Bachelor Thesis: Quantitative Finance

MULTIVARIATE RESIDUAL BASED MISSPECIFICATION TESTS APPLIED ON ASYMMETRIC DISTRIBUTIONS

Pepijn Harsveldt, 498787

Supervisor: dr. M. Zhelonkin

Second assessor: dr. M. Welz

Erasmus School of Economics

Erasmus University Rotterdam, Rotterdam

July 3, 2022

Abstract

Multivariate Linear Regression is a highly recommended and one of the most used models to model and predict financial time-series. However, diagnostic tests are sparse for the MLR model. This paper investigates several diagnostic tests and applies them to the MLR model. Multiple standard assumptions are rejected, such as the normality assumption. This paper tries to indicate which portmanteau test should be applied and which model including the error-term distribution suits the financial time-series appropriately. A 3-factor Fama-French model with skew t-distribution is assumed to be the highest performer for the model. Several tests such as the Engle and Ljung-Box lack power in the multivariate case.

Contents

1 Introduction

Multivariate Linear Regression (MLR) is a highly popular model in finance and econometrics, according to Dufour et al. (2010) and all the references therein. The reason for its popularity is that the MLR model is relatively easy to compute. This paper researches several MLR residual based misspecification tests. The starting point is the research done by Dufour et al. (2010).

The structure of the MLR model is similar to the univariate regression, however, the dimensions differ. The MLR model contains several dependent variables (in this paper 25) and the same independent variables applied in a linear fashion. The independent variables in this paper are the variables of the 3-factor Fama-French model (Fama and French, 1993), including an intercept.

MLR is a potentially high dimension model and therefore the specification is important. The model uses the same framework (3-factor Fama-French model) for the independent variables and assumptions for the error-terms for each equation. But, there might be cross-correlation or linear serial dependence in the actual data. The relevance of misspecification tests is significant in MLR models. Misspecification might lead to biased parameters and possible over-rejection (Dufour et al., 2004). Misspecification is a wide term and therefore several applications are applicable into this matter. The specification of the residuals is researched in this paper, other specifications such as linearity assumption is a topic of further research. Linear serial dependence and ARCH effects are discussed, together with the distribution of the error-terms. A combination between no linear serial dependence and no ARCH effects, result in white-noise errors. The assumption of white-noise errors (Heij et al., 2004), is widely used in modelling. However, the validity of the white-noise error-terms assumption might not hold when the MLR model model is applied to financial time-series.

The number of error-term distributions is numerous. However, the normal distribution is the default option in model specifications. Within the Ordinary Least Squares (OLS) assumption framework (Heij et al., 2004), the error-terms are assumed to be normally distributed (Heij et al., 2004). But, in financial time-series the error-terms might not follow a pattern according to the normal distribution. A wider explanation into this topic is given in Section 4.

Both the white-noise assumption and the normal distribution assumptions, might not be valid. As stated above, this might result in over-rejection (Dufour and Khalaf, 2000) and biased parameters. When there is over-rejection in tests, wrong conclusions can be drawn. As a result, actions can be taken based on wrong information. The specification of the model is therefore vital for not only the test-statistics, but also for the interpretations taken up from those.

For the univariate linear regression, the research into misspecification is numerous. However, according to Kroner and Ng (1998), research into misspecification test applied to MLR models is sparse. This paper tries to improve and extend the research done by Dufour et al. (2010), this is done by researching asymmetric error-term distributions applied to the tests. In Section 4, evidence is given that the error-term distribution might be asymmetric. The purpose is two fold, namely, checking if the p-values obtain stronger p-values (stronger means higher) and if possible over-rejection can be resolved. In other words, the assumption of normally distributed error-terms is replaced by several other distributions including asymmetric ones. Therefore, the research question is,

What happens to the test performance of residual based multi-equation misspecification tests, when asymmetric distributions are applied?.

In order to answer this question, first it is important to replicate the research done by Dufour et al. (2010). Several portmanteau misspecification tests are used, such as the Ljung-Box test (Ljung and Box, 1978) and the Lee-King (Lee and King, 1993). However, these tests are designed for univariate linear regression. The model used in this paper is MLR, the dimensions differ. The paper by Dufour et al. (2010) uses two types of portmanteau tests, namely, equation-by-equation based and multi-equation.

A portmanteau test is one with an unspecified alternative hypothesis. The equation-by-equation portmanteau tests are designed for univariate linear regression models. The procedure to use them in a multivariate environment is given in Section 3. However, there are two multi-equation portmanteau tests used, namely, the Hosking test (Hosking, 1980) and the Hosking-ARCH test (Duchesne and Lalancette, 2003). This test is specially designed for a multivariate environment.

A Monte Carlo (after this MC) simulation study is used to obtain the p-values. Using similar MC techniques in the replication, with adjustments, the extension p-values are obtained. The performance of a test is measured by its p-value accordingly, a high p-value results in high power.

This paper is constructed as follows, first the literature section (Section 2). The reasoning and intuition behind certain choices is given in this section. Secondly, the methodology is presented. The test specification and testing procedure is given. Thirdly, the data used is discussed. The data section provides crucial information concerning the time-series. Namely, for certain sub-samples the data is non-stationary. Furthermore, the presence of non-normality is proved and the asymmetry is shown. After performing the replication and its extension towards it, the results are provided. Finally, a conclusion is drawn.

The findings in this paper are remarkable. As an example, for the Ljung-Box test the p-values do not differ extensively, when they are obtained by non-Gaussian error-term distributions. However, for the Hosking test, the results do change significantly, when skewed distributions are applied. Under the Gaussian assumption the test statistic follows a $\chi^2(n^2G)$ distribution. But, under other distributions this differs.

For the Lee-King testthere is a similar pattern, non-Gaussian results in non-normal distribution of the test-statistic.

2 Literature

Modelling a financial time-series is a complex process. The number of models is endless and finding the correct model is hard (Refenes et al., 1997). However, in order to verify if a model is specified correctly, misspecification tests have been developed. Misspecification tests are not only important for forecasting and modelling, but also for test-statistics. For example, tests are likely to over-reject when the model is wrongly specified (Richardson and Smith, 1993).

The general framework is equation-by-equation based. In this paper several misspecification tests are discussed, applied to multivariate data, such as the Ljung-Box (Ljung and Box, 1978) test and Hosking-ARCH test (Duchesne and Lalancette, 2003). These test are applied in several procedures, such as MC simulation.

The first test procedure is the asymptotic test procedure, resulting in Bonferroni p-values. However, in MLR models the null-hypothesis tends to be over-rejected when the equations within a model becomes numerous, (Dufour and Khalaf, 2000). Furthermore, these are asymptotic p-values, they can lose power when the sample is finite.

In order to adjust for this over-rejection, p-values constructed by Monte Carlo simulations are applied (Harrison, 2010). The reason for this simulation based testing is to eliminate possible errors in the testing procedure. However, there are two unknown aspects in the simulation of the error-terms. Namely, the scale of the covariance matrix and the exact distribution. Therefore, the tests are conducted using standardized residuals (Section 3.1), several distributions are researched.

The first error-term distribution is the standard normal distribution. Moreover, the assumption of normally distributed error-terms is a assumption in the OLS regression framework (Heij et al., 2004). The normality assumptions seem to fit the general framework of financial time-series. Furthermore, normally distributed error-terms are relatively easy to compute. Distributions such as the χ^2 distribution converges asymptotically towards the normal distribution, when the degrees of freedom become numerous. However, the normal distribution does not capture all the characteristics and there is also some critics of this assumption (Sewell, 2011).

The assumption of $N(0, \sigma^2)$ distributed error-terms is highly used as discussed in the paragraph above. However, according to the data Section (4) and several papers, such as Sewell (2011), the assumption of normal distributed error-terms might not be valid. The student-t-distribution (after this t-distribution) is used, with unknown degrees of freedom. The t-distribution does fit financial time-series well according to Bormetti et al. (2007). However, some tests are constructed based on the normality assumption, deviation from this assumption might result in biased results (Dufour and Khalaf, 2000). The simulation is done by Maximized Monte Carlo simulation. For further explanation see Section 3.2.2.

The assumption of symmetric error-term distributions is short sighted according to Miron and Tudor (2010). This paper uses a skew normal and skew t-distribution, an applied MMC procedure is used in order to obtain the p-values. A better fitting error-term distribution might resolve possible over-rejection, if the tests do have sufficient power under asymmetric circumstances.

