

Thi Huong An Nguyen; Ruiz-Gazen, Anne; Thomas-Agnan, Christine; Laurent, Thibault

## Article

# Multivariate student versus multivariate Gaussian regression models with application to finance

Journal of Risk and Financial Management

## Provided in Cooperation with:

MDPI – Multidisciplinary Digital Publishing Institute, Basel

**Suggested Citation:** Thi Huong An Nguyen; Ruiz-Gazen, Anne; Thomas-Agnan, Christine; Laurent, Thibault (2019) : Multivariate student versus multivariate Gaussian regression models with application to finance, Journal of Risk and Financial Management, ISSN 1911-8074, MDPI, Basel, Vol. 12, Iss. 1, pp. 1-21,  
<https://doi.org/10.3390/jrfm12010028>

This Version is available at:

<https://hdl.handle.net/10419/239029>

## Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

## Terms of use:

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



<https://creativecommons.org/licenses/by/4.0/>

## Article

# Multivariate Student versus Multivariate Gaussian Regression Models with Application to Finance

Thi Huong An Nguyen <sup>1,2,†</sup>, Anne Ruiz-Gazen <sup>1,\*,†</sup>, Christine Thomas-Agnan <sup>1,†</sup>  
and Thibault Laurent <sup>3,†</sup>

<sup>1</sup> Toulouse School of Economics, University of Toulouse Capitole, 21 allée de Brienne, 31000 Toulouse, France; huongan.nguyen@tse-fr.eu (T.H.A.N.); christine.thomas@tse-fr.eu (C.T.-A.)

<sup>2</sup> Department of Economics, DaNang Architecture University, Da Nang 550000, Vietnam

<sup>3</sup> Toulouse School of Economics, CNRS, University of Toulouse Capitole, 31000 Toulouse, France; thibault.laurent@tse-fr.eu

\* Correspondence: anne.ruiz-gazen@tse-fr.eu; Tel.: +33-561-12-8767

† These authors contributed equally to this work.

Received: 29 December 2018; Accepted: 31 January 2019; Published: 9 February 2019



**Abstract:** To model multivariate, possibly heavy-tailed data, we compare the multivariate normal model (N) with two versions of the multivariate Student model: the independent multivariate Student (IT) and the uncorrelated multivariate Student (UT). After recalling some facts about these distributions and models, known but scattered in the literature, we prove that the maximum likelihood estimator of the covariance matrix in the UT model is asymptotically biased and propose an unbiased version. We provide implementation details for an iterative reweighted algorithm to compute the maximum likelihood estimators of the parameters of the IT model. We present a simulation study to compare the bias and root mean squared error of the ensuing estimators of the regression coefficients and covariance matrix under several scenarios of the potential data-generating process, misspecified or not. We propose a graphical tool and a test based on the Mahalanobis distance to guide the choice between the competing models. We also present an application to model vectors of financial assets returns.

**Keywords:** multivariate regression models; heavy-tailed data; Mahalanobis distances; maximum likelihood estimator; independent multivariate Student distribution; uncorrelated multivariate Student distribution

## 1. Introduction

Many applications involving models for multivariate data underline the limitations of the classical multivariate Gaussian model, mainly due to its inability to model heavy tails. It is then natural to turn attention to a more flexible family of distributions, for example the multivariate Student distribution.

In one dimension, the generalized Student distribution encompasses the Gaussian distribution as a limit when the number of degrees of freedom or shape parameter tends to infinity, allowing for heavier tails when the shape parameter is small. As we will see, a first difficulty in higher dimensions is that there are several kinds of multivariate Student distributions; see for example [Johnson and Kotz \(1972\)](#) and more recently [Kotz and Nadarajah \(2004\)](#). A nice summary of the properties of the multivariate Student distribution that we will use later on in this paper, and its comparison with the Gaussian multivariate, can be found in [Roth \(2013\)](#).

Before going further, let us mention that it is not so easy to have a clear overview of the results in terms of Student regression models for at least three reasons. The first reason is that this topic is scattered, with some papers in the statistical literature and others in the econometrics literature,

sometimes without cross-referencing. The second reason is that the word “multivariate” is sometimes misleading since, as we will see, the multivariate Student is used to define a univariate regression model. At last, the distinction between models UT and IT (see below) is not always clearly announced in the papers. Other miscellaneous reasons are that some authors just fit the distribution without covariates and finally that some authors consider the degrees of freedom as fixed, whereas others estimate it. Our first purpose here is to lead the reader through this literature and gather the results concerning the maximum likelihood estimators of the parameters in the multivariate UT and IT models with a common notation. In the present paper, we consider a multivariate dependent vector and a linear regression model with different assumptions on the error term distribution. The most common and convenient assumption is the Gaussian distribution. For a Gaussian vector, the assumption of independent coordinates is equivalent to the assumption of uncorrelated coordinates. Such an equivalence is no longer true when considering a multivariate Student distribution. We thus consider two cases: uncorrelated (UT) on the one hand and independent Student (IT) error vectors on the other hand.

The purpose of this paper is to contribute to the UT and IT models as well as to their comparisons. First of all, for the UT model, we extend to the multivariate case the results of Zellner (1976) for the derivation of the maximum likelihood estimators and Zellner’s formula (Zellner (1976)) for the bias of the covariance matrix estimator, and we prove that it does not vanish asymptotically. For the multivariate IT model, in the same spirit as Lange and Sinsheimer (1993), we provide details for the implementation of an iterative reweighted algorithm to compute the maximum likelihood estimators of the parameters. We devise a simulation study to measure the impact of misspecification on the bias, variance, and mean squared error of these different parameters’ estimates under several data-generating processes (Gaussian, UT, and IT) and try to answer the question: what are the consequences of a wrong specification? Finally we introduce a new procedure for model selection based on the knowledge of the distribution of the Mahalanobis distances under the different data-generating processes (DGP).

One application attracted our attention in the finance literature. The work in Platen and Rendek (2008) identified the Student distribution with between three and five degrees of freedom, with a concentration around four, as the typical distribution for modeling the distribution of log-returns of world stock indices. They embedded the Student  $t$  in the class of generalized hyperbolic distributions, itself a subclass of the normal/independent family. For bivariate returns, the work in Fung and Seneta (2010) compared a multivariate Student IT model with an alternative model obtained by a more complex mixing representation from the point of view of asymptotic tail dependence. The work in Hu and Kercheval (2009) insisted on the fact that the choice of distribution matters when optimizing the portfolio. They found that the Student UT model performs the best in the class of symmetric generalized hyperbolic distributions. The work in Kan and Zhou (2017) advocated using a multivariate IT model for fitting the joint distribution of stock returns for a few fixed values of the degrees of freedom parameter and showed that this model outperforms the multivariate Gaussian.

In Section 2, after recalling the univariate results, we extend the results of Zellner (1976) for the derivation of the maximum likelihood estimators and its properties in the UT model and propose an iterative implementation for the IT model. We present the results of the simulation study in Section 3 and of the model selection strategy in Section 4 using a toy example and a dataset from finance. Section 5 summarizes the findings and gives recommendations.

## 2. Multivariate Regression Models

### 2.1. Literature Review

In order to define a Student regression model, even in the univariate case (single dependent variable), one needs to use the multivariate Student distribution to describe the joint distribution of the vector of observations for the set of statistical units. There are mainly two options, which were

described in [Kelejian and Prucha \(1985\)](#) for the case of univariate regression. Indeed, the property of the equivalence between the independence and uncorrelatedness for components of a Gaussian vector are not satisfied anymore for a multivariate Student vector. One option, which we will call the IT model (for independent  $t$ -distribution) in the sequel, considers that the components of the random disturbance vector of the regression model are independent with the same marginal Student distribution. The second option, which we will call the UT model (for uncorrelated  $t$ -distribution), postulates a joint multivariate Student distribution for the vector of disturbances. Note that in both models, the marginal distribution of each component still is Student univariate.

The work in [Zellner \(1976\)](#) introduced a univariate Student regression model of the type UT with known degrees of freedom and studied the corresponding maximum likelihood and Bayesian estimators (with some adapted priors). The work in [Singh \(1988\)](#) considered the case of univariate Student regression with the UT model and with unknown degrees of freedom and derived an estimator of the degrees of freedom and subsequent estimators of the other parameters. However, [Fernandez and Steel \(1999\)](#) showed that this estimator was not consistent. Using one possible representation of the multivariate Student distribution, [Lange and Sinsheimer \(1993\)](#) embedded univariate Student regression with the UT model in a larger family of regression models (with normal/independent error distributions) and developed EM algorithms to compute their maximum likelihood estimates, as in [Dempster et al. \(1978\)](#).

In the framework of the spherical error distribution, which includes the Student error model as a special case, the work in [Fraser and Ng Kai \(1980\)](#) proved an extension to the multivariate case of Zellner's result stating that inference about the parameters corresponds closely to that under normal theory. Motivated by a financial application, the work in [Sutradhar and Ali \(1986\)](#) used a multivariate UT Student regression model with moment estimators instead of maximum likelihood and allowing the degrees of freedom to be unknown.

The univariate IT model was introduced in [Fraser \(1979\)](#) and compared to the UT model in [Kelejian and Prucha \(1985\)](#).

Concerning multivariate IT Student distributions, there was first a collection of results or applications for the case without regressors. The work in [McNeil et al. \(2005\)](#) used a representation of the multivariate IT Student distribution to derive an algorithm of the EM type for computing the maximum likelihood parameter estimators. They used the framework of normal mixture distributions in which the Student distribution can be expressed as a combination of a Gaussian random variable and an inverse gamma random variable. More recently, the work in [Dogru et al. \(2018\)](#) proposed a more robust extension, replacing maximum likelihood by a kind of M-estimation method based on the minimization of a  $q$ -entropy criterion. For the multivariate Student IT model, the work in [Prucha and Kelejian \(1984\)](#) derived the normal equations for the maximum likelihood estimators and their asymptotic properties with known degrees of freedom in a framework that encompasses our multivariate Student regression case. The work in [Lange et al. \(1989\)](#) illustrated this multivariate IT model on several examples. The work in [Lange and Sinsheimer \(1993\)](#) considered the framework of normal/independent error distributions (same as normal variance mixtures) and derived the EM algorithm for the maximum likelihood estimators in a model with covariates. The works in [Liu and Rubin \(1995\)](#) and [Liu \(1997\)](#) developed extensions of the EM algorithm for the multivariate IT model with known or unknown degrees of freedom, with or without covariates and with or without missing data. The work in [Katz and King \(1999\)](#) fit a multivariate IT distribution to multiparty electoral data. The work in [Fernandez and Steel \(1999\)](#) attracted attention to the fact that maximum likelihood inference can encounter problems of unbounded likelihood when the number of degrees of freedom is considered unknown and has to be estimated. Before engaging in the use of the multivariate Student distribution, it is wise to read [Hofert \(2003\)](#), which explained some traps to be avoided. One difficulty indeed is to be aware that some authors parametrize the multivariate Student distribution using the covariance matrix, while others use the scatter matrix, sometimes with the same notation for either one.

We consider the following version of the Student  $p$ -multivariate distribution denoted by  $T_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$  with  $\boldsymbol{\mu}$  being the  $p$ -vector of means,  $\boldsymbol{\Sigma}$  being the  $p \times p$  covariance matrix, and  $\nu > 2$  the degrees of freedom. It is defined, for a  $p$ -vector  $\mathbf{z}$ , by the probability density function:

$$p(\mathbf{z}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) = \frac{f(\nu)}{\det(\boldsymbol{\Sigma})^{1/2}} \left[ 1 + \frac{1}{\nu-2} (\mathbf{z} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{z} - \boldsymbol{\mu}) \right]^{-(\nu+p)/2}, \quad (1)$$

where  $^T$  denotes the transpose operator,  $f(\nu) = \frac{\Gamma[(\nu+p)/2]}{\Gamma(\nu/2)(\nu-2)^{p/2}\pi^{p/2}}$  and  $\Gamma$  is the usual Gamma function.

