



Article

# Model Selection and Post Selection to Improve the Estimation of the ARCH Model

Marwan Al-Momani <sup>1,\*</sup>  and Abdaljbar B. A. Dawod <sup>2</sup> 

<sup>1</sup> Department of Mathematics, University of Sharjah, Sharjah P.O. Box 27272, United Arab Emirates

<sup>2</sup> Department of Forest Management, Faculty of Forestry, Shambat Complex, University of Khartoum, Khartoum 13314, Sudan; abdaljbbar@gmail.com

\* Correspondence: malmomani@sharjah.ac.ae

**Abstract:** The Autoregressive Conditionally Heteroscedastic (ARCH) model is useful for handling volatilities in economical time series phenomena that ARIMA models are unable to handle. The ARCH model has been adopted in many applications that contain time series data such as financial market prices, options, commodity prices and the oil industry. In this paper, we propose an improved post-selection estimation strategy. We investigated and developed some asymptotic properties of the suggested strategies and compared with a benchmark estimator. Furthermore, we conducted a Monte Carlo simulation study to reappraise the relative characteristics of the listed estimators. Our numerical results corroborate with the analytical work of the study. We applied the proposed methods on the S&P500 stock market daily closing prices index to illustrate the usefulness of the developed methodologies.

**Keywords:** ARCH; heteroscedastic; financial markets; residuals bootstrapping; pretest; shrinkage



**Citation:** Al-Momani, Marwan, and Abdaljbar B. A. Dawod. 2022. Model Selection and Post Selection to Improve the Estimation of the ARCH Model. *Journal of Risk and Financial Management* 15: 174. <https://doi.org/10.3390/jrfm15040174>

Academic Editors: S. Ejaz Ahmed and Nataša Šarljija

Received: 1 March 2022

Accepted: 4 April 2022

Published: 10 April 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Modeling and forecasting financial markets are a challenging activities for both investors and researchers equally. Generally, financial markets are extremely manipulated by a number of factors such as interest rates, political issues, inflation rates and foreign exchange rates. More precisely, the uncertainty of stock markets produces high volatility that makes the forecasting stage very complex. Volatility forecasting is an important financial matter, and a precise and accurate volatility forecast is important to traders, investors and financial analysts. Before the 1980s, researchers were relying on ARIMA models, however, many financial time series violate the assumptions of the ARIMA model (cf. Brockwell and Davis 2016; Teräsvirta 2009).

Fortunately, Engle (1982) suggested a stationary non-linear model for economical time series and introduced the Autoregressive Conditionally Heteroscedastic (ARCH) model, wherein the conditional variance of a series  $\{y_k\}$  changes according to an autoregressive-type process. Subsequently, Francq and Zakoian (2012); Grublytė et al. (2017) discussed the properties of maximum likelihood (MLE) and ordinary least squares (OLS) estimates of ARCH model parameters. They also investigated the consistency and the asymptotic normality of the OLS estimator for the ARCH model.

In this article, we are interested in estimating the parameter vector of the ARCH model when some prior information is available in the form of potential linear restrictions on the parameters in the parameter space. Practically, an ample number of variables may be collected and included in the model in an initial stage. However, due to model complexity (in terms of both interpretation and variation), estimation when a subset of parameters are under linear restrictions is an important problem in such scenarios. In order to form such constraints, one requires some prior information about the parameter space under consideration. One possible source of prior information may be distinguishing which predictors are of most interest and which are not. An alternative source of prior information,

specifically uncertain prior information (UPI), might be obtained from previous studies or expert knowledge that search for some specified patterns.

This paper is organized as follows: Section 2 discuss the recent findings of modeling time series data using ARCH family models. Section 3 discusses the parameters' estimation of the ARCH model. Section 4 is dedicated to introducing the concept of restricted, pretest and shrinkage estimations of the ARCH model. We derive the asymptotic properties of the estimators and compare their performances using risk analysis and the mean squared error in Section 5. We conduct an extensive simulation study for our selected model and demonstrate the application of the proposed estimators in real-life problems in Section 6. In Section 7 we give some conclusions.,

## 2. Literature Review

Various studies have investigated the dilemma of having a financial time series with high volatility. For instance, Peiris and Peiris (2011) examined the volatility of different sectors in the Colombo Stock Exchange (CSE) and they applied ARCH/GARCH models on the monthly time series data of 20 sectors in CSE from 2005 to 2010. They investigated the impact of macroeconomic factors on volatility. As a result, they found that sixteen out of twenty sectors in CSE had significant volatility and both ARCH and GARCH terms on the fitted models for individual sectors were significant. Subsequently, Rathnayaka et al. (2013) carried out a study to understand the trends and cyclic patterns in CSE in order to predict future behaviors during seven years since January 2007. They investigated the causal relationships between market performances and economic growth conditions related to Sri Lanka. Their results revealed that both microeconomic and macroeconomic conditions had a direct impact on stock market volatility.

Recently, Wang (2021) utilized GARCH models to analyze Bitcoin's returns and volatility. As the GARCH (1,1) model was adopted, the outcome found that the returns and volatility of Bitcoin have clustering characteristics and returns and the volatility of Bitcoin is a persistent process; however, its effect gradually reduces with time. To overcome the limitations of the GARCH (1,1) model, researchers had used TAR and EGARCH models to overcome the *Leverage Effect* of the returns and volatility of Bitcoin.

In practice, regression models usually do not possess a pre-defined UPI; thus, model selection criterion such as *Akaike's Information Criterion* (AIC) Akaike (1974), *Bayesian Information Criterion* (BIC) Schwarz (1978) or any other technique can be used to construct a sub-model. Now, it is up to the practitioner to use a sub-model based on fewer predictors (under-fitted model) or to lean towards a so-called full or over-fitted model. Alternatively, one may utilize the prior information to test whether some parameters are indeed zero, or more generally, whether the full vector of parameters are under linear restrictions. To do this, we will explore the pretesting strategy to improve the post-estimation inference of the ARCH model. Furthermore, we will implement the use of the Stein-type shrinkage estimator as an alternative to pretesting which shrinks the full model estimator in the direction of the restrictions. This leads to more efficient estimators when the shrinkage is adaptive. In the first stage, we select a sub-model by variable selection method or impose a linear restriction on the parameter space to obtain a submodel. In the second stage, we combine the sub-model with the full model via a test statistic to improve the estimation efficiency.