This paper uses two tests concerning linear serial dependence and three for ARCH-effects. The tests for linear serial dependence are the Ljung-Box (Ljung and Box, 1978) and Hosking (Hosking, 1980). In Dufour et al. (2010), the Variance-Ratio test (Ayadi and Pyun, 1994) is also discussed. However, the tests are seemingly similar and according to Ayadi and Pyun (1994), the Ljung-Box test has a higher tolerance towards different distributions regarding the error-terms. But, the Ljung-Box is still sensitive for non-Gaussian errors (Burns, 2002). The Hosking test is a multi-equation portmanteau test. For further explanation see Section 3.3.2.

The tests for ARCH-effects are the Engle (Engle, 1982) , Lee-King (Lee and King, 1993) and Hosking-ARCH (Duchesne and Lalancette, 2003) test. The buildup within these tests is logical. The Engle's test performs well under normality (Raunig, 2004), but loses power under non-Gaussian distributions. Using this test, the p-values for the MC latter should be accurate. The Lee-King statistic is robust for the error-term distribution, used for its estimation (Dufour et al., 2004).

3 Methods

The methodology section is divided into three major parts. First, the test procedure and secondly, the simulation techniques. Thirdly, the different tests. All the testing and simulation is done in R. The MLR model is given by

$$Y = XB + U, \tag{1}$$

where T is the time (in months) and n the number of independent variables (in this paper equal to 25). Y is a $T \times n$ matrix . X represents a $T \times (s+1)$ matrix, X includes an intercept. s is the number of independent variables used per equation (3). B is the coefficient matrix which is estimated by least squares, the rank is $(s+1) \times n$, (Heij et al., 2004). Finally, the U is the residual matrix, $T \times n$.

However, the distribution of U is unknown. This paper follows the framework set by Dufour et al. (2010) for the U distribution and its assumptions. The following framework is used for the error-term,

$$U = JW, \tag{2}$$

where W is a distribution and J a lower triangular non-singular matrix. This paper uses several different distributions for W. The white-noise assumption is used for the construction and application of W. As an example the standard normal distribution is used, $W_i \sim N(0,1)$, where i is the i^{th} distribution. A popular transformation is,

$$W = U(J^{-1})', \tag{3}$$

this allows us to further specify the distribution W. J determines the scale of the distribution, in order to obtain this lower triangular matrix, the variance matrix of U is computed,

$$\Sigma = \frac{1}{T}\hat{U}'\hat{U}, \tag{4}$$

$$\Sigma = S'_{\hat{U}} S_{\hat{U}}. \tag{5}$$

Σ is the covariance matrix of U with dimensions $n \times n$, the $\frac{1}{T}$ term is included due to the fact that these are supposed to be regression error-terms. Equation 5, shows the Cholesky decomposition. The Cholesky decomposition divides the matrix in upper triangular matrices, the combination of those two (identical) matrices form the covariance matrix. This property is used to obtain the J matrix. Using the Cholesky decompositon,

$$W = (S_U^{-1})'U. \tag{6}$$

3.1 Test procedure

The general test procedure framework is applicable for each test conducted in this paper. This framework provides a p-value, namely, the Bonferroni p-value with Tippett's procedure (Tippett et al., 1931) application.

The tests are conducted based on the observed data, this gives either one (Hosking(-ARCH)) or n values. These values are converted into p-values by their asymptotic distribution. The minimum of the p-values is called the Bonferroni p-value, p_B.

The extended Tippett's procedure states that the combined test-statistic is given by, 1-min(p-values). The intuition behind this statistic is that when one statistic rejects the null-hypothesis, all tests are significant. For further explanation into this framework or procedure see Dufour et al. (2010).

However, in order to obtain the p-values of the simulated cases, the error-term needs to be simulated. This paper uses several distributions for W and tests which distribution best fits the observed data. However, U is obtained by regression and this is not convenient for simulation, because it consists out of the scale and distribution. Therefore, this paper makes use of the $Invariance\ of\ Cholesky - standardized\ multivariate\ residuals$ property (Dufour et al., 2010),

$$\tilde{W} = \hat{U}S_{\hat{U}}^{-1} = \hat{W}S_{\hat{W}}^{-1}. \tag{7}$$

The W matrix is relatively easy to compute, because it follows a (semi-)known distribution. However, in order to obtain $\tilde{W}$, a modified W matrix needs to be made. This matrix $\hat{W}$ is computed by,

$$M = I - X(X'X)^{-1}X', \tag{8}$$

$$\hat{W} = MW. \tag{9}$$

The reason for using $\hat{W}$ instead of W, is that tests use estimated error-terms and are not perfectly simulated. The M matrix transforms the W so that it fits the actual residuals better, the rank is $T \times T$.

$\tilde{W}$ is used in the several tests and is the standardized multivariate residual. Furthermore, $\tilde{W}$ is independent from B and J.

3.2 Simulation

This section is divided in several sub-sections containing different MC techniques. Additionally, the constructing of p-values under these different simulation techniques is discussed. The total number of simulations is equal to 999, similar to the paper by Dufour et al. (2010).

3.2.1 MC

The MC simulation, concerns the W matrix. This matrix is simulated and follows a certain distribution. For the MC case this is the standard normal distribution ($N(0,1)$), there are no nuisance parameters and therefore it is fairly easy to simulate. Furthermore, for the construction of W the assumptions of $i.i.d$ holds.

Based on the simulated W, the $\tilde{W}$ is constructed. As mentioned in Section 3.1 , $\tilde{W}$ is used to compute the test statistics. Computing the p-value for the MC simulation consists of two steps. The first step is to compute the p-value (using the asymptotic distribution) for each simulation, if applicable using Tippett's procedure.

Using the Bonferroni p-value, a benchmark statistic is created . This is done by subtracting the Bonferroni p-value from 1, benchmark =1-p_B. For each simulation the smallest p-value is subtracted from 1, resulting in N values. If this value is larger than the benchmark, $LRGST$ adds with 1. $LRGST$ is the number of times a simulated combined value exceeds the benchmark. The p_{mc} value is constructed by,

$$p_{mc} = \frac{1 + LRGST}{N + 1}. \tag{10}$$

Note that the smallest possible p-value is equal to 0.001, p_{mc} is the p-value for the MC simulation.

3.2.2 MMC

Maximized Monte Carlo (MMC) is a MC simulation with an extra dimension, due to the nuisance in parameters for the distribution. The t-distribution with ν is used for W, the possible values for ν range between 2 and T-n-2. This results for the full-sample in 453 possible degrees of freedom. However, if the degrees of freedom (after this DOF) for the t-distribution increases, the distribution converges to the normal distribution. To decrease the running time and memory issues a function for the possible degrees of freedom is used. For the full-sample this function has the following form:

$$\nu = \mathbb{1}_{2<=z<=20}*z+\mathbb{1}_{21<=z<=35}*\{20+2*(z-20)\}+\mathbb{1}_{36<=z<=48}*\{50+4*(z-35)\}+\mathbb{1}_{49<=z<=60}*\{102+30*(z-48)\}, \tag{11}$$

where z is the number of MC simulations already past. In the full-sample there would normally be 453 MC simulations, the function described in equation 11, decreases the number of MC simulations to 59. This is done by using an indicator function ($\mathbb{1}$) and an increasing step pattern. Similar to the MC case the statistics for the tests are calculated, the dimension is equal (full-sample) to 480x25xNx59. For each MC simulation the p-value is calculated according to the method set in Section 3.2.1. The p_{mmc} is the maximum (supremum) p-value across all MC simulations (you maximize step-value across all possible DOF).

3.2.3 CSMMC

Confidence Set MMC (CSMMC) is similar to the MMC method. However, instead of taking all possible DOF, a confidence set is created for the DOF. The exact construction of this confidence can been found in Dufour et al. (2010). Furthermore, the p-value should be taken more careful. The rejection criteria of 5% should be adjusted to 2.5%. For further explanation into this matter see Dufour et al. (2010).

3.2.4 MMMC

Maximized Maximized MC (MMMC) is used for the extension upon the paper by Dufour et al. (2010). In Section 1 and 2, the presence of possible skew error distributions is discussed. The data Section (4) suggests that the data is partially non-normal and potentially asymmetric. Therefore, the skew normal (Azzalini and Valle, 1996) and skew t-distribution (Gupta, 2003) are examined. The structure of the skew normal pdf is,

$$\frac{2}{\omega\sqrt{2\pi}}e^{-\frac{(x-\xi)^2}{2\omega^2}}\int_{-\infty}^{\alpha\left(\frac{x-\xi}{\omega}\right)}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}\,dt. \tag{12}$$

α is the Skewness parameter, this parameter is taken from -5.5 till 4.5 with steps of one for the estimation. These bounds for the skewness seems to be a proper working frame according to the data Section (4) A negative value indicates a left skewed-distribution. The location parameter is ξ, this represents the mean of the distribution. This is calculated by taking the mean of the observed distribution for the sub-sample. Thirdly, the scale parameter, ω, the scale is assumed to be equal to one. However, in this paper we take values for ω from 0.5 till 1.5, this is in line with the results obtained in the data Section (4).

