Erasmus University Rotterdam

Erasmus School of Economics

Bachelor Thesis BSc2 Econometrics and Economics

Asymmetric GARCH effects and residual diagnostics: A finite-sample testing approach for multivariate linear models

Author:

Nancy Gjini (499332)

Supervisor:

Dr. Mikhail Zhelonkin

Second assessor:

M.F.O. (Max Felix Orson) Welz

1 July 2022

Abstract

This paper investigates the finite-sample residual-based tests for ARCH effects and serial dependency designed by Dufour, Khalaf, and Beaulieu (2010), through an application to the Fama-French Three-Factor Model (Fama & French, 1993), and a simulation experiment on assessing the performance of the tests against asymmetric GARCH effects for Gaussian errors. The empirical application reveals an acceptable model for 5-year sub-periods when allowing both Gaussian and non-Gaussian errors, whereas the full sample model exhibits problematic residual diagnostics. The simulation study shows that the asymmetry does not affect the performance of the tests when compared to the symmetrical case. It however indicates a low power of the proposed tests when allowing both symmetric and asymmetric GARCH effects, especially EGARCH, which are commonly present in asset pricing models such as Fama-French, suggesting the case of under-rejection in the empirical application. Coming into such problematic conclusions on residual diagnostics can be prevented by first testing for (asymmetric) GARCH effects in the residuals.

Table of Contents

1 Introduction

Multivariate linear regression analysis is critical for modeling relationships between variables in various disciplines in quantitative finance. It aids asset valuation and understanding of correlations between variables such as commodity prices and the stocks of companies trading in those commodities, for investment and financial managers. The application of such models in asset pricing, specifically through factor models, is the subject of this research. Factor models are financial models that use statistical, fundamental, or macroeconomic explanatory factors to define a group of securities' risks and returns. The most well-known factor model is the Fama-French Three-Factor Model (Fama & French, 1993), which incorporates three effects when describing returns: (1) market excess return, (2) the outperformance of small versus big companies, and (3) the outperformance of high book/market versus low book/market companies.

The model fit and the explanatory power of all three factors have limitations. For instance, among the drawbacks of the Fama-French Three-Factor Model are the country-specificity (Griffin, 2002) and endogeneity (Allen & McAleer, 2021). On the other hand, research also reveals that, due to time-varying volatility, residuals of the model typically exhibits asymmetric generalized autoregressive conditional heteroskedasticity (GARCH) effects (Kalev & Zolotoy, 2012; Zhao, Stasinakis, Sermpinis, & Fernandes, 2019). The asymmetry results from the observance of the asymmetric volatility phenomenon (AVP), which explains that market volatility tends to be higher in declining markets than in rising markets due to the 'leverage effect' (Bekaert & Wu, 2000; Wu, 2001).

Discussions of this kind highlight the necessity for proper model misspecification techniques applicable to financial multivariate linear models that exhibit asymmetric conditional volatility among residuals, i.e. asymmetric GARCH effects. As a result, this research intends to aid in developing valid residual-based tests for multivariate linear regressions, based on the work of Dufour et al. (2010). Dufour et al. (2010) describe how to properly apply existing tests in the multivariate regression scenarios, by considering the assumption of cross-equation correlated disturbances. Prior to their study, there were no valid residual-based tests of high power for multivariate linear regressions, as either combined equation-by-equation tests or multivariate tests based on asymptotic distributions were applied to such models, without considering correlation among error terms and the finite sample size. The former tests were based on Bonferroni-type bounds, thus reducing of the test for a large number of equations, which is usually the case in practice. The existing multivariate portmanteau criteria indeed incorporate the cross-equation dependence, however, this only holds in the asymptotic case where there are no nuisance parameters in the disturbances' distribution — parameters which must be accounted for but are not of immediate interest.

The procedure utilizes the Cholesky-standardized residuals, which do not depend on the unknown variance matrix from the residual distribution (i.e. do not depend on unknown regression coefficients or the error covariance matrix), therefore solving the nuisance parameter problem. These residuals are applied to Bonferroni-type tests, asymptotic-based tests, and (augmented) Monte Carlo simulated tests. The latter's implementation derives exact p-values for each test on finite samples, based on asymptotic distributions.

This paper aims to evaluate the adequate diagnostic tools to examine the underlying assumptions of financial models with asymmetric conditional volatility of the disturbances by reproducing and extending on the paper of Dufour et al. (2010). The main research question is therefore as follows:

"*What impact does the addition of asymmetry on ARCH effects have on the performance of the multivariate residual-based finite-sample tests?*"

In the first part, the proposed tests are applied to the Fama-French Three-Factor Model in the 1966-2000 period and 5-year sub-periods. The residual diagnostics results using Bonferroni-type, asymptotic and simulation-based reveal that the Fama-French model is only a good fit for the sub-periods, as the regression over the full sample violates the two assumptions on the disturbances.

Secondly, a simulation study which extends that of Dufour et al. (2010) is executed, by introducing asymmetric GARCH effects. Accordingly, this simulation study examines the behavior of the constructed test statistics under more realistic financial scenarios, thus answering the research question. In particular, the size and the power of the tests in the presence of asymmetric GARCH effects are observed in sample designs similar to the empirical application of the Fama-French Three-Factor models. The volatility models are based on asymmetric power GARCH (APARCH) (Ding, Granger, & Engle, 1993) GJR-GARCH (Glosten, Jagannathan, & Runkle, 1993), and exponential GARCH model (EGARCH) (Nelson, 1991). The study reveals a similar performance of the tests for asymmetric volatility models as in the symmetric models. However, in all cases, the tests are not sufficiently powerful, especially in the EGARCH case, implying a lower performance in more realistic financial scenarios. This also suggests the dubious performance of the tests in the Fama-French model.

The remainder of the paper develops as follows. Section 2 provides a comprehensive overview of the relevant literature. Section 3 outlines the data used in the empirical application, and Section 4 describes the statistical methodology used in the application and the simulation study. Section 5 reports the findings, which are discussed and concluded in Section 6.

2 Literature Review

2.1 Asset Pricing Models and Asymmetric Volatility

The capital asset pricing model (CAPM), introduced by Sharpe (1964) and Lintner (1975) is the classical product of modern portfolio theory, which gave rise to these models. It uses market return as the basis for explaining stock returns. By employing the cross-section, Fama and French (1992) demonstrate the empirical shortcomings of the CAPM and how size, earning-price, and book-to-market ratios supplement the explanation of predicted stock returns offered by market beta. They propose the Three-Factor Model (Fama & French, 1993), the primary substitute for CAPM. Several extensions have been made to this model, such as the addition of the 'momentum' factor by Carhart (1997), yet the three factors are regarded as standard in academic research, and empirical investigations frequently adopt their definitions (Blitz, 2020). Celik (2012) provides a more thorough review of the literature on asset pricing models.

Zhou and Li (2016) prove a higher explanatory power of the model when incorporating GARCH-type volatility as opposed to the original version. Alternatively, Lozano (2017) studies a new explanatory factor additional to the ones of Fama and French (1993) — the volatility of the return — and shows a significant negative effect. The latter is explained by the asymmetric volatility phenomenon, which has been proven to occur when there is a negative correlation between contemporaneous returns and conditional return volatility (Wu, 2001). Accordingly, it has been shown that non-linear asymmetric extensions of volatility models outperform the traditional GARCH (Engle, 1982) in explaining and forecasting conditional volatility (Franses, Van Dijk, et al., 2000; Pagan & Schwert, 1990; Brailsford & Faff, 1996; Loudon, Watt, & Yadav, 2000). Such models include the Exponential GARCH (EGARCH) model by Nelson (1991), the so-called GJR model (GJR-GARCH) by Glosten et al. (1993) and the Asymmetric Power GARCH (APARCH) model by Ding et al. (1993). For the above reasons, it is believed that the Fama-French Three-Factor model best describes the returns when incorporating asymmetric GARCH effects.

2.2 Multivariate Residual Diagnostics

The analysis of regression residual diagnostics has become a common procedure in model specification, as it is of high importance to look into the assumptions on residuals to validate the model in use. Consequently, tests for normality, homoskedasticity and serial independence among residuals have been proposed in the literature for decades. Such tests include the Jarque-Bera normality test (Jarque & Bera, 1980), Ljung-Box test (Ljung & Box, 1978) for serial autocorrelation, Breusch-Pagan Lagrange Multiplier test (Breusch & Pagan, 1979) for detecting heteroskedasticity along with many others. They mainly apply to univariate data, i.e. when a single outcome variable is present in the regression, leaving no valid tests of high power for the multivariate case.