Note that the assumption  $\nu > 2$  implies the existence of the first two moments of the distribution and that the above density function is parametrized in terms of the covariance matrix. In most of the literature on multivariate Student distributions, the density is rather parametrized as a function of the scatter matrix  $((\nu-2)/\nu)\boldsymbol{\Sigma}$ . Using the covariance matrix parametrization facilitates the comparison with the Gaussian distribution. We first recall some results in the univariate regression context.

## 2.2. Univariate Regression Case Reminder

In the univariate regression case and for a sample of size  $n$ , we have a one-dimensional dependent variable  $\mathcal{Y}_i, i = 1, \dots, n$ , whose values are stacked in a vector  $\mathcal{Y}$ , and  $K$  explanatory variables defining a  $n \times (K+1)$  design matrix  $\mathcal{X}$  including the constant.

The regression model is written as  $\mathcal{Y} = \mathcal{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ , where  $\boldsymbol{\beta} = (\beta_0, \dots, \beta_K)^T$  is a  $(K+1)$ -dimensional vector of parameters and the error term  $\boldsymbol{\epsilon} = (\epsilon_1, \dots, \epsilon_n)^T$  is an  $n$ -dimensional vector. If we consider that the design matrix is fixed with rank  $K+1$  or look at the distribution of  $\boldsymbol{\epsilon}$  conditional on  $\mathcal{X}$ , the usual assumptions are the following. The errors  $\epsilon_i, i = 1, \dots, n$ , are independent and identically distributed (i.i.d.) with expectation zero and equal variance  $\sigma^2$ . In this context, it is well known that the least squares estimator of  $\boldsymbol{\beta}$  is equal to:

$$\hat{\boldsymbol{\beta}} = (\mathcal{X}^T \mathcal{X})^{-1} \mathcal{X}^T \mathcal{Y} \quad (2)$$

while the classical  $\sigma^2$  estimator is  $\hat{\sigma}^2 = \hat{\boldsymbol{\epsilon}}^T \hat{\boldsymbol{\epsilon}} / (n - K - 1)$  where  $\hat{\boldsymbol{\epsilon}} = \mathcal{Y} - \mathcal{X}\hat{\boldsymbol{\beta}}$ . These estimators are unbiased. In the case of a Gaussian error distribution, the estimator  $\hat{\boldsymbol{\beta}}$  coincides with the maximum likelihood estimator of  $\boldsymbol{\beta}$ , while the maximum likelihood estimator of  $\sigma^2$  is equal to  $\hat{\sigma}^2$  multiplied by  $(n - K - 1)/n$  and is only asymptotically unbiased. In the Gaussian case, there is an equivalence between the  $\epsilon_i$  being independent or uncorrelated. However, this property is no longer true for a Student distribution. This means that one should distinguish the case of uncorrelated errors from the case of independent errors. The case where the errors  $\epsilon_i, i = 1, \dots, n$ , follow a joint  $n$ -dimensional Student distribution with diagonal covariance matrix and equal variance is called the UT model, and its coordinates are uncorrelated, but not independent. Interestingly, the maximum likelihood method for the UT model with known degrees of freedom leads to the least squares estimator (2) of  $\boldsymbol{\beta}$  (Zellner (1976)). This property is true for more general distributions as long as the likelihood is a decreasing function of  $\boldsymbol{\epsilon}^T \boldsymbol{\epsilon}$ . Concerning the error variance, the maximum likelihood estimator is  $(n - K - 1)\nu \hat{\sigma}^2 / (n(\nu - 2))$  and is biased even asymptotically Zellner (1976). For the independent case, we assume that the errors  $\epsilon_i, i = 1, \dots, n$ , are i.i.d. with a Student univariate distribution and known degrees of freedom. The maximum likelihood estimators belong to the class of M-estimators, which are studied in detail in Chapter 7 of Huber and Ronchetti (2009). These estimators are defined through implicit equations and can be computed using an iterative reweighted algorithm.

In what follows, we consider the case of a multivariate dependent variable and propose to gather and complete the results from the literature. As we will see, the results derived in the multivariate case are very similar to their univariate counterpart. In particular, the maximum likelihood estimator of the error covariance matrix is biased for the uncorrelated Student model, while there is a need to define an iterative algorithm for the independent Student model.

### 2.3. The Multivariate Regression Model

Let us consider a sample of size  $n$ , and for  $i = 1, \dots, n$ , let us denote the  $L$ -dimensional dependent vector by:

$$\mathcal{Y}_i = (y_{i1}, \dots, y_{iL})^T.$$

For  $K$  explanatory variables, the design matrix is of size  $L \times (K + 1)L$  and is given by:

$$\mathcal{X}_i = \mathbf{I}_L \otimes \mathbf{x}_i^T$$

for  $i = 1, \dots, n$ , with the  $(K + 1)$ -vector  $\mathbf{x}_i = (1, x_{i1}, \dots, x_{iK})^T$ ,  $\mathbf{I}_L$  the identity matrix with dimension  $L$ , and  $\otimes$  the usual Kronecker product. The parameter of interest is a  $(K + 1)L$  vector given by:

$$\boldsymbol{\beta} = (\boldsymbol{\beta}_1^T, \dots, \boldsymbol{\beta}_L^T)^T,$$

where  $\boldsymbol{\beta}_j = (\beta_{0j}, \dots, \beta_{Kj})^T$ , for  $j = 1, \dots, L$ , and the  $L$ -vector of errors is denoted by:

$$\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \dots, \epsilon_{iL})^T$$

for  $i = 1, \dots, n$ . We consider the linear model:

$$\mathcal{Y}_i = \mathcal{X}_i \boldsymbol{\beta} + \boldsymbol{\epsilon}_i \quad (3)$$

with  $\mathbb{E}(\boldsymbol{\epsilon}_i) = 0$  and  $i = 1, \dots, n$ . Using matrix notations, we can write Model (3) as:

$$\mathcal{Y} = \mathcal{X} \boldsymbol{\beta} + \boldsymbol{\epsilon} \quad (4)$$

with the  $nL$ -vectors:

$$\begin{aligned} \mathcal{Y} &= (\mathcal{Y}_1^T, \dots, \mathcal{Y}_n^T)^T, \\ \boldsymbol{\epsilon} &= (\boldsymbol{\epsilon}_1^T, \dots, \boldsymbol{\epsilon}_n^T)^T \end{aligned}$$

and the  $nL \times (K + 1)L$  matrix:

$$\mathcal{X} = (\mathcal{X}_1^T, \dots, \mathcal{X}_n^T)^T.$$

In what follows, we make different assumptions on the distribution of  $\boldsymbol{\epsilon}$  and recall (for Gaussian and IT) or derive (for UT) the maximum likelihood estimators of the parameter  $\boldsymbol{\beta}$  and of the covariance matrix of  $\boldsymbol{\epsilon}$ .

### 2.4. Multivariate Normal Error Vector

Let us first consider Model (4) with independent and identically distributed error vectors  $\boldsymbol{\epsilon}_i$ ,  $i = 1, \dots, n$ , following a multivariate normal distribution  $\mathcal{N}_L(\mathbf{0}, \boldsymbol{\Sigma})$  with an  $L$ -vector of means equal to zero and an  $L \times L$  covariance matrix  $\boldsymbol{\Sigma}$ . This model is denoted by  $N$ , and the subscript  $N$  is used to denote the error terms  $\boldsymbol{\epsilon}_{Ni}$ ,  $i = 1, \dots, n$ , and the parameters  $\boldsymbol{\beta}_N$  and  $\boldsymbol{\Sigma}_N$  of the model. The maximum likelihood estimators of  $\boldsymbol{\beta}_N$  and  $\boldsymbol{\Sigma}_N$  are:

$$\hat{\boldsymbol{\beta}}_N = (\mathcal{X}^T \mathcal{X})^{-1} \mathcal{X}^T \mathcal{Y}, \quad (5)$$

$$\hat{\boldsymbol{\Sigma}}_N = \frac{\sum_{i=1}^n \hat{\boldsymbol{\epsilon}}_{Ni} \hat{\boldsymbol{\epsilon}}_{Ni}^T}{n}, \quad (6)$$

where  $\hat{\boldsymbol{\epsilon}}_{Ni} = \mathcal{Y}_i - \mathcal{X}_i \hat{\boldsymbol{\beta}}_N$  (see, e.g., Theorem 8.4 from [Seber \(2008\)](#)).

The estimator  $\hat{\boldsymbol{\beta}}_N$  is an unbiased estimator of  $\boldsymbol{\beta}_N$  while the bias of  $\hat{\boldsymbol{\Sigma}}_N$  is equal to  $-((K + 1)/n)\boldsymbol{\Sigma}_N$  and tends to zero when  $n$  tends to infinity (see, e.g., Theorems 8.1 and 8.2 from [Seber \(2008\)](#)).

For data such as financial data, it is well known that the Gaussian distribution does not fit the error term well. Student distributions are known to be more appropriate because they have heavier



tails than the Gaussian. As for the univariate case, for Student distributions, the independence of coordinates is not equivalent to their uncorrelatedness, and we consider below two types of Student distributions for the error term. In Section 2.5, the error vector  $\epsilon$  is assumed to follow a Student distribution with  $nL$  dimensions and a particular block diagonal covariance matrix. More precisely, we assume that the error vectors  $\epsilon_i, i = 1, \dots, n$ , are identically distributed and uncorrelated but are not independent. In Section 2.6, however, we consider independent and identically distributed error vectors  $\epsilon_i, i = 1, \dots, n$ , with an  $L$ -dimensional Student distribution.

### 2.5. Uncorrelated Multivariate Student (UT) Error Vector

Let us consider Model (4) with uncorrelated and identically distributed error vectors  $\epsilon_i, i = 1, \dots, n$ , such that the vector  $\epsilon$  follows a multivariate Student distribution  $T_{nL}(\mathbf{0}, \mathbf{\Omega}, \nu)$  with known degrees of freedom  $\nu > 2$  and covariance matrix  $\mathbf{\Omega} = \mathbf{I}_n \otimes \mathbf{\Sigma}$ . The  $L \times L$  matrix  $\mathbf{\Sigma}$  is the common covariance matrix of the  $\epsilon_i, i = 1, \dots, n$ . This model is denoted by UT, and the subscript UT is used to denote the error terms  $\epsilon_{UTi}, i = 1, \dots, n$ , and the parameters  $\beta_{UT}, \mathbf{\Omega}_{UT}$ , and  $\mathbf{\Sigma}_{UT}$  of the model. This model generalizes the model proposed by Zellner (1976) to the case of multivariate  $\epsilon_i$ s. We derive the maximum likelihood estimators of  $\beta_{UT}$  and  $\mathbf{\Sigma}_{UT}$  in Proposition 1 and give the bias of the covariance estimator in Proposition 2. The proofs of the propositions are given in the Appendix A.

**Proposition 1.** *The maximum likelihood estimators of  $\beta_{UT}$  and  $\mathbf{\Sigma}_{UT}$  are given by:*

$$\begin{aligned}\hat{\beta}_{UT} &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}, \\ \hat{\mathbf{\Sigma}}_{UT} &= \frac{\nu}{\nu - 2} \frac{\sum_{i=1}^n \hat{\epsilon}_{UTi} \hat{\epsilon}_{UTi}^T}{n},\end{aligned}\quad (7)$$

where  $\hat{\epsilon}_{UTi} = \mathbf{Y}_i - \mathbf{X}_i \hat{\beta}_{UT}$ .