In order to find the maximized p-value a maximizing over 2 nuisance parameters needs to be done, namely, α and ω. The procedure for obtain the p-value, p_{mmmc-N}, is similar to the MMC case.

The second distribution that is examined by the MMMC simulation is the skew t-distribution. The reasoning follows the same pattern as for the distinguishing between a normal and t-distribution. But, applied to the skewed latter. The pdf of the skew t-distribution,

$$\frac{2}{\gamma+\frac{1}{\gamma}}\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)}\left\{1+\frac{(\gamma x)^2}{\nu}\right\}^{-\frac{\nu+1}{2}} \qquad x<0, \tag{13}$$

$$\frac{2}{\gamma+\frac{1}{\gamma}}\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)}\left\{1+\frac{(\frac{\gamma}{x})^2}{\nu}\right\}^{-\frac{\nu+1}{2}} \qquad x\geq 0. \tag{14}$$

γ represents the scale of the Skewness, for the MMMC-t the range is between -5.5 and 4.5 (with step length equal to one). Furthermore, similar to the t-distribution there are DOF. ν represents the DOF, for computational matters they range from 8 till 34. Similar to the MMMC case, the p-value is obtained by the double MMC case. This results in the p_{MMMC-t}. However, the running time is excessive for this method.

Therefore, not all p-values for this application are given. The package *sn* (Azzalini, 2022) is used to compute the skew-normal and skew t-distribution.

3.3 Linear Serial Dependence tests

Linear serial dependence is the relation between certain periods in time (Heij et al., 2004). The number of lags is set and equal to twelve in this paper, this is equal to one year.

Many models assume that a model does not have linear serial dependence, this assumption (Heij et al., 2004) is crucial in constructing models. In perspective, the random terms do not follow a certain pattern of auto-correlation, this due to the fact that it is hard to measure and highly irregular (Chan and Hsiao, 2014). The assumption of no linear serial dependence is important in the construction of an *i.i.d.* time-series. This section is divided in two sub-sections. The first consist of the Ljung-Box test and the second the Hosking test. The null-hypothesis is for each Portmanteau test, that there is no linear serial dependence.

3.3.1 Ljung-Box

The Ljung-Box test is designed by Ljung and Box (1978), is easy to compute and is robust compared to the other alternatives. However, there are some drawbacks. Namely, if the number of lags is exceeds 5% of the series length, the power reduces significantly, (Burns (2002)). Furthermore, if the error-terms are non-Gaussian, but t-distributed, then the Ljung-Box test suffers when the nuissance parameters is larger than 10.

The Ljung-Box test uses the auto-correlation in order to compute the statistic,

$$\hat{\rho}_{ig} = \frac{\sum_{t=g+1}^{T} \hat{u}_{it}\hat{u}_{i,t-g}}{\sum_{t=1}^{T} \hat{u}_{it}^2}. \tag{15}$$

$\hat{u}$ are the residuals. For the MC cases this should be the standardized residuals combined with the Cholesky decomposition theory, $\tilde{W}$. The dimension per simulation for the auto-correlation, ρ_{ig}, is 25×12. The Ljung-Box test has the following form,

$$LB_i = T(T+2)\sum_{g=1}^{G} \frac{\hat{\rho}_{ig}^2}{T-g}. \tag{16}$$

where LB is the Ljung-Box statistic, per simulation there are 25 values. i stands for the i^{th} equation where the Ljung-Box is conducted on. Note that the test is equation-by-equation based and throughout the procedure given in Section 3.1, a p-value is obtained.

3.3.2 Hosking

The Hosking linear serial dependence test (Hosking, 1980) is a multi-equation portmanteau test, the general framework is explained first and than the application towards the Hosking statistic. This framework is also used in the Hoskings-ARCH case, within the ARCH effect section.

The multi-equation portmanteau is special in the case that it uses all (in this paper) 25 equation residuals simultaneously. This results in one statistic per simulation instead of 25 as previous discussed in the equation-by-equation framework.

In order to compute a multi-equation test statistic, a portmanteau multi-equation framework is used,

$$C_Z(g) = T^{-1} \sum_{t=g+1}^{T} Z_t Z'_{t-g}, \quad g = 0, 1, \ldots, G. \tag{17}$$

Z is a $T \times n$ matrix, the decomposition of Z is $[Z_1, ..., Z_t, ..., Z_T]$. To obtain $C_Z(g)$ multiplications based on the rows of the Z matrix are done, the rank is equivalent to 25×25. g is the number of lags used in this statistic, in this paper with a maximum of twelve. The framework of $C_Z(g)$ is used in the Hosking statistic,

$$HM = T^2 \sum_{g=1}^{G} (T-g)^{-1} \operatorname{tr}\left[C_{\hat{U}}(0)^{-1} C_{\hat{U}}(g) C_{\hat{U}}(0)^{-1} C_{\hat{U}}(g)'\right]. \tag{18}$$

Instead of Z, U (in the simulation $\tilde{W}$) is used. A similar invariance property is applicable as Equation (7). For further explanation see Dufour et al. (2010). HM is the statistic value, the distribution of this statistic is $\chi^2(n^2G)$. tr is the trace-function, predefined in the *pracma* (Borchers and Borchers, 2022) package. g represents the lag used, the statistic takes all lags into account (twelve).

The Hosking statistic loses performance in high dimension spaces, (Li et al., 2019). This paper uses for the full-sample latter a sample size of 480×25. Therefore, an assumption can be made that the prediction is not highly accurate. Furthermore, according to Li et al. (2019) the Hosking test is conservative and loses power under other than normal distributions. A possible over-rejection or under rejection is expected in the MMC latter and equivalents.

3.4 ARCH tests

This section is divided in three parts, namely, three different tests for ARCH effects. ARCH stands for Auto Regressive Condition Heteroskedasticity. ARCH tests check for possible volatility clustering, the variance in each (sub-)sample should be constant. This is important for maintaining the *i.i.d.* assumption, used in construction of the model and tests. The first test discussed is the Engle test (Engle, 1982). Secondly, the Lee-King (Lee and King, 1993) and the thirdly is the multi-equation portmanteau Hosking-ARCH test (Duchesne and Lalancette, 2003). The null-hypothesis applied to all tests is that there are no ARCH-effects.

3.4.1 Engle

The Engle's ARCH test, commonly known as the ARCH-LM test is the standard test for ARCH effects (Sjölander, 2011). However, the ARCH-LM test over-rejects the null-hypothesis in finite samples. MC results should eliminate this problem. An auxiliary test regression is used to obtain the Engle's test criteria,

$$\hat{u}_{t,i}^2 = \hat{\delta}_{0,i} + \sum_{s=1}^{q} \hat{\delta}_s \hat{u}_{t-s,i}^2 + v_{t,i}. \tag{19}$$

Note that in the simulation latter, the $\tilde{W}$ is used. $\hat{\delta}_{0,i}$ is the regression constant and $v_{t,i}$ the auxiliary regression disturbance term. The Engle's test statistic is obtained by, T x R^2, where R^2 is from the auxiliary regression. The distribution is $\chi^2(g)$, note that this is an equation-by-equation test and therefore produces 25 statistics per simulation.

3.4.2 Lee-King

The Lee-King test statistic has a high power under Leptokurtic symmetric distributions (Lee and King, 1993). The Lee-King test is developed by Lee and King (1993) and the test is formulated as follows,

$$LK_i = \frac{\left[(T-G)\sum_{t=G+1}^{T}\left\{\left(\hat{u}_{it}^2/\hat{\sigma}_i^2\right)-1\right\}\sum_{g=1}^{G}\hat{u}_{i,t-g}^2\right] / \left[\sum_{t=G+1}^{T}\left\{\left(\hat{u}_{it}^2/\hat{\sigma}_i^2\right)-1\right\}^2\right]^{1/2}}{\left\{(T-G)\sum_{t=G+1}^{T}\left(\sum_{g=1}^{G}\hat{u}_{i,t-g}^2\right)^2-\left(\sum_{t=G+1}^{T}\sum_{g=1}^{G}\hat{u}_{i,t-g}^2\right)^2\right\}^{1/2}}. \tag{20}$$

The Lee-King test is an equation-by-equation test, the statistic is given by LK_i, with distribution $N(0,1)$. G is the maximum number of lags. Note that in the simulation latter, $\tilde{W}$ is used instead of $\hat{U}$. Furthermore, the estimated variance, $\hat{\sigma}$, is used in this equation,

$$\hat{\sigma}_i^2 = \frac{1}{T}\sum_{t=1}^{T}\hat{u}_{it}^2. \tag{21}$$

For the simulation latter, $\tilde{W}$ is used.