Many studies have considered incorporating the UPI in the estimation process to obtain efficient estimators for many statistical models. Recently, Ahmed et al. (2015) proposed efficient estimators for the regression coefficients of the spatial conditional autoregressive model under the availability of uncertain auxiliary information about these coefficients. Al-Momani et al. (2016) proposed shrinkage and penalty estimators for the spatial error model. Thomson et al. (2016) investigated the relative performances of pretest and shrinkage estimators for time series following generalized linear models. In all these cases, shrinkage estimators outperformed classical estimators. Dawod et al. (2018) introduced Bayesian estimation strategies for jointly monitoring the linear profile. Al-Momani et al. (2019) proposed the use of the pretest, shrinkage and positive shrinkage in estimating the large-scale

regression parameter vector in the spatial moving average and showed that the positive shrinkage dominated all other estimators in terms of the relative efficiency of the mean squared error with respect to the classical maximum likelihood estimator. For more details, the reader is referred to (Ahmed and Raheem 2012; Emmert-Streib and Dehmer 2019; Yüzbaşı and Ahmed 2020; Ejaz Ahmed and Yüzbaşı 2016, 2017; Ahmed 2014) for detailed information on the subject.

### 3. Estimating ARCH(q) Parameters

Following Francq and Zakoïan (2010), we introduce the ARCH model and consider the existence of a strictly stationary solution to this model.

$$\sqrt{y_t} = \sigma_t \epsilon_t, \tag{1}$$

$$\sigma_t^2 = \omega + \sum_{i=1}^q \alpha_i y_{t-i}, \tag{2}$$

where  $\epsilon_t$  is the error term, independently and identically distributed with a mean 0 and variance 1,  $\omega > 0$ ,  $\alpha_i \geq 0$ ,  $\beta_j \geq 0$  are unknown constants,  $\forall i = 1, \dots, q$  and  $j = 1, \dots, p$ , and  $\sigma_t^2 = Var(\sqrt{y_t} | \sqrt{y_{t-1}})$ .

If the ARCH(q) model holds the conditions  $\omega > 0$  and  $\sum_{i=1}^q \alpha_i < 1$ , then the uniquely strictly stationary solution of the model is a weak white noise.

The Ordinary Least Squares (OLS) method will be used to estimate the parameters of ARCH(q). The OLS method uses the autoregressive representation on the squares of the observed process and no distributional assumptions are needed for the error term ( $\epsilon_t$ ).

The autoregressive AR(q) representation can be obtained by applying some mathematical transformations as follows

$$u_t = y_t - \sigma_t^2, \tag{3}$$

where  $(u_t, \mathcal{F}_t)$  is the sequence containing a martingale difference when  $E(y_t) = \sigma_t^2 < \infty$ , denoting by  $\mathcal{F}_t$  the  $\sigma$ -field generated by  $\{y_s : s \leq t\}$ .

By substituting  $\sigma_t^2$  from Equation (3) in Equation (2), we obtain

$$y_t = \omega + \sum_{i=1}^q \alpha_i y_{t-i} + u_t. \tag{4}$$

The true parameter will be denoted by  $\theta_0$ , where  $\theta_0 = (\omega, \alpha_1, \dots, \alpha_q)'$ .

Assuming that we observe  $\sqrt{y_1}, \dots, \sqrt{y_n}$ , observations of length  $n$  from a process  $\{Y_t\}$  and considering  $\sqrt{y_0}, \dots, \sqrt{y_{1-q}}$  as initial values of the process, these initial values can be chosen to be zeros. By introducing the vector  $\mathbf{Y}_{t-1} = (1, y_{t-1}, \dots, y_{t-q})'$ , we can rewrite Equation (4) as a linear system as follows

$$y_t = \mathbf{Y}'_{t-1} \theta_0 + u_t, \quad t = 1, \dots, n, \tag{5}$$

and in a matrix format as

$$\mathbf{Y} = \mathbf{X} \theta_0 + \mathbf{U}, \tag{6}$$

where  $X'_i = (1, y_{i-1}, y_{i-2}, \dots, y_{i-q})$ .

#### 3.1. Estimation of the Parameter

Assuming  $\mathbf{X}$  is of full rank,  $X'X$  is invertible, and the OLS estimator is given by

$$\hat{\theta} = \operatorname{argmin} \|\mathbf{Y} - \mathbf{X}\theta\|^2 = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y}. \tag{7}$$

In the forthcoming sections, we will refer to this estimator as *unrestricted estimator* (UE) or simply by  $\hat{\theta}^U$ .

### 3.2. Estimation of $\sigma_0^2$

Assuming that  $\epsilon_t$  follows normal distribution with mean 0 and variance  $\sigma_0^2$  and with the following conditions:

1.  $\{Y_t\}$  is non-anticipative strictly stationary solution of the model in Equation (1).
2.  $E(y_t) < +\infty$ .
3.  $P(\epsilon_t^2 = 1) \neq 1$ .
4.  $E(y_t^2) < +\infty$ .

Then,  $\sigma_0^2$  is estimated by  $\hat{\sigma}_0^2$ , where

$$\begin{aligned} \hat{\sigma}_0^2 &= \frac{1}{n - q - 1} \|\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\theta}}\|^2 \\ &= \frac{1}{n - q - 1} \sum_{t=1}^n \left\{ y_t - \hat{\omega} - \sum_{i=1}^q \hat{\alpha}_i y_{t-i} \right\}^2, \end{aligned} \tag{8}$$

where  $\hat{\omega}, \hat{\alpha}_1, \dots, \hat{\alpha}_q$  are estimated by Equation (7).

### 3.3. Estimation of the Information Matrices

Accordingly, Francq and Zikoian (2010) define  $A$  and  $B$  as  $A = E(\mathbf{Y}_{t-1}\mathbf{Y}'_{t-1})$ ,  $B = E(\sigma_t^4 \mathbf{Y}_{t-1}\mathbf{Y}'_{t-1})$ , respectively, where

1.  $A$  and  $B$  have the same length  $q \times q$ .
2.  $A$  and  $B$  are invertible.