The next proposition gives the bias of the maximum likelihood estimators and generalizes Zellner's result (Zellner (1976), p. 402) to the multivariate UT model. The maximum likelihood estimator of  $\beta_{UT}$  coincides with the least squares and with the method of moment estimators and is unbiased. This is no longer the case for the maximum likelihood estimator of  $\mathbf{\Sigma}_{UT}$ , which is biased even asymptotically. This gives an example of a maximum likelihood estimator that is not asymptotically unbiased in a context where the random variables are not independent. It illustrates that the independence assumption is crucial to derive the usual properties of the maximum likelihood estimators. Note that the method of moments estimator is a consistent estimator of  $\mathbf{\Sigma}_{UT}$  (see Sutradhar and Ali (1986)).

**Proposition 2.** *The estimator  $\hat{\beta}_{UT}$  is unbiased for  $\beta_{UT}$ . The estimator  $\hat{\mathbf{\Sigma}}_{UT}$  is biased for  $\mathbf{\Sigma}_{UT}$  even asymptotically. More precisely,*

$$\mathbb{E}(\hat{\mathbf{\Sigma}}_{UT}) = \frac{n-K}{n} \frac{\nu}{\nu-2} \mathbf{\Sigma}_{UT}$$

A consequence of Proposition 2 is that an asymptotically unbiased estimator of  $\mathbf{\Sigma}_{UT}$  is given by

$$\tilde{\mathbf{\Sigma}}_{UT} = \sum_{i=1}^n \hat{\epsilon}_{UTi} \hat{\epsilon}_{UTi}^T / n.$$

### 2.6. Independent Multivariate Student Error Vector

Let us consider Model (4) using the notations of Section 2.3 with i.i.d.  $\epsilon_i, i = 1, \dots, n$ , following a Student distribution with  $L$  dimensions and known degrees of freedom  $\nu > 2$ . We denote this model by IT and the parameters of the model by  $\beta_{IT}$  and  $\mathbf{\Sigma}_{IT}$ . The IT model is a particular case of Prucha and Kelejian (1984) where the  $B$  matrix in Expression (2.1) in Prucha and Kelejian (1984) is equal to zero.

Following [Prucha and Kelejian \(1984\)](#), we derive the maximum likelihood estimators for the IT model.

**Proposition 3.** *The maximum likelihood estimators of  $\beta_{IT}$  and  $\Sigma_{IT}$  in the IT regression model satisfy the following implicit equations:*

$$\begin{aligned}\hat{\beta}_{IT} &= \left( \sum_{i=1}^n \hat{w}_{ITi} \mathbf{x}_i^T \hat{\Sigma}_{IT}^{-1} \mathbf{x}_i \right)^{-1} \sum_{i=1}^n \hat{w}_{ITi} \mathbf{x}_i^T \hat{\Sigma}_{IT}^{-1} \mathbf{y}_i \\ \hat{\Sigma}_{IT} &= \frac{1}{n} \sum_{i=1}^n \hat{w}_{ITi} \hat{\epsilon}_{ITi} \hat{\epsilon}_{ITi}^T\end{aligned}\quad (8)$$

$$\text{with: } \hat{\epsilon}_{ITi} = \mathbf{y}_i - \mathbf{x}_i \hat{\beta}_{IT} \text{ and } \hat{w}_{ITi} = \frac{\nu + L}{\nu - 2 + \hat{\epsilon}_{ITi}^T \hat{\Sigma}_{IT}^{-1} \hat{\epsilon}_{ITi}}.$$

These estimators are consistent estimators of  $\beta_{IT}$  and  $\Sigma_{IT}$  (see Theorem 3.2 in [Prucha and Kelejian \(1984\)](#)). In order to compute them, we propose to implement the following iterative reweighted algorithm in the same spirit as in [Huber and Ronchetti \(2009\)](#) for the univariate case (see also [Lange et al. \(1989\)](#)).

Step 0: Let:

$$\begin{aligned}\hat{\beta}_{IT}^{(0)} &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \\ \hat{\epsilon}_{IT}^{(0)} &= \mathbf{y} - \mathbf{X} \hat{\beta}_{IT}^{(0)} \\ \hat{\Sigma}_{IT}^{(0)} &= \frac{1}{n} \sum_{i=1}^n \hat{\epsilon}_{ITi}^{(0)} \hat{\epsilon}_{ITi}^{(0)T}\end{aligned}$$

Step  $k \rightarrow$  Step  $(k+1)$ ,  $k > 0$ :

$$\begin{aligned}\hat{w}_{ITi}^{(k+1)} &= \frac{\nu + L}{\nu - 2 + \hat{\epsilon}_{ITi}^{(k)T} \hat{\Sigma}_{IT}^{(k)-1} \hat{\epsilon}_{ITi}^{(k)}} \\ \hat{\beta}_{IT}^{(k+1)} &= \left( \sum_{i=1}^n \hat{w}_{ITi}^{(k+1)} \mathbf{x}_i^T \hat{\Sigma}_{IT}^{(k)-1} \mathbf{x}_i \right)^{-1} \sum_{i=1}^n \hat{w}_{ITi}^{(k+1)} \mathbf{x}_i^T \hat{\Sigma}_{IT}^{(k)-1} \mathbf{y}_i \\ \hat{\epsilon}_{IT}^{(k+1)} &= \mathbf{y} - \mathbf{X} \hat{\beta}_{IT}^{(k+1)} \\ \hat{\Sigma}_{IT}^{(k+1)} &= \frac{1}{n} \sum_{i=1}^n \hat{w}_{ITi}^{(k+1)} \hat{\epsilon}_{ITi}^{(k+1)} \hat{\epsilon}_{ITi}^{(k+1)T}\end{aligned}$$

The process is iterated until convergence. Note that this algorithm is given in detail in Section 7.8 of [Huber and Ronchetti \(2009\)](#) for a general class of univariate regression M-estimators. It is also sometimes called IRLS for iteratively-reweighted least squares and can be seen as a particular case of the EM algorithm ([Dempster et al. \(1978\)](#)).

Table 1 gathers the likelihoods and thus summarizes the three models of interest.



**Table 1.** Distribution of the error vector  $\epsilon$  in the Gaussian, UT, and IT models.

Model	Distribution
$N(\epsilon_1, \dots, \epsilon_n)$	$\mathcal{N}_{nL}(\mathbf{0}, \mathbf{I}_n \otimes \Sigma_N) = \prod_{i=1}^n \mathcal{N}_L(\mathbf{0}, \Sigma_N)$
$UT(\epsilon_1, \dots, \epsilon_n)$	$T_{nL}(\mathbf{0}, \mathbf{I}_n \otimes \Sigma_{UT}, \nu)$
$IT(\epsilon_1, \dots, \epsilon_n)$	$\prod_{i=1}^n T_L(\mathbf{0}, \Sigma_{IT}, \nu)$

### 3. Simulation Study

#### 3.1. Design

This study aims at comparing the properties of the estimators of  $\beta$  and  $\Sigma$  as defined in the previous section for the multivariate Gaussian (N), the uncorrelated multivariate Student (UT), and the independent multivariate Student (IT) error distributions, under several scenarios for the DGP. Note that for the UT model, we used the asymptotically unbiased estimator  $\tilde{\Sigma}_{UT}$  to estimate  $\Sigma_{UT}$ . We considered a variety of degrees of freedom  $\nu_{DGP}$  for the Student IT and UT models with a focus on values between three and five. We used the function `rmvt` from the R package `mvnfast` to simulate the Student distributions. For a sample size  $n = 1000$  and a number of replications  $N = 10,000$ , we simulated an explanatory variable  $\mathcal{X}$  following a Gaussian distribution  $\mathcal{N}(45, 10)$ . The parameter vector  $\beta$  and the covariance matrix  $\Sigma$  are respectively chosen to be:

$$\beta = \begin{bmatrix} \beta_{01} \\ \beta_{11} \\ \beta_{02} \\ \beta_{12} \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \\ 4 \\ -3 \end{bmatrix}; \Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} = \begin{bmatrix} 2 & 0.5 \\ 0.5 & 1 \end{bmatrix}.$$

Note that similar results are obtained with other choices of parameters.

For each DGP, we calculate a number of Monte Carlo performance measures of the estimators proposed in Section 2. The performances are measured by the Monte Carlo relative bias (RB) and the mean squared error (MSE), which are defined for an estimator  $\hat{\theta}$  of a parameter  $\theta$  by:

$$\begin{aligned} \text{Bias}(\hat{\theta}) &= \frac{1}{N} \sum_{i=1}^N \hat{\theta}^{(i)} - \theta \\ \text{RB}(\hat{\theta}) &= 100 \frac{\text{Bias}(\hat{\theta})}{\theta} \\ \text{MSE}(\hat{\theta}) &= \frac{1}{N} \sum_{i=1}^N (\hat{\theta}^{(i)} - \theta)^2. \end{aligned} \quad (9)$$

We also compute a relative root mean squared error (RRMSE) with respect to a baseline estimator  $\tilde{\theta}$  as:

$$\text{RRMSE}(\hat{\theta}) = \left( \frac{\text{MSE}(\hat{\theta})}{\text{MSE}(\tilde{\theta})} \right)^{1/2}.$$

In our case, the baseline estimator is the maximum likelihood estimator (MLE) corresponding to the DGP. For example, in Table 2, the RRMSE of the  $\hat{\beta}_{IT}$  for the Gaussian DGP is the ratio of the MSE of  $\hat{\beta}_{IT}$  with the degrees of freedom  $\nu_{MLE}$  and the MSE of  $\hat{\beta}_N$ . Note that if  $\hat{\theta} = \tilde{\theta}$ , then the RRMSE of  $\hat{\theta}$  is equal to one.

**Table 2.** Relative bias and relative root mean squared error of the estimators of  $\beta$  ( $\hat{\beta}_N$ ,  $\hat{\beta}_{UT}$ ,  $\hat{\beta}_{IT}$ ) for the corresponding DGP (Gaussian, UT, and IT).

DGP		N		UT ( $\nu_{DGP} = 3$ )		IT ( $\nu_{DGP} = 3$ )	
Methods	Estimators	RB (%)	RRMSE	RB (%)	RRMSE	RB (%)	RRMSE
$\hat{\beta}_N, \hat{\beta}_{UT}$	$\hat{\beta}_{01}$	−0.07	1.00	−0.06	1.00	−0.09	1.48
	$\hat{\beta}_{02}$	0.00	1.00	0.00	1.00	0.00	1.48
	$\hat{\beta}_{11}$	−0.02	1.00	−0.01	1.00	−0.07	1.46
	$\hat{\beta}_{12}$	−0.00	1.00	−0.00	1.00	−0.00	1.46
$\hat{\beta}_{IT} (\nu_{MLE} = 3)$	$\hat{\beta}_{01}$	−0.09	1.04	−0.09	1.09	−0.03	1.00
	$\hat{\beta}_{02}$	0.00	1.04	0.00	1.09	0.00	1.00
	$\hat{\beta}_{11}$	−0.04	1.07	−0.02	1.08	−0.03	1.00
	$\hat{\beta}_{12}$	−0.00	1.07	−0.00	1.08	−0.00	1.00

### 3.2. Estimators of the $\beta$ Parameters

Table 3 reports the bias and the MSE of the Gaussian MLE estimator  $\hat{\beta}_N$ , the UT MLE estimator  $\hat{\beta}_{UT}$  ( $\nu_{DGP} = 3$ ), and the IT MLE estimator  $\hat{\beta}_{IT}$  ( $\nu_{DGP} = 3$ ) when the model is well specified, i.e., under the corresponding DGP. The bias and MSE of the estimators of  $\beta$  are small and comparable under the Gaussian and the UT DGP, but smaller for the IT DGP. Note that, in our implementation, the results of the algorithm for the IT estimators are very similar to those obtained using the function heavyLm from the R package heavy.