3.4.3 Hosking-ARCH

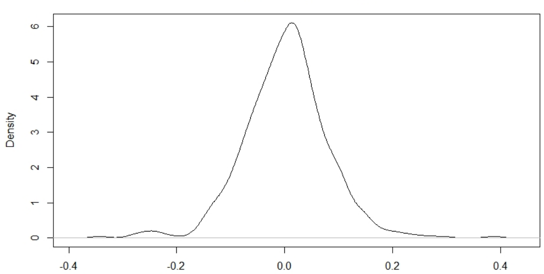
The Hosking-ARCH test (Duchesne and Lalancette, 2003) is a multi-equation portmanteau test. The framework of a multi-equation Portmanteau test is described in Section 3.3.2. The Hosking-ARCH test is an extension to the Hosking test (Hosking, 1980), this test is developed by Duchesne and Lalancette (2003). Similar use of the Hosking-Invariance as discussed in Section 3.3.2 and Dufour et al. (2010), is applied. The test is formulated as,

$$HM_2 = T^2\sum_{g=1}^{G}(T-g)^{-1}\,\mathrm{tr}\left\{C_{\hat{U}^2}(0)^{-1}C_{\hat{U}^2}(g)C_{\hat{U}^2}(0)^{-1}C_{\hat{U}^2}(g)'\right\}. \tag{22}$$

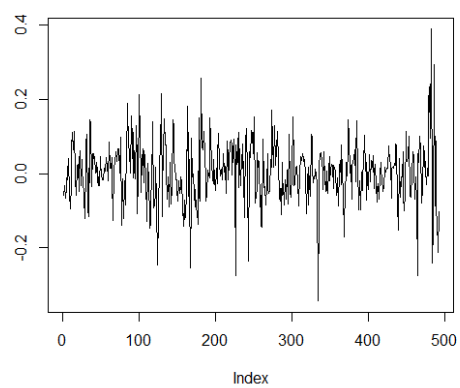
Similar to Hosking test, the Hosking-ARCH test is $\chi^2(n^2G)$ distributed. However, instead of using $\tilde{W}$, the squared is used.

4 Data

The data used in this paper is similar to the data in Dufour et al. (2010), there are in total 492 observations from January 1960 till December 2000. In total there are 28 columns, 25 columns containing data for the dependent variables (Y) and three for the independent variable (X). The data concerning the dependent variables consists out of monthly returns (in US Dollars) of portfolios, originating from the Fama and French database. The portfolios are constructed by taking five size portfolios and five ratio book-equity to market-equity, for further information see Dufour et al. (2010). This data section first discusses the statistics and characteristics of the dependent variables (portfolios, Y). Secondly, the construction of the independent variables is discussed. Thirdly, the characteristics of the error-terms is given. A plot of the distribution of $portfolio_1$ is given, together with the time lapse.



(a) Figure 1: Kernel plot portfolio 1, full-sample



(b) Figure 1: Time lapse portfolio 1, full sample

The first plot shows the distribution of the first portfolio. The plot shows a non-Gaussian behaviour, a left skewed distribution. Furthermore, in the second image there is some volatility clustering. The time-series shows locally non-stationarity. The descriptive statistics concerning the dependent variables (Y) is discussed below, note that these statistics are based on an average across all portfolio's.

	Full Sample	61-65	66-70	71-75	76-80	81-85	86-90	91-95	96-00
Mean	0.007	0.010	0.003	-0.002	0.015	0.007	0.002	0.013	0.009
St Dev	0.056	0.044	0.062	0.068	0.058	0.048	0.059	0.037	0.063
Skewness	-0.234	-0.196	-0.084	0.753	-0.601	0.415	-1.656	0.087	-0.371
Kurtosis	0.715	3.889	2.829	4.958	5.848	2.962	9.002	3.200	4.729
Jarque-Bera	174.078*	3.69	0.705	19.007*	29.947 *	2.847	135.502*	2.826	13.948*

Table 1: Combined descriptive statistics of portfolio's (Y), $*$ is significant at 5%

The mean of the returns is approximately zero, which is expected for returns. Furthermore, based on the

Jarque-Bera test statistic (Jarque and Bera, 1987) the null-hypothesis of normally distributed error-terms needs to be rejected for the full-sample and four sub-samples. The Jarque-Bera (JB) test is $\chi^2(2)$ distributed for 5%. The assumption of a left skewed distribution is moderately confirmed by the negative skewness. The standard deviations (St Dev) do differ across the several sub-samples compared to the full-sample. *moments* package (Komsta and Novomestky, 2022) is used for the computation of the skewness and kurtosis.

In order to check the suspicion of non-stationarity, an Augmented Dickey Fuller test (Dickey and Fuller, 1979)(after this ADF test) is conducted to check for the presence of a unit root in the returns. If a unit root is present, then the time-series is stationary. This is important for the misspecification tests, namely, if a test is non-stationary then the results might be unreliable and spurious (Stewart et al., 2007). The ADF test is conducted over the full-sample and sub-samples for each dependent variable. The null-hypothesis is that the time-series is non-stationary. *tseries* package (Trapletti and Hornik, 2022) is used with a default value of 0.01.

	1	2	3	4	5	6	7	8	9	10	11	12
full-sample	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
61-65	0.01	0.023	0.01	0.036	0.028	0.01	0.028	0.036	0.015	0.039	0.023	0.037
66-70	0.116	0.186	0.110	0.146	0.269	0.09	0.105	0.202	0.138	0.063	0.052	0.139
71-75	0.039	0.048	0.018	0.041	0.016	0.070	0.045	0.034	0.045	0.021	0.037	0.078
76-80	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
81-85	0.098	0.074	0.255	0.267	0.267	0.044	0.096	0.193	0.091	0.061	0.065	0.116
86-90	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
91-95	0.034	0.01	0.01	0.01	0.012	0.01	0.01	0.022	0.011	0.01	0.01	0.01
96-00	0.033	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.023	0.01

13	14	15	16	17	18	19	20	21	22	23	24	25
0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
0.073	0.059	0.029	0.099	0.083	0.062	0.079	0.124	0.232	0.210	0.107	0.113	0.177
0.164	0.179	0.134	0.055	0.223	0.071	0.232	0.196	0.043	0.043	0.031	0.195	0.084
0.025	0.090	0.012	0.082	0.058	0.080	0.050	0.021	0.149	0.102	0.155	0.141	0.084
0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.011	0.01	0.01	0.01	0.01
0.106	0.042	0.048	0.080	0.038	0.083	0.031	0.062	0.088	0.034	0.064	0.032	0.01
0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
0.01	0.01	0.01	0.01	0.01	0.01	0.020	0.015	0.026	0.026	0.024	0.064	0.020
0.01	0.052	0.018	0.025	0.01	0.01	0.01	0.01	0.015	0.01	0.01	0.035	0.028

Table 2: Augmented Dickey Fuller p-values

The full sample is stationary, therefore predicting the full does not lose power while predicting. However, over the sub-samples there is non-stationarity. The power of the estimation decreases when there is stationarity, this affects the residuals as well.

As independent variables this paper uses the 3-factor Fama-French model. The first factor is the market return over all markets, this factor is calculated by taking all the returns of the following exchanges, NYSE, AMEX and Nasdaq. These excess returns are subtracted by the one-month treasury bill rate. The second dependent variable is Small Minus Big, this variable is constructed by subtracting the average return of three large portfolios from the average of three small portfolio's. The formula for obtaining SMB is,

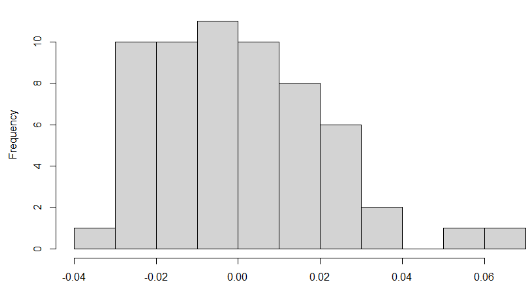
$$SMB = AverageSmallPortfolio - AverageLargePortfolio. \tag{23}$$

Note that these portfolios need to be measured monthly, due to the fact that the SMB figure is monthly.