Then, the estimates of  $A$  and  $B$  are respectively, given by

$$\hat{A} = \frac{1}{n} \sum_{t=1}^n \mathbf{Y}_{t-1}\mathbf{Y}'_{t-1}, \tag{9}$$

$$\hat{B} = \frac{1}{n} \sum_{t=1}^n \hat{\sigma}_t^4 \mathbf{Y}_{t-1}\mathbf{Y}'_{t-1}, \tag{10}$$

where  $\hat{\sigma}_t^2 = \mathbf{Y}'_{t-1}\hat{\boldsymbol{\theta}}^U$ . The fourth-order moment of the process  $\epsilon_t = \frac{\sqrt{y_t}}{\sigma_t}$  is  $E(\epsilon_t^4)$ ; that is also consistently estimated by  $\hat{\mu}_4 = \frac{1}{n} \sum_{t=1}^n \frac{y_t^2}{\hat{\sigma}_t^4}$ .

### 3.4. Asymptotic Distribution of OLS Estimator

Weiss (1986) was the pioneer who discussed the properties of maximum likelihood and the least squares estimates of the parameters of both the regression and ARCH models in parallel with the properties of various tests of the model that are available. He did not assume that the errors are normally distributed. Rich et al. (1991) introduced another attractive way to estimate the parameters of the ARCH model without assuming normality condition. They used the generalized method of moments of Hansen (1982) and showed that, under fairly weak conditions, the estimator is consistent and asymptotically normally distributed. Francq and Zikoian (2004, 2012) proved the consistency and asymptotic normality of OLS. In this subsection, we list two theorems by Francq and Zikoian (2010) about the consistency and asymptotic normality of the OLS estimator for  $\boldsymbol{\theta}$ .

**Theorem 1 (Francq and Zikoian 2010).** Consistency of OLS estimates: If  $\hat{\boldsymbol{\theta}}^U$  is a sequence of estimators satisfying the OLS solution for ARCH under the assumptions (1)–(4) in Section 3.2, then

$$\hat{\boldsymbol{\theta}}^U \xrightarrow{P} \boldsymbol{\theta}, \hat{\sigma}^2 \xrightarrow{P} \sigma^2 \text{ as } n \rightarrow \infty, \tag{11}$$

as  $\hat{\boldsymbol{\theta}}^U$  is a consistent estimator for  $\boldsymbol{\theta}$  and where  $p$  denotes convergence in probability.

**Theorem 2** (Francq and Zikoian 2010). Referring to  $A$  and  $B$  given in Equations (9) and (10), we have

$$\sqrt{n}(\hat{\theta}^U - \theta) \xrightarrow{L} \mathcal{N}_q\left(\mathbf{0}, (\hat{\mu}_4 - 1)A^{-1}BA^{-1}\right), \tag{12}$$

where  $\hat{\mu}_4 = E(\epsilon_t^4)$ ,  $\hat{\theta}^U$  has asymptotic multivariate normal distribution and  $L$  denotes convergence in distribution.

#### 4. Efficient Estimation Strategies

Usually in the case of  $\hat{\theta}^U$ , the corresponding model is recognized as a *full model* because all parameters are included even though some of them may not have a significant effect. In this section, we will consider different estimation methods of  $\theta$  when some UPIs are available.

##### 4.1. Restricted Estimator

UPI(s) can be formulated as a linear hypothesis in which some of the given parameters are zeros or there is a restriction on some parameters. Then, the estimated parameters under such UPI is known as the *restricted estimator* (RE) and is simply denoted by  $\hat{\theta}^R$ . The derivation idea of this estimator is given below:

Suppose that the UPI is formulated in the form of the null hypothesis:

$$H_0 : R\theta = \mathbf{r}, \tag{13}$$

where  $R$  is  $m \times q$  known matrix of rank( $m$ ) ( $m \leq q$ ) (cf. Neter et al. 1996) and  $\mathbf{r}$  is an  $m \times 1$  vector of known constants.

Under the restrictions given in Equation (13), the method uses the Lagrange Multiplier for each restriction. The method minimizes the following function,

$$f(X, \theta) = (\mathbf{Y} - X\theta)'(\mathbf{Y} - X\theta) - \lambda'(r - R\theta), \tag{14}$$

with respect to  $\theta$  and  $\lambda$  to obtain the restricted estimator. This estimator is denoted by  $\hat{\theta}^R$  and defined by

$$\hat{\theta}^R = \hat{\theta}^U - (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}(R\hat{\theta}^U - r). \tag{15}$$

$\hat{\theta}^R$  given by Equation (15) is a biased estimator for  $\theta$  unless the restriction given in Equation (13) is true.

**Theorem 3.** The Wald test statistic for testing the hypothesis in Equation (13) is given by

$$\begin{aligned} \mathcal{L}_n &= (R\hat{\theta}^U - r)'[R(\text{Var}(\hat{\theta}^U))R']^{-1}(R\hat{\theta}^U - r) \\ &= (R\hat{\theta}^U - r)'[\sigma^2R(X'X)^{-1}R']^{-1}(R\hat{\theta}^U - r) \\ &= \frac{(R\hat{\theta}^U - r)'[R(X'X)^{-1}R']^{-1}(R\hat{\theta}^U - r)}{\hat{\sigma}^2}, \end{aligned} \tag{16}$$

where  $\hat{\sigma}^2$  is estimated in Equation (8) and it can be shown that  $\mathcal{L}_n \xrightarrow{L} \chi^2(m)$ .

We will use  $\alpha = 0.05$  as a level of significance for testing purposes.

##### 4.2. Pretest Estimator

The pretest estimate of  $\theta$  denoted by  $\hat{\theta}^{PT}$  is defined by:

$$\hat{\theta}^{PT} = \begin{cases} \hat{\theta}^U, & \text{if } \mathcal{L}_n \geq \mathcal{L}_{n,\alpha}, \\ \hat{\theta}^R, & \text{if } \mathcal{L}_n < \mathcal{L}_{n,\alpha}, \end{cases} \tag{17}$$

where  $\mathcal{L}_n$  is given in Equation (16) and  $\mathcal{L}_{n,\alpha}$  is the  $\alpha$ -critical value from the distribution. For more details, the reader can refer to Bancroft (1944); Saleh (2006); Stein (1956).

The pretest estimator is a binary choice function which chooses  $\hat{\theta}^U$  if the null hypothesis is rejected and  $\hat{\theta}^R$  if the test fails to reject the null hypothesis.