**Table 3.** Bias and MSE of the maximum likelihood estimators of  $\beta$  for the corresponding DGP (Gaussian, UT, and IT).

DGP		N		UT ( $\nu_{DGP} = 3$ )		IT ( $\nu_{DGP} = 3$ )	
Estimators		Bias	MSE	Bias	MSE	Bias	MSE
$\hat{\beta}_{01}$	$-1.39 \times 10^{-3}$	$4.57 \times 10^{-2}$	$-1.27 \times 10^{-3}$	$3.72 \times 10^{-2}$	$6.65 \times 10^{-4}$	$1.99 \times 10^{-2}$	
$\hat{\beta}_{02}$	$2.41 \times 10^{-5}$	$2.18 \times 10^{-5}$	$1.47 \times 10^{-5}$	$1.76 \times 10^{-5}$	$9.90 \times 10^{-6}$	$9.50 \times 10^{-6}$	
$\hat{\beta}_{11}$	$-6.62 \times 10^{-4}$	$2.16 \times 10^{-2}$	$3.23 \times 10^{-4}$	$2.05 \times 10^{-2}$	$-1.02 \times 10^{-3}$	$9.84 \times 10^{-3}$	
$\hat{\beta}_{12}$	$1.87 \times 10^{-5}$	$1.02 \times 10^{-5}$	$3.90 \times 10^{-6}$	$9.60 \times 10^{-6}$	$2.14 \times 10^{-5}$	$4.70 \times 10^{-6}$	

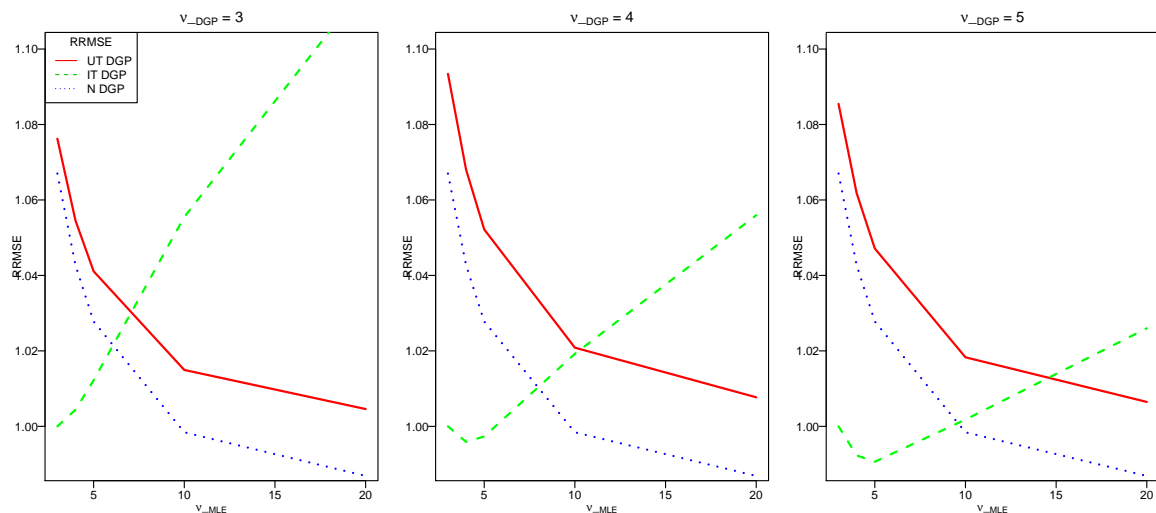
In Table 2, we start considering misspecifications and report the corresponding relative values RB and RRMSE of the same estimators and the same DGP as in Table 3 with all possible combinations of DGP and estimation methods. The results indicate that the RB of  $\hat{\beta}$  are all very small. If the DGP is Gaussian and the estimator is IT, the RRMSE of coordinates of  $\hat{\beta}$  is about 1.09. However, if the DGP is IT and the estimator is Gaussian, the RRMSE of coordinates of  $\hat{\beta}$  is higher (from 1.46–1.48). Hence for the Gaussian DGP, we do not lose too much efficiency using the IT estimator  $\hat{\beta}_{IT}$  with three degrees of freedom. Inversely, we lose much more efficiency when using  $\hat{\beta}_N$  for the IT DGP with three degrees of freedom.

In order to consider more degrees of freedom (3, 4, and 5), we now drop the bias and focus on the RRMSE. Table 4 indicates that the RRMSE of  $\hat{\beta}$  is very similar and close to one, with a maximum of 1.09, except for the case of the N estimator under the IT DGP, where it can reach 1.48. The work in Maronna (1976) provided theoretical asymptotic efficiencies of the Student versus the Gaussian estimators, the ratio of asymptotic variances being equal to  $\frac{(\nu - 2)(\nu + L + 2)}{\nu(\nu + L)}$ . The values obtained in Table 2 are very similar to these asymptotic values.

**Table 4.** The root relative mean squared errors of  $\hat{\beta}$ .

Methods	DGP	N	UT			IT		
			$\nu_{DGP} = 3$	$\nu_{DGP} = 4$	$\nu_{DGP} = 5$	$\nu_{DGP} = 3$	$\nu_{DGP} = 4$	$\nu_{DGP} = 5$
N	RRMSE	$\hat{\beta}_{01}$	1.00	1.00	1.00	1.48	1.22	1.14
		$\hat{\beta}_{02}$	1.00	1.00	1.00	1.48	1.23	1.14
		$\hat{\beta}_{11}$	1.00	1.00	1.00	1.46	1.22	1.13
		$\hat{\beta}_{12}$	1.00	1.00	1.00	1.46	1.22	1.13
IT ( $\nu_{MLE} = 3$ )	RRMSE	$\hat{\beta}_{01}$	1.04	1.09	1.08	1.00	1.00	1.01
		$\hat{\beta}_{02}$	1.04	1.09	1.08	1.00	1.00	1.01
		$\hat{\beta}_{11}$	1.07	1.08	1.08	1.00	1.00	1.01
		$\hat{\beta}_{12}$	1.07	1.08	1.09	1.00	1.00	1.01
IT ( $\nu_{MLE} = 4$ )	RRMSE	$\hat{\beta}_{01}$	1.02	1.07	1.06	1.00	1.00	1.00
		$\hat{\beta}_{02}$	1.01	1.06	1.05	1.00	1.00	1.00
		$\hat{\beta}_{11}$	1.04	1.06	1.06	1.00	1.00	1.00
		$\hat{\beta}_{12}$	1.04	1.05	1.06	1.00	1.00	1.00
IT ( $\nu_{MLE} = 5$ )	RRMSE	$\hat{\beta}_{01}$	1.00	1.05	1.04	1.01	1.00	1.00
		$\hat{\beta}_{02}$	1.00	1.05	1.04	1.01	1.00	1.00
		$\hat{\beta}_{11}$	1.03	1.04	1.05	1.01	1.00	1.00
		$\hat{\beta}_{12}$	1.03	1.04	1.05	1.01	1.00	1.00

Figure 1 shows the performances in terms of RRMSE of the IT estimators  $\hat{\beta}_{12}^{IT}$  under different DGP as a function of the degrees of freedom of the IT estimator ( $\nu_{MLE}$ ). The considered DGP are the Gaussian, UT, and IT DGP with the degrees of freedom  $\nu_{DGP} = 3$  (respectively,  $\nu_{DGP} = 4$ ,  $\nu_{DGP} = 5$ ) on the left (respectively, middle, right) plot. Overall, the RRMSE of  $\hat{\beta}_{12}^{IT}$  for the IT DGP has a down trend and then an up trend, while for the Gaussian and the UT DGP, the RRMSE are decreasing when  $\nu_{MLE}$  increases. The maximum RRMSE of  $\hat{\beta}_{12}^{IT}$  is around 1.09 under the UT DGP and is around 1.08 under the Gaussian DGP. It decreases then to one when  $\nu_{MLE}$  increases to twenty under the Gaussian and the UT DGP; thus, the risk under misspecification is not very high. The curve is U-shaped under the IT DGP with a minimum when  $\nu_{MLE} = \nu_{DGP}$ . The worst performance is when  $\nu_{DGP}$  is small and  $\nu_{MLE}$  is large. The RRMSE of  $\hat{\beta}_{12}^{IT}$  with  $\nu_{DGP} = 4$  is similar than the one with  $\nu_{DGP} = 5$ .



**Figure 1.** The RRMSE of the IT estimator of  $\hat{\beta}_{12}$  for the UT DGP in solid line, for the IT DGP in dashed line, and for the Gaussian DGP in dotted line with  $\nu_{DGP} = 3$  (respectively,  $\nu_{DGP} = 4$ ,  $\nu_{DGP} = 5$ ) on the left (respectively, middle, right) plot.

### 3.3. Estimators of the Variance Parameters

Table 5 reports the biases and the MSE of  $\hat{\rho}$ ,  $\hat{\sigma}_1^2$ ,  $\hat{\sigma}_2^2$  for the Gaussian DGP, the UT ( $\nu_{DGP} = 3$ ) DGP, and the IT ( $\nu_{DGP} = 3$ ) DGP. The bias and the MSE of  $\hat{\rho}$  are very similar and small for all cases. The MSE of the Gaussian estimators  $\hat{\sigma}_1^2$  and  $\hat{\sigma}_2^2$  are small under the Gaussian DGP, but they are higher under the UT and IT DGP. The biases and MSE of the IT estimator  $\hat{\sigma}_1^2$  and  $\hat{\sigma}_2^2$  are small under the IT DGP, but high under the Gaussian and the UT DGP. Besides, Table 5 also indicates that there is no method that estimates the variances well under the UT DGP.

**Table 5.** The bias and the MSE of  $\hat{\rho}$ ,  $\hat{\sigma}_1^2$ ,  $\hat{\sigma}_2^2$ .

Methods	DGP	N		UT ( $\nu_{DGP} = 3$ )		IT ( $\nu_{DGP} = 3$ )	
		Bias	MSE	Bias	MSE	Bias	MSE
N	$\hat{\rho}$	$-4.85 \times 10^{-4}$	$9.46 \times 10^{-4}$	$-2.08 \times 10^{-4}$	$7.68 \times 10^{-4}$	$-3.99 \times 10^{-3}$	$1.17 \times 10^{-2}$
	$\hat{\sigma}_1^2$	$-3.89 \times 10^{-3}$	$8.33 \times 10^{-3}$	$-1.05 \times 10^{-1}$	58	$6.94 \times 10^{-3}$	3.17
	$\hat{\sigma}_2^2$	$-1.75 \times 10^{-3}$	$2.01 \times 10^{-3}$	$-5.17 \times 10^{-2}$	14.93	$-1.77 \times 10^{-2}$	$2.85 \times 10^{-1}$
IT $\nu_{MLE} = 3$	$\hat{\rho}$	$-1.70 \times 10^{-4}$	$8.94 \times 10^{-4}$	$-2.18 \times 10^{-4}$	$9.05 \times 10^{-4}$	$-2.03 \times 10^{-4}$	$1.07 \times 10^{-3}$
	$\hat{\sigma}_1^2$	2.00	4.06	1.80	244.87	$-1.43 \times 10^{-2}$	$1.54 \times 10^{-2}$
	$\hat{\sigma}_2^2$	1.00	1.02	0.91	64.75	$-7.30 \times 10^{-3}$	$3.94 \times 10^{-3}$

As before, we now consider misspecified cases and focus on relative bias in Table 6. We observe that the relative bias for  $\hat{\rho}$  is negligible in all situations. The RB for  $\hat{\sigma}_1^2$  and  $\hat{\sigma}_2^2$  are also quite small (less than around 5%) when using the Gaussian estimator for all DGP. This is also true when using the IT estimator for the IT DGP with the same degrees of freedom  $\nu_{MLE} = \nu_{DGP}$ . There are some biases for  $\hat{\sigma}_1^2$  and  $\hat{\sigma}_2^2$  if the DGP is Gaussian or UT and the estimator is IT. For this estimator, the relative bias of  $\hat{\sigma}_1^2$ ,  $\hat{\sigma}_2^2$  is around 100% for the Gaussian DGP, 96% for the UT DGP with  $\nu_{DGP} = 5$  and  $\nu_{MLE} = 3$ , and 22% for the UT DGP with  $\nu_{DGP} = 5$  and  $\nu_{MLE} = 5$ . The RB for  $\hat{\sigma}_1^2$  and  $\hat{\sigma}_2^2$  are also quite high (up to 50%) for the IT estimator when the DGP is IT with  $\nu_{MLE} \neq \nu_{DGP}$ . To summarize, in terms of the RB of the variance estimators, the Gaussian estimator yields better results than the IT estimator.