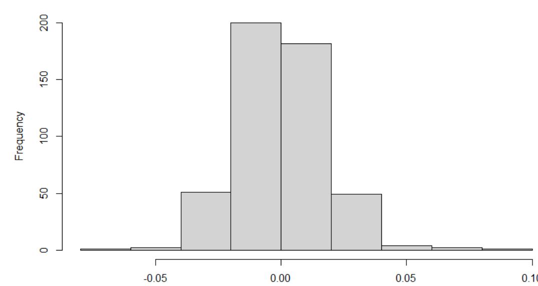
The third independent variable used is the High Minus Low (HML) variable. This variable is constructed by subtracting two growth portfolios from two value-weighted portfolios. For further explanation and assumptions, see Dufour et al. (2010).

The residuals are obtained by OLS regression, assumed to be normally distributed (Heij et al., 2004). The Jarque-Bera test (Jarque and Bera, 1987) is used to check if this assumption is correct.

The results of the Jarque-Bera test are given in the Appendix, Section 7.1. The results indicate that for certain portfolios in different sub-samples the normality assumption is rejected. For the full-sample there is a strong rejection towards normality. The pattern among the different sub-samples seems to be irregular. Therefore, it is not possible to state that the distribution of the error-term is Gaussian. In the Appendix (Section 7), the results concerning the skewness and kurtosis is given to indicate the asymmetry.



(a) Figure 2: Histogram, 61-65, portfolio1



(b) Figure 2: Histogram, full-sample, portfolio1

The figure left indicates lack kurtosis and left skewness in the residuals. Asymmetry is visible in this sub-sample. Furthermore over the full-sample (when looking at portfolio 1) excess kurtosis combined with left skewness is visible. Together with the results in Section 7, there is evidence of asymmetry within the distribution of the error-terms.

5 Results

The result section is divided into two parts. Namely, the linear serial dependence and the ARCH effect results. Furthermore, the Bonferroni p-values have a rejection criteria of $\frac{\alpha}{n}$, which results in 0.002.

5.1 Linear Serial Dependence

The results for the linear serial dependence are shown in the table below. First the Ljung-Box (Ljung and Box, 1978) results are discussed and then the Hosking test (Hosking, 1980) results. Afterwards a comparison between the two tests is drawn.

	Ljung-Box					
Sample	p_B	p_{mc}	p_{mmc}	p_{csmmc}	p_{mmmc-N}	p_{mmmc-t}
full	2.88*10ˆ(-5)*	0.002*	0.002*	0.001*	0.001*	0.001*
61-65	9.46*10ˆ(-6)*	0.039*	0.054	0.054	0.044*	0.05*
66-70	0.00014*	0.112	0.153	0.153	0.143	0.129
71-75	0.0052 *	0.495	0.528	0.528	0.514	0.510
76-80	0.00047*	0.203	0.228	0.228	0.231	0.226
81-85	0.0098	0.649	0.673	0.673	0.663	0.678
86-90	0.0017 *	0.301	0.356	0.356	0.359	0.349
91-95	0.000*	0.016*	0.019 *	0.019*	0.023*	0.019*
96-00	0.003	0.154	0.189	0.189	0.186	0.176

Table 3: Ljung-Box p-values, * indicates significant at 5%

The Bonferroni indicate that the hypothesis of no linear serial dependence for the full-sample, 61-65, 66-70, 76-80, 86-90, 91-95, is rejected. These results are in line with the results presented in Dufour et al. (2010). The MC based p-values are similar across the different cases, namely, that for the full-sample, 61-65 (MC,MMMC-N) and 91-95 the hypothesis of no linear serial dependence is rejected.

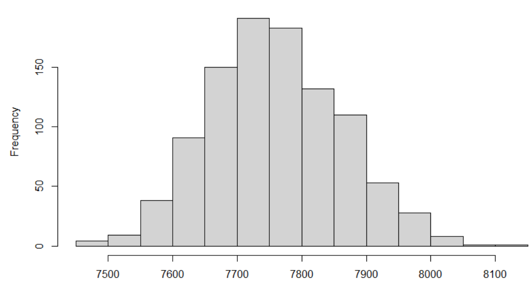
The difference between the p-values across the different MC cases is subtle. The Ljung-Box test performs well under normal distributions, only lacking power when other distributions are used (Burns, 2002). The difference between normally and t-distributed W is minimal, possibly due to performance loss under (asymmetric) non-Gaussian distributions.

The Hosking test result in similar pattern when comparing p_B. There is linear serial dependence for the full sample and the 96-00 sub-sample, according to the Hosking test ($0.000 < 0.002$). Furthermore, the difference for different W is remarkable. The p-values are in increasing order, p_{MMMC-t} is the highest. Under different distributions the performance of the Hosking statistic decreases according to Li et al. (2019). Remarkable result is the high performance under asymmetric distributions, this might indicate either stronger

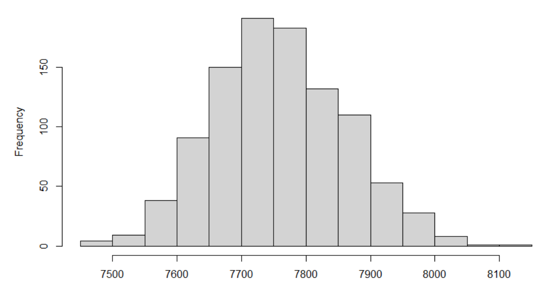
	Hosking					
Sample	p_B	p_{mc}	p_{mmc}	p_{csmmc}	p_{mmmc-N}	p_{mmmc-t}
full	0.000*	0.001*	0.001*	0.001*	1	1
61-65	0.040	0.681	0.566	0.560	0.710	0.773
66-70	0.104	0.872	0.705	0.705	0.878	0.873
71-75	0.263	0.918	0.741	0.740	1	1
76-80	0.325	0.998	0.938	0.994	1	0.998
81-85	0.325	0.998	0.938	0.994	1	0.998
86-90	0.082	0.767	0.610	0.746	1	1
91-95	0.146	0.847	0.677	0.677	0.862	0.905
96-00	.000*	0.01*	0.004*	0.006*	0.017*	0.023*

Table 4: Hosking p-values, * indicates significant at 5%

framework or under-rejection. The distribution of the Hosking's statistic is supposed to be $\chi^2(n^2G)$, the plots below shows the behaviour of the Hosking distribution.

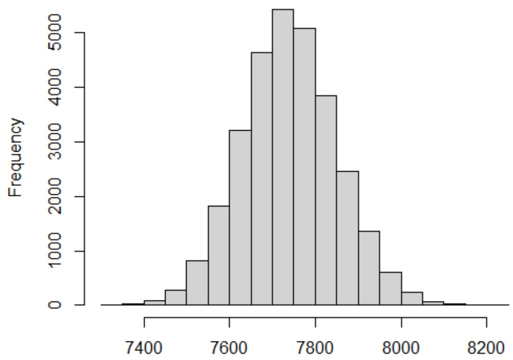


(a) Figure 3: 91-95, MC Hosking statistic

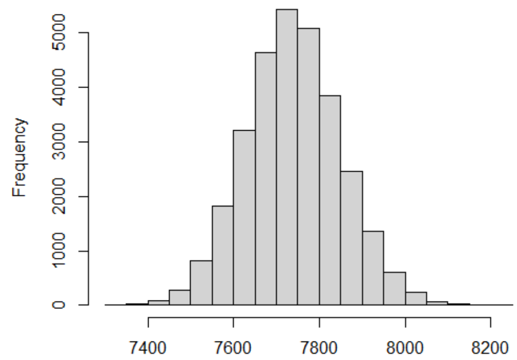


(b) Figure 3: 91-95, MMC Hosking statistic

The plot of the $\chi^2(n^2G)$ distribution is given in Section 5.2. A noticeable result is that when the distribution for W is asymmetric the test-statistic distribution suffers from excess kurtosis. The relatively low rejection is based around this result. Namely, due to relatively small tails and excess kurtosis, the distribution does not reject the null-hypothesis based on distribution. The MC simulation follows the smoothest pattern concerning distribution resemblance.



(a) Figure 4: 91-95, MMMC-N Hosking statistic



(b) Figure 4: 91-95, MMMC-t Hosking statistic

5.2 ARCH-effect

As discussed in Section 1, there are three different ARCH-effect tests. Each test has its own application and performance under specifications. The first test to be discussed is the Engle test (Engle, 1982).