In the paper "Modeling Asymmetric Comovements of Asset Returns", Kroner and Ng (1998) point out the scarcity of the literature on multivariate linear regressions (MLR), although being one of the most widely used models in various areas. This leads to the deficient use of univariate tests on MLR, by simply applying them across equations in the presence of cross-equation correlations. Consequently, Dufour et al. (2003; 2010) identify the problems that arise in the case when asymptotic approximations or Bonferroni-type bounds are used, and propose joint specification testing on (i) normality, (ii) serial autocorrelation, and (iii) heteroskedasticity, using Monte Carlo simulation methods and Cholesky- standardized residuals to extend on the already existing tests. Since then, there have been developments in the field of multivariate testing using simulation methods, such as in detecting (symmetric) ARCH/GARCH effects via Lagrange Multiplier tests (Gel & Chen, 2012; Catani & Ahlgren, 2017), or developing serial dependence and predictability tests (Dufour, Khalaf, & Voia, 2015).

2.3 Specification Tests for Asymmetric GARCH effects

Despite their widespread application in financial models, there is little research on how well multivariate tests function when applied to models with asymmetric GARCH effects. Only specification tests for the detection of asymmetric GARCH effects against symmetric GARCH effects have been constructed, such as the Sign Bias, Negative Size Bias, Positive Size Bias tests, as well as joint versions of the three (Engle & Ng, 1993). Hagerud et al. (1997) then develop two univariate Lagrange Multiplier specification tests with a null of symmetric against asymmetric GARCH with more power advantages. Akpan and Moffat (2017) support the latter, by showing that LM-type tests perform better in detecting asymmetric GARCH effects in discrete time-series models. On the other hand, multivariate portmanteau goodness-of-fit tests using the autocovariances of the squared residuals with asymmetric volatilities have been constructed (Carbon & Francq, 2011; Francq, Wintenberger, & Zakoïan, 2018), but the simulation experiment of Brooks and Henry (2000) demonstrates that the tests are unable to detect GJR- and EGARCH effects. Additionally, Ben and Jiang (2020) note that the limiting distribution of the test does not hold in the presence of boundary parameters. Moreover, reasonable power is only obtained using very large sample sizes.

All the above tests face the issues mentioned in Section 2.2, and, instead of having the null of no ARCH effect whatsoever, they test against the null of symmetric GARCH effect. Consequently, valid multivariate tests for asymmetric volatilities in multivariate models have yet to be developed, therefore this study applies the finite-sample tests suggested by Dufour et al. (2010) in detecting asymmetric GARCH effects.

3 Data

The dataset of this study, employed in the empirical application, consists of panel data extracted directly from the Fama and French website (French, n.d.). The CRSP database, managed by the University of Chicago's Booth School of Business, provides the Fama-French variables.

The Fama-French data contains historical monthly returns of 25 portfolios, from January 1961 to December 2000, and three explanatory factors expressed in returns, namely (i) the excess market return (market), (ii) size premium (SMB), and (iii) value premium (HML). All returns, denoted in U.S. dollars, are not continuously compounded and include dividends and capital gains.

The 25 portfolios derive from the intersections of five portfolios sorted on size with market equity (ME) as a proxy, and five portfolios based on value, measured by the book equity over market equity (BE/ME). The construction of these portfolios occurs every June. The book equity is measured at the end of the previous fiscal year, whereas market equity is equal to the price multiplied by the number of shares outstanding at the end of December of the previous year. The size and value breakpoints are the New York Stock Exchange (NYSE) 25th, 50th, 75th, and 100th percentiles of ME and BE/ME, respectively.

Figure 1 displays the equally weighted average portfolio return for all 25 against the three factors in the whole sample size. All four of them show similar characteristics; they are stationary processes centered around 0, which may point to a high explanatory power of the three factors.

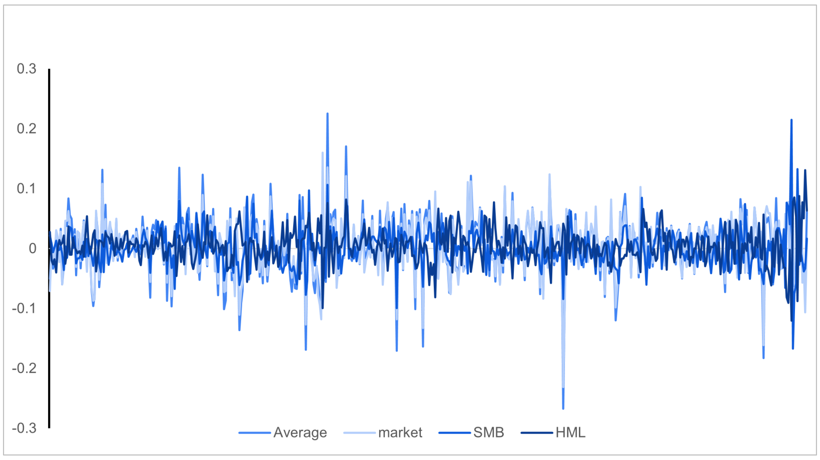

Figure 1: Average portfolio return from Jan. 1961 to Dec. 2000

The first factor amounts to the difference between the market return (R_{tM}) and the one-month Treasury bill rate at month t (R_{tf}), which is risk-free, retrieved from Ibbotson Associates. The market return equals the value-weighted return on all NYSE, AMEX, and NASDAQ stocks at month t.

French constructs the following two factors based on six portfolios, two formed on book size and three on market value. The size breakpoint is the median NYSE market equity at the end of June, whereas the BE/ME breakpoints are the 30th and 70th NYSE percentiles. The resulting portfolios are: Small Value (SV), Small Neutral (SN), Small Growth (SG), Big Value (SV), Big Neutral (SN), and Big Growth (SG).

The second factor, known as the size factor (SMB_t), is the average of the returns on the three small stock portfolios minus the average of the returns on the three big stock portfolios at month t. The size factor is essentially the excess return that smaller market capitalization companies return versus larger companies, also known as the “small firm effect” - smaller companies outperform larger ones over the long-term (De Moor & Sercu, 2013). Mathematically, it is defined as:

$$SMB_t = \frac{SV_t + SN_t + SG_t}{3} - \frac{BV_t + BN_t + BG_t}{3}. \tag{1}$$

Finally, the third factor, the value factor (HML_t), is defined as the equal-weight average of the returns for the two high BE/ME portfolios minus the average of the returns for the two low BE/ME portfolios at month t. It argues that companies with high book-to-market ratios, also known as value stocks, outperform those with lower book-to-market values, known as growth stocks. It amounts to

$$HML_t = \frac{SV_t + BV_t}{2} - \frac{SG_t + BG_t}{2}. \tag{2}$$

4 Methodology

The first part of this study is similar to the approach presented in Dufour et al. (2010), containing the introduction of the multivariate residual-based residual tests applied to the Fama-French Three-Factor Model. The second part consists of the extension of the simulation study for the performance of the ARCH effects tests when considering asymmetric conditional volatilities among residuals.

4.1 Fama-French Three-Factor Model

Using the three already constructed Fama-French factors ($R_{tM} - R_{tf}, SMB_t, HML_t$), perform a Multivariate Linear Regression (MLR) on the (excess) monthly returns of the 25 portfolios, formulated as follows:

$$R_{ti} - R_{tf} = \alpha_{ti} + \beta_1(R_{tM} - R_{tf}) + \beta_2 SMB_t + \beta_3 HML_t + \varepsilon_{ti}, \tag{3}$$

where R_{ti} denotes the historical portfolio return, R_{tf} the risk-free return, R_{tM} the market return, $\beta_1, \beta_2, \beta_3$ the factor coefficients, and ε_{ti} the error term for $t = 1, ..., T$ observations and $i = 1, ..., 25$ portfolios. The general form of the multivariate regression of the model, referred to from this point on, is

$$Y = XB + U, \tag{4}$$

where $Y = [R_1, R_2, ..., R_{25}]$ is a $T \times 25$ matrix with each column representing the i-th portfolio return for T observations across time, $X = [\iota, R_M - Rf, SMB, HML]$ is a $T \times 4$ full-column rank matrix containing vector of ones and the three factors, B is a 4×25 matrix of unknown factor coefficients, and $U = [U_1, ..., U_T]^\top$ is a $T \times 25$ matrix of random error terms.

The model is estimated by system OLS, as opposed to equation-by-equation, as the disturbances across the system can be correlated (Heij et al., 2004). The Gaussian-based quasi-maximum likelihood estimators for the coefficient matrix B and the error variance-covariance matrix Σ are

$$\hat{B} = (X^\top X)^{-1} X^\top Y \text{ and} \tag{5}$$

$$\hat{\Sigma} = \frac{1}{T} \hat{U}^\top \hat{U}, \tag{6}$$

where $\hat{U} = [\hat{u}_1, ..., \hat{u}_n] = [\hat{U}_1, ..., \hat{U}_T]$ represents the residual matrix.

The distribution of $\hat{U}$ depends on the unknown covariance matrix Σ, which sets the 'scale', allowing the presence of nuisance parameters. More details on the problems arising from the latter can be found in Appendix A. The following sections introduce a new method for residual standardization that deals with nuisance parameters, followed by the application of these residuals to obtain exact joint tests.