**Table 6.** The RB of  $\hat{\rho}$ ,  $\hat{\sigma}_1^2$ ,  $\hat{\sigma}_2^2$  with  $\nu = 3, 4, 5$ .

Methods	DGP	N	UT			IT		
			$\nu_{DGP} = 3$	$\nu_{DGP} = 4$	$\nu_{DGP} = 5$	$\nu_{DGP} = 3$	$\nu_{DGP} = 4$	$\nu_{DGP} = 5$
N	$\hat{\rho}$	−0.14	−0.06	−0.06	−0.06	−1.13	−0.24	0.02
	$\hat{\sigma}_1^2$	−0.21	−5.23	−3.34	−2.31	0.35	−0.08	−0.12
	$\hat{\sigma}_2^2$	−0.18	−5.17	−3.33	−2.20	−1.77	−0.30	−0.09
IT, $\nu_{MLE} = 3$	$\hat{\rho}$	−0.05	−0.06	−0.06	−0.06	−0.06	−0.04	−0.02
	$\hat{\sigma}_1^2$	99.99	90.25	93.89	95.80	−0.72	32.79	50.12
	$\hat{\sigma}_2^2$	100.05	90.60	93.90	96.03	−0.73	32.79	50.13
IT, $\nu_{MLE} = 4$	$\hat{\rho}$	−0.05	−0.06	−0.06	−0.06	−0.06	−0.04	−0.01
	$\hat{\sigma}_1^2$	42.62	35.80	38.32	39.68	−24.66	−0.24	11.18
	$\hat{\sigma}_2^2$	42.66	36.01	38.34	39.85	−24.67	−0.23	11.19
IT, $\nu_{MLE} = 5$	$\hat{\rho}$	−0.06	−0.06	−0.06	−0.06	−0.06	−0.04	−0.00
	$\hat{\sigma}_1^2$	24.71	18.85	21.03	22.23	−31.75	−10.13	−0.14
	$\hat{\sigma}_2^2$	24.74	19.02	21.04	22.38	−31.76	−10.13	−0.14

Finally, Table 7 presents the RRMSE in the same cases. It shows that the RRMSE of  $\hat{\rho}$  varies from 0.94–1.09 for all DGP except for the case of the IT DGP with the Gaussian estimator, which ranges between 1.42 and 3.21. Besides, if the DGP is Gaussian and the estimator is IT or if the DGP is IT and the estimator is Gaussian, the RRMSE of  $\hat{\sigma}_1^2$  and  $\hat{\sigma}_2^2$  are high in particular for  $\nu_{DGP} = 3$  or  $\nu_{MLE} = 3$ : we lose a lot of efficiency in these misspecified cases. To conclude, we have seen from Table 6 that the RB of  $\hat{\sigma}_1^2$  and  $\hat{\sigma}_2^2$  are smaller for the Gaussian estimator than for the IT estimator. However, in terms of RRMSE, there is no clear advantage in using the Gaussian estimator with respect to the IT estimator.

It should be noted that for  $\nu \leq 4$ , the Student distribution has no fourth-order moment, which may explain the fact that the covariance estimators have large MSE.

**Table 7.** The RRMSE of  $\hat{\rho}$ ,  $\hat{\sigma}_1^2$ ,  $\hat{\sigma}_2^2$  in the Gaussian DGP, the UT DGP ( $\nu_{DGP} = 3, 4, 5$ ), and the IT DGP ( $\nu_{DGP} = 3, 4, 5$ ).

Methods	DGP	N	UT			IT		
	RRMSE		$\nu_{DGP} = 3$	$\nu_{DGP} = 4$	$\nu_{DGP} = 5$	$\nu_{DGP} = 3$	$\nu_{DGP} = 4$	$\nu_{DGP} = 5$
N	$\hat{\rho}$	1.00	1.00	1.00	1.00	3.21	1.91	1.42
	$\hat{\sigma}_1^2$	1.00	1.00	1.00	1.00	14.33	2.65	1.64
	$\hat{\sigma}_2^2$	1.00	1.00	1.00	1.00	8.50	2.24	1.78
IT, $\nu_{MLE} = 3$	$\hat{\rho}$	0.97	1.09	1.09	1.09	1.00	1.00	1.01
	$\hat{\sigma}_1^2$	22.07	2.05	2.11	2.16	1.00	5.89	9.18
	$\hat{\sigma}_2^2$	22.45	2.08	2.11	2.16	1.00	5.77	9.13
IT, $\nu_{MLE} = 4$	$\hat{\rho}$	0.95	1.06	1.06	1.06	1.01	1.00	1.00
	$\hat{\sigma}_1^2$	9.49	1.46	1.47	1.48	4.04	1.00	2.31
	$\hat{\sigma}_2^2$	9.65	1.48	1.47	1.48	4.00	1.00	2.30
IT, $\nu_{MLE} = 5$	$\hat{\rho}$	0.94	1.05	1.05	1.05	1.01	1.00	1.00
	$\hat{\sigma}_1^2$	5.58	1.27	1.27	1.28	5.16	1.99	1.00
	$\hat{\sigma}_2^2$	5.68	1.28	1.28	1.27	5.10	1.95	1.00

In order to allow the reproducibility of the empirical analyses contained in the present and the following sections, some Supplementary Material is available at the following link: <http://www.thibault.laurent.free.fr/code/jrfm/>.

#### 4. Selection between the Gaussian and IT Models

In this section, we propose a methodology to select a model between the Gaussian and independent Student models and to select the degrees of freedom for the Student in a short list of possibilities. Following the warnings of Fernandez and Steel (1999) and the empirical results of Katz and King (1999), Platen and Rendek (2008), and Kan and Zhou (2017), we decided to focus on a small selection of degrees of freedom and fit our models without estimating this parameter, considering that a second step of model selection will make the choice. Indeed, there is a limited number of interesting values, which are between three and eight (for larger values, the distribution gets close to being Gaussian). The work in Lange et al. (1989), p.883, proposed the likelihood ratio test for the univariate case. In what follows, we use the fact that the distribution of the Mahalanobis distances is known under the two DGP, which allows building a Kolmogorov–Smirnov test and using Q-Q plots. Unfortunately, this technique does not apply to the UT model for which the  $n$  observations are a single realization of the multivariate distribution. One advantage of this approach is that the Mahalanobis distance is a one-dimensional variable, whereas the original observations have  $L$  dimensions.

##### 4.1. Distributions of Mahalanobis Distances

For an  $L$ -dimensional random vector  $\mathbf{Y}$ , with mean  $\boldsymbol{\mu}$ , and covariance matrix  $\boldsymbol{\Sigma}$ , the squared Mahalanobis distance is defined by:

$$d^2 = (\mathbf{Y} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{Y} - \boldsymbol{\mu})$$

If  $\mathbf{Y}_1, \dots, \mathbf{Y}_n$  is a sample of size  $n$  from the  $L$ -dimensional Gaussian distribution  $\mathcal{N}_L(\boldsymbol{\mu}_N, \boldsymbol{\Sigma}_N)$ , the squared Mahalanobis distance of observation  $i$ , denoted by  $d_{Ni}^2$ , follows a  $\chi_L^2$  distribution. If  $\boldsymbol{\mu}_N$  and  $\boldsymbol{\Sigma}_N$  are unknown, then the squared Mahalanobis distance of observation  $i$  can be estimated by:

$$\hat{d}_{Ni}^2 = (\mathbf{Y}_i - \hat{\boldsymbol{\mu}}_N)^T \hat{\boldsymbol{\Sigma}}_N^{-1} (\mathbf{Y}_i - \hat{\boldsymbol{\mu}}_N)$$

where  $\hat{\boldsymbol{\mu}}_N = \bar{\mathbf{Y}} = \frac{1}{n} \sum_{i=1}^n \mathbf{Y}_i$  and  $\hat{\boldsymbol{\Sigma}}_N$  is the sample covariance matrix. The work in [Gnanadesikan and Kettenring \(1972\)](#) (see also [Bilodeau and Brenner \(1999\)](#)) proved that this square distance follows a Beta distribution, up to a multiplicative constant:

$$\frac{n}{(n-1)^2} (\mathbf{Y}_i - \hat{\boldsymbol{\mu}}_N)^T \hat{\boldsymbol{\Sigma}}_N^{-1} (\mathbf{Y}_i - \hat{\boldsymbol{\mu}}_N) \sim \text{Beta} \left( \frac{L}{2}, \frac{n-L-1}{2} \right)$$

where  $L$  is the dimension of  $\mathbf{Y}$ . For large  $n$ , this Beta distribution can be approximated by the chi-square distribution  $d_{Ni}^2 \sim \chi_L^2$ . According to [Gnanadesikan and Kettenring \(1972\)](#) (p. 172),  $n = 25$  already provides a sufficiently large sample for this approximation, which is the case in all our examples below. If we now assume that  $\mathbf{Y}_1, \dots, \mathbf{Y}_n$  is a sample of size  $n$  from the  $L$ -dimensional Student distribution  $\mathbf{Y}_i \sim \mathbf{T}(\boldsymbol{\mu}_{IT}, \boldsymbol{\Sigma}_{IT}, \nu)$ , then the squared Mahalanobis distance of observation  $i$ , denoted by  $d_{ITi}^2$  and properly scaled, follows a Fisher distribution (see [Roth \(2013\)](#)):

$$\frac{1}{L} \frac{\nu}{\nu-2} d_{ITi}^2 \sim \mathcal{F}(L, \nu)$$

If  $\boldsymbol{\mu}_{IT}$  and  $\boldsymbol{\Sigma}_{IT}$  are unknown, then the squared Mahalanobis distance of observation  $i$  can be estimated by:

$$\hat{d}_{ITi}^2 = (\mathbf{Y}_i - \hat{\boldsymbol{\mu}}_{IT})^T \hat{\boldsymbol{\Sigma}}_{IT}^{-1} (\mathbf{Y}_i - \hat{\boldsymbol{\mu}}_{IT}),$$

where  $\hat{\boldsymbol{\mu}}_{IT}$  and  $\hat{\boldsymbol{\Sigma}}_{IT}$  are the MLE of  $\boldsymbol{\mu}_{IT}$  and  $\boldsymbol{\Sigma}_{IT}$ . Note that in the IT model,  $\hat{\boldsymbol{\mu}}_{IT}$  is no longer equal to  $\bar{\mathbf{Y}}$ . Up to our knowledge, there is no result about the distribution of  $\hat{d}_{ITi}^2$ .