	Engle					
Sample	p_B	p_{mc}	p_{mmc}	p_{csmmc}	p_{mmmc-N}	p_{mmmc-t}
full	0.000*	0.001*	0.001*	0.001*	0.001*	none
61-65	0.111	0.591	0.731	0.635	0.629	none
66-70	0.141	0.743	0.778	0.778	0.756	none
71-75	0.160	0.799	0.834	0.834	0.830	none
76-80	0.002*	0.004*	0.182	0.181	0.016	none
81-85	0.098	0.542	0.696	0.696	0.546	none
86-90	0.124	0.670	0.755	0.741	0.688	none
91-95	0.082	0.414	0.664	0.664	0.463	none
96-00	0.001*	0.001*	0.124	0.118	0.438	none

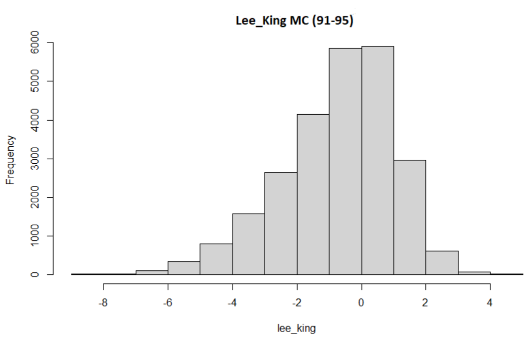
Table 5: Engle p-values, * indicates significant at 5%

The Engle test p_B reject the null-hypothesis of no ARCH effects for the full-sample and 96-00 sub-sample. The power pattern among MC and MMMC-N is similar, however, the power under MMMC-N is slightly higher. Furthermore, the p-values under the assumption of a t-distributed W seems to lack power. This is in accordance with the literature Section (2) stating that the performance under normality is superior when compared to other distributions. However, noticeable that for the skew normal distribution the power is higher. But, the difference is fairly low.

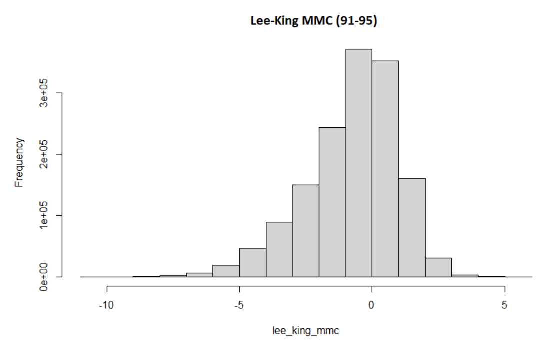
	Lee-King					
Sample	p_B	p_{mc}	p_{mmc}	p_{csmmc}	p_{mmmc-N}	p_{mmmc-t}
full	0.000*	0.001*	0.001*	0.001*	0.001 *	0.001*
61-65	0.038	0.631	0.665	0.625	1	1
66-70	0.022	0.414	0.453	0.397	1	1
71-75	0.094	0.88	0.883	0.866	1	1
76-80	0.000*	0.009*	0.013*	0.008*	0.821	0.874
81-85	0.054	0.721	0.767	0.748	1	1
86-90	0.06	0.767	0.819	0.779	1	1
91-95	0.002*	0.081	0.112	0.095	0.981	0.988
96-00	0.001*	0.015*	0.027*	0.023*	0.967	0.988

Table 6: Lee-King p-values, * indicates significant at 5%

The Lee-King p_B show a significantly different pattern compared to the Engle p_B. Only for the following sub-samples the null-hypothesis is not rejected, 71-75,81-85 and 86-90. The Lee-King performs well under a symmetric non-Gaussian distribution, see Section 2. There is a slight improvement in p-values for the MMC compared to MC. However, the MMMC-N and MMMC-t are large compared to the MC and MMC. The Lee-King distribution might under-reject the null-hypothesis when encountered for skewed distributions. A plot of the different distributions of the Lee-King statistic is given under different simulation techniques.

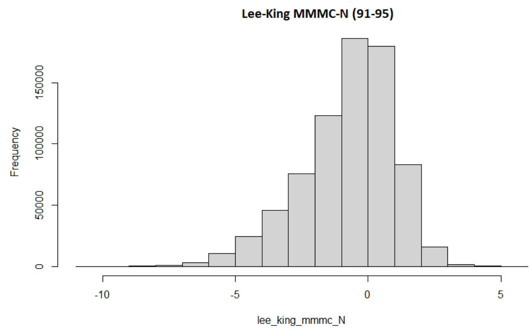


(a) Figure 5: 91-95, Lee-King MC

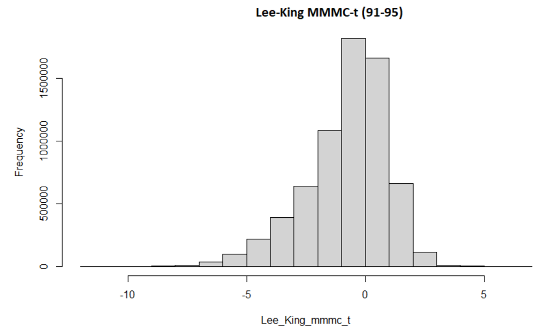


(b) Figure 5: 91-95, Lee-King MMC

The distribution seems to be left skewed. Lee-King should be $N(0, 1)$ distributed. This is not the case when comparing the plots. Possible under rejection is assumed in the non-skewed latter. The Lee-King statistic does not seem to fit the financial time-series used in this paper. The distribution does not follow a standard normal pattern. The results obtained concerning the Lee-King are in line with the results obtained in Dufour et al. (2010), therefore assumed to be correct.



(a) Figure 6: 91-95, Lee-King MMMC-N



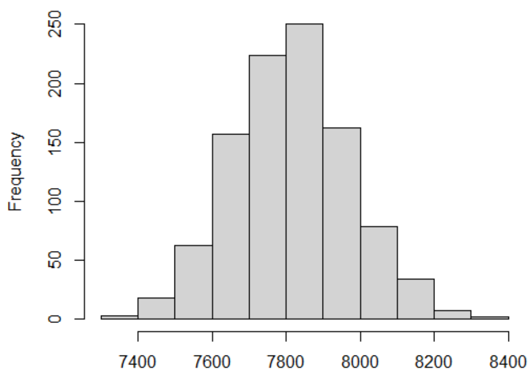
(b) Figure 6: 91-95, Lee-King MMMC-t

	Hoskings-ARCH					
Sample	p_B	p_{mc}	p_{mmc}	p_{csmmc}	p_{mmmc-N}	p_{mmmc-t}
full	0.000*	0.001*	0.001*	0.001*	0.001*	none
61-65	0.438	0.990	1	1	0.990	none
66-70	0.106	0.911	0.681	0.681	0.913	none
71-75	0.022	0.616	0.451	0.448	0.615	none
76-80	0.513	0.998	1	1	1	none
81-85	0.001*	0.281	0.330	0.330	0.310	none
86-90	0.005	0.469	0.509	0.509	0.492	none
91-95	0.091	0.848	0.844	0.844	1	none
96-00	0.000*	0.027*	0.046*	0.046	0.067	none

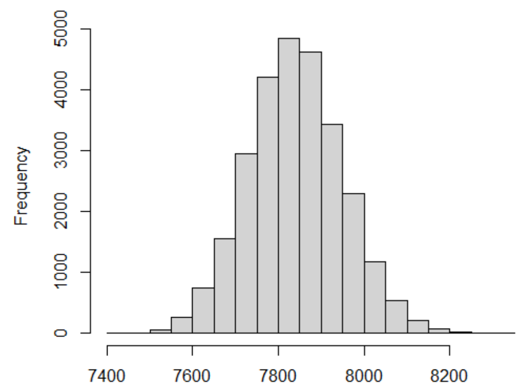
Table 7: Hosking-ARCH p-values, * indicates significant at 5%

The last test is the Hosking-ARCH test. The properties are fairly similar to the properties of the Hosking test (Hosking, 1980). According the p_B, the null-hypothesis is rejected for the full-sample and for the following sub-samples, 71-75, 81-85, 86-90 and 96-00. The different MC techniques provide the same pattern as for the Hosking test. Namely, the p-values is higher for the MC latter compared to the MMC. Furthermore, the MMMC-N provides the highest p-values. The plots of the different simulation techniques, statistical values is given below.

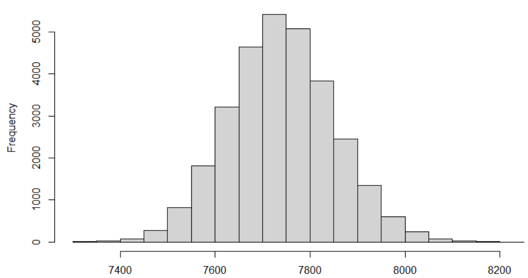
The Hosking-ARCH test is $\chi^2(n^2G)$ distributed. The distribution of the MMMC-N latter seems to follow the pattern the best. For the MC a slight right-skewness is detected. However, the simulation process is sparse in this case an empirical simulation study might fix this issue. The MMMC-N seems to follow the $\chi^2(n^2G)$ framework the most appropriately. Therefore, there is no evidence to question the test results based on the distribution of the test-statistics.