$\hat{\theta}^{PT}$  can be rewritten in a more attractive way as follows:

$$\begin{aligned} \hat{\theta}^{PT} &= \hat{\theta}^R I(\mathcal{L} \leq \mathcal{L}_\varnothing) + \hat{\theta}^U I(\mathcal{L} > \mathcal{L}_\varnothing) \\ &= \hat{\theta}^U - (\hat{\theta}^U - \hat{\theta}^R) I(\mathcal{L} \leq \mathcal{L}_\varnothing), \end{aligned} \tag{18}$$

where  $I(A)$  is the indicator function of the set  $A$ .

#### 4.3. Shrinkage Estimator

The shrinkage estimator of Stein (1956) denoted by  $\hat{\theta}^S$  is defined by:

$$\hat{\theta}^S = \hat{\theta}^R + \left[ 1 - \frac{m-2}{\mathcal{L}_n} \right] (\hat{\theta}^U - \hat{\theta}^R), m \geq 3. \tag{19}$$

It is clear that  $\hat{\theta}^S$  is no longer a binary choice regardless of whether  $H_0$  is rejected. The shrinkage estimator is a smoothed function of the two choices.  $\hat{\theta}^S$  does not represent a convex combination of  $\hat{\theta}^U$  and  $\hat{\theta}^R$  and suffers from a phenomenon known as over-shrinkage which occurs when  $\mathcal{L}_n$  is smaller than  $(m - 2)$  and hence, an unexpected sign for some of the estimated parameters may be obtained.

#### 4.4. Positive Shrinkage Estimator

A modified version of James–Stein estimator was proposed by Stein (1966) to overcome the phenomenon of the over-shrinkage estimator known as the positive part shrinkage estimator. This estimator is denoted by  $\hat{\theta}^{S+}$  and defined as:

$$\hat{\theta}^{S+} = \hat{\theta}^U + \left[ 1 - \frac{m-2}{\mathcal{L}_n} \right]^+ (\hat{\theta}^U - \hat{\theta}^R), m \geq 3, \tag{20}$$

where  $Z^+ = \max(0, Z)$ .

### 5. Asymptotic Results

In this section, we will study the asymptotic behavior of the proposed estimators  $\hat{\theta}^U, \hat{\theta}^R, \hat{\theta}^{PT}, \hat{\theta}^S, \hat{\theta}^{S+}$ . We will show that the restricted and unrestricted estimators are jointly asymptotically normal. In addition, we will define and extract expressions for the asymptotic distributional quadratic bias and the asymptotic quadratic risk of the estimators relying on the joint normality of  $\hat{\theta}^U$  and  $\hat{\theta}^R$ .

#### 5.1. Joint Normality of the Unrestricted and Restricted Estimators

The asymptotic distribution of all the estimators under hypothesis (13) are the same. Hence, we will study the asymptotic properties under a class of local alternatives that is given by

$$H_{(n)} : R\theta = r + \frac{\zeta}{\sqrt{n}}, \tag{21}$$

where  $\zeta$  is a  $q \times 1$  fixed vector in  $\mathbb{R}^q$ . If we set  $\zeta = 0$ , the local alternative becomes as in (13) which is a linear hypothesis representing the candidate null subspace.

Some distributional results involving the estimators  $\hat{\theta}^U$  and  $\hat{\theta}^R$  are given in the following theorem.

**Theorem 4.** Under the local alternatives in (21) and the regularity conditions (1)–(4) appearing in Section 3.2 and assuming that

$$\left(\frac{X'_{q \times n} X_{n \times q}}{n}\right) \xrightarrow{P} C_{q \times q}, \tag{22}$$

as  $n \rightarrow \infty$  and  $C$  is a positive definite matrix (p.d.m), then we have

$$\begin{aligned} (1) \quad T_n^{(1)} &= \sqrt{n}(\hat{\theta}^U - \theta) \xrightarrow{L} T^{(1)} \sim \mathcal{N}_q\left(0, \sigma^2 C^{-1}\right) \\ (2) \quad T_n^{(2)} &= \sqrt{n}(\hat{\theta}^R - \theta) \xrightarrow{L} T^{(2)} \sim \mathcal{N}_q\left(-\delta, \sigma^2 [C^{-1} - A]\right) \\ (3) \quad T_n^{(3)} &= \sqrt{n}(\hat{\theta}^U - \hat{\theta}^R) \xrightarrow{L} T^{(3)} \sim \mathcal{N}_q\left(\delta, \sigma^2 A\right) \\ (4) \quad \sqrt{n} \begin{bmatrix} T_n^{(1)} \\ R\hat{\theta}^U - r \end{bmatrix} &\xrightarrow{L} \begin{bmatrix} T^{(1)} \\ R\theta - r \end{bmatrix} \sim \mathcal{N}_{2q}\left(\begin{pmatrix} 0 \\ R\theta - r \end{pmatrix}, \sigma^2 \begin{bmatrix} C^{-1} & C^{-1}R' \\ RC^{-1} & RC^{-1}R' \end{bmatrix}\right) \\ (5) \quad \begin{bmatrix} T_n^{(1)} \\ T_n^{(3)} \end{bmatrix} &\xrightarrow{L} \begin{bmatrix} T^{(1)} \\ T^{(3)} \end{bmatrix} \sim \mathcal{N}_{2q}\left(\begin{bmatrix} 0 \\ \delta \end{bmatrix}, \sigma^2 \begin{bmatrix} C^{-1} & A \\ A & A \end{bmatrix}\right) \\ (6) \quad \begin{bmatrix} T_n^{(2)} \\ T_n^{(3)} \end{bmatrix} &\xrightarrow{L} \begin{bmatrix} T^{(2)} \\ T^{(3)} \end{bmatrix} \sim \mathcal{N}_{2q}\left(\begin{bmatrix} -\delta \\ \delta \end{bmatrix}, \sigma^2 \begin{bmatrix} C^{-1} - A & 0 \\ 0 & A \end{bmatrix}\right), \end{aligned}$$

where  $A = C^{-1}R'[RC^{-1}R']^{-1}RC^{-1}$ ,  $\delta = C^{-1}R'[RC^{-1}R']^{-1}(R\theta - r)$ .

**Proof.** The proof of the theorem is located in the Appendix A.  $\square$

### 5.2. Asymptotic Bias and Quadratic Bias

Assuming local alternatives in (21), and under the assumptions of Theorem (4), the asymptotic distributional bias  $b_i(\hat{\theta}^*)$  and the quadratic bias  $B_i(\hat{\theta}^*)$  where  $\hat{\theta}^* \in \{\hat{\theta}^U, \hat{\theta}^R, \hat{\theta}^{PT}, \hat{\theta}^S, \hat{\theta}^{S+}\}$  are given in the following theorem.