In the elliptical distribution family, the distribution of Mahalanobis distances characterizes the distribution of the observations. Thus, in order to test the normality of the data, we can test whether the Mahalanobis distances follow a chi-square distribution. Similarly, testing the Student distribution is equivalent to testing whether the Mahalanobis distances follow the Fisher distribution. There are two difficulties with the approach. The first one is that the estimated Mahalanobis distances are not a sample from the chi-square (respectively, the Fisher) distribution because there is dependence due to the estimation of the parameters. The second one is that, in our case, we not only estimate  $\boldsymbol{\mu}$  and  $\boldsymbol{\Sigma}$ , but we are in a regression framework where  $\boldsymbol{\mu}$  is linear combination of regressors, and we indeed estimate its coefficients. In what follows, we will ignore these two difficulties and consider that, for large  $n$ , the distributions of the estimated Mahalanobis distances behave as if  $\boldsymbol{\mu}$  and  $\boldsymbol{\Sigma}$  were known.

We propose to implement several Kolmogorov–Smirnov tests in order to test different null hypothesis: Gaussian, Student with three degrees of freedom, and Student with four degrees of freedom. As an exploratory tool, we also propose drawing Q-Q plots of the Mahalanobis distances with respect to the chi-square and the Fisher distribution [Small \(1978\)](#).

#### 4.2. Examples

This section illustrates some applications of the proposed methodology for selecting a model. We use a real dataset from finance and three simulated datasets with the same DGP as in Section 3.

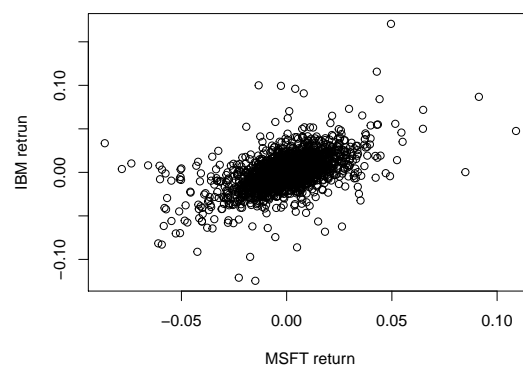
The real dataset is the daily closing share price of IBM and MSFT, which are imported from Yahoo Finance from 3 January 2007–27 September 2018 using the quantmod package in R. It contains  $n = 2955$  observations. Let  $\mathbf{S}_t$ ,  $t = 1, \dots, n$  be the daily share price of IBM and MSFT and  $\mathbf{Y}_t$  be the log-price increment (return) (see [Fung and Seneta \(2010\)](#)) over a day period, then:

$$\mathbf{Y}_t = \log \mathbf{S}_t - \log \mathbf{S}_{t-1}.$$



The three other datasets are simulated using the same model as in Section 3 with the Gaussian DGP, the IT DGP with  $\nu_{DGP} = 3$ , and the IT DGP with  $\nu_{DGP} = 4$  and with sample size  $n = 1000$ . Figure 2 (respectively, Figure 3) displays the scatterplots of the financial data (respectively, the three toy data).

We compute the Gaussian and the IT estimators as in Section 3. We then calculate the squared Mahalanobis distances of the residuals and use a Kolmogorov–Smirnov test for deciding between the models. For the financial data, we have no predictor. We test the Gaussian (respectively the Student with three degrees of freedom, the Student with four degrees of freedom) null hypothesis. When testing one of the null hypotheses, we use the estimator corresponding to the null. Moreover, when the null hypothesis is Student, we use the corresponding degrees of freedom for computing the maximum likelihood estimator. We do reject the null hypothesis if the  $p$ -value is smaller than  $\alpha = 5\%$ . Note that we could adjust the level of  $\alpha$  by taking into account multiple testing.

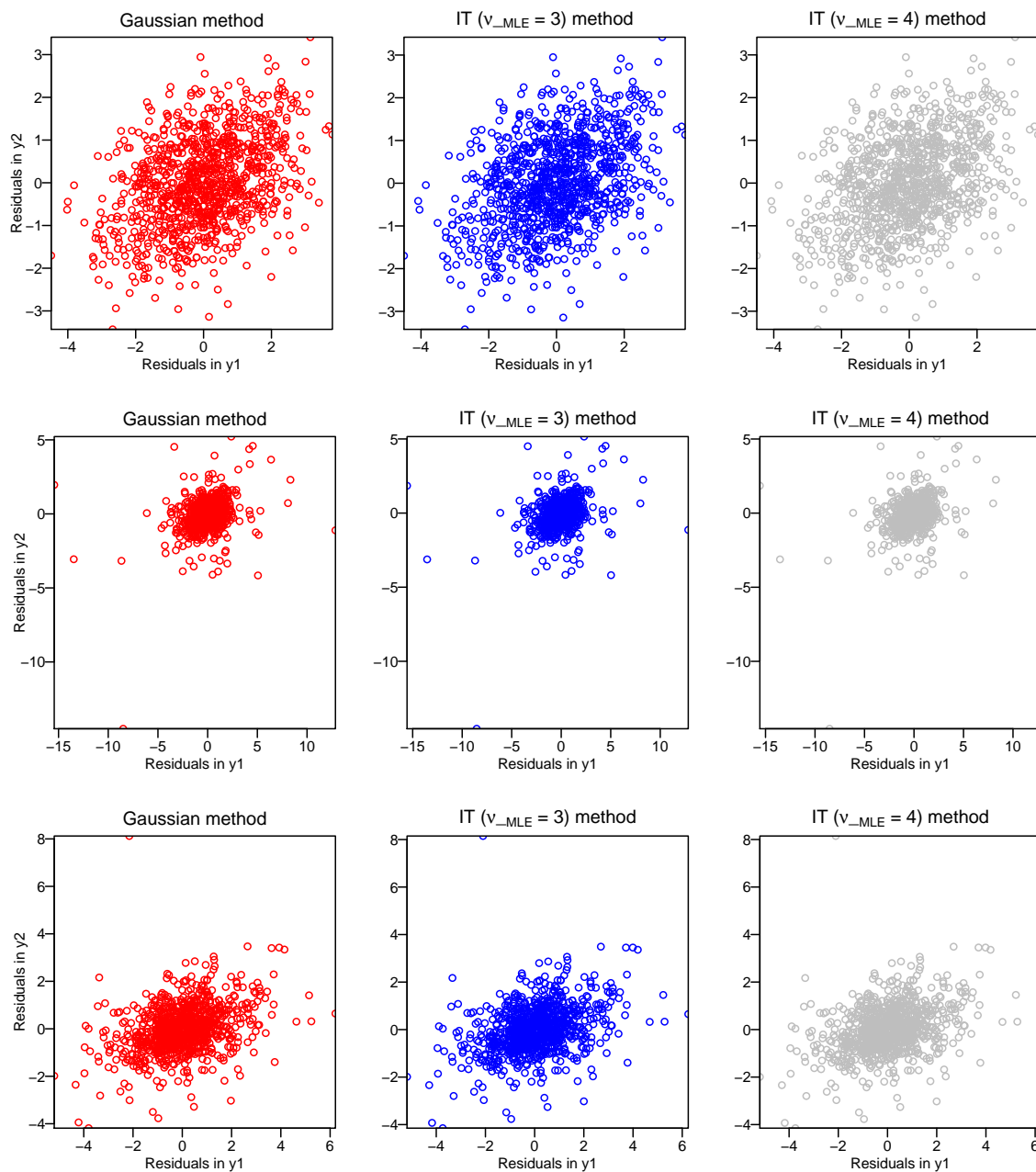


**Figure 2.** Financial data: scatterplot of returns.

Table 8 shows the  $p$ -values of these tests. For the simulated data, at the 5% level, we do not reject the Gaussian assumption when the DGP is Gaussian. Similarly, we do not reject the Student distribution with three (respectively, four) degrees of freedom when the DGP is the IT with degrees of freedom  $\nu_{DGP} = 3$  (respectively,  $\nu_{DGP} = 4$ ). For the financial data, we do not reject the Student distribution with three degrees of freedom, but we do reject the Gaussian distribution and the Student distribution with four degrees of freedom.

**Table 8.** All datasets: the  $p$ -values of the Mahalanobis distances tests with the null hypothesis and the corresponding estimators.

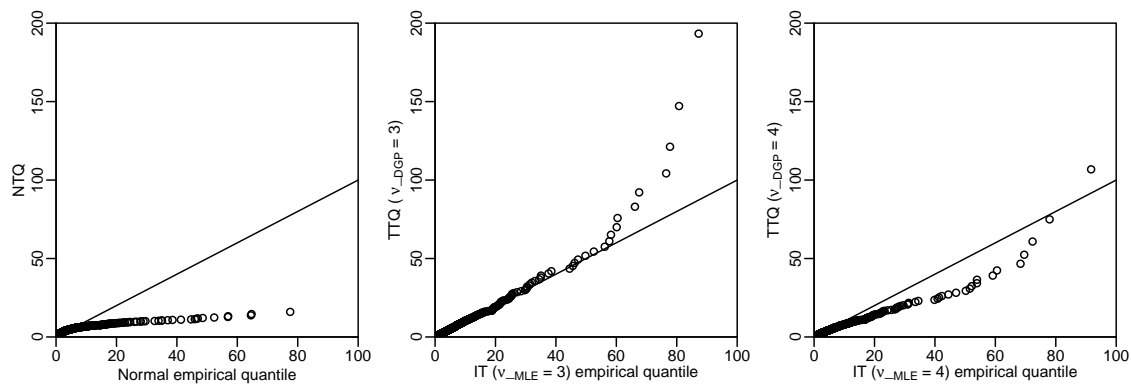
Hypothesis $H_0$		Toy DGP		Financial Data
Methods	N	IT, $\nu_{DGP} = 3$	IT, $\nu_{DGP} = 4$	
N	0.546	$2.2 \times 10^{-16}$	$2.2 \times 10^{-16}$	$2.2 \times 10^{-16}$
IT, $\nu_{MLE} = 3$	$2.2 \times 10^{-16}$	0.405	0.033	0.882
IT, $\nu_{MLE} = 4$	$2.2 \times 10^{-16}$	0.023	0.303	0.049



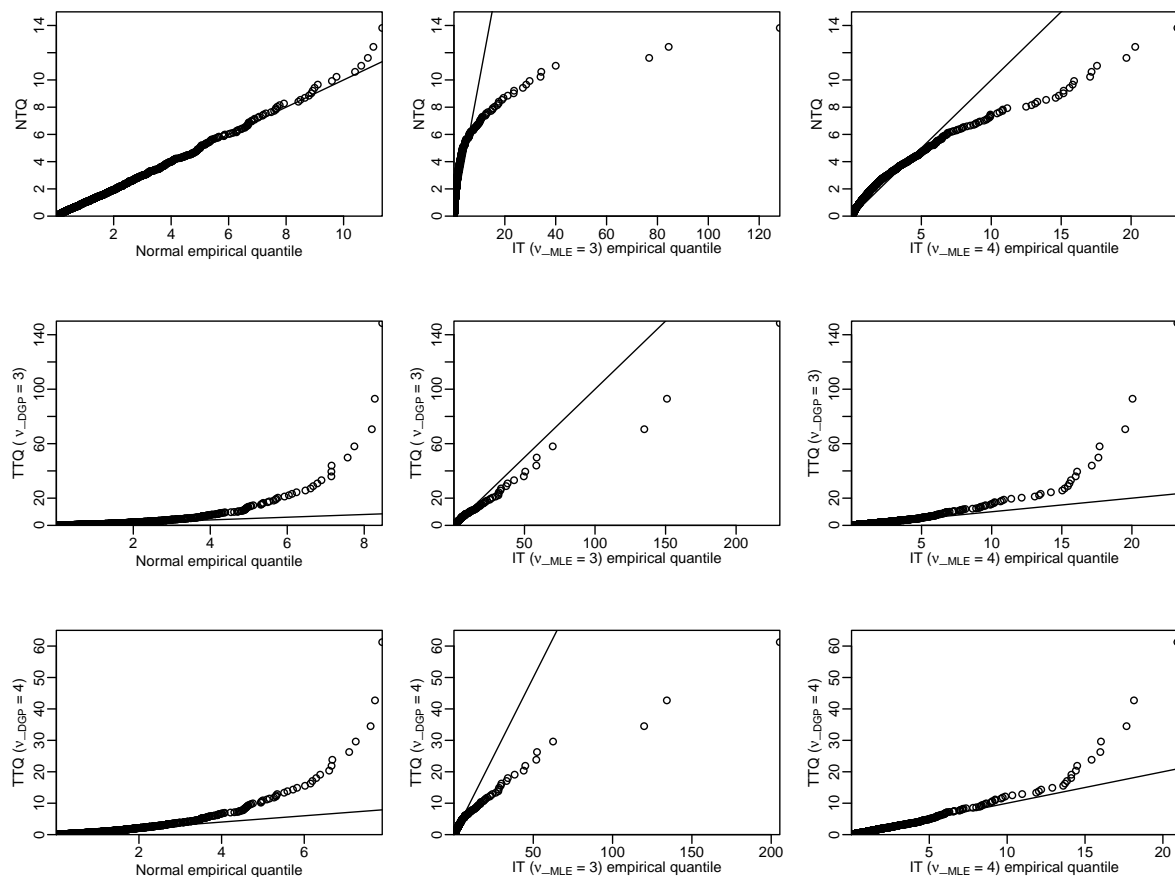
**Figure 3.** Toy data: scatterplots of residuals in the Gaussian DGP (respectively, the IT DGP with  $\nu_{DGP} = 3$ , the IT DGP with  $\nu_{DGP} = 4$ ) on the first row (respectively, the second row, the third row).

