. It is worth noting that power increases with the level of persistence. As an example, for distribution $A6$ and $n = 1000$, power increases from 0.90 ($\rho = 0$) to 0.99 ($\rho = 0.8$), which points to the benefit of using HAC estimators to compute the standard error of $\hat{\kappa}$. Despite the high power exhibited by both estimators, the test results are based on severely biased estimates from $\hat{\kappa}$. Interpretation of these results must take the difficulties of accurately estimating κ into account.

Table 3. Size and Power of GMM Test of Kurtosis

		Sample size															
		100				500				1000				5000			
	κ	Size		Power (vs 3)		Size		Power (vs 3)		Size		Power (vs 3)		Size		Power (vs 3)	
$\rho = 0$		$\hat{\kappa}'$	GMM	$\hat{\kappa}'$	GMM	$\hat{\kappa}'$	GMM	$\hat{\kappa}'$	GMM	$\hat{\kappa}'$	GMM	$\hat{\kappa}'$	GMM	$\hat{\kappa}'$	GMM	$\hat{\kappa}'$	GMM
S4	3	1,00	0,13	0,72	0,53	0,04	0,09	1,00	0,83	0,04	0,08	1,00	0,91	0,04	0,05	1,00	0,90
S5	35,67	1,00	0,99	0,85	0,74	1,00	0,96	1,00	0,93	1,00	0,94	1,00	0,91	1,00	0,93	1,00	0,88
S6	60,84	1,00	0,99	0,88	0,81	1,00	0,98	1,00	0,95	1,00	0,98	1,00	0,90	1,00	0,97	1,00	0,90
S7	56,61	1,00	1,00	0,93	0,86	1,00	0,99	1,00	0,96	1,00	0,99	1,00	0,92	1,00	0,96	1,00	0,86
S8	39,51	1,00	0,97	0,92	0,87	1,00	0,91	1,00	0,97	1,00	0,84	1,00	0,89	1,00	0,61	1,00	0,89
A4	74,77	1,00	1,00	0,90	0,62	1,00	1,00	1,00	0,96	1,00	0,99	1,00	0,92	1,00	0,95	1,00	0,85
A5	53,38	1,00	1,00	0,90	0,69	1,00	0,99	1,00	0,94	1,00	0,97	1,00	0,90	1,00	0,92	1,00	0,89
A6	44,20	1,00	0,99	0,85	0,71	1,00	0,97	1,00	0,92	1,00	0,97	1,00	0,90	1,00	0,93	1,00	0,87
A7	92,52	1,00	1,00	0,90	0,70	1,00	1,00	1,00	0,95	1,00	1,00	1,00	0,92	1,00	0,96	1,00	0,84
A8	48,87	1,00	1,00	0,95	0,71	1,00	1,00	1,00	0,99	1,00	1,00	1,00	0,97	1,00	0,99	1,00	0,88
$\rho = 0.5$																	
S4	3	1,00	1,00	0,89	0,80	1,00	1,00	1,00	0,98	1,00	1,00	1,00	0,98	1,00	0,99	1,00	0,94
S5	35,67	1,00	1,00	0,89	0,82	1,00	1,00	1,00	0,98	1,00	1,00	1,00	0,95	1,00	0,91	1,00	0,89
S6	60,84	1,00	1,00	0,93	0,84	1,00	1,00	1,00	0,98	1,00	1,00	1,00	0,95	1,00	0,93	1,00	0,90
S7	56,61	1,00	0,99	0,95	0,87	1,00	1,00	1,00	0,99	1,00	1,00	1,00	0,98	1,00	0,92	1,00	0,90
S8	39,51	1,00	0,99	0,96	0,88	1,00	0,99	1,00	0,99	1,00	1,00	1,00	0,98	1,00	0,91	1,00	0,88
A4	74,77	1,00	1,00	0,93	0,71	1,00	0,99	1,00	0,98	1,00	1,00	1,00	0,97	1,00	0,93	1,00	0,89
A5	53,38	1,00	1,00	0,91	0,74	1,00	0,99	1,00	0,98	1,00	1,00	1,00	0,95	1,00	0,90	1,00	0,89
A6	44,20	1,00	1,00	0,88	0,76	1,00	1,00	1,00	0,98	1,00	1,00	1,00	0,97	1,00	0,91	1,00	0,88
A7	92,52	1,00	1,00	0,91	0,77	1,00	1,00	1,00	0,98	1,00	1,00	1,00	0,97	1,00	0,88	1,00	0,89
A8	48,87	1,00	0,98	0,96	0,76	1,00	1,00	1,00	0,99	1,00	0,99	1,00	0,99	1,00	0,91	1,00	0,92
Rho 0.8																	
S4	3	1,00	1,00	0,95	0,83	1,00	1,00	1,00	0,96	1,00	1,00	1,00	0,96	1,00	0,93	1,00	0,97
S5	35,67	1,00	0,99	0,95	0,85	1,00	1,00	1,00	0,99	1,00	1,00	1,00	0,98	1,00	0,92	1,00	0,95
S6	60,84	1,00	0,98	0,93	0,84	1,00	1,00	1,00	0,99	1,00	1,00	1,00	0,99	1,00	0,94	1,00	0,96
S7	56,61	1,00	0,99	0,95	0,87	1,00	0,99	1,00	0,99	1,00	0,99	1,00	0,98	1,00	0,95	1,00	0,95
S8	39,51	1,00	0,98	0,97	0,90	1,00	0,98	1,00	0,98	1,00	0,99	1,00	0,98	1,00	0,93	1,00	0,93
A4	74,77	1,00	0,99	0,93	0,76	1,00	0,98	1,00	0,99	1,00	0,99	1,00	0,98	1,00	0,95	1,00	0,95
A5	53,38	1,00	0,99	0,92	0,80	1,00	0,99	1,00	0,98	1,00	0,99	1,00	0,98	1,00	0,92	1,00	0,96
A6	44,20	1,00	0,97	0,94	0,84	1,00	0,99	1,00	0,98	1,00	0,99	1,00	0,99	1,00	0,93	1,00	0,95
A7	92,52	1,00	0,98	0,94	0,83	1,00	0,99	1,00	0,99	1,00	0,99	1,00	0,98	1,00	0,91	1,00	0,93
A8	48,87	1,00	0,98	0,97	0,80	1,00	0,98	1,00	0,98	1,00	0,99	1,00	0,99	1,00	0,94	1,00	0,96

4.3.4. Empirical application

Accurately estimating skewness and kurtosis of financial returns is relevant for risk adjustment of investment strategies. An argument commonly put forward is that rational investors should prefer strategies with positive asymmetry and no excess kurtosis. Incorporating sound estimates into risk models may thus lead to safer and more efficient capital allocation. We applied the proposed GMM estimators to the logarithmic daily returns of four major currency pairs from January 1st 2010 to December 31st 2020 (EUR, JPY, GBP and CHF, against 1 USD), comprising ten years of data (2870 observations for each series). We have also computed the tests using the traditional estimators and their respective standard errors. All series showed statistical significance for the presence of persistence and conditional heteroskedasticity (Ljung-Box and ARCH tests). Table 5 shows the results for $H_0 : \tau = 0$ and $H_0 : \kappa = 3$ at $\alpha = 0,05$.

Table 4. Test results (Skewness)

	$\hat{\tau}$	t-stat (g1)	p-value	Std. Error	t-stat (G1)	p-value	Std. Error	t-stat (GMM)	p-value	Std. Error
EUR	0,09	2,17	1,51E-02	0,05	2,17	1,50E-02	0,05	0,88	0,19	0,11
JPY	0,11	2,44	7,29E-03	0,05	2,44	7,26E-03	0,05	0,64	0,26	0,17
GBP	1,03	22,45	6,16E-112	0,05	22,46	4,73E-112	0,05	1,22	0,11	0,84
CHF	-5,4	-118,16	0,00E+00	0,05	-118,22	0,00E+00	0,05	-1,19	0,12	4,54

Table 5. Test results (Kurtosis)

	$\hat{\kappa}$	t-stat (G2)	p-value	Std. Error	t-stat (GMM)	p-value	Std. Error
EUR	5,25	24,65	2,02E-134	0,09	0,50	0,31	22,9
JPY	7,01	43,90	0,00E7	0,09	0,88	0,19	22,5
GBP	18,57	170,23	0,00E8	0,09	3,43	0,00	31,62
CHF	181,75	1954,40	0,00E9	0,09	39,39	0,00	100,75

The null hypothesis of symmetry was rejected for all series when applying traditional estimators. Applying the GMM estimator, the same hypothesis was never rejected, which suggests that the stylized fact of asymmetric financial returns may be too readily accepted, especially when dealing with serially correlated and heteroskedastic series. Even for USDCHF, which contains a severe outlier (-18% on Jan. 16th 2015), the GMM does not reject the null hypothesis of symmetry. This corroborates the benefit of using robust (HAC) estimates of the estimator's (GMM) standard error.

The null hypothesis of mesokurtosis was also rejected for all series when applying the traditional estimator. The GMM estimator does not reject the null hypothesis for the USDEUR and USDJPY series, and the dimension of the consistent standard error plays a significant role, leading to the non-rejection. Although this may point to the absence of excess kurtosis in these series and the refutability of the stylized fact, the erratic behavior of $\hat{\kappa}$, as discussed in the previous section, makes it difficult to obtain solid conclusions. It is also noteworthy that the Swiss franc, despite its typical stability, as the Swiss National Bank actively maintains exchange rates, exhibits a very large sample kurtosis ($\hat{\kappa} = 181.75$), where the aforementioned severe outlier plays a dominant role. This points once more to the difficulty of accurately estimating the true kurtosis of financial return distributions in the presence of outliers and brings into question the option of non-parametric estimators of higher moments.

4.4. Conclusion

The estimation of higher order moments has attracted increasing attention in the financial literature. Fully capturing the risk of financial assets is paramount to robust asset management strategies. Namely, the standard Value-at-Risk models used by financial institutions rely on the Normal distribution to estimate expected portfolio losses over different time periods. As stated, it is a commonly accepted fact amongst both practitioners and researchers that financial return distributions are negatively skewed and leptokurtic, and empirical analyses of returns are often performed that calculate their respective skewness and kurtosis to demonstrate the departures from normality (**zudietal2010**; **Ivanovski2015**). Employing robust tests of skewness and kurtosis in such situations may prove a valuable tool for verification of the aforementioned departures. In this paper, we start by evaluating the standard estimators of skewness and kurtosis against an alternative set of estimators by replacing the sample mean with more robust central tendency measures in their calculation. Using MSE as a loss function, we conclude that employing the trimmed mean, the median and an adjusted median provide more accurate estimates of higher order moments. Additionally, we proposed two GMM tests (of skewness and kurtosis) for time series where serial correlation and conditional heteroskedasticity are present, applying a robust procedure to estimate standard errors. Their performance was assessed using Monte Carlo simulation, and compared to the standard alternative that assumes normality and independence. The GMM test of skewness exhibited good size and seems a better alternative to estimate the true asymmetry of returns. Regarding kurtosis, none of the estimators considered provides reliable estimates of the fourth moment. Sample kurtosis behaves erratically, rendering the estimation inaccurate even for sample sizes as large as $n = 5000$.