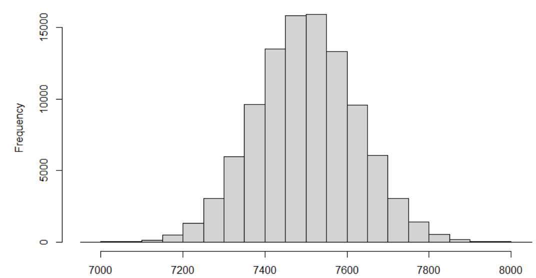
(a) Figure 7: Hosking-ARCH MC, 91-95



(b) Figure 7: Hosking-ARCH MMC, 91-95



(a) Figure 8: Hosking-ARCH MMMC-N, 91-95



(b) Figure 8: Distribution of the test statistic

6 Conclusion

A conclusion drawn in the Dufour et al. (2010) paper, is that the 3-factor Fama-French model with a t-distribution is an appropriate model for modelling the sub-samples. The results based on replication and extension differ from this conclusion.

Similar to the Dufour et al. (2010) paper, this paper uses the 3-factor Fama-French model as its linear independent variable model. This model seems to work appropriately and there is no need to differ from this model. However, a potential extension to this paper would be to replicate the data only under a 4-factor Carhartt model. The 4-factor Carhartt model is assumed to have a higher performance, according to Rehnby (2016).

The Bonferroni p-values seem to lack power, this approximation is poor. Furthermore, a noticeable difference is found between the Ljung-Box and Hosking test. Both these tests, test for the presence of linear serial dependence . However, the results change remarkably when asymmetric error-term distributions are being applied. The Ljung-Box test does not perform well under non-Gaussian conditions. This assumption is confirmed by this paper. However, under asymmetric conditions the p-value for the Hosking statistic do differ. The excess kurtosis might give an explanation for these high p-values.

For the ARCH-effect tests the results follow a similar pattern as the linear serial dependence test. The Engle test is a fairly simple and basic test and does not perform well under (asymmetric) non-Gaussian conditions.The Lee-King statistic results do differ extensively among different W. The Lee-King test performs well under non-Gaussian conditions, but the difference along MC and MMC is small. When asymmetric distributions are applied, the p-values change noticeably. The highest p-values are obtained for the skew t-distribution. However, when plotting the test statistic values, the distribution does not seem to follow a $N(0, 1)$ pattern. The flight from this property, might indicate that this test is not appropriate.

Finally, the Hosking-ARCH test, this test follows a similar pattern as the Engle test. The distribution of the test-statistic is in line with the assumption towards this issue. In order to answer the research question,

What happens to the test performance of residual based multi-equation mis-specification tests, when skewed distributions are applied?.

All the tests perform better when an asymmetric distribution is applied. However, the tests are not all designed for all asymmetric distributions or even MLR models. The test results should not be taken as definitive. Furthermore, the data used in this paper follows a non-stationary pattern. This does affect the results and should be solved first, before the residuals are applied on the data.

The conclusion drawn in Kroner and Ng (1998) is still valid. Currently, there are no appropriate tests for MLR models applied to financial time-series. Furthermore, a more extensive research can been done in the construction of a linear serial dependence and ARCH-effect test when skewed data is applied. This paper shows that, the test statistics either do not change or changes significantly, without a clear explanation.

Possible extensions on this paper might be the test provided in Li et al. (2019). Because, this test has sufficient power when n becomes large. Furthermore, a more advanced regression method can be applied to find the parameter coefficient, OLS is the most basic method to use. Finally, a possible non-linear approach can be applied. Several references provide research into possible non-linear time-series applications.

References

Ayadi, O. F., & Pyun, C. S. (1994). An application of variance ratio test to the korean securities market. *Journal of Banking & Finance*, *18*(4), 643–658.

Azzalini, A. (2022). *The R package `sn`: The skew-normal and relateddistributions such as the skew-t and the SUN (version 2.0.2).* Università degli Studi di Padova, Italia.

Azzalini, A., & Valle, A. D. (1996). The multivariate skew-normal distribution. *Biometrika*, *83*(4), 715–726.

Borchers, H. W., & Borchers, M. H. W. (2022). Package 'pracma'. *accessed on*, *4*.

Bormetti, G., Cisana, E., Montagna, G., & Nicrosini, O. (2007). A non-Gaussian approach to risk measures. *Physica A: Statistical Mechanics and its Applications*, *376*, 532–542.

Burns, P. (2002). Robustness of the Ljung-Box test and its rank equivalent. *SSRN Electronic Journal*.

Chan, J. C., & Hsiao, C. Y. (2014). Estimation of stochastic volatility models with heavy tails and serial dependence. *Bayesian Inference in the Social Sciences*, 155–176.

Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association*, *74*(366a), 427–431.

Duchesne, P., & Lalancette, S. (2003). On testing for multivariate ARCH effects in vector time series. *Canadian Journal of Statistics*, *31*, 275–292.

Dufour, J.-M., Khalaf, L., & Beaulieu, M.-C. (2010). Multivariate residual-based finite-sample tests for serial dependence and ARCH effects with applications to asset pricing models. *Journal of Applied Econometrics*, *25*(2), 263–285.

Dufour, J.-M., Khalaf, L., Bernard, J.-T., & Genest, I. (2004). Simulation-based finite-sample tests for heteroskedasticity and ARCH effects. *Journal of Econometrics*, *122*(2), 317–347.

Dufour, J.-M., & Khalaf, L. (2000). *Simulation-Based Finite and Large Sample Tests in Multivariate Regressions* (Cahiers de recherche No. 2000-10). Centre Interuniversitaire de Recherche en Economie Quantitative, CIREQ.

Engle, R. F. (1982). A general approach to Lagrange multiplier model diagnostics. *Journal of Econometrics*, *20*(1), 83–104.

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, *33*(1), 3–56.

Gupta, A. (2003). Multivariate skew t-distribution. *Statistics: A Journal of Theoretical and Applied Statistics*, *37*(4), 359–363.

Harrison, R. L. (2010). Introduction to Monte Carlo simulation. *AIP conference proceedings*, *1204*(1), 17–21.

Heij, C., de Boer, P., Franses, P., Kloek, T., & van Dijk, H. (2004). Econometric methods with applications in business and economics. *Oxford University Press*.

Hosking, J. R. M. (1980). The multivariate portmanteau statistic. *Journal of the American Statistical Association*, *75*(371), 602–608.

Jarque, C. M., & Bera, A. K. (1987). A test for normality of observations and regression residuals. *International Statistical Review/Revue Internationale de Statistique*, 163–172.

Komsta, L., & Novomestky, F. (2022). *Moments: Moments, cumulants, skewness, kurtosis and related tests.*

Kroner, K. F., & Ng, V. K. (1998). Modeling asymmetric comovements of asset returns. *The Review of Financial Studies*, *11*(4), 817–844.

Lee, J., & King, M. (1993). A locally most mean powerful based score test for ARCH and GARCH regression disturbances. *Journal of Business Economic Statistics*, *11*, 17–27.

Li, Z., Lam, C., Yao, J., & Yao, Q. (2019). On testing for high-dimensional white-noise. *The Annals of Statistics*, *47*, 3382–3412.

Ljung, G., & Box, G. (1978). On a measure of lack of fit in Time-Series models. *Biometrika*, *65*.

Miron, D., & Tudor, C. (2010). Asymmetric conditional volatility models: Empirical estimation and comparison of forecasting accuracy. *Romanian Journal of Economic Forecasting*, *13*(3), 74–92.

Raunig, B. (2004). Detecting Autoregressive Conditional Heteroskedasticity in Non-Gaussian Time-Series. *SSRN*.

Refenes, A., Bentz, Y., Bunn, D. W., Burgess, A. N., & Zapranis, A. D. (1997). Financial time-series modelling with discounted least squares backpropagation. *Neurocomputing*, *14*(2), 123–138.

Rehnby, N. (2016). Does the fama-french three-factor model and carhart four-factor model explain portfolio returns better than CAPM? : - A study performed on the swedish stock market.

Richardson, M., & Smith, T. (1993). A test for multivariate normality in stock returns. *Journal of Business*, 295–321.

Sewell, M. (2011). Characterization of financial time series. *Rn*, *11*(01), 01.