**Theorem 5.** Under the assumptions of Theorem (4) and the local alternatives in (21), we have

$$\begin{aligned} (1) \quad b_1(\hat{\theta}^U) &= 0, B_1(\hat{\theta}^U) = 0. \\ (2) \quad b_2(\hat{\theta}^R) &= -C^{-1}R'[RC^{-1}R']^{-1}(R\theta - r) = -\delta, \\ B_2(\hat{\theta}^R) &= \frac{\delta C \delta}{\sigma^2} = \Delta^2. \\ (3) \quad b_3(\hat{\theta}^R) &= C^{-1}R'[RC^{-1}R']^{-1}(R\theta - r)G_{m+2}(\chi_m^2(\alpha); \Delta^2) \\ &= -\delta G_{m+2}(\chi_m^2(\alpha); \Delta^2), \\ B_3(\hat{\theta}^{PT}) &= \Delta^2[G_{m+2}(\chi_m^2(\alpha); \Delta^2)]^2. \\ (4) \quad b_4(\hat{\theta}^S) &= -(m-2)(C^{-1}R'[RC^{-1}R']^{-1}(R\theta - r))E(\chi_{m+2}^{-2}(\Delta^2)) \\ &= -(m-2)\delta E(\chi_{m+2}^{-2}(\Delta^2)), \\ B_4(\hat{\theta}^S) &= (m-2)^2\Delta^2[E(\chi_{m+2}^{-2}(\Delta^2))]^2. \\ (5) \quad b_5(\hat{\theta}^{S+}) &= C^{-1}R'[RC^{-1}R']^{-1}(R\theta - r) \\ &= -\delta\{(m-2)E[[\chi_m^{-2}(\Delta^2)]^{-1}I(\chi_m^{-2}(\Delta^2) \leq (m-2))] \\ &\quad - d_1 E[\chi_m^{-2}(\Delta^2)]^{-1} - G_{m+2}(\chi_m^2(\alpha); \Delta^2)\}, \\ B_5(\hat{\theta}^{S+}) &= \Delta^2\{(m-2)E[[\chi_m^{-2}(\Delta^2)]^{-1}I(\chi_m^{-2}(\Delta^2) \leq (m-2))] \\ &\quad - (m-2)E[\chi_m^{-2}(\Delta^2)]^{-1} - G_{m+2}(\chi_m^2(\alpha); \Delta^2)\}, \end{aligned}$$

where  $\Delta^2$  is the non-centrality parameter and  $G_m(\mathcal{L}_\alpha; \Delta^2)$  is the non-central chi-square distribution function with  $q$ -degrees of freedom and non-centrality parameter  $\Delta^2$ .

The proof can be found in Appendix B.

### 5.3. Quadratic Weighted Risks

For any estimator  $\hat{\theta}^*$  of  $\theta$ , define the quadratic loss as

$$\begin{aligned} L(\hat{\theta}^*, \theta) &= n(\hat{\theta}^* - \theta)'W(\hat{\theta}^* - \theta), \\ &= \text{tr}\left\{W(n(\hat{\theta}^* - \theta)(\hat{\theta}^* - \theta)')\right\}, \end{aligned} \tag{23}$$

where  $W$  is a positive semidefinite matrix of order  $q \times q$ , and  $\text{tr}(\mathbf{A})$  is the trace of the matrix  $\mathbf{A}$ .

The asymptotic mean squared error matrix  $M(\hat{\theta}^*)$  is given by

$$M(\hat{\theta}^*) = E(n(\hat{\theta}^* - \theta)'(\hat{\theta}^* - \theta)), \tag{24}$$

and the asymptotic quadratic risk (AQR) is defined as

$$\begin{aligned} R(\hat{\theta}^*, W) &= E[n(\hat{\theta}^* - \theta)'W(\hat{\theta}^* - \theta)] \\ &= \text{tr}[WM(\hat{\theta}^*)]. \end{aligned} \tag{25}$$

The asymptotic weighted quadratic risk expressions are given in the following Theorem.

**Theorem 6.** Under the assumptions of Theorem 4, we have

- (1)  $R_1(\hat{\theta}^U, W) = \sigma^2 \text{tr}(WC^{-1})$ .
- (2)  $R_2(\hat{\theta}^R, W) = \sigma^2 \text{tr}(WA) + \delta W \delta'$ .
- (3)  $R_3(\hat{\theta}^{PT}, W) = \sigma^2 \text{tr}(WC^{-1}) - \sigma^2 \text{tr}(WA)G_{m+2}(\chi_m^2(\alpha); \Delta^2) + 2\delta W \delta' \{G_{m+2}(\chi_m^2(\alpha); \Delta^2) - G_{m+4}(\chi_m^2(\alpha); \Delta^2)\}$ .
- (4)  $R_4(\hat{\theta}^S, W) = \sigma^2 \text{tr}(WC^{-1}) - \sigma^2(m-2)\text{tr}(WA) \times \{2E[\chi_{m+4}^{-4}(\Delta^2)] - (m-2)E[\chi_{m+2}^{-4}(\Delta^2)]\} + (m-2)(m+2)\delta W \delta' E[\chi_{m+4}^{-4}(\Delta^2)]$ .
- (5)  $R_5(\hat{\theta}^{S+}, W) = R_4(\hat{\theta}^S, W) - \sigma^2(C^{-1} - A)E[(1 - (m-2)\chi_{m+2}^{-2}(\Delta^2))^2 I(\chi_{m+2}^{-2}(\Delta^2) < (m-2)) + \delta W \delta' \{2E[(1 - (m-2)\chi_{m+4}^2(\Delta^2)) I(\chi_{m+4}^2(\Delta^2) < (m-2))] - E[(1 - (m-2)\chi_{m+4}^{-2}(\Delta^2))^2 I(\chi_{m+4}^2(\Delta^2) < (m-2))]\}$ .

The proof can be found in Appendix C.