We applied the estimators to a 10-year series of logarithmic returns of four major currency pairs. Testing skewness using GMM and traditional estimators led to opposite conclusions. For all series, the former does not reject the null hypothesis of symmetry, whereas the latter points to its rejection ($\alpha = 0.05$). This suggests that the stylized fact of asymmetry of financial returns may be too readily accepted when persistence and heteroskedasticity are unaccounted for. Kurtosis estimates of the currency pairs lead to the non-rejection of mesokurtosis in two series of returns, contradicting the conclusions obtained from the traditional test.

4.5. Appendix: Distributions used in simulation

All distributions used for simulation in Section 3 were chosen from the SGT family, with two exceptions (the standard normal and t-student distributions). The SGT density requires 5 parameters $f_{SGT}(x;\mu,\sigma,\lambda,p,q)$ for location, scale, skewness and tail behavior (kurtosis), respectively. Parametrizations used in the simulations are provided below.

Skewness

S4: $N(0,1)$
S5: $f_{SGT}(0,1,0,1.7,2.5)$
S6: $f_{SGT}(0,1,0,1.6,2.6)$
S7: $f_{SGT}(0,1,0,1.5,2.8)$
S8: $f_{SGT}(0,1,0,1.9,2.2)$
A4: $f_{SGT}(0,1,-0.5,1.7,2.5)$
A5: $f_{SGT}(0,1,-0.3,1.7,2.5)$
A6: $f_{SGT}(0,1,0.2,1.7,2.5)$
A7: $f_{SGT}(0,1,0.3,1.6,2.6)$
A8: $f_{SGT}(0,1,0.5,1.2,4)$

Kurtosis

S4: $N(0,1)$ for size, $t(5)$ for power.
S5: $f_{SGT}(0,1,0,1.7,2.5)$
S6: $f_{SGT}(0,1,0,1.6,2.6)$
S7: $f_{SGT}(0,1,0,1.5,2.8)$
S8: $f_{SGT}(0,1,0,1.9,2.2)$
A4: $f_{SGT}(0,1,-0.5,1.7,2.5)$
A5: $f_{SGT}(0,1,-0.3,1.7,2.5)$
A6: $f_{SGT}(0,1,0.2,1.7,2.5)$
A7: $f_{SGT}(0,1,0.3,1.6,2.6)$
A8: $f_{SGT}(0,1,0.5,1.2,4)$

CHAPTER 5

The impact of COVID-19 on S&P500 sector indices and FATANG stocks volatility: An expanded APARCH model

ABSTRACT

In this paper we hypothesize that not all stocks and sectors are affected equally by COVID-19 in terms of return volatility. Specifically, we hypothesize that at least some sectors (Information Technology, Consumer Discretionary, Telecom Services, Consumer Staples and Energy) must show statistically significant differences. We analyze eleven SP500 sectors and FATANG stocks, estimating an Asymmetric Power GARCH model including a dummy variable to account for the outbreak. Results reveal an exacerbation of volatility after February 2020 and validate our hypothesis with few exceptions. Based on a likelihood ratio test, the null hypothesis is rejected in most cases in favor of our APARCH(1, 1).

5.1. Introduction

Prior to the breakout of COVID-19, the USA witnessed a strong economic expansion, which had a large positive impact on US stocks and originated the longest bull market in US history. As the effects of pandemic transferred to the US economy and financial markets, the S&P 500 dropped 25% during March 2020. The year 2020 was thus characterized by unusual variations in stock prices leading to a period of extremely high volatility.

This paper concerns the impact of COVID-19 on the volatility of financial returns. While there is a general perception of impact, econometrical approaches are relevant to quantify the statistical significance of such perception. **Baietal2020** applied an extended GARCH-MIDAS model to financial returns and a newly developed volatility tracker (EMV-ID) is used to investigate the effects of COVID-19 on volatility of several markets, between January 2005 and April 2020. Results show that, up to a 24-month lag, the pandemic had significant positive impacts on the permanent volatility of international stock markets, even after controlling for the influences of realized volatility, global economic policy uncertainty and the volatility leverage effect. **Shehzadetal2020** found, based on the Asymmetric Power GARCH model, that COVID-19 has significantly harmed the US and Japan's market returns to a greater extent than the Global Financial Crisis (GFC).

Assessing the impact of events on volatility has a long tradition in financial literature. **Gribisch2016** generalizes the basic Wishart multivariate stochastic volatility model by allowing for state-dependent (co)variance and correlation levels and state-dependent volatility spillover effects. The model is applied to five European stock index return series. Results show that the proposed regime-switching specification substantially improves the fit to persistent covariance dynamics relative to the basic model. **BrixLunde2015** investigate the finite sample performance of the Prediction-based estimating functions (PBEFs) – based estimator and compare its performance to that of the Generalized Method of Moments (GMM). **Banerjee2021** analyzes time-varying volatility spillovers between index future markets by applying an ADCC EGARCH (to address dynamic conditional correlations) on the residuals of a VAR model, concluding that most markets that had trade relations with China witnessed volatility contagion due the COVID-19 breakout as well as for the presence of asymmetric volatility.

Researchers have been prolific with respect to the impact of COVID-19 on financial markets. Apart from detecting statistical significance, several authors have attempted to provide an economic rationale underlying those relations. **Baker2020** resorted to automated and human readings of newspapers, looking back to the year 1900, and found that no other extreme event has impacted the markets as strongly as the current pandemic. They suggest that such unprecedented impact may be due to the role of government restrictions on a service-oriented economy. **Baek2020** applied a Markov-switching model to assess the impact of COVID-19 on total, systematic and idiosyncratic risk. While total and idiosyncratic risks have increased across all industries of the S&P500, systematic risk appears to have increased in defensive industries such as telecom and utilities and decreased in their aggressive counterparts such as automobiles and business equipment. The authors attribute the difference between industries to their respective price elasticities. Additional contributions on the topic have been made by **Mazur2020**, **Verma2021**, **Malgorzata2020** and **Zaremba2020**.

As the volatility of financial asset returns remains an important investigation topic in finance, and due to the lack of empirical analyses involving S&P500, FATANG stocks and S&P500 sector indices, the motivation of this study is to address and compare the impact of COVID-19 on return volatility of these classes of US stocks and indices. We wish to address how the three main stylized facts (see **Cont2001**) return volatility (clustering, persistence and asymmetry) compare differently among the eleven S&P500 sector indices and the six FATANG stocks, and to conclude if the impact of COVID-19 has been different in size and direction of financial returns.

Our contribution is threefold. Firstly, because Tesla has become an important player in technology sector, we include it in the set of FAANG stocks and refer to the new set jointly as FATANG. Second, we analyze COVID-19 impact on the volatility of all the 11 S&P 500 sector indices (Information Technology (IT), Health Care (HC), Financials (FI), Consumer Discretionary (CD), Telecom Services (TS), Industrials (ID), Consumer Staples (CS), Energy (EN), Utilities (UT), Real Estate (RE), and Materials (MT)) and on the volatility of FATANG stocks' returns (Facebook, Amazon, Tesla, Apple, Netflix and Google). Considering all sectoral indices, and not just the global index, as is common, also contributes to differentiate this investigation. Secondly, we include six stocks in the empirical analysis to conclude more generally about the impact of COVID-19 on the largest capitalization tech stocks. Finally, an extended Asymmetric Power GARCH (APARCH) model is proposed to assess the statistical significance of COVID-19 on stock volatility. Extreme events are, by definition, rare. We believe that taking the opportunity to test the econometrical toolbox in stressful conditions is a valuable contribution to the scientific community, not only in terms of descriptive but also predictive abilities.

Although it may be unsurprising that not all stocks/indices are affected in the same manner by COVID-19, it is important to assess if differences are statistically significant, and not merely by chance. Our research hypothesis is that at least some sectoral indices (IT, CD, TS, CS, EN) must show significantly different behavior in terms of volatility. As governments impose lockdown restrictions, IT and TS companies providing remote work tools saw a surge in demand, while more idling time may increase online purchases and therefore bring higher demand to Consumer Discretionary products and services. Spending more time at home should also significantly increase the purchase of Consumer Staples. We also hypothesize that the EN sector should be negative and significantly affected, as confinements halt the consumption of fuel for automobiles and airplanes.

The dataset includes prices and returns from March 9, 2009 (a market bottom after the 2008 subprime mortgage crisis) to May 24th, 2021. The time period was selected to include the longest bull market in the history of the US financial markets and to include different clusters of volatility (not just the one resulting from COVID-19) in the estimation process to better model, and describe the conditional heteroskedasticity of financial returns. Specifically, we estimate the APARCH model twice, including a dummy variable (0 before March 2020, 1 otherwise): once ending in December 2020 and another ending in May 24th 2021. In the second estimation, we include an additional dummy (from January 2021 through May 24th 2021) to anticipate, although it may be premature, and assess the impact of vaccination programs on volatility.

The paper is organized as follows: In Section 5.2 we discuss the econometrical methods used in the paper and describe our dataset. Time varying volatility is modeled in Section 5.3 by using an AR-APARCH specification. Inspection of the estimated models point to significant differences in parameter estimates, volatility clustering, volatility persistence and to an asymmetric effect. Section 5.4 concludes the analysis and sheds light on the volatility of returns for the financial assets under scrutiny.

5.2. Methodology

5.2.1. Econometrical framework

We start by establishing the typical representation of financial asset returns as a time series with a predictable and a random component:

$$r_t = E\left[r_t \mid \Phi_{t-1}\right] + u_t, \tag{5.1}$$

where Φ_{t-1} is the relevant (past) information set until (and including) time period $t-1$. A natural assumption for the conditional mean ($E\left[r_t \mid \Phi_{t-1}\right]$) is to model its dynamics as a white noise process, because the empirical distributions of returns under study relate to the most liquid and efficient global equity markets. However, anticipating our findings in the data analysis section, we specify the conditional mean equation as a fourth-order autoregressive process, AR(4), in order to remove the observed linear dependency in returns:

$$r_t = c + \phi_1 r_{t-1} + \phi_2 r_{t-2} + \phi_3 r_{t-3} + \phi_4 r_{t-4} + u_t, \tag{5.2}$$

where $u_t = z_t \sigma_t$ and the standardized innovations (z_t) are assumed to be independently and identically distributed (i.i.d.) following a Student's t distribution (see **Bollerslev1987**):

$z_t \sim t\,(v)$, where $t\,(v)$ is the zero-mean Student's t distribution with v degrees of freedom. This statistical distribution has a long tradition in econometrics as a popular choice from the set of fat-tailed distributions, because it has finite second moment, of its mathematical tractability, and is often found capable of capturing the excess of kurtosis observed in financial time-series. Other non-Normal alternative distributions have also been used in econometrical literature. **nelson91** proposed the Generalized Error distribution (GED), while the Laplace distribution has been employed by **GrangerDing1995**. **Hsieh1989** applied both the Student's t and GED as alternative distributional models for innovations. Stable Paretian distributions have been investigated, among others, by **LiuBrorsen1995**, **MittnikPaolellaRachev1998** and **CurtoPintoTavares2009**.