Sjölander, P. (2011). A stationary unbiased finite sample ARCH-LM test procedure. *Applied Economics*, *43*(8), 1019–1033.

Stewart, C., Kelly, T., & Zhang, A. (2007). Exploiting nonstationarity for performance prediction. *SIGOPS Oper. Syst. Rev.*, *41*(3), 31–44.

Tippett, L. H. C. et al. (1931). The methods of statistics. *The Methods of Statistics.*

Trapletti, A., & Hornik, K. (2022). *Tseries: Time series analysis and computational finance.*

7 Appendix

7.1 Jarque-Bera

	1	2	3	4	5	6	7	8	9	10	11	12
full-sample	0.00*	0.00*	0.00*	0.00*	0.00*	0.00*	0.00*	0.00 *	0.00*	0.00*	0.00*	0.00*
61-65	0.02*	0.11	0.70	0.42	0.80	0.99	0.85	0.37	0.35	0.69	0.21	0.99
66-70	0.75	0.73	0.02*	0.52	0.78	0.57	0.00 *	0.85	0.99	0.12	0.61	0.61
71-75	0.87	0.96	0.37	0.87	0.32	0.57	0.39	0.76	0.91	0.15	0.93	0.84
76-80	0.07	0.84	0.30	0.32	0.32	0.09	0.63	0.80	0.05*	0.84	0.37	0.29
81-85	0.02*	0.48	0.47	0.05*	0.36	0.09	0.44	0.26	0.34	0.76	0.65	0.20
86-90	0.49	0.72	0.950	0.00 *	0.20	0.18	0.57	0.39	0.80	0.89	0.62	0.03 *
91-95	0.19	0.62	0.16	0.77	0.27	0.30	0.46	0.83	0.20	0.79	0.31	0.94
96-00	0.01*	0.10	0.96	0.28	0.72	0.29	0.85	0.25	0.03*	0.94	0.22	0.24

13	14	15	16	17	18	19	20	21	22	23	24	25
0.00*	0.00*	0.00*	0.00*	0.00*	0.00*	0.00*	0.00*	0.00 *	0.00*	0.00*	0.00*	0.00*
0.56	0.62	0.56	0.57	0.80	0.30	0.15	0.00*	0.33	0.39	0.53	0.66	0.79
0.71	0.82	0.00*	0.00*	0.77	0.73	0.86	0.63	0.71	0.42	0.72	0.22	0.93
0.21	0.22	0.54	0.18	0.45	0.63	0.82	0.47	0.05*	0.20	0.26	0.63	0.88
0.47	0.12	0.00*	0.05*	0.26	0.99	0.00*	0.00 *	0.01*	0.63	0.00*	0.70	0.73
0.99	0.05 *	0.97	0.86	0.56	0.00*	0.28	0.61	0.44	0.75	0.99	0.83	0.50
0.57	0.46	0.41	0.78	0.74	0.15	0.51	0.37	0.66	0.72	0.22	0.79	0.24
0.76	0.31	0.79	0.38	0.34	0.47	0.51	0.32	0.67	0.58	0.64	0.89	0.88
0.83	0.21	0.92	0.06	0.22	0.58	0.23	0.43	0.10	0.16	0.57	0.00*	0.04*

Table 8: JB, test statistic, * indicates significant at 5%

7.2 Skewness

	1	2	3	4	5	6	7	8	9	10	11	12
full-sample	0.22	0.45	-0.04	0.30	0.02	0.05	0.39	-0.02	-0.02	0.14	0.07	0.33
61-65	0.53	0.65	-0.26	-0.05	-0.21	-0.05	-0.11	0.29	0.45	0.25	0.56	0.00
66-70	0.19	-0.16	-0.79	-0.28	0.22	-0.21	1.10	0.13	0.03	0.54	0.25	-0.27
71-75	0.05	-0.00	0.43	0.10	0.30	0.20	-0.33	-0.21	-0.05	-0.14	0.11	0.09
76-80	0.64	0.03	-0.09	0.48	0.45	0.59	0.15	0.17	-0.40	-0.19	-0.39	-0.08
81-85	0.62	0.36	0.06	0.69	0.38	0.62	-0.03	-0.43	-0.23	0.21	0.07	-0.57
86-90	-0.32	-0.23	-0.02	-0.59	-0.30	0.58	0.30	-0.33	0.18	0.01	0.18	0.56
91-95	0.58	0.25	0.61	0.23	0.45	0.33	-0.06	0.06	0.19	0.21	-0.47	-0.01
96-00	-0.06	0.64	0.03	0.30	-0.14	-0.23	0.15	0.38	0.31	-0.02	0.54	0.22

13	14	15	16	17	18	19	20	21	22	23	24	25
-0.06	0.33	0.55	0.51	0.45	0.56	0.10	0.07	-0.01	-0.06	0.41	0.37	0.09
-0.26	-0.18	-0.01	-0.32	-0.06	0.45	0.46	-0.66	-0.29	0.38	-0.12	0.05	-0.20
-0.25	0.20	1.44	-0.67	0.09	0.16	0.18	-0.30	0.26	-0.42	-0.23	0.41	0.09
-0.56	0.32	0.21	0.51	0.34	0.28	0.17	0.10	-0.67	0.56	-0.33	-0.18	0.11
0.04	-0.29	0.93	0.63	0.32	-0.02	-1.22	0.71	0.67	0.11	1.03	0.06	0.00
0.02	0.75	-0.08	0.04	-0.34	0.80	-0.47	-0.04	0.38	-0.06	-0.03	0.19	0.18
0.22	-0.38	0.35	-0.10	-0.14	-0.07	0.28	-0.30	0.28	0.24	0.51	-0.22	-0.27
-0.09	0.20	0.02	-0.11	0.44	0.21	-0.29	0.48	0.16	-0.04	-0.18	-0.11	0.04
0.08	0.41	0.11	0.63	0.54	0.31	0.49	0.40	-0.47	-0.15	-0.19	0.83	0.46

Table 9: Skewness

7.3 Kurtosis

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
full-sample	6.72	4.84	3.94	3.56	4.55	4.37	4.51	5.68	5.24	3.74	4.00	7.96	5.89	6.60	4.23	6.08	4.90	6.69	4.41	3.70	3.59	4.87	6.11	5.25	4.27
61-65	4.48	3.24	3.07	2.17	3.00	2.99	2.73	3.68	3.16	2.77	3.02	2.94	3.43	3.50	3.68	3.19	2.59	2.60	3.83	4.96	3.75	3.42	2.33	3.56	2.81
66-70	2.71	3.39	3.91	2.55	2.94	2.48	6.33	3.26	3.07	3.75	2.62	2.66	2.86	2.95	7.33	4.78	3.42	2.62	3.01	3.11	2.92	2.98	2.78	3.73	2.83
71-75	2.68	2.82	3.21	2.73	3.75	3.54	3.57	2.79	3.26	4.20	3.12	2.67	3.11	3.90	2.43	2.42	2.58	2.76	2.79	2.25	3.77	2.95	3.81	3.48	2.77
76-80	3.71	2.62	2.04	2.96	3.32	3.76	3.52	2.75	4.28	2.99	2.57	2.02	3.77	4.17	4.18	3.86	3.81	3.10	7.54	4.51	4.39	2.43	5.80	2.49	2.50
81-85	4.27	2.74	2.23	3.65	2.52	3.67	2.19	3.58	2.19	2.79	2.43	2.95	3.05	3.44	3.01	3.34	2.87	4.91	3.36	2.38	2.74	2.53	2.95	2.90	2.35
86-90	3.38	3.22	3.20	4.84	3.97	3.16	3.30	2.44	2.79	3.30	2.49	4.30	2.49	2.78	3.47	3.41	3.39	4.22	2.53	3.67	2.86	3.17	2.58	3.02	3.92
91-95	3.03	2.64	3.03	3.10	2.53	2.26	2.23	2.63	1.94	3.11	2.72	3.23	3.44	2.12	2.56	3.85	3.32	3.66	2.53	3.05	2.54	2.35	2.52	2.81	3.31
96-00	5.02	3.44	2.83	3.81	3.43	3.88	2.79	3.72	4.59	3.22	3.20	3.97	2.65	3.77	2.85	3.78	2.84	3.22	3.46	2.79	3.96	4.18	2.45	6.39	4.34

Table 10: Kurtosis

7.4 Code

The Zip-file containing the code is included. Several files containing code are present, each file has its own simulation process. The first and upmost file is the $lb-setting$ file, here the regression based on the observed data is done and the residuals are created. Each simulation or Bonferroni p-value is constructed by separate files accordingly. Note that R version 4.0.5 or higher is needed in order for packages to run.