Figure 4 shows the Q-Q plots comparing the empirical quantiles of the Mahalanobis distances for the normal (respectively, the IT ( $\nu_{MLE} = 3$ ), the IT ( $\nu_{MLE} = 4$ )) estimators on the horizontal axis to the theoretical quantiles of the Mahalanobis distances for the normal (respectively, the IT ( $\nu_{MLE} = 3$ ), the IT ( $\nu_{MLE} = 4$ )) on the vertical axis for the financial data. These Q-Q plots are coherent with the results of the tests in Table 8. The IT model with three degrees of freedom fits our financial data well.

Figure 5 displays the Q-Q plots for the toy DGP: the Gaussian DGP in the first column, the IT DGP with  $\nu_{DGP} = 3$  in the second column, and the IT DGP with  $\nu_{DGP} = 4$  in the third column. The first row compares the empirical quantiles to the normal case quantiles, the second to the Student case quantiles with  $\nu_{DGP} = 3$ , and the third row to Student case quantiles with  $\nu_{DGP} = 4$ . The Q-Q plots on the diagonal confirm that the fit is good when the model is correct. The other Q-Q plots outside the diagonal correctly reveal a clear deviation from the hypothesized model.



**Figure 4.** Financial data: Q-Q plots of the Mahalanobis distances for the normal, IT ( $v_{MLE} = 3$ ), and IT ( $v_{MLE} = 4$ ) estimators.



**Figure 5.** Toy data: Q-Q plots of the Mahalanobis distances of the residuals for the normal (respectively, the IT with  $v_{DGP} = 3$ , the IT with  $v_{DGP} = 4$ ) case empirical quantiles against the normal (respectively, the IT with  $v_{MLE} = 3$ , the IT with  $v_{MLE} = 4$ ) case theoretical quantiles in the first row (respectively, the second row, the third row).

To summarize the findings of this study, let us first say that there may be an abusive use of the Gaussian distribution in applications due to its simplicity. We have seen that considering the Student distribution instead is just slightly more complex, but feasible, and that one can test this choice. Concerning the two Student models, we have seen that the UT model is simpler to fit than the IT model, but has limitations due to the fact that it assumes a single realization, which restricts the properties of the maximum likelihood estimators and prevents the use of tests against the other two models.

## 5. Conclusions

We have compared three different models: the multivariate Gaussian model and two different multivariate Student models (uncorrelated or independent). We have derived some theoretical properties of the Student UT model and proposed a simple iterative reweighted algorithm to compute the maximum likelihood estimators in the IT model. Our simulations show that using a multivariate Student IT model instead of a multivariate Gaussian model for heavy tail data is simple and can be viewed as a safeguard against misspecification in the sense that there is more to lose if the DGP is Student and one uses a Gaussian model than in the reverse situation. Finally, we have proposed some graphical tools and a test to choose between the Gaussian and the IT models. The IT model fits our finance dataset quite well. There is still work to do in the direction of improving the model selection procedure to overcome the fact that the parameters are estimated and hence the hypothetical distribution is only approximate. Let us mention that it is also possible to adapt our algorithm for the IT model to the case of missing data. We intend to work in the direction of allowing different degrees of freedom for each coordinate. It may be also relevant to consider an alternative estimation method by generalizing the one proposed in Kent et al. (1994) to the multivariate regression case. Finally, another perspective is to consider multivariate errors-in-variables models, which allow incorporating measurement errors in the response and the explanatory variables. A possible approach is proposed in Croux et al. (2010).

**Supplementary Materials:** In order to allow the reproducibility of the empirical analyses contained in the present paper, some Supplementary Material is available at the following link: <http://www.thibault.laurent.free.fr/code/jrfm/>.

**Author Contributions:** T.H.A.N., C.T.-A. and A.R.-G., Methodology, analysis, review, and editing; T.H.A.N., writing, original draft preparation; C.T.-A. and A.R.-G. supervision and validation; T.L. and T.H.A.N., data curation.

**Funding:** This research received no external funding

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

EM	Expectation-maximization
MLE	Maximum likelihood estimator
N	Normal (Gaussian) model
IT	Independent multivariate Student
UT	Uncorrelated multivariate Student
RB	Relative bias
MSE	Mean squared error
RRMSE	Root relative mean squared error
DGP	Data-generating process

## Appendix A

**Proof of Proposition 1.** Using Expression (1), the joint density function of  $\hat{\epsilon}_{UT}$  is:

$$\begin{aligned}
 p(\epsilon_{UT} | 0, \Omega_{UT}, \nu) &= \frac{f(\nu)}{\det(\mathbf{I}_n \otimes \Sigma_{UT})^{1/2}} \left[ 1 + \frac{1}{\nu - 2} \epsilon_{UT}^T (\mathbf{I}_n \otimes \Sigma_{UT})^{-1} \epsilon_{UT} \right]^{-\frac{\nu + nL}{2}} \\
 &= \frac{f(\nu)}{\det(\Sigma_{UT})^{n/2}} \left[ 1 + \frac{1}{\nu - 2} \epsilon_{UT}^T (\mathbf{I}_n \otimes \Sigma_{UT})^{-1} \epsilon_{UT} \right]^{-\frac{\nu + nL}{2}} \\
 &= \frac{f(\nu)}{\det(\Sigma_{UT})^{n/2}} \left[ 1 + \frac{1}{\nu - 2} \sum_{i=1}^n \epsilon_{UTi}^T \Sigma_{UT}^{-1} \epsilon_{UTi} \right]^{-\frac{\nu + nL}{2}}
 \end{aligned}$$

Therefore, the logarithm of  $p(\epsilon_{UT}|\mathbf{0}, \mathbf{\Sigma}_{UT}, \nu)$  is:

$$\log p(\epsilon_{UT}|\mathbf{0}, \mathbf{\Sigma}_{UT}, \nu) = \log f(\nu) - \frac{n}{2} \log \mathbf{\Sigma}_{UT} - \frac{\nu + nL}{2} \log \left[ 1 + \frac{1}{\nu - 2} \sum_{i=1}^n \epsilon_{UTi}^T \mathbf{\Sigma}_{UT}^{-1} \epsilon_{UTi} \right]. \quad (\text{A1})$$

In order to maximize  $\log p(p(\epsilon_{UT}|\mathbf{0}, \mathbf{\Sigma}_{UT}, \nu))$  as a function of  $\beta_{UT}$ , we follow the same argument as in Theorem 8.4 from [Seber \(2008\)](#) for the Gaussian case and obtain that the minimum of  $\sum_{i=1}^n \epsilon_{UTi}^T \mathbf{\Sigma}_{UT}^{-1} \epsilon_{UTi}$  is obtained for:

$$\hat{\beta}_{UT} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}.$$

Besides, taking the partial derivative of (A1) as a function of  $\mathbf{\Sigma}_{UT}$ , we obtain:

$$\begin{aligned} \frac{\partial \log(p(\epsilon_{UT}|\mathbf{0}, \mathbf{\Sigma}_{UT}, \nu))}{\partial \mathbf{\Sigma}_{UT}} &= -\frac{n \mathbf{\Sigma}_{UT}^{-1}}{2} - \frac{(\nu + nL)}{2} \frac{\partial \log(\nu - 2 + \sum_{i=1}^n \epsilon_{UTi}^T \mathbf{\Sigma}_{UT}^{-1} \epsilon_{UTi})}{\partial \mathbf{\Sigma}_{UT}} \\ &= -\frac{n \mathbf{\Sigma}_{UT}^{-1}}{2} - \frac{(\nu + nL)}{2} \frac{\partial(\nu - 2 + \sum_{i=1}^n \epsilon_{UTi}^T \mathbf{\Sigma}_{UT}^{-1} \epsilon_{UTi}) / \partial \mathbf{\Sigma}_{UT}}{\nu - 2 + \sum_{i=1}^n \epsilon_{UTi}^T \mathbf{\Sigma}_{UT}^{-1} \epsilon_{UTi}}. \end{aligned}$$

Let:

$$w_{UT} = \frac{1}{\nu - 2 + \sum_{i=1}^n \epsilon_{UTi}^T \mathbf{\Sigma}_{UT}^{-1} \epsilon_{UTi}}. \quad (\text{A2})$$

We have:

$$\begin{aligned} \frac{\partial \log(p(\epsilon_{UT}|\mathbf{0}, \mathbf{\Sigma}_{UT}, \nu))}{\partial \mathbf{\Sigma}_{UT}} &= -\frac{n \mathbf{\Sigma}_{UT}^{-1}}{2} - \frac{(\nu + nL)}{2} w_{UT} \partial(\nu - 2 + \sum_{i=1}^n \epsilon_{UTi}^T \mathbf{\Sigma}_{UT}^{-1} \epsilon_{UTi}) / \partial \mathbf{\Sigma}_{UT} \\ &= -\frac{n \mathbf{\Sigma}_{UT}^{-1}}{2} + \frac{(\nu + nL)}{2} w_{UT} \sum_{i=1}^n \mathbf{\Sigma}_{UT}^{-1} \epsilon_{UTi} \epsilon_{UTi}^T \mathbf{\Sigma}_{UT}^{-1} \end{aligned}$$

Solving  $\frac{\partial \log(p(\epsilon_{UT}|\mathbf{0}, \mathbf{\Sigma}_{UT}, \nu))}{\partial \mathbf{\Sigma}_{UT}} = 0$  and letting  $\mathbf{E} = \sum_{i=1}^n \epsilon_{UTi} \epsilon_{UTi}^T$ , we have:

$$\begin{aligned} \mathbf{\Sigma}_{UT}^{-1} &= \frac{\nu + nL}{n} w_{UT} \sum_{i=1}^n \mathbf{\Sigma}_{UT}^{-1} \epsilon_{UTi} \epsilon_{UTi}^T \mathbf{\Sigma}_{UT}^{-1} \\ \mathbf{\Sigma}_{UT} \mathbf{\Sigma}_{UT}^{-1} \mathbf{\Sigma}_{UT} &= \frac{\nu + nL}{n} w_{UT} \sum_{i=1}^n \mathbf{\Sigma}_{UT} \mathbf{\Sigma}_{UT}^{-1} \epsilon_{UTi} \epsilon_{UTi}^T \mathbf{\Sigma}_{UT}^{-1} \mathbf{\Sigma}_{UT} \\ \mathbf{\Sigma}_{UT} &= (\nu + nL) w_{UT} \frac{\mathbf{E}}{n} \end{aligned} \quad (\text{A3})$$