To model the conditional variance of u_t: $E\left[u_t^2 \mid \Phi_{t-1}\right] = \sigma_t^2$, we apply the Asymmetric Power ARCH (APARCH) model proposed by **DingGrangerEngle1993** in which the power of the conditional heteroskedasticity equation is estimated from the data:

$$\sigma_t^\delta = \omega + \sum_{i=1}^{q} \alpha_i \left(|u_{t-i}| - \gamma_i u_{t-i}\right)^\delta + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^\delta. \tag{5.3}$$

This model couples the flexibility of a varying exponent with the asymmetry coefficient, therefore accounting for the well-known leverage effect (see for example **TavaresCurtoTavares2007** and **CurtoPinto2012**). APARCH is in fact a general class that encompasses seven other models, namely the GARCH in Standard Deviation $[\delta = 1,\ \gamma_i = 0\ (i = 1, \cdots, q)]$ (**Taylor1986** and **Schwert1990**), $GARCH(p,q)$ $[\delta = 2, \gamma_i = 0\ (i = 1, \cdots, p)]$ (**Bollerslev1986**) and GJR $[\delta = 2]$ (**Glosten1993**) models. As observed by **Black1976**, volatility responds asymmetrically to the sign of any change in the price of the financial asset, i.e., volatility increases are greater after negative changes than after positive changes of the same magnitude. This phenomenon has become known as the leverage effect. By estimating an asymmetrical model, we attempt to capture the presence of this leverage effect in the indices and stock returns under analysis. We show that the Utilities sector and Tesla stock returns exhibit no leverage effect, i.e. positive and negative news have the same impact on volatility in the selected time period. However, the leverage effect is detected in the remaining sectors and FATANG stocks.

APARCH models are defined by their order, that is, the number of relevant dependence lags, given by parameters (p, q). Despite the theoretical interest of general (p, q) models, the $(1, 1)$ specification is, in general, satisfactory when modeling financial returns' volatility (see **Bollerslev1992** and more recently **hansenlunde05**). To analyze the impact of COVID-19 on volatility we expand the standard APARCH(1,1) model by including a dummy (D_t) as exogenous variable that assumes the value 1 after the end of February, 2020 and 0 otherwise:

$$\sigma_t^\delta = \omega + \alpha \left(|u_{t-i}| - \gamma u_{t-1}\right)^\delta + \beta_i \sigma_{t-1}^\delta + \theta_1 D_t. \tag{5.4}$$

The model is estimated using maximum likelihood (MLE) and the APARCH model in (5.4) is used to analyze how the three stylized facts of the returns' volatility (clustering, persistence and asymmetry) compare differently among the eleven S&P 500 sector indices and the six FATANG stocks.

By comparing the MLE values from the unrestricted model (APARCH) with those from restricted models: $\delta = 1$ and $\gamma = 0$ (Taylor/Schwert), $\delta = 2$ and $\gamma = 0$ (GARCH) and $\delta = 2$ (GJR), a nested Likelihood Ratio (LR) test can be constructed to compare the in-sample goodness-of-fit of APARCH against either of the other three models, that is, the LR test serves as a procedure to select which model is likely to provide the best fit. Let l_0 be the maximum log-likelihood value under the null hypothesis that the true model is a Taylor/Schwert's GARCH in standard deviation, a GARCH or a GJR, and l_1 be the maximum log-likelihood value under the alternative that the true model is APARCH, then:

$$LR = 2(l_1 - l_0) \tag{5.5}$$

should have a χ^2 distribution with 2, 2 and 1 degrees of freedom, respectively, when the null hypothesis is true. Rejecting the null hypothesis in favor of the alternative that the true model is the proposed APARCH structure provides evidence in support of our model as the data generating process. Our analysis focuses therefore on detecting this asymmetric effect and testing the significance of the COVID-19 dummy variable when APARCH-modeled asymmetry is present.

5.2.2. Data

Our dataset consists of daily closing prices of the S&P 500, the eleven S&P 500 sector indices and six American stocks, five of which go under the acronym FAANG (Facebook, Amazon, Apple, Netflix and Google/Alphabet). We add Tesla as a sixth element and refer to the set of six stocks as FATANG. The period under analysis starts on March 9, 2009 (the market bottom after the 2008 crash. We note that this local minimum is not concurrent to all series – see Table 1). The two exceptions to the time period are Tesla and Facebook stocks, in which prices go back to their first trading day, which occurred later than the otherwise starting date of March 9, 2009. The end date is May 24, 2021 for all series. We performed two distinct analysis: one until December 31st 2020 and another until May 24th 2021, as per referee suggestion, to evaluate the impact of vaccination programs. All data was obtained from `https://www.investing.com`.

We compute the continuously compounded percentage rates of return as:

$$r_t = 100 \times [\ln(P_t) - \ln(P_{t-1})] . \tag{5.6}$$

where P_t is the closing value for each index or stock at time t. Table 1 summarizes the basic statistical properties of the data. All results, with the exception of skewness (which is positive for two series), corroborate the stylized facts of financial returns.

TABLE 1. Summary statistics of r_t

Index/Stock	Starting date	# Obs	Mean	Median	Min	Max	St Dev	Skew	Kurt	J-B
IT	9 Mar '09	2981	0.08	0.12	-14.98	11.30	1.34	-0.47	12.19	0.000
HC	9 Mar '09	2981	0.05	0.08	-10.53	7.31	1.07	-0.36	8.82	0.000
FI	9 Mar '09	2981	0.05	0.07	-15.07	16.33	1.72	0.31	15.42	0.000
CD	9 Mar '09	2981	0.08	0.13	-12.88	8.29	1.23	-0.59	9.99	0.000
TS	9 Mar '09	2981	0.03	0.07	-11.03	8.80	1.14	-0.41	8.46	0.000
ID	9 Mar '09	2981	0.06	0.08	-12.16	12.00	1.33	-0.41	10.84	0.000
CS	9 Mar '09	2981	0.04	0.05	-9.69	8.07	0.88	-0.38	15.17	0.000
EN	9 Mar '09	2981	-0.00	0.02	-22.42	15.11	1.73	-0.90	19.63	0.000
UT	9 Mar '09	2981	0.03	0.09	-12.27	12.32	1.12	-0.28	19.54	0.000
RE	9 Mar '09	2981	0.05	0.09	-18.09	16.24	1.70	0.07	16.93	0.000
MT	9 Mar '09	2981	0.05	0.09	-12.15	11.00	1.41	-0.46	7.22	0.000
SP500	9 Mar '09	2981	0.05	0.07	-12.77	8.97	1.14	-0.64	14.32	0.000
FACEBOOK	18 May '12	2170	0.09	0.11	-21.02	25.94	2.34	0.34	15.11	0.000
AMAZON	9 Mar '09	2981	0.13	0.10	-13.53	23.74	2.07	0.64	10.81	0.000
TESLA	29 Jun '10	2646	0.19	0.12	-23.65	21.83	3.53	-0.04	6.05	0.000
APPLE	9 Mar '09	2981	0.13	0.11	-13.77	11.32	1.79	-0.27	5.98	0.000
NETFLIX	9 Mar '09	2981	0.15	0.05	-42.92	35.22	3.20	-0.30	23.34	0.000
GOOGLE	9 Mar '09	2981	0.08	0.07	-11.77	14.89	1.62	0.29	9.67	0.000

S&P 500 sectors: Information Technology (IT), Health Care (HC), Financials (FI), Consumer Discretionary (CD), Telecom Services (TS), Industrials (ID), Consumer Staples (CS), Energy (EN), Utilities (UT), Real Estate (RE) and Materials (MT). Skew: Coeff. of Skewness, Kurt: Coeff. of Kurtosis and J-B is the $p-$value associated to the Jarque-Bera test.

We note that average returns are all positive but close to zero (higher means correspond to the FATANG stocks), with the exception of the Energy sector. The distribution of returns appears to be asymmetric as reflected by negative and positive skewness estimates. All series exhibit heavy tails and show a strong departure from normality (the skewness and kurtosis coefficients are all statistically different from those of the Normal distribution). The Jarque-Bera normality test statistic is highly significant, which points to the departure of r_t from normality for all series.

In 2020, some sectors have seen a more definitive shock than others (positive or negative). The EN sector was down more than 37%, primarily because crude-oil prices entered a bear market, and have been battered by fears that the outbreak could hurt uptake of crude from China. Three other sectors stand out for their negative returns and impact of COVID-19: RE (-5.17%), FI (-4.10%) and UT (-2.83%). Returns of the remaining sectors have been positive. The best performers among the S&P 500's sectors were IT (+42%), CD (+32%) and TS (+22%). The CS sector showed the weakest growth: 7.63%. Regarding FATANG stocks, the growth in price has been substantial, with Tesla standing out with a remarkable 743.44%. Apple, Amazon, Netflix, Facebook and Google fill the next ranks in terms of stock price growth, contributing significantly to the positive performance of the IT sector.

TABLE 2. Growth of Stocks and S&P 500 Sector Indices

Index/Stock	Prices			Growth	
	Mar 3 '09	Dec 31 '19	Dec 31 '20	Dec 31 '19	Dec 31 '20
IT	199.62	1611.17	2291.28	707.12%	42.21%
HC	253.27	1188.20	1324.01	369.14%	11.43%
FI	83.77	511.39	490.43	510.47%	-4.10%
CD	125.72	986.29	1302.56	684.51%	32.07%
TS	88.10	181.64	221.92	106.17%	22.18%
ID	132.83	687.60	749.54	417.65%	9.01%
CS	199.80	646.97	696.32	223.81%	7.63%
EN	310.92	456.46	286.14	46.81%	-37.31%
UT	113.81	328.36	319.07	188.52%	-2.83%
RE	44.42	240.32	227.90	441.02%	-5.17%
MT	108.82	385.85	455.71	254.58%	18.11%
SP500	676.53	3230.78	3756.0701	377.55%	16.26%
FACEBOOK	38.23	205.25	272.00	436.88%	32.52%
AMAZON	60.49	1847.84	3256.93	2954.79%	76.26%
TESLA	4.77	83.67	705.67	1655.48%	743.44%
APPLE	2.56	72.78	130.20	2743.12%	78.89%
NETFLIX	5.50	323.57	540.73	5783.09%	67.11%
GOOGLE	144.50	1337.02	1752.64	825.27%	31.09%

S&P 500 sectors: Information Technology (IT), Health Care (HC), Financials (FI), Consumer Discretionary (CD), Telecom Services (TS), Industrials (ID), Consumer Staples (CS), Energy (EN), Utilities (UT), Real Estate (RE) and Materials (MT).