4.2 Location-scale Invariant Residuals

Dufour et al. (2010) construct a rescaled residual matrix, referred to as the Cholesky-standardized residual matrix ($\tilde{W}$), which they prove to not be influenced by the unknown regression coefficients (B) nor the error covariance matrix (Σ). This conclusion has crucial implications for the residual diagnostic tests investigated, such that, in the case of finite samples, simulation-based procedures for the joint tests, introduced in Sections 4.3 and 4.4, yield proper p-values. The steps taken to attain $\tilde{W}$ are as follows.

Define the residuals as $\hat{W} = MW$, where M denotes the projection matrix and W the fundamental data generating process (DGP). This result derives from $U = Y - \hat{Y} = Y - (X^\top X)^{-1}X^\top Y = (I - (X^\top X)^{-1}X^\top)Y$. The Cholesky-standardized residual matrix equals

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

where $S_{\hat{W}}$ is the Cholesky factor of $\frac{1}{T}\hat{W}^\top\hat{W}$. The DGP for all observations, denoted as $vec(W_1, ..., W_T)$, follows a multivariate random distribution that is either (1) fully specified, or (2) only dependent on one nuisance parameter κ. This study examines the cases for the following two distributions of $vec(W_1, ..., W_T)$, respectively:

$$\begin{aligned} &vec(W_1, ..., W_T) \sim N(0, 1), \text{ and} \\ &vec(W_1, ..., W_T) \sim t(\kappa). \end{aligned} \tag{8}$$

The invariance of $\tilde{W}$ holds in the sense that Equation 7 is equivalent to the multivariate Cholesky-standardized residual matrix using the estimated variance-covariane matrix of the residuals $\hat{\Sigma}$:

$$\tilde{W} = \hat{U} S_{\hat{U}}^{-1}, \tag{9}$$

where $S_{\hat{U}}$ refers to the upper triangular Cholesky factor of $\hat{U}$. $\tilde{W}$, as defined in Equation 7, is not a function of $\hat{U}$, but rather of the hypothesized distribution matrix. This implies that the Cholesky-standardized multivariate residuals are location-scale invariant, as they do not depend on neither J nor B through the estimated residual matrix $\hat{U}$.

4.3 Multivariate Specification Tests

4.3.1 Joint Equation-by-equation Tests

The model specification analysis of this paper first considers univariate tests, which are applied to the residuals of each equation i, for $i = 1, ..., n$. The tests in consideration measure (i) serial autocorrelation, and (ii) autoregressive conditional heteroscedasticity (ARCH) effects.

The procedure can be applied to any relevant univariate test, however, for illustration purposes, the two test statistics included to possibly detect the presence of serial autocorrelation are Ljung-Box (Ljung & Box, 1978) and Variance Ratio (Lo & MacKinlay, 1988). The Ljung-Box test statistic is

$$LB_i = T(T+2)\sum_{g=1}^{G}\frac{\hat{\rho}_{gi}^2}{T-g}, \quad \hat{\rho}_{gi} = \frac{\sum\limits_{t=g+1}^{T}\hat{u}_{ti}\hat{u}_{t-g,1}}{\sum\limits_{t=1}^{T}\hat{u}_{ti}^2}, \tag{10}$$

where T refers to the sample size, G the maximum number of lags, and $\hat{\rho}_{gi}$ is the g-th order autocorrelation of the residuals in equation i. The variance ratio two-sided test statistics takes the form:

$$VR_i = 1 + 2\sum_{g=1}^{G}\left(1-\frac{g}{G}\right)\hat{\rho}_{gi}. \tag{11}$$

The null hypothesis H_0 for both tests is the absence of such autocorrelation, and the asymptotic distributions of the statistics are: $LB_i \overset{\text{asy}}{\sim} \chi^2(G)$, and $VR_i \overset{\text{asy}}{\sim} N[1\ ,2(2G-1)(G-1)/3G]$.

Subsequently, the univariate test statistics for ARCH effects considered in this study are Engle (Engle, 1982) and Lee King (Lee & King, 1993), which test against the null of no ARCH effects. The test based on the Engle-type procedure for equation i is given by

$$E_i = T \ \times \ (R^2 \text{ of the regression of } \hat{u}_{ti}^2 \text{ on a constant and } \hat{u}_{t-g,1}^2), \quad g = 1, \dots G. \tag{12}$$

Lastly, Lee-King's one-sided statistic for equation i is determined as follows:

$$LK_i = \frac{\left\{(T-G)\sum\limits_{t=G+1}^{T}[(\hat{u}_{ti}^2/\hat{\sigma}_i^2)-1]\sum\limits_{g=1}^{G}\hat{u}_{t-g,1}^2\right\} \Big/ \left\{\sum\limits_{t=G+1}^{T}[(\hat{u}_{ti}^2/\hat{\sigma}_i^2)-1]^2\right\}^{1/2}}{\left\{(T-G)\sum\limits_{t=G+1}^{T}\left(\sum\limits_{g=1}^{G}\hat{u}_{t-g,i}^2\right)^2 - \left(\sum\limits_{t=G+1}^{T}\sum\limits_{g=1}^{G}\hat{u}_{t-g,i}^2\right)^2\right\}^{1/2}}, \tag{13}$$

where $\hat{\sigma}_i^2 = \frac{1}{T}\sum\limits_{t=1}^{T}\hat{u}_{ti}^2$ refers to the variance of the error terms in equation i.

The location-scale invariance property of Cholesky-standardized residuals allows obtaining standardized versions of these statistics, which are then combined to form joint test statistics. Algorithm 1 summarizes this procedure. For presentation clarity, the algorithm refers to any test statistics for equation i as TS_i.

Algorithm 1: Joint test statistic algorithm

1. Compute the standardized test statistic TS_i for each equation $i = 1, ..., n$ by replacing u_{ti} with the standardized residuals $\tilde{w}_{ti}$.
2. Obtain the p-value for each equation either through their asymptotic distribution, e.g. $\tilde{LB}_i \overset{\text{asy}}{\sim} \chi^2(G)$, or via the MC methods as described in Section 4.4.
3. Construct the combined statistics as $\tilde{TS} = 1 - \min_{1 \leq i \leq n} [p(\tilde{TS}_i)]$. The statistic rejects H_0 if at least one of the individual standardized tests is significant.

4.3.2 Multivariate Portmanteau Criteria

The study extends by considering two existing multivariate portmanteau statistics additional to the univariate test statistics, which are commonly applied to multivariate linear models. The first test is the serial-dependence statistic derived by Hosking (Hosking, 1980), and the second is an extension to Hosking's statistic proposed by Duchesne and Lalancette (2003), designed to detect ARCH effects. Both tests make use of the autocovariance matrix for a given $T \times n$ matrix $Z = [Z_1, ..., Z_T]^\top$ for all lags g:

$$C_{\hat{Z}}(g) = T^{-1} \sum_{t=g+1}^{G} \hat{U}_t \hat{U}_{t-g}^\top. \tag{14}$$

The Hosking statistic (HM) and its extension, referred to as Hosking-ARCH (HM_2), are:

$$HM = T^2 \sum_{g=1}^{G} (T-g)^{-1} \text{tr}\{C_{\hat{U}}^{-1}(0) C_{\hat{U}}(g) C_{\hat{U}}^{-1}(0) C_{\hat{U}}^\top(g)\}, \quad \text{and} \tag{15}$$

$$HM_2 = T^2 \sum_{g=1}^{G} (T-g)^{-1} \text{tr}\{C_{\hat{U}^2}^{-1}(0) C_{\hat{U}^2}(g) C_{\hat{U}^2}^{-1}(0) C_{\hat{U}^2}^\top(g)\} \tag{16}$$

Dufour et al. (2010) prove the location-scale invariance of the Hosking statistic (HM) when replacing the estimated residual matrix $\hat{U}$ with the Cholesky-standardized residual matrix $\tilde{W}$. Moreover, the invariance of the extended Hosking statistic holds when $\hat{U}^2$ replaces $\tilde{W}^2$, which contains the squared standardized residuals. Therefore, I make use of the modified location-scale invariant multivariate portmanteau criteria:

$$\tilde{HM} = T^2 \sum_{g=1}^{G} (T-g)^{-1} \mathrm{tr}\{C_{\tilde{W}}^{-1}(0) C_{\tilde{W}}(g) C_{\tilde{W}}^{-1}(0) C_{\tilde{W}}^{\top}(g)\}, \text{ and} \tag{17}$$

$$\tilde{HS}_2 = T^2 \sum_{g=1}^{G} (T-g)^{-1} \mathrm{tr}\{C_{\tilde{W}^2}^{-1}(0) C_{\tilde{W}^2}(g) C_{\tilde{W}^2}^{-1}(0) C_{\tilde{W}^2}^{\top}(g)\}. \tag{18}$$

The following standard asymptotic null distributions hold: $HM \overset{\text{asy}}{\sim} \chi^2(n^2G)$ and $HS_2 \overset{\text{asy}}{\sim} \chi^2(n^2G)$.