The expression of  $w_{UT}$  in (A3) can be simplified by noting that:

$$\begin{aligned} \mathbf{\Sigma}_{UT}^{-1} &= n((\nu + nL)w_{UT})^{-1} \mathbf{E}^{-1} \\ \sum_{i=1}^n \epsilon_{UTi}^T \mathbf{\Sigma}_{UT}^{-1} \epsilon_{UTi} &= n((\nu + nL)w_{UT})^{-1} \sum_{i=1}^n \epsilon_{UTi}^T \mathbf{E}^{-1} \epsilon_{UTi} \\ &= \frac{n}{(\nu + nL)w_{UT}} \text{tr} \left( \sum_{i=1}^n \epsilon_{UTi} \epsilon_{UTi}^T \mathbf{E}^{-1} \right) \\ \sum_{i=1}^n \epsilon_{UTi}^T \mathbf{\Sigma}_{UT}^{-1} \epsilon_{UTi} &= \frac{nL}{(\nu + nL)w_{UT}}. \end{aligned} \quad (\text{A4})$$

Replacing the expression of  $\sum_{i=1}^n \epsilon_{UTi}^T \Sigma_{UT}^{-1} \epsilon_{UTi}$  from (A4) into  $w_{UT}$ , we get:

$$w_{UT} = \frac{\nu}{(\nu - 2)(\nu + nL)}.$$

Finally,

$$\hat{\Sigma}_{UT} = \frac{\nu}{\nu - 2} \frac{\sum_{i=1}^n \hat{\epsilon}_{UTi} \hat{\epsilon}_{UTi}^T}{n}.$$

□

**Proof of Proposition 2.** The property  $\mathbb{E}(\hat{\beta}_{UT}) = \beta_{UT}$  is immediate. In order to facilitate the derivation of the proof for  $\hat{\Sigma}_{UT}$ , we write Model (4) as:

$$\mathbf{Y} = \mathbf{X}\mathbf{B} + \boldsymbol{\epsilon} \quad (\text{A5})$$

where:

$$\mathbf{Y} = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1L} \\ \vdots & \vdots & \vdots & \vdots \\ y_{n1} & y_{n2} & \cdots & y_{nL} \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1K} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & \cdots & x_{nK} \end{bmatrix}, \mathbf{B} = \begin{bmatrix} \beta_{01} & \beta_{0L} \\ \beta_{11} & \beta_{1L} \\ \vdots & \vdots \\ \beta_{K1} & \beta_{KL} \end{bmatrix}$$

$$\boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_{11} & \epsilon_{12} & \cdots & \epsilon_{1L} \\ \vdots & \vdots & \vdots & \vdots \\ \epsilon_{n1} & \epsilon_{n2} & \cdots & \epsilon_{nL} \end{bmatrix}, \hat{\mathbf{B}}_{UT} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} \text{ and } \hat{\boldsymbol{\epsilon}}_{UT} = \mathbf{Y} - \mathbf{X} \hat{\mathbf{B}}_{UT}.$$

Let  $\mathbf{E} = \hat{\boldsymbol{\epsilon}}_{UT}^T \hat{\boldsymbol{\epsilon}}_{UT}$  and  $\mathbf{M} = \mathbf{I}_n - \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$ . We have  $\mathbf{M}\mathbf{X}\mathbf{B} = 0$ , and following Seber (2008), Theorem 8.2,

$$\begin{aligned} \mathbf{E} &= (\mathbf{Y} - \mathbf{X} \hat{\mathbf{B}}_{UT})^T (\mathbf{Y} - \mathbf{X} \hat{\mathbf{B}}_{UT}) = (\mathbf{M}\mathbf{Y})^T \mathbf{M}\mathbf{Y} = \mathbf{Y}^T \mathbf{M}\mathbf{Y} \\ &= (\mathbf{Y} - \mathbf{X}\mathbf{B})^T \mathbf{M}(\mathbf{Y} - \mathbf{X}\mathbf{B}) = \boldsymbol{\epsilon}^T \mathbf{M}\boldsymbol{\epsilon} = \sum_h \sum_i M_{hi} \epsilon_{hi} \epsilon_i^T. \end{aligned}$$

$$\text{Since } \mathbb{E}(\epsilon_{hi} \epsilon_i^T) = \begin{cases} \boldsymbol{\Sigma} & \text{if } h = i \\ 0 & \text{otherwise} \end{cases}, \text{ for } h, i = 1, \dots, n, \mathbb{E}(\mathbf{E}) = \sum_h M_{hh} \boldsymbol{\Sigma} = \text{tr}(\mathbf{M}) \boldsymbol{\Sigma} = (n - K) \boldsymbol{\Sigma}$$

and:

$$\mathbb{E}(\hat{\Sigma}_{UT}) = \mathbb{E}\left(\frac{\nu}{\nu - 2} \frac{\mathbf{E}}{n}\right) = \frac{\nu}{\nu - 2} \frac{\mathbb{E}(\mathbf{E})}{n} = \frac{\nu}{\nu - 2} \frac{n - K}{n} \boldsymbol{\Sigma}_{UT}.$$

□

## References

- Bilodeau, Martin, and David Brenner. 1999. *Theory of Multivariate Statistics (Springer Texts in Statistics)*. Berlin: Springer. ISBN 978-0-387-22616-3.
- Croux, Christophe, Mohammed Fekri, and Anne Ruiz-Gazen. 2010. Fast and robust estimation of the multivariate errors in variables model. *Test* 19: 286–303. [CrossRef]
- Dempster, Arthur P., Nan M. Laird, and Donald B. Rubin. 1978. Iteratively Reweighted Least Squared for Linear Regression when Errors are Normal/Independent distributed. *Multivariate Analysis V* 5: 35–37.
- Dogru, Fatma Zehra, Y. Murat Bulut, and Olcay Arslan. 2018. Double Reweighted Estimators for the Parameters of the Multivariate t distribution. *Communications in Statistics-Theory and Methods* 47: 4751–71. [CrossRef]



- Fernandez, Carmen, and Mark F. J. Steel. 1999. Multivariate Student  $t$ -Regression Models: Pitfalls and Inference. *Biometrika Trust* 86: 153–67. [\[CrossRef\]](#)
- Fung, Thomas, and Eugene Seneta. 2010. Modeling and Estimating for Bivariate Financial Returns. *International Statistical Review* 78: 117–33. [\[CrossRef\]](#)
- Fraser, Donald Alexander Stuart. 1979. *Inference and Linear Models*. New York: McGraw Hill. ISBN 9780070219106.
- Fraser, Donald Alexander Stuart, and Kai Wang Ng. 1980. Multivariate regression analysis with spherical error. *Multivariate Analysis* 5: 369–86.
- Gnanadesikan, Ram, and Jon R. Kettenring. 1972. Robust estimates, residuals, and outlier detection with multiresponse data. *Biometrics* 28: 81–124. [\[CrossRef\]](#)
- Hofert, Marius. 2003. On Sampling from the Multivariate  $t$  Distribution. *The R Journal* 5: 129–36. [\[CrossRef\]](#)
- Hu, Wenbo, and Alec N. Kercheval. 2009. Portfolio optimization for Student  $t$  and skewed  $t$  returns. *Quantitative Finance* 10: 129–36. [\[CrossRef\]](#)
- Huber, Peter J., and Elvezio M. Ronchetti. 2009. *Robust Statistics*. Hoboken: Wiley. ISBN 9780470129906.
- Johnson, Norman L., and Samuel Kotz. 1972. Student multivariate distribution. In *Distribution in Statistics: Continuous Multivariate Distributions*. Michigan: Wiley Publishing House. ISBN-13: 9780471443704.
- Kan, Raymond, and Guofu Zhou. 2017. Modeling non-normality using multivariate  $t$ : implications for asset pricing. *China Finance Review International* 7: 2–32. [\[CrossRef\]](#)
- Katz, Jonathan N., and Gary King. 1999. A Statistical Model for Multiparty Electoral Data. *American Political Science Review* 93: 15–32. [\[CrossRef\]](#)
- Kelejian, Harry H., and Ingmar R. Prucha. 1985. Independent or Uncorrelated Disturbances in Linear Regression. *Economics Letters* 19: 35–38. [\[CrossRef\]](#)
- Kent, John T., David E. Tyler, and Yahuda Vard. 1994. A curious likelihood identity for the multivariate  $t$ -distribution. *Communications in Statistics-Simulation and Computation* 23: 441–53. [\[CrossRef\]](#)
- Kotz, Samuel, and Saralees Nadarajah. 2004. *Multivariate  $t$  Distributions and Their Applications*. Cambridge: Cambridge University Press. ISBN 9780511550683.
- Lange, Kenneth, Roderick J. A. Little, and Jeremy Taylor. 1989. Robust Statistical Modeling Using the  $t$ -Distribution. *International Statistical Review* 84: 881–96. [\[CrossRef\]](#)
- Lange, Kenneth, and Janet S. Sinsheimer. 1993. Normal/Independent Distributions and Their Applications in Robust Regression. *Journal of Computational and Graphical Statistics* 2: 175–98. [\[CrossRef\]](#)
- Liu, Chuanhai, and Donald B. Rubin. 1995. ML estimation of the  $t$  distribution using EM and its extensions, ECM and ECME. *Statistica Sinica* 5: 19–39.
- Liu, Chuanhai. 1997. ML Estimation of the Multivariate  $t$  Distribution and the EM Algorithm. *J. Multivar. Anal.* 63: 296–312. [\[CrossRef\]](#)
- Maronna, Ricardo Antonio. 1976. Robust M-Estimators of Multivariate Location and Scatter. *The Annals of Statistics* 4: 51–67. [\[CrossRef\]](#)
- McNeil, Alexander J., Rüdiger Frey, and Paul Embrechts. 2005. *Quantitative Risk Management: Concepts, Techniques and Tools*. Vol. 3. Princeton: Princeton University Press.
- Platen, Eckhard, and Renata Rendek. 2008. Empirical Evidence on Student- $t$  Log-Returns of Diversified World Stock Indices. *Journal of Statistical Theory and Practice* 2: 233–51. [\[CrossRef\]](#)
- Prucha, Ingmar R., and Harry H. Kelejian. 1984. The Structure of Simultaneous Equation Estimators: A generalization Towards Nonnormal Disturbances. *Econometrica* 52: 721–36. [\[CrossRef\]](#)
- Roth, Michael. 2013. *On the Multivariate  $t$  Distribution*. Report Number: LiTH-ISY-R-3059. Linköping: Department of Electrical Engineering, Linköping University.
- Seber, George Arthur Frederick. 2008. *Multivariate Observations*. Hoboken: John Wiley & Sons. ISBN 9780471881049.
- Singh, Radhey. 1988. Estimation of Error Variance in Linear Regression Models with Errors having Multivariate Student  $t$ -Distribution with Unknown Degrees of Freedom. *Economics Letters* 27: 47–53. [\[CrossRef\]](#)
- Small, N. J. H. 1978. Plotting squared radii. *Biometrics* 65: 657–58. [\[CrossRef\]](#)

- Sutradhar, Brajendra C., and Mir M. Ali. 1986. Estimation of the Parameters of a Regression Model with a Multivariate t Error Variable. *Communication Statistics Theory and Method* 15: 429–50. [[CrossRef](#)]
- Zellner, Arnold. 1976. Bayesian and Non-Bayesian Analysis of the Regression Model with Multivariate Student-t Error Terms. *Journal of the American Statistical Association* 71: 400–5. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).