In terms of correlation analysis, Figure 1[1] shows positive correlations between all stocks and sectoral indices, i.e. all returns tend to move in the same direction. We note that correlations remain positive and generally increase (larger, darker circles) after March 2020 (after COVID-19). This corroborates the well known concern with portfolio theory that diversification benefits erode in face of extreme events (correlations increase and tend to +1).

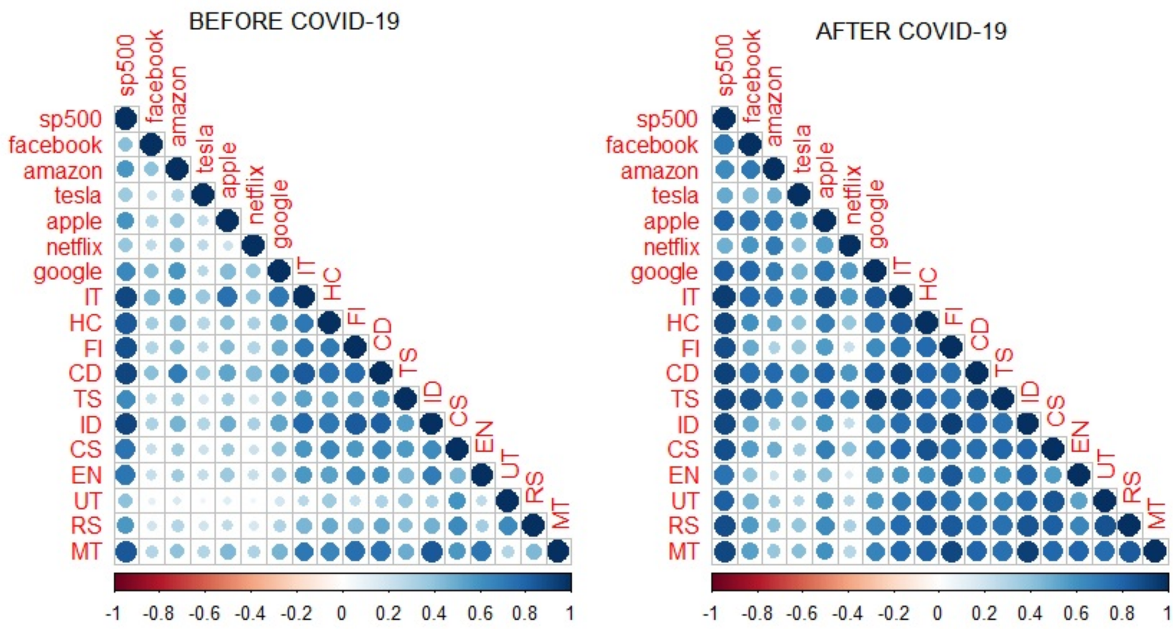
5.3. Empirical Results

In Table 3, we report the MLE APARCH(1,1) parameter estimates for the 18 return series. Volatility clustering is evident in all series, from the signal and significance of estimates for α and β. By 'volatility clustering' we mean persistence of high and low volatility, i.e. large (absolute) returns tend to be followed by returns of similar (large) magnitude.

Inspection of parameter estimates reveals that the ARCH effect (volatility clustering) is present in all series (α and β are significant across the board with only slight differences). For most series, APARCH(1,1) processes are highly persistent (almost-integrated: Facebook and Netflix) with the estimate for $\alpha + \beta$ ranging from 0.9492 to 0.9999, i.e. shocks on volatility not only cluster but also cancel out slowly, in line with the extant literature. These results are of practical relevance to practitioners, pointing to a slow decay of large (absolute) returns when dealing with FATANG stocks and S&P500 sectoral indices.

[1]As per referee suggestion.

Figure 1. Pearson correlation between FATANG stocks, S&P 500 and sectoral indices



S&P 500 sectors: Information Technology (IT), Health Care (HC), Financials (FI), Consumer Discretionary (CD), Telecom Services (TS), Industrials (ID), Consumer Staples (CS), Energy (EN), Utilities (UT), Real Estate (RE) and Materials (MT). We use the R `corrplot` package. Larger, darker circles represent stronger absolute value correlations.

The leverage effect (γ) is statistically significant for all series, except for Utilities sector and Tesla. By leverage effect we mean the asymmetric impact of good and bad news on the volatility of financial returns. For the remaining series, where leverage is present, negative shocks (bad news) have a stronger impact than their positive counterpart (good news).

Considering the effect of COVID-19 on return volatility, results show that the estimate for the coefficient (θ_1) of the dummy variable is always positive, revealing an increase in volatility after February 1st, 2020. However, the estimate is statistically significant only for the IT, CD, TS, ID, CS and EN sectors, in spite of the significance of the asymmetric γ estimates. The statistical significance of $\hat{\theta}_1$ reinforces the impact of shocks (due to bad news) on volatility, i.e. COVID-19 exacerbated the leverage effect. From the performance of these sectoral indices in 2020 we can conclude that higher volatility had a positive impact on the returns of the first five, extending their uptrend in price, and a strong negative impact on the returns of the EN sector (see Table 2). For FATANG stocks, the estimate is statistically significant for Apple and Google only, and the volatility had a strong positive impact on their prices. For the remaining sectoral indices and prices, the impact of COVID-19 may be already included in the asymmetric γ estimate, reinforcing, but not increasing statistically, the effect of bad news on volatility.

In terms of the Likelihood Ratio test results, the rejection of the null in favor of the APARCH model occurs in most of the eighteen series under study. However, in the case of Tesla, we do not reject that the data may have been generated by the symmetric Taylor/Schwert or GARCH models. This conclusion is in accordance with the aforementioned insignificance of the γ estimate. Non-rejection of the null in favor of the asymmetric GJR model occurs in the IT, CD, UT and RE time series.

Our hypothesis is generally validated – sectors that seemed *a priori* candidates for a statistically significant impact of COVID-19 (IT, CD, TS, CS and EN) on volatility show in fact significance, most likely due to lockdown-induced behaviors such as remote work and increased consumption (the one exception is the ID sector, where we did not expect significance). Our results are generally in line with the financial COVID-19 literature: while most authors describe a relevant impact of COVID-19 on financial markets, **Baek2020** finds that systematic risk has increased significantly for telecoms. **Verma2021** and **Mazur2020** observe significant downturns in the energy sector (specifically crude oil companies). The latter author also establishes the impact of COVID-19 in the food and software industries.

We have re-estimated our APARCH model to include the period from January 1, 2021 through May 24th 2021 and evaluate the impact of vaccination programs on volatility as well as leverage of the same stocks and sectoral indices[2]. To that effect, we introduced an additional dummy variable V_t in Equation 5.4 (1 from January 1st 2021 through May 24th 2021, 0 otherwise):

$$\sigma_t^{\delta} = \omega + \alpha \left(|u_{t-i}| - \gamma u_{t-1}\right)^{\delta} + \beta_i \sigma_{t-1}^{\delta} + \theta_2 D_t + \phi V_t \tag{5.7}$$

Table 3 shows the estimate for θ_1, which represents the effect of COVID-19 until December 31st, 2020, while the estimate for θ_2 that represents the effect of COVID-19 until May 24, 2021 (the sample is increased by five months). The estimate $\hat{\phi}$ represents the impact of the vaccination on the US stock markets volatility. Re-estimating the model necessarily affected the initial estimates for the remaining parameters. New estimates show very small deviations, which, for clarity, we do not show. These results are available upon request to the authors.

The estimates for θ_2 (which include data from the first five months of 2021) are very close to those of θ_1 (which included data up until December 2020 only). Re-estimated values are slightly different, signs are maintained and only those estimates that were statistically significant so remain. Thus, it seems that COVID-19 impact on FATANG

[2]As per referee suggestion.

and S&P500 sectors did not change substantially in the first five months of vaccination. With regards to ϕ, estimates are almost all negative (except for EN sector and Google), indicating a negative impact of vaccination on volatility. However, none of the estimates is statistically significant. Therefore, we cannot yet conclude that vaccination had a statistically significant impact on US stock market volatility.

TABLE 3. APARCH(1,1) estimates and Likelihood Ratio test

Indices/Stocks	α	β	$\alpha+\beta$	δ	leverage(γ)	covid (θ_1)	covid (θ_2)	vacc. (ϕ)	t−df	T/S	GARCH	GJR
IT	0.1130*	0.8696*	0.9826	0.9440*	1.000*	0.0458*	0.0478*	-0.0247	5.5099*	134.79*	127.55*	38.05*
HC	0.0886*	0.8858*	0.9744	1.0649*	0.925*	0.0084	0.0102	-0.0087	7.2675*	90.12*	73.75*	14.87*
FI	0.1225*	0.8731*	0.9956	1.1347*	0.5598*	0.0263	0.0317	-0.0023	5.8516*	71.61*	55.05*	8.84*
CD	0.1010*	0.8675*	0.9685	1.4660*	0.6516*	0.0399**	0.0565*	-0.0184	7.2462*	88.03*	65.72*	4.90***
TS	0.0641*	0.8858*	0.9499	1.5165*	0.5887*	0.0422**	0.0448**	-0.0083	5.8375*	37.78*	28.48*	2.28
ID	0.0826*	0.8950*	0.9776	1.4260*	0.7606*	0.0285**	0.0338***	-0.0126	6.6689*	93.03*	73.2*	6.23**
CS	0.0960*	0.8597*	0.9557	1.4183*	0.6098*	0.0191***	0.0184***	-0.0017	7.4517*	62.64*	44.57*	4.2
EN	0.0753*	0.9203*	0.9956	1.0362*	0.6413*	0.0319***	0.0376***	0.0042	8.535*	60.00*	51.84*	12.51*
UT	0.0593*	0.8899*	0.9492	2.5707*	0.1103	0.0428	0.0536	-0.0252	8.3229*	30.47*	7.55*	2.05
RE	0.0994*	0.8828*	0.9822	1.9573*	0.1821*	0.0283	0.0316	-0.0229	7.9654*	35.47*	10.77*	0.02
MT	0.0824*	0.9148*	0.9972	1.0543*	0.7910*	0.0132	0.0165	-0.0007	8.811*	81.83*	72.06*	19.36*
FACEBOOK	0.0765*	0.9333*	0.9999	0.9564*	0.4153*	0.0100	0.0121	-0.0033	3.8841*	16.46*	40.38*	22.53*
AMAZON	0.1003*	0.8888*	0.9891	0.793*	0.5156*	0.0225	0.0229	-0.0239	4.2747*	33.33*	90.49*	43.59*
TESLA	0.0785*	0.9106*	0.9891	0.9182*	0.1584	0.0747	0.0819	-0.0346	3.6068*	2.14	2.90	8.73**
APPLE	0.1145*	0.8493*	0.9638	0.8571*	0.7450*	0.0654**	0.0661**	-0.0434	4.986*	70.44*	92.99*	29.86*
NETFLIX	0.0716*	0.9285*	0.9999	0.5416*	0.5121*	0.0020	0.0024	-0.0131	3.3392*	28.56*	88.95*	68.62*
GOOGLE	0.0868*	0.8993*	0.9861	0.848*	0.6391*	0.0219***	0.0237***	0.0042	3.9208*	36.73*	74.42*	32.31*
SP500	0.1268*	0.8712*	0.9980	0.9487*	0.9999*	0.0251**	0.0267*	-0.0137	5.6529*	153.01*	140.31*	38.51*