Table 1 provides an overview of all multivariate and univariate tests used to test for ARCH effects and serial dependence among residuals.

Table 1: Overview of the residual-based multivariate and univariate tests for serial dependence and ARCH effects

	Notation	**Type**	**Rejection type**	**Asymptotic distribution**
H_0: No ARCH effects				
Ljung-Box	LB	univariate	one-sided	$\chi^2(n^2G)$
Variance ratio	VR	univariate	one-sided	N [1, 2(2G - 1)(G - 1)/ 3G]
Hosking	HM	multivariate	one-sided	$\chi^2(n^2G)$
H_0: No serial-correlation				
Engle	E	univariate	one-sided	$\chi^2(n^2G)$
Lee-King	LK	univariate	two-sided	N [0,1]
Hosking-Arch	HS_2	multivariate	one-sided	$\chi^2(n^2G)$

4.4 Simulation-based Testing

4.4.1 Monte Carlo

I first assume a multivariate standard normal distribution for $vec(W_1, ..., W_T)$, in which nuisance parameters are absent, as is fully specified. I implement Algorithm 2 outlined by Dufour et al. (2010), to obtain the Monte Carlo (MC) p-value for each test. This technique provides exact simulation-based tests based on any statistic that follows an unknown finite-sample distribution. The number of replications N only influences the power of the tests and not the validity. However, the power gains are typically rather small for lengthy simulations (Dufour, 2006).

Algorithm 2: Monte Carlo test

1. Compute $\tilde{TS}$ through the given data and denote it $\tilde{TS}^{(0)}$.
2. Generate N standard normal realizations of W; denote the drawn variates $W^{(j)}, j = 1, ..., N$. Ensure $\alpha(N+1)$ is an integer, e.g. $N = 999$.
3. For each realization, calculate $\hat{W}^{(j)} = MW^{(j)}$, $S_{\hat{W}}^{(j)}$, the Cholesky factor of $\frac{1}{T}\hat{W}^{(j)\top}\hat{W}^{(j)}$, and $\tilde{W}^{(j)} = \hat{W}^{(j)}(S_{\hat{W}}^{(j)})^{-1}$.
4. Compute

$$\tilde{TS}^{(j)} = 1 - \min_{1 \leq i \leq n} [p(\tilde{TS}_i^{(j)})],$$

where $TS_i^{(j)}$ refers to the simulated statistic for equation i and the MC draw j.
5. Compute $P_N(\tilde{TS}^{(0)})$, which refers to the number of simulated values greater than or equal to $\tilde{TS}^{(0)}$. The MC p-value is

$$\hat{p}_N(\tilde{TS}) = [P_N(\tilde{TS}^{(0)}) + 1] \ / \ (N+1).$$

6. The null hypothesis is rejected at level α when $\hat{p}_N(\tilde{TS}) \leq \alpha$.

Note. This is the procedure for obtaining MC-based p-values as described by Dufour et al. (2010).

4.4.2 Maximized Monte Carlo

The underlying assumption of the MC method is the absence of nuisance parameters in the distribution, thus limiting its applicability. Consequently, in a more realistic scenario, one can consider the altered MC algorithm, namely Maximixed Monte Carlo (MMC), with the advantage of allowing for the presence of one nuisance parameter in the null distribution of the test statistic (Dufour, 2006). Here, the errors follow a Student-t distribution, i.e. $W_1, W_2, ..., W_N \sim \tau(\kappa)$ so that the nuisance parameter is the total degrees of freedom (κ) accounted for.

This procedure simulates functions for p-values dependant on a single nuisance parameter under the null. Dufour (2006) shows that, no matter the sample size T or the number of replications N utilized, taking the maximum p-value for all nuisance parameter results in a test with an exact level.

The Maximized Monte Carlo procedure highly resembles Monte Carlo as described in Algorithm 3. The procedure for the MMC algorithm, in short, is as follows:

Algorithm 3: Maximized Monte Carlo test

1. Perform the Monte Carlo algorithm for a given test statistic $\tilde{TS}$, where the observations $W_1, ..., W_T \sim t(\kappa_0 = 2)$ instead of standard normal.
2. Obtain the modified MC p-value $\hat{p}_N(\tilde{TS}|\kappa_0)$.
3. Repeat the process for $\kappa_j \in \{3, ..., T - n - 2\}$, and obtain the respective MC p-values.
4. Compute the MMC p-value as

$$\sup_{\kappa}[\hat{p}_N(\tilde{TS}|\kappa)], \tag{19}$$

and reject the null hypothesis when this value is $\leq \alpha$.

Note. This is the procedure for obtaining MC-based p-values as described by Dufour et al. (2010).

Lastly, I apply an additional modification to the MMC procedure to obtain exact test statistics, particularly the Confidence Set MMC (CSMMC) method, proposed by Dufour and Khalaf (2003), which incorporates goodness-of-fit of the error distribution under the hypothesis.

This modification consists of two parts:

1. Run the MMC algorithm only for κ that are included in the exact confidence set $CS(\kappa)$ with significance level α_1, which is essentially a subset of $\{2, 3, ..., T - n - 2\}$. Please refer to Appendix B for the procedure to obtain the confidence set.

2. Reject the null hypothesis when $\sup_{\kappa \in CS(\kappa)} [\hat{p}_N(\tilde{TS}|\kappa)] \leq \alpha - \alpha_1$.

4.5 Simulation Study for Asymmetric GARCH Effects

Dufour et al. (2010) present a simulation experiment to assess the power and the size of their proposed ARCH effects joint tests. They do so by first introducing ARCH(1), and GARCH(1,1) effects to the covariance and residual matrix with (1) standard normal and (2) Student-t errors, and applying MC- and MMC-based ARCH effect tests.

I extend their simulation methodology by presenting the asymmetric GARCH effect. More specifically, the GJR-GARCH (Glosten et al., 1993), APARCH (Ding et al., 1993), and EGARCH (Nelson, 1991) models are used for volatility h_{ti} of the residuals u_{ti}, which are assumed to be normally distributed. The purpose of this extended simulation study is to assess its robustness and consider more realistic scenarios, especially when dealing with financial data.

The simulation experiment is relatively small-scale, in order to keep execution time within manageable ease. I follow a similar approach to the one of Dufour et al. (2010) to construct the MC Gaussian p-values. The number of observations is the same as in the sub-periods of the empirical application, i.e. $T = 60$. The numbers of equations considered are (1) $n_1 = 12$, and (2) $n_2 = 20$. The number of replications for the MC tests is $N = 99$, and each experiment contains 1000 simulations.[1] Such combinations are chosen to represent the financial models better, as it often occurs to have a great number of equations n relative to the sample size T.

The $T \times n$ regressor matrix X for $n = n_1$, n_2 takes the form

$$X = [\,\imath \quad X_1], \tag{20}$$

where $\imath$ is a vector of ones and X_1 is a multivariate standard normal variate. I then derive the location-scale invariant disturbances matrix $\hat{U} = MW$, where W is drawn from a standard normal distribution.

To study the size of the ARCH effects test, I compute the probability of incorrectly rejecting the null hypothesis if it is true, i.e. there are no ARCH effects in the residuals (Type I error). Therefore, the residuals u_{ti} are simply equal to w_{ti}.

I incorporate GJR-GARCH, APARCH, and EGARCH effects in the first $m = \lfloor n/3 \rfloor$ equations. All these models are essentially a modification to the widely known GARCH model when taking the 'leverage effect' into account, leading to asymmetric volatility. Note that, even if this assumption of normality is incorrect, as long as the conditional means and variances are correctly specified, the quasi-maximum likelihood estimates will be consistent and asymptotically normal (Glosten et al., 1993).

More precisely, the residuals u_{ti} with GJR-GARCH(1,1) effects are determined as

$$\begin{aligned} u_{ti} &= w_{ti} h_{ti}^{1/2}, \\ h_{ti} &= 1 + (\alpha_1 w_{t-1,i}^2 + \alpha_2 + \phi w_{t-1,i}^2 1_{[u_{t-1,i}<0]}) h_{t-1,i}^2, \quad i = 1, \dots m,\ t = 1, \dots T. \end{aligned} \tag{21}$$

When $\phi = 0$, the GJR-GARCH(1,1) model is equivalent to a GARCH(1,1) model. For positive values of ϕ, the conditional volatility $h_{i,t}$ is influenced at a higher degree by negative shocks.

[1] The value of N is lower than the one proposed by Dufour et al. (2010), only for runtime execution purposes, as the available computational power is much lower. Nevertheless, for $\alpha = 0.05$, it fulfills the criteria of $\alpha(N+1)$ being an integer.