### 5.4. Risk Analysis of the Estimators

In this section, all estimators will be compared based on their asymptotic quadratic risk. We will not carry out all derivations; instead, we will give a summary of our results as follows:

- i. Comparison of  $\hat{\theta}^U$  and  $\hat{\theta}^R$ : The risk of  $\hat{\theta}^U$  is constant, whereas the risk of  $\hat{\theta}^R$  depends on  $\delta'W\delta$ ; hence, the difference in their risks is

$$\begin{aligned}
 R_1(\hat{\theta}^U, W) - R_2(\hat{\theta}^R, W) &= \sigma^2 tr(WC^{-1}) - \sigma^2 tr[WC^{-1}R'[RC^{-1}R']^{-1}RC^{-1}] \\
 &\quad + \delta'W\delta. \\
 &= \sigma^2 tr(WC^{-1}) - tr(A_{11}) + \eta_1' A_{11} \eta_1,
 \end{aligned}$$

where  $C^{-1/2}R'[RC^{-1}R']^{-1}RC^{-1/2}$  is a symmetric idempotent matrix with rank  $m (\leq q)$ . Therefore, by Courant's theorem—see Saleh (2006)—there exists an orthogonal matrix  $\Gamma$  such that

$$\Gamma C^{-1/2}R'[RC^{-1}R']^{-1}RC^{-1/2}\Gamma' = \begin{pmatrix} I_m & 0 \\ 0 & 0 \end{pmatrix}, \text{ and}$$

$$\Gamma C^{-1/2}WC^{-1/2}\Gamma' = \begin{pmatrix} A_{11} & A_{12} \\ A_{12}' & A_{22} \end{pmatrix}.$$

Then

$$tr \left[ W \{ C^{-1}R'[RC^{-1}R']^{-1}RC^{-1} \} \right] = tr(A_{11}). \tag{26}$$

$$\delta'W\delta = \eta_1' A_{11} \eta_1, \tag{27}$$

where  $\eta = \Gamma C^{1/2}\theta - \Gamma C^{-1/2}R'[RC^{-1}R']^{-1}r = \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix}$ .

By Courant's theorem, see Saleh (2006), we have

$$\sigma^2 \Delta^2 Ch_{min}(A_{11}) \leq \eta_1' A_{11} \eta_1 \leq \sigma^2 \Delta^2 Ch_{max}(A_{11}),$$

where  $Ch_{min}(A_{11}), Ch_{max}(A_{11})$  are, respectively, the minimum and the maximum characteristic roots of  $A_{11}$ , and  $\Delta^2 = \eta_1' \eta_1 / \sigma^2$ , so,  $\hat{\theta}^R$  performs better than  $\hat{\theta}^U$  when  $\Delta^2 \leq \frac{tr(A_{11})}{Ch_{max}(A_{11})}$ , whereas  $\hat{\theta}^U$  performs better than  $\hat{\theta}^R$  whenever  $\Delta^2 \geq \frac{tr(A_{11})}{Ch_{min}(A_{11})}$ .

ii. Comparison of  $\hat{\theta}^{PT}$  and  $\hat{\theta}^U$ :  $\hat{\theta}^{PT}$  performs better than  $\hat{\theta}^U$  when

$$\Delta^2 \leq \frac{tr(A_{11})}{Ch_{min}(A_{11})} \frac{G_{m+2}(\chi_m^2(\alpha); \Delta^2)}{\{2G_{m+2}(\chi_m^2(\alpha); \Delta^2) - G_{m+4}(\chi_m^2(\alpha); \Delta^2)\}},$$

where the opposite holds whenever

$$\Delta^2 \geq \frac{tr(A_{11})}{Ch_{min}(A_{11})} \frac{G_{m+2}(\chi_m^2(\alpha); \Delta^2)}{\{2G_{m+2}(\chi_m^2(\alpha); \Delta^2) - G_{m+4}(\chi_m^2(\alpha); \Delta^2)\}}.$$

iii. Comparison of  $\hat{\theta}^S$  and  $\hat{\theta}^U$ :  $\hat{\theta}^S$  performs better than  $\hat{\theta}^U$  whenever

$$\frac{tr(A_{11})}{Ch_{max}(A_{11})} \geq \frac{m+2}{2}.$$

Note that  $A_{11}$  involves the matrix  $W$ , hence,  $\hat{\theta}^S$  dominates  $\hat{\theta}^U$ . As  $\Delta^2 \rightarrow \infty$ , the risk difference approaches 0 from below.

iv. Comparison of  $\hat{\theta}^S$  and  $\hat{\theta}^{S+}$ : The risk difference is non-negative for all  $\Delta^2$  so we have  $R_5(\hat{\theta}^{S+}, W) \leq R_4(\hat{\theta}^S, W)$ . As a result, we can conclude that  $R_5(\hat{\theta}^{S+}, W) \leq R_4(\hat{\theta}^S, W) \leq R_1(\hat{\theta}^U, W)$ . which means that  $\hat{\theta}^{S+}$  uniformly dominates the unrestricted estimator.

### 6. Numerical Studies

In this section, we will carry out a numerical study to investigate the performance of the proposed estimators. In the first subsection, we aim to examine the relative performance of the restricted, pretest and shrinkage estimators while appointing the unrestricted estimator as a benchmark for comparison. A real dataset from the S&P500 stock market will be used to compare the performance of the estimators to confirm the analytical results obtained in the previous section.

6.1. Monte Carlo Simulation Experiments

The Monte Carlo simulation experiments will be conducted to compare the restricted, pretest and shrinkage estimators with respect to the unrestricted estimator. The following algorithm is used for the Monte Carlo simulation

1. We consider the model in Equation (5), we partition  $\theta$  as  $\theta = (\theta_1, \theta_2)$ , where  $\theta_1$  is a  $(q - m + 1) \times 1$  of non-zeros and  $\theta_2$  is  $m \times 1$  vector of zeros. We define the parameter  $\Delta^2 = \|\delta\|$  where  $\delta = \theta - \theta_0, \theta_0 = (\theta_1, \mathbf{0}), \theta = (\theta_1, \mathbf{0} + \delta)$  and  $\|\cdot\|$  denotes the Euclidean norm.  $\Delta^2$  values were chosen to vary from 0 to 0.55 and  $m = 3, 4, 5, 9, 12$  and  $15$ .
2. Generate an error term ( $\eta_t$ ) from standard normal distribution.
3. Generate the  $X$  matrix of size  $n \times (q + 1)$  with initial values estimated from standard normal distribution with  $n = 30, 50, 75, 100$  and  $150$ .
4. Estimate a matrix  $U_{n \times 1} = (\eta^2 - 1) * X\theta_0$ .
5. Estimate the vector  $Y = X\theta_0 + U$
6. Estimate the unrestricted, restricted, pretest, shrinkage and positive shrinkage estimators.
7. Compute the simulated mean squared errors (SMSE) for each estimator using the following formula

$$SMSE(\hat{\theta}^*) = \sum_{i=1}^{q+1} (\hat{\theta}^* - \theta)^2, \tag{28}$$

where  $\hat{\theta}^*$  denotes any one of  $\{\hat{\theta}^U, \hat{\theta}^R, \hat{\theta}^{PT}, \hat{\theta}^S, \hat{\theta}^{S+}\}$ .