S&P 500 sectors: Information Technology (IT), Health Care (HC), Financials (FI), Consumer Discretionary (CD), Telecom Services (TS), Industrials (ID), Consumer Staples (CS), Energy (EN), Utilities (UT), Real Estate (RS) and Materials (MT). *, **, *** denote statistically significant at the 1%, 5% and 10% significance levels, respectively. $\alpha+\beta=1$ is the measure of volatility persistence. `t-df` represents the Student's t degrees of freedom. The LR test is shown under the names T/S (Taylor/Schwert), GARCH and GJR, respectively. θ_1 represents the effect of COVID-19 until December 31st, 2020 and θ_2 represents the effect of COVID-19 until May 24th, 2021. ϕ covers the period from January through May 24th 2021, to measure the impact of vaccination programs on volatility.

5.4. Conclusion

We compared the impact of COVID-19 on US stock prices and volatility by analyzing the continuously compounded returns of the eleven S&P 500 sectors and the six FATANG stocks from March 9th, 2009 to May 24th, 2021.

We note the remarkable growth of US stocks since the subprime crisis in 2008 until December, 2019. In 2020, however, we observed mixed results, as some sectors were more heavily impacted than others. All year-to-year returns of FATANG stocks have exceeded 30%, and a similar performance was only achieved by the IT and CD sectors.

To assess the effect of COVID-19 on return volatility we proposed an expanded Asymmetric Power GARCH (APARCH) model (by introducing a dummy variable controlling for the breakout of COVID-19). Results show that the estimate for its coefficient is always positive, revealing an increase in volatility after February 2020. The estimate is statistically significant for the IT, CD, TS, ID, CS and EN sectors. For FATANG stocks, the estimate is statistically significant for Apple and Google. These results generally validate our research hypothesis (with the exception of increased volatility of Apple and Google), and serve as reminder to portfolio and risk managers that while it may be unsurprising that not all sectors of the economy are affected equally in terms of volatility, there may be certain regularities in population behavior that impact specific sectors. COVID-19 should therefore serve as a learning experience to identify which sectors may be most affected by similar circumstances in the future.

Based on the likelihood ratio test, we conclude that the symmetrical GARCH model is rejected and the asymmetrical GJR model is the particular case of APARCH that is accepted more often in our data set. For most series, therefore, we conclude for the presence of the volatility leverage effect. Exceptions are the UT sector and Tesla, where good and bad news have the same level of impact. APARCH processes are highly persistent for most series. The exceptions are the UT, TS, CS, CD sectors, and Apple, where the effect of shocks on volatility is less persistent, canceling out quickly.

Concluding, we can retrospectively state that the COVID-19 outbreak did not hit all the US sectors and all the US stock prices in the same manner, and our analysis provides an attempt at quantifying those differences. The higher volatility has favored mainly FATANG stocks and the IT, CD, TS, ID and CS sectors, while the EN sector was the most negatively affected. Promising, albeit simple, explanations are the significant increase in remote work (IT and TS) and the stay-at-home population behavior (and needs), which became prevalent as a result of lockdown restrictions (CD and CS).

We have focused on the S&P 500 due to its dimension, recognition and importance for portfolio managers. Further research can be done by replicating our method to indices from different countries (e.g. EUROSTOXX 600) and asset classes (e.g. fixed income). As countries are still rolling out their vaccination programs, the available sample size to assess their impact may be too small. We did not find that vaccination has impacted return volatility in this time period. Re-estimating models when more data is available may also prove a valuable contribution to better understand the impact of pandemic events on stock market returns and volatility.

CHAPTER 6

Conclusion

As stated in the Introduction, the main goal of this thesis is to tackle known estimation issues pertaining to the unreliability of standard estimators of distributional moments when traditional regularity conditions are violated, such as the normality of random variables and the identical and independent distribution of observations, as well as the absence of collinearity between explanatory variables in linear regression models. The author concludes that, generally, the ongoing quest for more robust alternatives proves valuable as it increases the reliability of estimation procedures.

In Chapter 2, the main conclusion is that the geometric mean provides a better fit to portfolio realized price-to-earnings values. While several authors argue that the harmonic mean is preferable to average ratios due to the arithmetic mean's sensitivity to in-sample ratio sizes, it is also true that the harmonic mean tends strongly towards the smallest observations, mitigating the impact of large outliers and amplifying the impact of smaller ones. This conclusion is confirmed by applying this theoretical result to real financial time series. In this regard, further research may be conducted by extending the results to other financial ratios and quantities.

In Chapter 3, the issue of variance partitioning and relative importance of independent variables in linear regression is addressed. While a wide set of relative importance metrics is available to researchers, drawing meaningful conclusions from data requires the careful analysis of correlation matrices must be performed. As relative importance is a construct rather than an actual quantity to be estimated, it is argued that no measure should be applied as a general solution. Notwithstanding, our measure tackles the known issue of single analysis methods being strongly affected by the presence of collinearity, while being intuitive and easy to compute. It accounts for this interdependence by incorporating the Variance Inflation Factor as a weighting criterium when assigning shares of the coefficient of determination to explanatory variables to determine their relative importance. On this topic, further research may be conducted by either extending the simulation settings to different distributions, as well as to the presence of heteroskedasticity and autocorrelation, or incorporating the Variance Inflation Factor into other existing single analysis methods.

In Chapter 4, using Mean Squared Error as a loss function, we conclude that employing the trimmed mean, the median and an adjusted median provide more accurate estimates of higher order moments. Regarding the GMM tests and confidence intervals, we conclude that, to estimate skewness in the presence of heteroskedasticity and autocorrelation, our solution is a more robust procedure to estimate standard errors. With respect to kurtosis, none of the estimators considered provides reliable estimates of the fourth moment. Sample kurtosis behaves erratically, rendering the estimation inaccurate even for sample sizes as large as $n = 5000$. Further research may be conducted by extending the results to evaluate the behavior of hypothesis tests and confidence intervals to temporal aggregation and/or the multivariate case.

In Chapter 5, to assess the effect of COVID-19 on return volatility we proposed an expanded Asymmetric Power GARCH (APARCH) model by introducing a dummy variable controlling for the breakout of COVID-19. Results show that the estimate for its coefficient is always positive, revealing an increase in volatility after February 2020. The estimate showed statistical significance for several sectors of the S&P500 index, namely the Information Technology, Consumer Discretionary, Telecom, Industrial, Consumer Staples and Energy sectors. For FATANG stocks, the estimate showed statistical significance for Apple and Google. These results generally validate our research hypothesis, and serve as reminder to portfolio and risk managers that while it may be unsurprising that not all sectors of the economy are affected equally in terms of volatility, there may be certain regularities in population behavior that impact specific sectors. COVID-19 should therefore serve as a learning experience to identify which sectors may be most affected by similar circumstances in the future.

CHAPTER 7

Appendix (code)

This appendix contains samples of the `R` code written for simulation in the first three articles. Code for the last article is excluded because no simulation was performed.

7.1. Averaging ratios: geometric vs. harmonic mean

```
library(moments)
library(rmutil)

# Mean definition
# Geometric mean
gm_mean <- function(a){prod(a)^(1/length(a))}
# Harmonic mean
hm_mean <- function(b){length(b)/(sum(1/b))}
## CUBIC ROOT AVERAGE
dados <- c(300, 20, 10)
hm_mean(dados)
CRAve  <- function(x){
  len <- length(x)
  invs <- vector(length=len)
  weigs <- vector(length=len)
  for(j in 1:len){
    invs[j] <- 1/(x[j]^(1/2))
  }
  for(j in 1:len){
    weigs[j] <-invs[j]/sum(invs)
  }
  CRAve <- t(x)%*%weigs
  return(CRAve)
}

CRAve(dados)

# Sample sizes
size <- 10
smplsize <- c(10, 20, 30, 40, 50, 60, 70, 80, 90, 100)
library(rmutil)
numbout <- matrix(ncol=3, nrow=size)
for(j in 1:size) {
  n <- smplsize[j]
  reps <- 10000
  SM1 <- matrix(NA,reps)
  for(i in 1:reps) {
    #(rat <- rbeta(n, 1, 1))
    (rat <- runif(n, 0, 1))
    # Symmetric distributions
    #(rdx <- rnorm(n, 2, 1))
    #(rdx <- rnorm(n, 10, 5))
    #(rdx <- rlogis(n, 2, 1))
    #(rdx <- rlogis(n, 10, 5))
    # (rdx <- rt(n, 10))
    #(rdx <- rt(n, 10))
    ## BETA(3, 3)
    v <- 3
    w <- 3
    #(rdx <- rbeta(n, v, w))
    v <- 2
    w <- 5
    #(rdx <- rbeta(n, v, w))
```

```
  v <- 1
  w <- 10
  #(rdx <- rbeta(n, v, w))
  ## LOGNORMAL
  # (rdx <- rlnorm(n, 0, 1))
  # (rdx <- rlnorm(n, 2, 1))
  # (rdx <- rlnorm(n, 4, 1))
  ## WEIBULL
  lam <- 1 # scale
  k <- 2 # shape
  # (rdx <- rweibull(n, k, lam))
  #(rdx <- rweibull(n, 1, 1))
  lam <- 2 # scale
  k <- 5 # shape
  #(rdx <- rweibull(n, k, lam))
  ## GAMMA
  b <- 1 # scale
  c <- 6 # shape
  #(rdx <- rggamma(n, c, b, 1))
  b <- 2 # scale
  c <- 5 # shape
  #(rdx <- rggamma(n, c, b, 1))
  b <- 3 # scale
  c <- 3 # shape
  #(rdx <- rggamma(n, c, b, 1))

  ## EXPONENCIAL
  # (rdx <- rexp(n))
  # (rdx <- rexp(n, 5))
  # (rdx <- rexp(n, 10))
  ## CHI-SQUARED(1)
  # v <- 1
  # (rdx <- rchisq(n, v))
  v <- 5
  (rdx <- rchisq(n, v))
  #v <- 15
  #(rdx <- rchisq(n, v))

  (rdy <- rdx/rat)
  (rat <- rdx/rdy )
  (realrat <- sum(rdx)/sum(rdy))
  (len <- length(rat))
  powermean <- function(k) { (((1/len)*sum(rat^k))^(1/k)-realrat)^2}
  opt <- optimize(powermean, interval=c(-100, 100), maximum=FALSE)
  (SM1[i] <- opt$minimum)
  ## MSE
  #MSEHA <- ((hm_mean(rat)-realrat)^2)/len
  #MSEGA <- ((gm_mean(rat)-realrat)^2)/len
  #MSECRA <- ((CRAve(rat)-realrat)^2)/len
  #MSEAA <- ((mean(rat)-realrat)^2)/len
}
numbout[j, 1] <- quantile(SM1, 0.025)
numbout[j, 2] <- mean(SM1)
numbout[j, 3] <- quantile(SM1, 0.975)
#numbout[j, 4] <- MSEHA
#numbout[j, 5] <- MSEGA
#numbout[j, 6] <- MSECRA
#numbout[j, 7] <- MSEAA
}
```