Otherwise, the inverse is true. The indicator function allows for asymmetric properties in the model. The different cases considered are:

1. $\alpha_1 = 0.4,\ \alpha_2 = 0.5,\ \phi = 0$ in the case of no asymmetry,
2. $\alpha_1 = 0.8,\ \alpha_2 = 0.1,\ \phi = 0.6,$
3. $\alpha_1 = 0.2,\ \alpha_2 = 0.6,\ \phi = 0.3.$

The residuals u_{ti} with the incorporated APARCH(1,1) effects for any δ take the following form:

$$\begin{aligned} u_{ti} &= w_{ti} h_{ti}^{1/2}, \\ h_{ti}^{\delta/2} &= 1 + \beta_1 h_{t-1,i}^{\delta/2} + \beta_2 (|h_{t-1,i} w_{t-1,i}| - \phi h_{t-1,i} w_{t-1,i})^{\delta/2}, \quad i = 1, \ldots m,\ t = 1, \ldots T. \end{aligned} \tag{22}$$

The study considers the power term parameter $\delta = 1$, which makes the model alike GJR-GARCH(1,1), but using absolute residuals instead of squared residuals. Note that GJR-GARCH is a special case of APARCH with $\delta = 2$. Similarly, the study for APARCH(1,1) includes three cases, namely (1) $\beta_1 = 0.4,\ \beta_2 = 0.5,\ \phi = 0$ in the case of no asymmetry, (2) $\beta_1 = 0.8,\ \beta_2 = 0.1,\ \phi = 0.6$, and (3) $\beta_1 = 0.2,\ \beta_2 = 0.6,\ \phi = 0.3$.

The study concludes with EGARCH(1,1) effects among residuals:

$$\begin{aligned} u_{ti} &= w_{ti} h_{ti}^{1/2}, \\ \ln h_{ti} &= 1 + \gamma_1 \ln h_{t-1,i} + \gamma_2 |w_{t-1,i}| + \phi w_{t-1,i}, \quad i = 1, \ldots m,\ t = 1, \ldots T. \end{aligned} \tag{23}$$

for (1) $\gamma_1 = 0.4,\ \gamma_2 = 0.5,\ \phi = 0$ where there is of no asymmetry, (2) $\gamma_1 = 0.8,\ \gamma_2 = 0.1,\ \phi = 0.6$, and (3) $\gamma_1 = 0.2,\ \gamma_2 = 0.6,\ \phi = 0.3$. The advantage of the EGARCH model is the logarithmic specification, which enables the parameters' positive constraints to be relaxed.

The simulation procedure includes testing for ARCH effects using the Bonferroni-type tests and the MC-type tests, with the significance level of α/n, and α, respectively. I compute the probability of acceptance of the null hypothesis over the 1000 simulations (1 - P(Type II error)) to assess the power of the tests.

5 Results

5.1 Fama-French Three-Factor Model

The first part focuses on the application of the multivariate residual-based tests on the considered asset pricing model, i.e. for the Fama-French Three-Factor Model, on the 25 portfolio returns. The approach is a replication of the procedure applied in Dufour et al. (2010). The regression is performed on the full sample, as well as eight 5-year intervals, meaning that T is either equal to 492 for the full sample regression or 60 for the sub-periods. The tests are (1) Bonferroni-based and asymptotic, (2) MC Gaussian-based, and (3) MMC Student-t-based, applied on $G = 12$ lags. The significance levels α and α_1 are 5% and 2.5%, respectively. The following tables report the p-values of the tests on serial correlation and ARCH effects presence on the residuals. [2]

Table 2 displays the Bonferroni p-values for the Engle, Lee-King, Ljung-Box, and Variance ratio tests as well as the asymptotic p-values for the multivariate portmanteau tests — Hosking and Hosking-ARCH. Note that the p-values for the Bonferroni tests need to be referred to the $\alpha/25$ = 0.2% cutoff level. Generally, the tests do not reject the null of no ARCH effects in most 5-year sub-periods, except for the last one (1996-2000). Similarly, there is no serial correlation detected among the residuals in any sub-periods for most tests, except for the periods 1976-1980 and 1996-2000. Consequently, the Bonferroni- and asymptotic-based tests suggest a good model specification for the sub-periods. This does not hold for the regression on the full sample, as all tests reject the null. Such a result implies inconsistent and biased estimators of the factor coefficients.

[2] Note that the reported results do not match those of Dufour et al. (2010), albeit the same dataset and procedure were applied. Hence, I suspect the authors did not disclose every detail of their study. For this reason, I will discuss the new results in detail. The general conclusion does not change however.

Table 2: Bonferroni-based and asymptotic p-value results for all subperiods and the full sample

	ARCH effects tests			**Serial-correlation tests**		
	Engle	Lee-King	Hosking-ARCH	Ljung-Box	Variance ratio	Hosking
61-65	0.111	0.038	0.839	0.000	0.972	0.278
66-70	0.141	0.022	0.466	0.019	0.960	0.464
71-75	0.160	0.094	0.198	0.011	0.932	0.700
76-80	0.002	0.000	0.882	0.014	0.933	0.759
81-85	0.098	0.054	0.019	0.028	0.941	0.042
86-90	0.124	0.063	0.083	0.025	0.946	0.409
91-95	0.082	0.002	0.433	0.002	0.946	0.544
96-00	0.001	0.001	0.000	0.002	0.946	0.001
ALL	0.001	0.001	0.001	0.001	0.001	0.000

Next, I present the results in the case of a multivariate normal distribution for the error covariance matrix. Table 3 summarizes the simulated MC p-values for all tests and regressions. The number of MC replications N is set to 999.

Table 3: Gaussian-based MC p-value results for all subperiods and the full sample

	ARCH effects tests			**Serial-correlation tests**		
	Engle	Lee-King	Hosking-ARCH	Ljung-Box	Variance ratio	Hosking
61-65	0.611	0.640	0.298	0.006	0.825	0.021
66-70	0.702	0.424	0.055	0.591	0.556	0.094
71-75	0.792	0.867	0.027	0.437	0.125	0.463
76-80	0.001	0.005	0.163	0.514	0.148	0.141
81-85	0.560	0.753	0.002	0.669	0.235	0.001
86-90	0.672	0.791	0.008	0.652	0.276	0.091
91-95	0.427	0.090	0.102	0.176	0.294	0.223
96-00	0.001	0.010	0.001	0.231	0.285	0.001
ALL	0.000	0.000	0.000	0.000	0.001	0.000

The MC-based p-values for the joint equation-by-equation tests in most cases are higher in comparison to the Bonferroni-based counterparts, especially in the case of Lee-King, whereas the

p-values for the Hosking and Hosking-ARCH tests tend to be lower compared to the asymptotic tests, resulting in significantly more rejections. There is no level adjustment required for the Gaussian MC p-values, therefore I refer to the 5% level. Again, the model for most sub-periods yields no autocorrelated nor heteroskedastic error terms according to the Engle, Lee-King, Ljung-Box, and Variance Ratio tests. In contrast, the portmanteau tests generally reject the null hypotheses of no serial dependence and homoskedasticity. Similar to the previous scenario, all tests are rejected for the full sample, indicating a poor model fit for the 1961-2000 sample period.

Lastly, Table 4 displays the MMC and CSMMC p-values when considering Student-t distribution maximized over the effective sample size, i.e. $\{2, \ldots, 33\}$ for MMC and $CS(\kappa)$ for CSMMC. The number of replications N_0 and N_1, relevant in calculating the confidence set $CS(\kappa)$, are both equal to 999. Note that the cutoff point for the CSMMC p-values equals $\alpha - \alpha_1 = 2.5\%$

The resulting p-values follow a similar pattern to that in the case of Gaussian errors. The Fama-French models applied to the full period yield errors with significant ARCH effects and serial correlations. According to the Ljung-Box, Variance Ratio, Engle and Lee-King statistics, the error terms of the 1976-1980 and 1996-2000 models exhibit significant ARCH effects but no serial dependence, whereas the rest suffer from none of the two. The portmanteau criteria again have notably lower p-values however, thus rejecting the null of no ARCH effects or serial dependence more often.