8. Repeat steps (2)–(7)  $K$  times. We found that  $K = 3000$  is a suitable choice to obtain stable results.
9. Compute the simulated relative efficiency (SRE) as follows

$$SRE(\hat{\theta}^U, \hat{\theta}^*) = \frac{SMSE(\hat{\theta}^U)}{SMSE(\hat{\theta}^*)}, \tag{29}$$

where  $\hat{\theta}^U$  is appointed as benchmark. A value greater than one of the  $SRE(\hat{\theta}^U, \hat{\theta}^*)$  indicates that  $\hat{\theta}^*$  performs better than  $\hat{\theta}^U$  and vice versa.

Results of these simulations are reported in Figures 1–4. The numerical results effectively assure our analytical results that the positive shrinkage estimator plays the role of a safeguard against the high risks associated with the reduced model that we obtained under the set of local alternatives.  $\hat{\theta}^R$  shows the best performance under the null space and it degrades towards zero as the value of  $\Delta^2$  goes way from the null space.

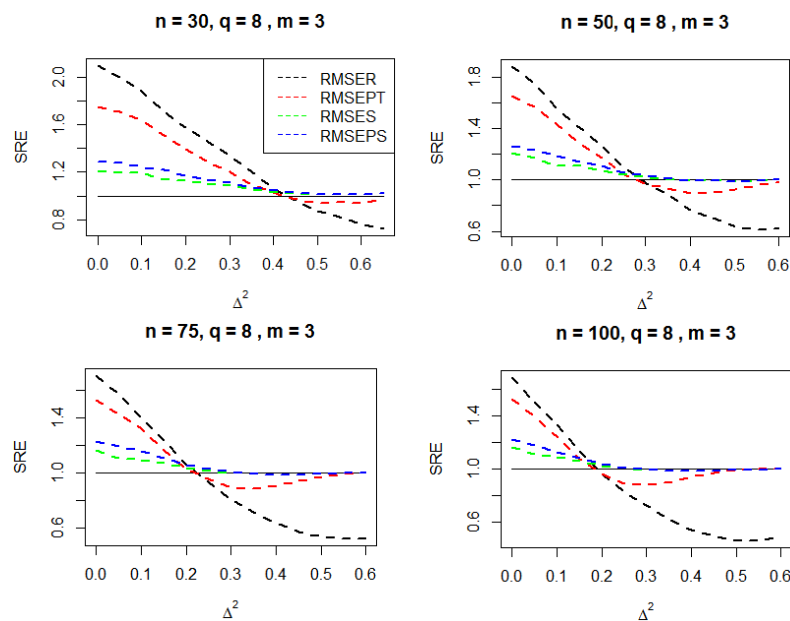
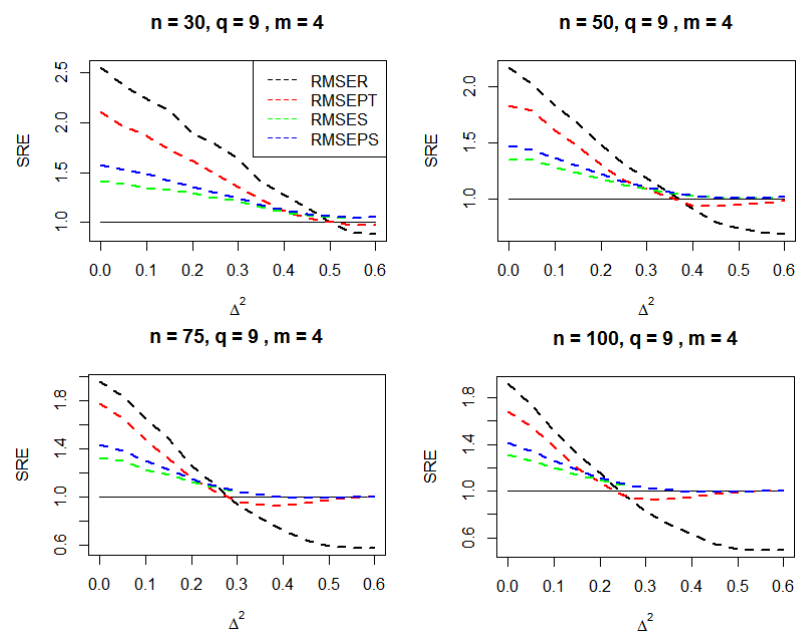
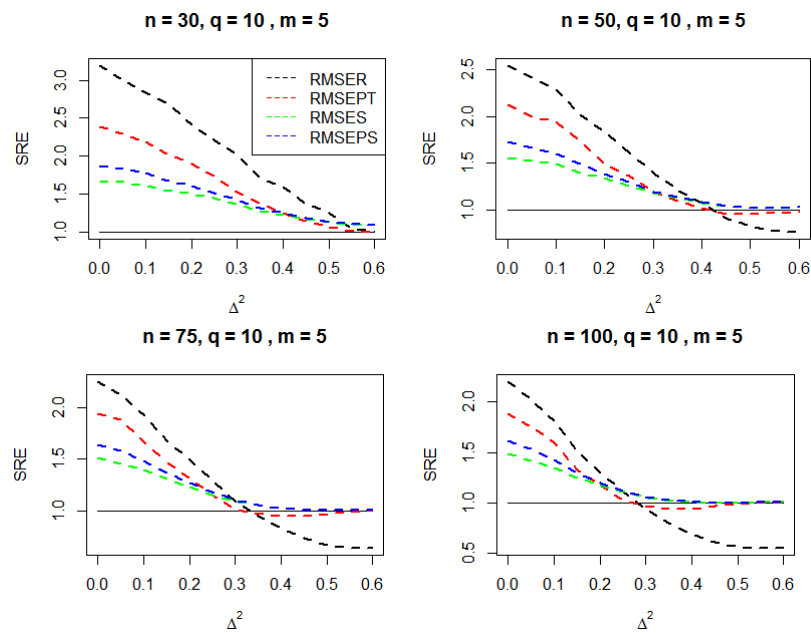