7.2. Revisiting relative importance: a VIF-based measure

```
library(MASS)
library(mvtnorm)
library(car)
library(lmSupport)
library(ppcor)
library(regclass)
library(QuantPsyc)
library(matrixStats)
library(gridExtra)

#number of observations
k = 0
n <- 100
nsims <- 500 #number of simulations

#Create matrix to store simulation results

simres <- matrix(, nsims, 4)
simres2 <- matrix(, nsims, 4)
iqrs <-  matrix(, 19, 4)
rsmpl <- matrix(,19,4)
rhocount = 1
######SIMULATION

for (rho in seq(-0.9,0.9, by = 0.1)){

for (z in 1:nsims){

  j=0
#Generate covariance matrix
varX1 <- 1
varX2 <- 1
varX3 <- 1
varX4 <- 1

varvec <- cbind(varX1, varX2, varX3, varX4)
cormat <- matrix(, nrow = 4, ncol = 4)
covmat <- matrix(, nrow = 4, ncol = 4)

for (i in 1:4){
  for(j in 1:4){
    covmat[i,j] = rho^abs(i-j)*sqrt(varvec[i])*sqrt(varvec[j])
  }
}
#Cov matrix as described in the paper (see Table 2, first line)

#Simulate Multivariate Normal, 4 regressors
bcoeffs <- cbind(5, 1, 1, 1) # Coefficients
X <- rmvnorm(n, c(0, 0, 0, 0), sigma = covmat) # Regressors
intercept <- 0 #Intercept
truersq <- 0.90 #Desired true rsq
```

```
#Computing variance without error term to control R^2
woerror <- t(intercept + bcoeffs %*% t(X)) #True model
Xvar <- var(woerror) #Variance of true model
u <- rnorm(n, sd = sqrt((Xvar/truersq)-Xvar)) #Setting the variance of u to control rsq

#Generate Y and check Rsq
Y <- t(intercept + bcoeffs %*% t(X) + u) #Y
sum((woerror - mean(Y))^2)/sum((Y-mean(Y))^2) #Check Rsq (should be close to 'truersq', difference due to sampling error)

#Estimate linear model
model <- lm(Y ~ scale(X[,1]) + scale(X[,2]) + scale(X[,3]) + scale(X[,4]))
rsq <- summary(model)$r.squared #storing rsquared

#Compute semipartial correlations from output
a <- cor(lm(scale(X[,1]) ~ scale(X[,2]) + scale(X[,3]) + scale(X[,4]))$residuals, Y, method="pearson")^2
b <- cor(lm(scale(X[,2]) ~ scale(X[,1]) + scale(X[,3]) + scale(X[,4]))$residuals, Y, method="pearson")^2
c <- cor(lm(scale(X[,3]) ~ scale(X[,2]) + scale(X[,1]) + scale(X[,4]))$residuals, Y, method="pearson")^2
d <- cor(lm(scale(X[,4]) ~ scale(X[,2]) + scale(X[,3]) + scale(X[,1]))$residuals, Y, method="pearson")^2
sqpart <- rbind(a,b,c,d)

#Building the measure

sumindividual <- a + b + c + d #Sum individual squared part correlations
vifs <- VIF(model) #Compute VIFs for each regressor
vifweights <- vifs/sum(vifs) #Compute VIF weights

#Generate matrices to store shares
  unexpshare <- matrix(,4,1)
  expshare <- matrix(,4,1)
  totalshare <- matrix(,4,1)

#Fill matrices

  unexpshare = vifweights %*% (rsq-sumindividual)
  expshare = sqpart
  totalshare = unexpshare + expshare

simres[z,] <- t(totalshare)

#print(all.equal(sum(totalshare), rsq)) #check if shares add up to r-squared
#sumindividual <= rsq #Check supression
}

iqrs[rhocount,] <- cbind(iqr(simres[,1]),iqr(simres[,2]), iqr(simres[,3]), iqr(simres[,4]))
rhocount = rhocount + 1
}
```