Table 4: Student-t MMC and CSMMC p-value results for all subperiods and the full sample

	ARCH effects						**Serial-correlation**						**CS(κ)**
	Engle		Lee-King		Hosking-ARCH		Ljung-Box		Variance ratio		Hosking		
	$\hat{p}_a$	$\hat{p}_i$	$\hat{p}_a$	$\hat{p}_i$	$\hat{p}_a$	$\hat{p}_i$	$\hat{p}_a$	$\hat{p}_i$	$\hat{p}_a$	$\hat{p}_i$	$\hat{p}_a$	$\hat{p}_i$	
61-65	0.635	0.634	0.646	0.645	0.562	0.412	0.010	0.010	0.852	0.852	0.225	0.060	8-33
66-70	0.759	0.756	0.414	0.414	0.343	0.130	0.631	0.623	0.588	0.576	0.340	0.172	11-33
71-75	0.813	0.824	0.864	0.864	0.251	0.072	0.466	0.466	0.144	0.144	0.542	0.503	11-33
76-80	0.038	0.005	0.011	0.011	0.490	0.295	0.568	0.568	0.165	0.165	0.400	0.205	11-33
81-85	0.581	0.556	0.752	0.752	0.061	0.005	0.720	0.720	0.265	0.265	0.072	0.008	8-33
86-90	0.690	0.690	0.783	0.783	0.159	0.017	0.678	0.678	0.306	0.297	0.340	0.136	22-33
91-95	0.512	0.448	0.105	0.105	0.392	0.172	0.197	0.194	0.327	0.327	0.430	0.296	16-33
96-00	0.030	0.003	0.028	0.028	0.006	0.001	0.230	0.230	0.316	0.316	0.011	0.001	4-15
ALL	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	6-11

Note. $\hat{p}_a$ refers to the MMC p-value over all degrees of freedom. $\hat{p}_i$ refers to the CSMMC p-value over $\kappa \in CS(\kappa)$

5.2 Simulation Study for Asymmetric GARCH Effects

This section focuses on the simulation study results when taking asymmetric conditional volatilities into account. In particular, I present the size and the power of the Bonferroni- and MC Gaussian-based tests for several scenarios: (1) $n = 12,\ G = 2$, (2) $n = 12,\ G = 12$, (3) $n = 20,\ G = 2$, and (4) $n = 20,\ G = 12$.

Table 5 displays the empirical rejections for 5% nominal significance test levels when errors are i.i.d. normally distributed, implying that a good test must have a size of 5% at most. Note, however, that the Bonferroni correction α/n is applied to the Bonferroni-type tests ($\tilde{LK}_B$ and $\tilde{E}_B$). In most cases, the tests perform better when the number of lags is 12, like in the empirical application. For instance, the size of $\tilde{E}$ for both cases of n is closer to 0.05 when considering 12 lags (0.105 versus 0.065 and 0.045 versus 0.044). Another observation is a higher size for smaller number of equations in all cases. Additionally, the table shows that MC Gaussian-based test sizes are closer to the significance level than Bonferroni, which, in most cases, severely over-reject the null as a consequence of the poor performance of the univariate tests, resulting in worse Type I errors. Lastly, both portmanteau tests fail the size tests, as they reject for almost all iterations.

Table 5: Size of diagnostic tests

n	**G**	$\tilde{LK}_B$	$\tilde{LK}$	$\tilde{E}_B$	$\tilde{E}$	HM_2	$\tilde{HS}_2$
12	2	0.087	0.105	0.094	0.096	0.728	1.000
	12	0.068	0.065	0.022	0.068	0.998	1.000
20	2	0.029	0.045	0.041	0.037	0.173	0.927
	12	0.058	0.044	0.004	0.043	0.668	0.969

Next, we turn to the power study of the tests against different types of ARCH effects. The conditional volatilities in the first $m = \lfloor n/3 \rfloor$ equations follow GJR-GARCH(1,1), APARCH(1,1) and EGARCH(1,1).

Table 6 summarizes the empirical rejections for 5% nominal significance test levels when the variance of the errors have a GJR-GARCH(1,1) structure in the first m equations. The power of the MC-based test statistics is always higher and closer to the desired level than the Bonferroni- and asymptotic-based counterparts, due to the superiority of the MC finite-sampling method.

Incorporating more lags reduces the power for all cases of GJR-GARCH(1,1), which is expected due to the higher number of restrictions, which increases the degrees of freedom. More precisely, the larger the number of restrictions, the larger the critical values for the null distribution will be, hence making it harder to reject (Mikusheva, 2013). Additionally, the power decreases when considering 20 observations instead of 12. It makes sense due to the acceptance region getting smaller and closer to the value of the null as (effective) sample size increases.

Statistics of the Lee-King type generally perform better than those of the Engle type in the presence of ARCH effects, asymmetric or not. Hosking-ARCH statistic ($\tilde{HS}_2$) exhibits a power of almost 1 every time, but such a result is of concern since it appears to be the case of over-rejection, as seen in the size tests.

When $\phi = 0$, which is the case of no asymmetry, the MC-based tests, in some cases, are less powerful compared to the asymmetrical ARCH. This occurs especially when considering the second case, i.e. $\beta_1 = 0.8,\ \beta_2 = 0.1,\ \phi = 0.6$. It could either be due to the large effect of β_1 on the autocorrelation function, which does not indicate anything regarding the asymmetry effect, or, more importantly, ϕ increases the autocorrelation between the squared error terms, thus leading to a high test statistics which goes beyond the critical value. The latter implies a higher rejection rate, therefore higher power.

On the other hand, the tests against asymmetric GARCH effects in the third case, i.e., $\beta_1 = 0.2,\ \beta_2 = 0.6,\ \phi = 0.3$, are generally less powerful compared to the symmetric scenario. This possibly implies that the magnitude of the effect of ϕ is low and the difference in power is mainly driven by the other coefficients rather than asymmetry.

We can conclude that the effect of the GJR-GARCH-type volatility on the power of the tests is ambiguous, and, since the values do not differ at a high scale, the tests perform relatively just as well in the asymmetric volatility scenario. An important thing to note is that the tests in most scenarios, regardless of the structure of the volatility, have a poor performance.

Table 6: Power of ARCH tests against GJR-GARCH(1,1) effects

		Parameters			**ARCH Effects Tests**					
n	G	α_1	α_2	ϕ	LK_B	$\tilde{LK}$	E_B	$\tilde{E}$	$HM2$	$\tilde{HS}_2$
12	2	0.4	0.5	0.00	0.796	0.793	0.683	0.674	0.908	1.000
		0.8	0.1	0.60	0.968	0.966	0.951	0.952	0.979	1.000
		0.2	0.6	0.30	0.563	0.564	0.421	0.395	0.884	1.000
	12	0.4	0.5	0.00	0.437	0.350	0.180	0.392	0.998	1.000
		0.8	0.1	0.60	0.327	0.275	0.355	0.632	0.996	1.000
		0.2	0.6	0.30	0.232	0.264	0.044	0.166	1.000	1.000
20	2	0.4	0.5	0.00	0.854	0.851	0.732	0.708	0.474	0.989
		0.8	0.1	0.60	0.946	0.941	0.938	0.921	0.626	0.991
		0.2	0.6	0.30	0.489	0.501	0.345	0.320	0.333	0.983
	12	0.4	0.5	0.00	0.467	0.379	0.171	0.404	0.857	0.989
		0.8	0.1	0.60	0.274	0.220	0.300	0.562	0.848	0.993
		0.2	0.6	0.30	0.296	0.309	0.031	0.147	0.823	0.989

Tables 7 and 8 display the results when considering APARCH(1,1) and EGARCH(1,1) effects in the residuals. We observe similar patterns to the GJR-GARCH(1,1) scenario. Most importantly, the tests exhibit a similar power when asymmetry is introduced in the conditional volatilities compared to the symmetric ARCH case, except for the third scenario ($\beta_1 = 0.2,\ \beta_2 = 0.6,\ \phi = 0.3$), where the powers are almost always below 0.300 (excluding Hosking-ARCH).

The powers of the tests in the EGARCH(1,1) case are notably lower in comparison to the APARCH(1,1) and GJR-GARCH(1,1) counterparts, implying a poor performance of the tests against exponential GARCH effects. In all scenarios, the Hosking-ARCH portmanteau tests almost always reject the null, suggesting the superiority of the individual-equation criteria because, as seen in the size tests, this is due to over-rejection.

The simulation study shows that, by taking into account asymmetric conditional volatilities, which are often present in the disturbances of the Fama-French model, the overall performance of the tests in correctly detecting ARCH effects does not decline, therefore the tests are robust to different ARCH structures, except for EGARCH.

Such finding has implications for the reliability of the misspecification tests conducted on the empirical application in the section before, which showed that the majority of the sub-periods lacked ARCH-type effects. According to the simulation study, tests tend to under-reject the null hypothesis of no ARCH effects when it does not hold when considering 12 lags. The above suggests that, given the high likelihood of a Type II error, the conclusion that no ARCH effects exist in the Fama-French model for the 5-year periods is doubtful, whether or not there are asymmetries.