7.3. Contributions to the diagnosis of Skewness and Kurtosis

```
library(e1071)
library(matlib)
library(muStat)

##########################################################
#ESTIMATORS

#Aux function adj_median
adj_median <- function(arr){
  med1 <- median(arr)
  sumleft <- sum(med1-arr[arr<med1])
  sumright <- sum(arr[arr>med1]-med1)
  tot <- sumleft + sumright
  sk <- (sumright-sumleft)/tot
  diff1 <- abs(mean(arr)-median(arr))
  diff2 <- sk*diff1

  if(mean(arr)< median(arr)) {
    aMED <- median(arr)-diff2 } else {
      aMED <- median(arr)+diff2 }
  return(aMED)
}

#Skewness

#Mean
emeans <- function(n, arr){
  mean((arr-mean(arr))^3)/(var(arr)^(3/2))
}
#Trimmed mean
tmeans <- function(n, arr){
  tmean = mean(arr, trim = 1/(2*sqrt(n-4)))
  mean((arr-tmean)^3)/(var(arr)^(3/2))
}

#Median
emedians <- function(n,arr){
  mean((arr-median(arr))^3)/(var(arr)^(3/2))
}

#Adjusted median
eadjmeds <- function(n,arr){
  mean((arr-adj_median(arr))^3)/(var(arr)^(3/2))
}

#Kurtosis
#Mean
emeank <- function(n, arr){
  mean((arr - mean(arr))^4)/(var(arr)^(2))
}
#Trimmed mean
tmeank <- function(n, arr){
  tmean = mean(arr, trim = 1/(2*sqrt(n-4)))
  mean((arr-tmean)^4)/(var(arr)^(2))
}
```

```
#Median
emediank <- function(n,arr){
  mean((arr-median(arr))^4)/(var(arr)^(2))
}
#Adjusted median
eadjmedk <- function(n,arr){
  mean((arr-adj_median(arr))^4)/(var(arr)^(2))
}
##########################################################

reps = 10000
n = 10
x = 0

Ssse <- matrix(0, 4,7)
Ksse <- matrix(0, 4,7)

repeat{
#DISTRIBUTIONS

  #1. Chi-square distribution

    #Parameters
      k = 1
    #true skewness and kurtosis
      scsq = sqrt(8/k)
      kcsq = (12/k) + 3
    #simulation
      csq <- rchisq(n,k)
      #plot(density(csq))

  #2. t-student distribution

    #Parameters
      v = 5
    #true skewness and kurtosis
      st = 0
      kt = 6/(v-4) + 3
    #simulation
      t <- rt(n, v)
      #plot(density(t))

  #3. Beta distribution

    #Parameters
      alphapar = 1
      betapar = 10
    #true skewness and kurtosis
      sbeta = (2*(betapar-alphapar)*sqrt(alphapar+betapar+1))/((alphapar+betapar+2)*sqrt(alphapar*betapar))
      kbeta = (6*((alphapar-betapar)^2 * (alphapar+betapar+1)-alphapar*betapar*(alphapar+betapar+2)))
      /(alphapar*betapar*(alphapar+betapar+2)*(alphapar+betapar+3)) + 3
    #simulation
      beta <- rbeta(n,alphapar,betapar)
      #plot(density(beta))
  kbeta
```

```
#4. Uniform distribution

  #Parameters
    low = 0
    high = 1
  #true skewness and kurtosis
    sunif = 0
    kunif = -6/5 + 3
  #simulation
    unif <- runif(n,low,high)
    #plot(density(unif))

#5. Standard normal distribution

    #Parameters
      m = 0
      sig = 1
    #true skewness and kurtosis
      snor = 0
      knor = 3
    #simulation
      norm <- rnorm(n,m,sig)
      #plot(density(norm))

#6. Gamma distribution

    #Parameters
      shape = 1
      rate = 6
    #true skewness and kurtosis
      sgam = 2/sqrt(shape)
      kgam = (6/shape) + 3
    #simulation
      gamma <- rgamma(n,shape,rate)
      #plot(density(gamma))

#7. Exponential distribution

    #Parameters
      lamb = 1
    #true skewness and kurtosis
      sexp = 2
      kexp = 6 + 3
    #simulation
      exp <- rexp(n,lamb)
      #plot(density(exp))

#Storing data
      truth <- matrix(c(scsq,kcsq,st,kt,sbeta,kbeta,sunif,kunif,snor,knor,sgam,kgam,sexp,kexp),7,2,byrow=TRUE)
      data <- cbind(csq,t,beta,unif,norm,gamma,exp)

      #for (i in 1:7){
       # data[10,i] <- 50 * quantile(data[,i], probs = 0.25)
      #}
```

```
  #Estimates
    #Skewness
    Ssse = Ssse + (sweep(rbind(
      apply(data,2, function(arr) emeans(n,arr)),
      apply(data,2, function(arr) tmeans(n,arr)),
      apply(data,2, function(arr) emedians(n,arr)),
      apply(data,2, function(arr) eadjmeds(n,arr))), 2, truth[,1]))^2

    #Kurtosis
    Ksse = Ksse + (sweep(rbind(
      apply(data,2, function(arr) emeank(n,arr)),
      apply(data,2, function(arr) tmeank(n,arr)),
      apply(data,2, function(arr) emediank(n,arr)),
      apply(data,2, function(arr) eadjmedk(n,arr))), 2, truth[,2]))^2

    x = x + 1
    if(x == reps) {
      rownames(Ssse) = c('Mean', 'Trimmed Mean', 'Median', 'Adj. Mean')
      rownames(Ksse) = c('Mean', 'Trimmed Mean', 'Median', 'Adj. Mean')
      break}
}
t(sqrt(Ssse/reps))
t(sqrt(Ksse/reps))

reps = 1000
  n1 = 50
  n2 = 250
  n3 = 500
  n4 = 1000
  n5 = 2500
  n6 = 5000

  set.seed(123)
  data1 <- matrix(runif(n1 * reps,0,1), nrow = reps)
  data2 <- matrix(runif(n2 * reps,0,1), nrow = reps)
  data3 <- matrix(runif(n3 * reps,0,1), nrow = reps)
  data4 <- matrix(runif(n4 * reps,0,1), nrow = reps)
  data5 <- matrix(runif(n5 * reps,0,1), nrow = reps)
  data6 <- matrix(runif(n6 * reps,0,1), nrow = reps)

#OUTLIER REPLACEMENT

for (i in 1:1000){

  data1[i,10] <- 50 * quantile(data1[i,], probs = 0.25)
  data2[i,10] <- 50 * quantile(data2[i,], probs = 0.25)
  data3[i,10] <- 50 * quantile(data3[i,], probs = 0.25)
  data4[i,10] <- 50 * quantile(data4[i,], probs = 0.25)
  data5[i,10] <- 50 * quantile(data5[i,], probs = 0.25)
  data6[i,10] <- 50 * quantile(data6[i,], probs = 0.25)
}
    a <- apply(data1,1, function(arr) eadjmedk(n1,arr))
    b <- apply(data2,1, function(arr) eadjmedk(n2,arr))
    c <- apply(data3,1, function(arr) eadjmedk(n3,arr))
    d <- apply(data4,1, function(arr) eadjmedk(n4,arr))
    e <- apply(data5,1, function(arr) eadjmedk(n5,arr))
    f <- apply(data6,1, function(arr) eadjmedk(n6,arr))
```

```
#SKEWNESS

library(gld)
library(tidyverse)
library(e1071)
library(sgt)

########CALCULATION OF SKEWNESS FOR SGT DISTRIBUTION

skewsgt <- function (mu, sigma, lambda, p, q){
  v = (q^(-1/p))*(1/(sqrt((((3*lambda^2)+1)*(beta(3/p,q-(2/p))/beta(1/p,q))
  -4*(lambda^2)*((beta(2/p, q-(1/p))/beta(1/p,q))^2)))))
  skewsgtc = ((2 * (q^(3/p)) * lambda * ((v*sigma)^3)) / (beta(1/p,q)^3))
  * ( 8 * lambda^2 * beta(2/p, q-(1/p))^3 - 3 * (1+3*lambda^2)
  * beta(1/p,q) * beta(2/p, q-(1/p)) * beta(3/p, q-(2/p)) + 2 * (1+lambda^2) * beta(1/p,q)^2 * beta(4/p,q-(3/p)))
  return(skewsgtc)
}

varsgt <- function(mu,sigma,lambda,p,q){
  v <- (q^(-1/p))*(1/(sqrt((((3*lambda^2)+1)*(beta(3/p,q-(2/p))/beta(1/p,q))
  -4*(lambda^2)*((beta(2/p, q-(1/p))/beta(1/p,q))^2)))))
  varsgtc = (((v*sigma)^2)*q^(2/p)) * (  (3*lambda^2 + 1)*(beta(3/p,q-(2/p))/beta(1/p,q))
  -(4*lambda^2) * ((beta(2/p,q-(1/p))/beta(1/p,q))^2))
  return(varsgtc)
}


################################TEST SIZE#############################################

s1 = 0
s2 = skewsgt(0 , 1, 0, 1.7, 2.5)/(varsgt(0 , 1, 0, 1.7, 2.5)^(3/2))
s3 = skewsgt(0,1, 0, 1.6, 2.6)/(varsgt(0,1, 0, 1.6, 2.6)^(3/2))
s4 = skewsgt(0, 1, 0, 1.5, 2.8)/(varsgt(0, 1, 0, 1.5, 2.8)^(3/2))
s5 = skewsgt(0, 1, 0, 1.9, 2.2)/(varsgt(0, 1, 0, 1.9, 2.2)^(3/2))
param <- cbind(s1, s2, s3, s4, s5)

nsims = 200

MATFINAL <- lapply(1:6, matrix, data = NA, nrow = nsims, ncol = ncol(param))
MATFINAL1 <- lapply(1:3, matrix, data = NA, nrow = nsims, ncol = ncol(param))

for (z in 1:nsims){
size = 500
MATS <- lapply(1:3, matrix, data = NA, nrow = size, ncol= 5)
j = 0

for (rho in c(0, 0.5, 0.8)){
    j = j + 1
      S = rnorm(size)
      MATS[[j]][1, 1] = median(S)
      for(ssize in 2:size){
        MATS[[j]][ssize, 1] = rho*MATS[[j]][ssize - 1, 1] + S[ssize]
    }
```

```
      S = rsgt(size,mu = 0, sigma = 1, lambda = 0, p = 1.7, q = 2.5, mean.cent = TRUE, var.adj = TRUE)
      MATS[[j]][1, 2] = median(S)
      for(ssize in 2:size){
        MATS[[j]][ssize, 2] = rho*MATS[[j]][ssize - 1, 2] + S[ssize]
      }

      S = rsgt(size,mu = 0, sigma = 1, lambda = 0, p = 1.6, q = 2.4, mean.cent = TRUE, var.adj = TRUE)
      MATS[[j]][1, 3] = median(S)
      for(ssize in 2:size){
        MATS[[j]][ssize, 3] = rho*MATS[[j]][ssize - 1, 3] + S[ssize]
      }

      S = rsgt(size,mu = 0, sigma = 1, lambda = 0, p = 1.5, q = 2.3, mean.cent = TRUE, var.adj = TRUE)
      MATS[[j]][1, 4] = median(S)
      for(ssize in 2:size){
        MATS[[j]][ssize, 4] = rho*MATS[[j]][ssize - 1, 4] + S[ssize]
      }

      S = rsgt(size,mu = 0, sigma = 1, lambda = 0, p = 1.4, q = 2.2, mean.cent = TRUE, var.adj = TRUE)
      MATS[[j]][1, 5] = median(S)
      for(ssize in 2:size){
        MATS[[j]][ssize, 5] = rho*MATS[[j]][ssize - 1, 5] + S[ssize]
      }
  }

i = 0
j = 0
T = size - 1
for (a in 1:3){
for(b in 1:ncol(param)){

  avgcubdev <- sum((MATS[[a]][, b] - mean(MATS[[a]][, b]))^3)/T
  skew = skewness(MATS[[a]][, b], na.rm = FALSE, type = 1)
  #MOMENT CONDITIONS
  mom1s <- MATS[[a]][, b] - mean(MATS[[a]][, b])
  mom2s <- mom1s^2 - var(MATS[[a]][, b])
  mom3s <- mom1s^3 - avgcubdev

  #NEWEY-WEST KERNEL ESTIMATION OF SKEWNESS ESTIMATOR ASYMPTOTIC VARIANCE
  momvec <- matrix(c(mom1s, mom2s, mom3s), ncol = 3)

  #Omega0
  omega0 <- cov(momvec)

  #Omega1
  omega1 <- matrix(0L, nrow = ncol(momvec), ncol = ncol(momvec))
  omega2 <- matrix(0L, nrow = ncol(momvec), ncol = ncol(momvec))
  #Truncation lag
  m <- 6
```

```
  #Computing sigmahat

  for(i in 1:ncol(momvec)){
    for(k in 1:ncol(momvec)){
      for(j in 1:m){
        w <- 1 - (j/(m+1)) #weights
        for(t in (j+1):T){
          omega1[i,k] <- omega1[i,k] + momvec[t,i]*momvec[(t-j),k] #VER OUTRA VEZ
        }
      }
    }
  }

  omega1 <- (1/T) * omega1
  omega2 = omega2 + w*(omega1 + t(omega1))
  sigmahat <- matrix(nrow = ncol(momvec), ncol = ncol(momvec))
  sigmahat = omega0 + omega2

  #VECTOR OF ESTIMATED DERIVATIVES
  ders <- ders <- matrix(c(-3/sqrt(var(MATS[[a]][, b])), -(3/2)*avgcubdev/(sqrt(var(MATS[[a]][, b]))^5),
  1/sqrt(var(MATS[[a]][, b]))), 3, 1)

  #H matrix

  r1 <- c(-1, 0, 0)
  r2 <- c(0, -1, 0)
  r3 <- c(-3*var(MATS[[a]][, b]), 0, -1)

  H <- rbind(r1, r2, r3)
  hinv <- solve(H)

  # Calculating asymptotic variance
  avars <- t(ders) %*% hinv %*% sigmahat %*% t(hinv) %*% ders
  avars = sqrt(avars/T)
  #lb = skew - qnorm(0.95, mean = 0, sd = 1, lower.tail = TRUE, log.p = FALSE) * avars
  #ub = skew + qnorm(0.95, mean = 0, sd = 1, lower.tail = TRUE, log.p = FALSE) * avars
  x = param[b]
  #c = between(x,lb,ub)
  test = (skew-x)/avars
  test2 = (skew-x)/sqrt(6/T)
  c = (abs(test) >= 1.96)
  d = (abs(test2) >= 1.96)
  MATFINAL[[a]][z,b] = c
  MATFINAL[[a+3]][z,b] = d
  #MATFINAL1[[a]][z,b] = ub - lb
}
}
}

sum(MATFINAL[[1]][,1])
sum(MATFINAL[[1]][,2])
sum(MATFINAL[[1]][,3])
sum(MATFINAL[[1]][,4])
sum(MATFINAL[[1]][,5])
```

```
#######################################TEST POWER########################################
#Asymmetric

a1 = skewsgt(0 , 1, -0.5, 1.7, 2.5)/(varsgt(0 , 1, -0.5, 1.7, 2.5)^(3/2))
a2 = skewsgt(0 , 1, -0.3, 1.7, 2.5)/(varsgt(0 , 1, -0.3, 1.7, 2.5)^(3/2))
a3 = skewsgt(0,1, 0.2, 1.7, 2.5)/(varsgt(0,1, 0.2, 1.7, 2.5)^(3/2))
a4 = skewsgt(0, 1, 0.3, 1.6, 2.6)/(varsgt(0, 1, 0.3, 1.6, 2.6)^(3/2))
a5 = skewsgt(0, 1, 0.5, 1.2, 4)/(varsgt(0, 1, 0.5, 1.2, 4)^(3/2))

param <- cbind(a1, a2, a3, a4, a5)
param

nsims = 200

MATFINAL <- lapply(1:6, matrix, data = NA, nrow = nsims, ncol = ncol(param))

for (z in 1:nsims){
  size = 1000 #+ 1
  MATS <- lapply(1:3, matrix, data = NA, nrow = size, ncol= 5)
  j = 0

  for (rho in c(0, 0.5, 0.8)){
    j = j + 1

    S = rsgt(size,mu = 0, sigma = 1, lambda = -0.5, p = 1.7, q = 2.5, mean.cent = TRUE, var.adj = TRUE)
    MATS[[j]][1, 1] = median(S)
    for(ssize in 2:size){
      MATS[[j]][ssize, 1] = rho*MATS[[j]][ssize - 1, 1] + S[ssize]
    }

    S = rsgt(size,mu = 0, sigma = 1, lambda = -0.3, p = 1.7, q = 2.5, mean.cent = TRUE, var.adj = TRUE)
    MATS[[j]][1, 2] = median(S)
    for(ssize in 2:size){
      MATS[[j]][ssize, 2] = rho*MATS[[j]][ssize - 1, 2] + S[ssize]
    }

    S = rsgt(size,mu = 0, sigma = 1, lambda = 0.2, p = 1.7 , q = 2.5, mean.cent = TRUE, var.adj = TRUE)
    MATS[[j]][1, 3] = median(S)
    for(ssize in 2:size){
      MATS[[j]][ssize, 3] = rho*MATS[[j]][ssize - 1, 3] + S[ssize]
    }

    S = rsgt(size,mu = 0, sigma = 1, lambda = 0.3, p = 1.6, q = 2.4, mean.cent = TRUE, var.adj = TRUE)
    MATS[[j]][1, 4] = median(S)
    for(ssize in 2:size){
      MATS[[j]][ssize, 4] = rho*MATS[[j]][ssize - 1, 4] + S[ssize]
    }

    S = rsgt(size,mu = 0, sigma = 1, lambda = 0.5, p = 1.5, q = 2.3, mean.cent = TRUE, var.adj = TRUE)
    MATS[[j]][1, 5] = median(S)
    for(ssize in 2:size){
      MATS[[j]][ssize, 5] = rho*MATS[[j]][ssize - 1, 5] + S[ssize]
    }
    }
```

```
#ESTIMATION

i = 0
j = 0

T = size - 1

for (a in 1:3){
  for(b in 1:(ncol(param))){

    avgcubdev <- sum((MATS[[a]][, b] - mean(MATS[[a]][, b]))^3)/T
    skew = skewness((MATS[[a]][, b]))

    #MOMENT CONDITIONS
    mom1s <- MATS[[a]][, b] - mean(MATS[[a]][, b])
    mom2s <- mom1s^2 - var(MATS[[a]][, b])
    mom3s <- mom1s^3 - avgcubdev

    #NEWEY-WEST KERNEL ESTIMATION OF SKEWNESS ESTIMATOR ASYMPTOTIC VARIANCE
    momvec <- matrix(c(mom1s, mom2s, mom3s), ncol = 3)

    #Omega0
    omega0 <- cov(momvec)

    #Omega1
    omega1 <- matrix(0L, nrow = ncol(momvec), ncol = ncol(momvec))
    omega2 <- matrix(0L, nrow = ncol(momvec), ncol = ncol(momvec))
    #Truncation lag
    m <- 6
    #Computing sigmahat

    for(i in 1:ncol(momvec)){
      for(k in 1:ncol(momvec)){
        for(j in 1:m){
          w <- 1 - (j/(m+1)) #weights
          for(t in (j+1):T){
            omega1[i,k] <- omega1[i,k] + momvec[t,i]*momvec[(t-j),k] #VER OUTRA VEZ
          }
        }
      }
    }

    omega1 <- (1/T) * omega1
    omega2 = omega2 + w*(omega1 + t(omega1))
    sigmahat <- matrix(nrow = ncol(momvec), ncol = ncol(momvec))
    sigmahat = omega0 + omega2
    #print(sigmahat)

    #VECTOR OF ESTIMATED DERIVATIVES
    ders <- ders <- matrix(c(-3/sqrt(var(MATS[[a]][, b])), -(3/2)*avgcubdev/(sqrt(var(MATS[[a]][, b]))^5),
    1/sqrt(var(MATS[[a]][, b]))), 3, 1)
    #print(ders)
```

```
      #H matrix

      r1 <- c(-1, 0, 0)
      r2 <- c(0, -1, 0)
      r3 <- c(-3*var(MATS[[a]][, b]), 0, -1)

      H <- rbind(r1, r2, r3)
      hinv <- solve(H)

 # Calculating asymptotic variance
      avars <- t(ders) %*% hinv %*% sigmahat %*% t(hinv) %*% ders
      avars = sqrt(avars/T)
      lb = skew - qnorm(0.95, mean = 0, sd = 1, lower.tail = TRUE, log.p = FALSE) * avars
      ub = skew + qnorm(0.95, mean = 0, sd = 1, lower.tail = TRUE, log.p = FALSE) * avars
      #x = kvec[b]
      x = 0
      #c = !(between(x,lb,ub))
      test = skew/avars
      test2 = skew/sqrt(6/T)
      c = (abs(test) >= 1.96)
      d = (abs(test2) >= 1.96)
      MATFINAL[[a]][z,b] = c
      MATFINAL[[a+3]][z,b] = d
      }
  }
}

sum(MATFINAL[[1]][,1])
sum(MATFINAL[[1]][,2])
sum(MATFINAL[[1]][,3])
sum(MATFINAL[[1]][,4])
sum(MATFINAL[[1]][,5])

#KURTOSIS

library(gld)
library(tidyverse)
library(e1071)
library(sgt)

######CALCULATION OF KURTOSIS FOR SKEWED GENERALIZED T-DISTRIBUTION##################
kurtsgt <- function(mu,sigma,lambda,p,q){
  v <- (q^(-1/p))*(1/(sqrt((((3*lambda^2)+1) * (beta(3/p,(q-2/p))/beta(1/p,q))
  - 4*(lambda^2) * ((beta(2/p, (q-1/p))/beta(1/p,q))^2)))))
  kurtsgtc = (((q^(4/p))*((v*sigma)^4))/beta(1/p,q)^4)*((-48*(lambda^4))*(beta(2/p,q-(1/p))^4)
  + (24*(lambda^2))*(1+3*(lambda^2))*beta(1/p,q)*(beta(2/p,q-(1/p))^2)*beta(3/p,q-(2/p))
  - (32*(lambda^2)) * (1 + lambda^2) * (beta(1/p,q)^2) * beta(2/p,q-(1/p)) * beta(4/p,q-(3/p))
  + (1+10*(lambda^2)+ (5*lambda^4))*( beta(1/p,q)^3) * beta(5/p,q-(4/p) ) )
  return(kurtsgtc)
}

varsgt <- function(mu,sigma,lambda,p,q){
  v <- (q^(-1/p))*(1/(sqrt((((3*lambda^2)+1) * (beta(3/p,q-(2/p))/beta(1/p,q))
  -4*(lambda^2)*((beta(2/p, q-(1/p))/beta(1/p,q))^2)))))
  varsgtc = (((v*sigma)^2)*q^(2/p)) * ((3*lambda^2 + 1)*(beta(3/p,q-(2/p))/beta(1/p,q))
  -4*(lambda^2) * ((beta(2/p,q-(1/p))/beta(1/p,q))^2))
  return(varsgtc)
}
```

```
#ESTIMATION

i = 0
j = 0

T = size

for (a in 1:3){
  for(b in 1:ncol(param)){

    avgcubdev <- sum((MATS[[a]][, b] - mean(MATS[[a]][, b]))^3)/T
    avgfdev <- sum((MATS[[a]][, b] - mean(MATS[[a]][, b]))^4)/T

    #skew = skewness((MATS[[a]][, b]))
    kurt = kurtosis((MATS[[a]][, b]), na.rm = FALSE, type = 1) + 3

    #MOMENT CONDITIONS
    mom1k <- MATS[[a]][, b] - mean(MATS[[a]][, b])
    mom2k <- mom1k^2 - var(MATS[[a]][, b])
    mom3k <- mom1k^4 - avgfdev

    #NEWEY-WEST KERNEL ESTIMATION OF SKEWNESS ESTIMATOR ASYMPTOTIC VARIANCE
    momvec <- matrix(c(mom1k, mom2k, mom3k), ncol = 3)

    #Omega0
    omega0 <- cov(momvec)

    #Omega1
    omega1 <- matrix(0L, nrow = ncol(momvec), ncol = ncol(momvec))
    omega2 <- matrix(0L, nrow = ncol(momvec), ncol = ncol(momvec))
    #Truncation lag
    m <- 6
    #Computing sigmahat

    for(i in 1:ncol(momvec)){
      for(k in 1:ncol(momvec)){
        for(j in 1:m){
          w <- 1 - (j/(m+1)) #weights
          for(t in (j+1):T){
            omega1[i,k] <- omega1[i,k] + momvec[t,i]*momvec[(t-j),k]
          }
        }
      }
    }

    omega1 <- (1/T) * omega1
    omega2 = omega2 + w*(omega1 + t(omega1))
    sigmahat <- matrix(nrow = ncol(momvec), ncol = ncol(momvec))
    sigmahat = omega0 + omega2
    #print(sigmahat)

    # Vector of estimated derivatives

    ders <- matrix(c(-4*avgcubdev/(sqrt(var(MATS[[a]][, b]))^4), -2*avgfdev/(sqrt(var(MATS[[a]][, b]))^6),
    1/(sqrt(var(MATS[[a]][, b]))^4)), 3, 1)
    #print(ders)
```

```
      #H matrix

      r1 <- c(-1, 0, 0)
      r2 <- c(0, -1, 0)
      r3 <- c(-4*avgcubdev, 0, -1)

      H <- rbind(r1, r2, r3)
      hinv <- solve(H)

      # Calculating asymptotic variance
      avark <- t(ders) %*% hinv %*% sigmahat %*% t(hinv) %*% ders
      avark = sqrt(avark/T)

      #lb = kurt - qnorm(0.95, mean = 0, sd = 1, lower.tail = TRUE, log.p = FALSE) * avark
      #ub = kurt + qnorm(0.95, mean = 0, sd = 1, lower.tail = TRUE, log.p = FALSE) * avark

      x = param[,b]
      #c = between(x,lb,ub)
      test1 = (kurt-x)/avark
      test2 = (kurt-x)/(sqrt(24/T))
      #print((kurt-x))
      #print(avark)
      c = (abs(test1) >= 1.96)
      d = (abs(test2) >= 1.96)
      MATFINAL[[a]][z,b] = c
      MATFINAL[[a+3]][z,b] = d
      #MATFINAL1[[a]][z,b] = ub - lb
    }
  }
}

sum(MATFINAL[[1]][,1])
sum(MATFINAL[[1]][,2])
sum(MATFINAL[[1]][,3])
sum(MATFINAL[[1]][,4])
sum(MATFINAL[[1]][,5])
sum(MATFINAL[[1]][,6])
sum(MATFINAL[[1]][,7])
sum(MATFINAL[[1]][,8])
sum(MATFINAL[[1]][,9])
sum(MATFINAL[[1]][,10])
```