Table 7: Power of ARCH tests against APARCH(1,1) effects

		Parameters			**ARCH Effects Tests**					
n	G	β_1	β_2	ϕ	LK_B	$\tilde{LK}$	E_B	$\tilde{E}$	$HM2$	$\tilde{HS}_2$
12	2	0.4	0.5	0.00	0.804	0.809	0.650	0.669	0.911	1.000
		0.8	0.1	0.60	0.614	0.628	0.639	0.629	0.794	1.000
		0.2	0.6	0.30	0.463	0.469	0.399	0.397	0.803	1.000
	12	0.4	0.5	0.00	0.372	0.291	0.136	0.324	0.998	1.000
		0.8	0.1	0.60	0.108	0.085	0.217	0.364	0.979	1.000
		0.2	0.6	0.30	0.166	0.136	0.075	0.216	0.992	1.000
20	2	0.4	0.5	0.00	0.865	0.866	0.749	0.812	0.477	0.989
		0.8	0.1	0.60	0.655	0.660	0.729	0.723	0.411	0.964
		0.2	0.6	0.30	0.185	0.175	0.102	0.288	0.816	0.966
	12	0.4	0.5	0.00	0.451	0.358	0.154	0.401	0.863	0.997
		0.8	0.1	0.60	0.122	0.103	0.258	0.426	0.744	0.971
		0.2	0.6	0.30	0.570	0.654	0.500	0.536	0.373	0.926

Table 8: Power of ARCH tests against EGARCH(1,1) effects

		Parameters			ARCH Effects Tests					
n	G	γ_1	γ_2	ϕ	LK_B	$\tilde{LK}$	E_B	$\tilde{E}$	$HM2$	$\tilde{HS}_2$
12	2	0.4	0.5	0.00	0.094	0.113	0.082	0.080	0.783	1.000
		0.8	0.1	0.60	0.071	0.082	0.048	0.046	0.771	1.000
		0.2	0.6	0.30	0.096	0.109	0.102	0.104	0.996	1.000
	12	0.4	0.5	0.00	0.080	0.068	0.001	0.039	0.995	1.000
		0.8	0.1	0.60	0.275	0.195	0.003	0.043	1.000	1.000
		0.2	0.6	0.30	0.090	0.064	0.004	0.044	0.996	1.000
20	2	0.4	0.5	0.00	0.077	0.091	0.078	0.075	0.251	0.964
		0.8	0.1	0.60	0.069	0.065	0.048	0.049	0.286	0.962
		0.2	0.6	0.30	0.082	0.089	0.104	0.101	0.245	0.959
	12	0.4	0.5	0.00	0.090	0.062	0.001	0.050	0.719	0.985
		0.8	0.1	0.60	0.295	0.205	0.003	0.046	0.847	0.995
		0.2	0.6	0.30	0.091	0.061	0.006	0.048	0.706	0.975

6 Discussion and Conclusion

When attempting to explain returns using three components, investors frequently refer to the Fama-French Three-Factor Model. According to recent research, there is a negative correlation between the return and the volatility, which causes the disturbances of these models to have asymmetric GARCH effects. These results underline the importance to develop appropriate diagnostic tests for a wide range of financial models with asymmetric volatilities. Due to this, this paper first examines the method suggested by Dufour et al. (2010), in which exact variations of the standard multivariate portmanteau tests and exact joint univariate tests are offered, and then conducts a simulation study to assess the size and power of the tests in the presence of asymmetric GARCH effects.

Throughout 1966–2000 and five separate 5-year subperiods, these diagnostic tests were used in the Fama–French Three-Factor model. The results point to a reasonable model for the subperiods, but the regression for the entire sample shows disturbances that go against least squares principles.

The research question on the effectiveness of the suggested tests, when introducing asymmetric conditional volatilities as opposed to symmetric, is then answered through a simulated study. It reveals that, when asymmetric GARCH effects exist, the tests' power does not differ compared to the symmetric scenario. When dealing with small effective sample sizes, the tests in all scenarios exhibit a low power, which raises the possibility of under-rejection. This experiment has implications for the application of the tests in the Fama-French model. It suggests that the latter is potentially not as good of a fit as the p-values suggest, therefore the extensions suggested in the literature, such that the additional 'momentum' factor, could provide better residual diagnostics (Carhart, 1997).

In a broader sense, the simulation study reveals that, when it comes to financial models with asymmetric GARCH effects, particularly EGARCH models, the tests proposed by Dufour et al. (2010) must be used with caution. It is recommended is to first find the best fit for the conditional volatility structure, by comparing different models via information criteria, to potentially detect models such as EGARCH, for which the tests do not perform well. Afterward, the tests are either employed or not, depending on the previous result. Lastly, the findings also reveal that, as the literature suggests, the portmanteau tests perform poorly in finite samples therefore their use should be limited. Contrary to the literature however, the tests over-reject instead of under-rejecting (Brooks & Henry, 2000).

There are a number of limitations to this study that other researchers can address in future research. For one, only residuals of the Gaussian type were taken into account in the simulation study, which limits the applicability because errors in financial models often have broader tails. Therefore, a similar simulation study using MMC-based p-values could be performed to investigate the effect of asymmetry, where the errors are non-Gaussian, as often occurs in financial models. Additionally, future research can look into the modification of the Monte Carlo tests such that they are applicable in the case of multivariate non-linear models. Lastly, a challenging open topic is to expand the proposed resampling approach to testing and modeling in a multivariate ARCH/GARCH model context, for example in GARCH Dynamic Conditional Correlation (DCC) models, while advancing the approach outlined by Dufour et al. (2010).

References

Akpan, E. A., & Moffat, I. U. (2017). Detection and modeling of asymmetric GARCH effects in a discrete-time series. *International Journal of Statistics and Probability*, *6*(6), 111–119.

Allen, D. E., & McAleer, M. (2021). Drawbacks in the 3-Factor Approach of Fama and French (2018). *Annals of Financial Economics*, 2240001.

Bekaert, G., & Wu, G. (2000). Asymmetric volatility and risk in equity markets. *The Review of Financial Studies*, *13*(1), 1–42.

Ben, Y., & Jiang, F. (2020). A note on portmanteau tests for conditional heteroscedastistic models. *Economics Letters*, *192*, 109159.

Blitz, D. (2020). Factor Performance 2010–2019: A Lost Decade? *The Journal of Index Investing*, *11*(2), 57–65.

Brailsford, T. J., & Faff, R. W. (1996). An evaluation of volatility forecasting techniques. *Journal of Banking & Finance*, *20*(3), 419–438.

Breusch, T. S., & Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrica: Journal of the Econometric Society*, 1287–1294.

Brooks, C., & Henry, O. T. (2000). Can portmanteau nonlinearity tests serve as general mis-specification tests?: Evidence from symmetric and asymmetric GARCH models. *Economics Letters*, *67*(3), 245–251.

Carbon, M., & Francq, C. (2011). Portmanteau goodness-of-fit test for asymmetric power GARCH models. *Austrian Journal of Statistics*, *40*(1&2), 55–64.

Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of finance*, *52*(1), 57–82.

Catani, P., & Ahlgren, N. (2017). Combined Lagrange multiplier test for ARCH in vector autoregressive models. *Econometrics and Statistics*, *1*, 62–84.

Celik, S. (2012). Theoretical and empirical review of asset pricing models: A structural synthesis. *International Journal of Economics and Financial Issues*, *2*(2), 141–178.

De Moor, L., & Sercu, P. (2013). The smallest firm effect: An international study. *Journal of International Money and Finance*, *32*, 129–155.

Ding, Z., Granger, C. W., & Engle, R. F. (1993). A long memory property of stock market returns and a new model. *Journal of Empirical Finance*, *1*(1), 83–106.

Duchesne, P., & Lalancette, S. (2003). On testing for multivariate ARCH effects in vector time series models. *Canadian Journal of Statistics*, *31*(3), 275–292.

Dufour, J.-M. (2006). Monte Carlo tests with nuisance parameters: A general approach to finite-sample inference and nonstandard asymptotics. *Journal of Econometrics*, *133*(2), 443–477.

Dufour, J.-M., & Khalaf, L. (2002). Simulation based finite and large sample tests in multivariate regressions. *Journal of Econometrics*, *111*(2), 303–322.

Dufour, J.-M., & Khalaf, L. (2003). Finite-sample simulation-based tests in seemingly unrelated regressions. In *Computer-aided econometrics* (pp. 44–69). CRC Press.

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., & Voia, M. (2015). Finite-sample resampling-based combined hypothesis tests, with applications to serial correlation and predictability. *Communications in Statistics-Simulation and Computation*, *44*(9), 2329–2347.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the Econometric Society*, 987–1007.

Engle, R. F., & Ng, V. K. (1993). Measuring and testing the impact of news on volatility. *The Journal of Finance*, *48*(5), 1749–1778.

Fama, E. F., & French, K. R. (1992). The cross-section of expected stock returns. *The Journal of Finance*, *47*(2), 427–465.

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.

Francq, C., Wintenberger, O., & Zakoïan, J.-M. (2018). Goodness-of-fit tests for Log-GARCH and EGARCH models. *Test*, *27*(1), 27–51.

Franses, P. H., Van Dijk, D., et al. (2000). *Non-linear time series models in empirical finance.* Cambridge University Press.

French, K. R. (n.d.). Retrieved from `https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`

Gel, Y. R., & Chen, B. (2012). Robust lagrange multiplier test for detecting ARCH/GARCH effect using permutation and bootstrap. *Canadian Journal of Statistics*, *40*(3), 405–426.

Glosten, L. R., Jagannathan, R., & Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance*, *48*(5), 1779–1801.

Griffin, J. M. (2002). Are the Fama and French factors global or country specific? *The Review of Financial Studies*, *15*(3), 783–803.

Hagerud, G. E., et al. (1997). *Specification tests for asymmetric GARCH.* Stockholm School of Economics, the Economic Research Inst.

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

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

Jarque, C. M., & Bera, A. K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters*, *6*(3), 255–259.

Kalev, P. S., & Zolotoy, L. (2012). Time–Varying Dynamics and Asymmetric Effects of the Fama–French Factor Betas. In *25th australasian finance and banking conference.*

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

Lee, J. H., & King, M. L. (1993). A locally most mean powerful based score test for ARCH and GARCH regression disturbances. *Journal of Business & Economic Statistics*, *11*(1), 17–27.

Lintner, J. (1975). The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. In *Stochastic optimization models in finance* (pp. 131–155). Elsevier.

Ljung, G. M., & Box, G. E. (1978). On a measure of lack of fit in time series models. *Biometrika*, *65*(2), 297–303.

Lo, A. W., & MacKinlay, A. C. (1988). Stock market prices do not follow random walks: Evidence from a simple specification test. *The Review of Financial Studies*, *1*(1), 41–66.

Loudon, G. F., Watt, W. H., & Yadav, P. K. (2000). An empirical analysis of alternative parametric ARCH models. *Journal of Applied Econometrics*, *15*(2), 117–136.

Lozano, C. C. (2017). Volatility and Fama-French factors in the European Market.

Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, *57*(3), 519–530.

Mikusheva, A. (2013). Survey on statistical inferences in weakly-identified instrumental variable models. *Applied Econometrics*(1), 117–131.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica: Journal of the Econometric Society*, 347–370.

Pagan, A. R., & Schwert, G. W. (1990). Alternative models for conditional stock volatility. *Journal of Econometrics*, *45*(1-2), 267–290.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, *19*(3), 425–442.

Wu, G. (2001). The determinants of asymmetric volatility. *The Review of Financial Studies*, *14*(3), 837–859.

Zhao, Y., Stasinakis, C., Sermpinis, G., & Fernandes, F. D. S. (2019). Revisiting fama–french factors' predictability with bayesian modelling and copula-based portfolio optimization. *International Journal of Finance & Economics*, *24*(4), 1443–1463.

Zhou, W., & Li, L. (2016). A New Fama-French 5-Factor Model Based on SSAEPD Error and GARCH-Type Volatility. *Journal of Mathematical Finance*, *6*(5), 711–727.

Appendices

A Nuisance Parameter Problem

Consider the generalized Multivariate Linear Model of the form

$$Y = XB + U, \tag{24}$$

where $Y = [Y_1, ..., Y_n]$ is a $T \times n$ matrix of observations on n dependent variables, X is a $T \times k$ full-column rank matrix, $B = [B_1, ..., B_n]$ is a $k \times n$ matrix of unknown fixed coefficients and $U = [U_1, ..., U_T]^\top$ is a $T \times n$ matrix of random errors. Suppose the errors have the following structure:

$$\begin{aligned} U_t &= JW_t, \quad t = 1, ..., T, \\ vec(W_1, ..., W_T) &\sim \mathscr{F}(\kappa), \end{aligned} \tag{25}$$

where $\mathscr{F}(\kappa)$ is a given distribution, which could be dependent on the parameter κ, and J is either (1) an unknown non-singular lower triangular matrix, or (2) an unknown non-singular matrix (Dufour & Khalaf, 2002).

The assumptions above, for $W = [W_1, \dots, W_T]^\top$, require that

$$W = U(J^{-1})^\top \tag{26}$$

When the degree-of-freedom parameter κ is fully known, for instance when $vec(W_1, ..., W_T) \sim N[0, I_n]$, particular, this condition will be satisfied.

Otherwise, let the covariance of U_t be $\Sigma = JJ^\top$. The least squares estimate of B, $\hat{B}$, and the corresponding residuals $\hat{U} = MU$, where

$$M = I - X(X^\top X)^{-1}X^\top, \tag{27}$$

show that the distribution of $\hat{U}$ depends on the unknown scaling matrix J (or Σ), which sets the 'scale'. Consequently, the test statistic based on the residual matrix $\hat{U}$ potentially includes J as a nuisance parameter and needs to be estimated from the data.

B Three-Stage MC

Algorithm 4: Three-stage MC algorithm

A1. Draw N_0 observations of W following the Student-t distribution with κ degrees of freedom.

A2. For each draw, construct the Mahalanobis matrix (Mardia, 1970) and the Cholesky-standardized residual matrix

$$\hat{D} = T\hat{W}(\hat{W}^\top \hat{W})^{-1}\hat{W}^\top, \quad \tilde{W} = \hat{W} S_{\hat{W}}^{-1},$$

which yield N_0 of the following multivariate skewness and kurtosis coefficients:

$$SK_M = \frac{1}{T^2}\sum_{s=1}^{T}\sum_{t=1}^{T}\hat{d}_{st}^3, KU_M = \frac{1}{T}\sum_{t=1}^{T}\hat{d}_{tt}^2.$$

A3. Obtain the desired estimates, referred to as '*reference simulated moments*' (RSM):

$$ESK_M(\kappa_0) = |SK_M - \overline{SK}_M(\kappa_0)|, \quad EKU_M(\kappa_0) = |KU_M - \overline{KU}_M(\kappa_0)|,$$

where $\overline{SK}_M(\kappa_0)$ and $\overline{KU}_M(\kappa_0)$ are simulation-based mean estimates of SK_M and KU_M. Define the initial vector of estimates:

$$E^{(0)} = [\overline{ESK}_M^{(0)}(\kappa_0), \overline{EKU}_M^{(0)}(\kappa_0)]^\top.$$

B1. Generate N i.i.d. realizations of W following the same distribution, and for each of these draws compute $\hat{D}$ and $\tilde{W}$.

B2. $\overline{ESK}_M(\kappa_0)$ and $\overline{EKU}_M(\kappa_0)$ are calculated from each of these MC samples via the same RSM as for the observed sample; referred to as the '*basic simulated statistics*' (BSS):

$$E^{(j)} = [\overline{ESK}_M^{(j)}(\kappa_0), \overline{EKU}_M^{(j)}(\kappa_0)]^\top, \quad j = 1, ..., N.$$

B3. Compute a simulated p-value, for any one of the test statistics in $E^{(0)}$: $\hat{p}_N[ESK_M|\kappa_0]$, $\hat{p}_N[EKU_M|\kappa_0]$, where

$$\hat{p}_N[\cdot] = \frac{NG_N(\cdot) + 1}{N + 1},$$

and $NG_N(\cdot)$ is the number of simulated values greater than or equal to each statistic in $E^{(0)}$.

C1. Draw N_1 independent additional i.i.d. realizations of W following the same distribution, and for each draw compute $\hat{D}$ and $\tilde{W}$.
C2. Using the RSM and the N_1 draws generated, compute the corresponding simulated statistics:

$$EE^{(l)} = [\overline{ESK}_M^{(l)}(\kappa_0), \overline{EKU}_M^{(l)}(\kappa_0)]^\top, \quad l = 1, ..., N_1.$$

C3. Compute the simulated p-values for the observed and the N_1 additional simulated statistics considering the p-value functions obtained at step B4:

$$\hat{p}_N^{(l)}[S] \equiv p_N(S^{(l)}; S), l = 1, ..., N_1, \quad \text{for } S = ESK_M(\kappa_0), EKU_M(\kappa_0).$$

C4. Compute the corresponding values:

$$CSK_M^{(l)}(\kappa_0) = 1 - \ \min\{\hat{p}_N(ESK_M^{(l)}(\kappa_0), \hat{p}_N(EKU_M^{(l)}(\kappa_0)\}, \quad l = 0, 1, ..., N_1.$$

C5. Obtain the p-value functions for the combined test $CSK_M(\kappa_0)$ $\hat{p}_N[CSK_M|\kappa_0]$, and include the corresponding κ_0 in the confidence set $CS(\kappa)$ if it is greater than α_1.

Note. This is the procedure for obtaining the interval $CS(\kappa)$ as described by Dufour and Khalaf (2003)

C Short Description of the Code

The code of the paper consists of two R scripts:

- `replication_code_thesis_499332.R`: Contains the replication code, i.e. the Fama-French application. It is designed such that the input changes per test and subperiod (1 to 8), therefore both should be changed and ran every time. It contains all Bonferroni, MC, MMC codes, finalized with the $CS(\kappa)$ interval computation via CSMMC.
- `extension_code_thesis_499332.R`: Contains the extension code, i.e. the simulation study for asymmetric GARCH effects. Similarly, it should be ran separately for the three different scenarios, two different values for n, and each GARCH effects' model. For the size tests, exclude the GARCH effects' computation by commenting the lines out. This file only includes the case of Gaussian error terms, therefore only the functions for Bonferroni and MC-based p-values are provided.