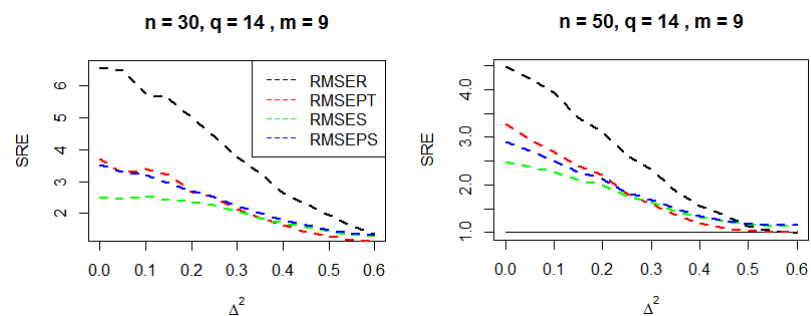
Figure 1. SRE of the restricted, pretest and shrinkage estimators with respect to  $\hat{\theta}^U$  when  $(m, q) = (3, 8)$  and for different sample sizes.



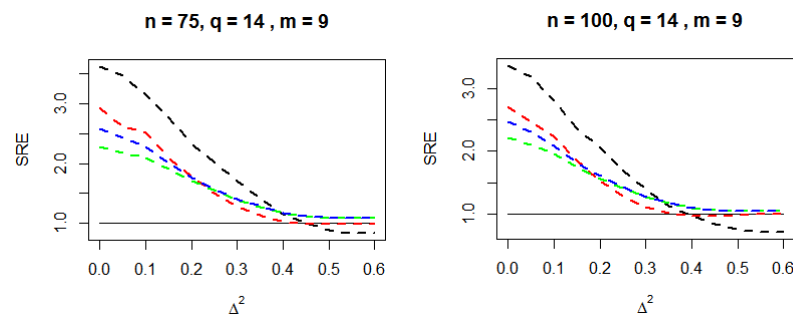
**Figure 2.** SRE of the restricted, pretest and shrinkage estimators with respect to  $\hat{\theta}^U$  when  $(m, q) = (4, 9)$  and for different sample sizes.



**Figure 3.** SRE of the restricted, pretest and shrinkage estimators with respect to  $\hat{\theta}^U$  when  $(m, q) = (5, 10)$  and for different sample sizes.



**Figure 4.** Cont.



**Figure 4.** SRE of the restricted, pretest and shrinkage estimators with respect to  $\hat{\theta}^U$  when  $(m, q) = (9, 14)$  and for different sample sizes.

As the value of  $\Delta^2$  increases, the superiority changes from  $\hat{\theta}^R$  to  $\hat{\theta}^{PT}$ ,  $\hat{\theta}^S$  and  $\hat{\theta}^{S+}$ , respectively, and  $\hat{\theta}^{S+}$  dominates other estimators, because it acts as a safeguard against the high risks associated with the reduced model.

### 6.2. Application on Standard & Poor 500 (SP500) Stock Market

The “sp500dgc” dataset contains daily closing prices of the Standard & Poor 500 (SP500) stock market that has been used by Ding et al. (1993). The dataset is also available in fGarch/R-package produced by Wuertz and Chalabi (2008). Following the illustrative example of Ding et al. (1993), we considered the most recent returns as our targeted subset from 3 December 1988 to 30 August 1991. This contains 1000 daily returns (i.e., the official working days in the financial market is 252).

To fit the ARCH model, we first conducted a Lagrange–Multiplier (LM) test to check the effect of ARCH; more details about this test can be found in Tsay (2005). Then, we fit an ARCH model with an adequate order. The order  $q = 12$  is an adequate selection for our data which represents the full model that given by Formula (30).  $\hat{\theta}^U$  is then obtained by fitting the full model.

$$\sqrt{y_t} = \sigma_t \epsilon_t, \epsilon_t \sim N(0, 1), \quad \sigma_t^2 = \omega + \alpha_1 y_{t-1} + \dots + \alpha_q y_{t-q}. \tag{30}$$

In order to obtain the UPI from the data, we used AIC and BIC selection criteria to pick the significant order under the forward selection strategy the selected order under the auxiliary information of AIC and BIC represents the reduced model given by Formula (31). Consequently, from the reduced model, we compute  $\hat{\theta}^R$ , the restricted estimator.

$$\sqrt{y_t} = \sigma_t \epsilon_t, \epsilon_t \sim N(0, 1), \quad \sigma_t^2 = \omega + \alpha_1 y_{t-1} + \dots + \alpha_{q-m+1} y_{t-q-m+1}. \tag{31}$$

To assess the performance of the estimators, we use the relative efficiency of the mean squared error (RMSE) with respect to the true parameters  $\theta$  which will be estimated by  $\hat{\theta}^*$ , where  $\hat{\theta}^*$  can be any of the estimators. The approach is based on the bootstrapping method which is similar to that introduced by Freedman (1981).

After fitting the full model on the original data, the procedure is conducted in two steps. The first step is as follows:

1. Select a sample of size  $n$  from the residuals of the full model, say  $E_1, \dots, E_n$  with replacement.
2. Compute the observations  $Y_1^*, \dots, Y_n^*$  as follows

$$Y_i^* = \hat{Y}_i + E_i, \quad i = 1, \dots, n, \tag{32}$$

where  $\hat{Y}_i$  is the  $i$ th fitted observation from the full model applied on the original data, and  $E_i$  is the  $i$ th residual in (1).

3. Fit the ARCH model on  $Y_i^*$  to obtain  $\hat{\theta}_{boot}^U(1)$ .
4. Repeat steps (1)–(3) a number of times  $K$  until stable results are obtained—we found that  $K = 3000$  worked well.

5. Compute the average of  $K$  iterations which will represent the true parameter  $\theta$ .  
 After the true parameters' vector has been estimated in the previous step, the second step is conducted as follows:

1. Select a sample of size  $n$  from the residuals of the full model, say  $E_1, \dots, E_n$  with replacement.
2. Compute  $Y_1^*, \dots, Y_n^*$  as follows

$$Y_i^* = \hat{Y}_i + E_i, \quad i = 1, \dots, n, \tag{33}$$

where  $\hat{Y}_i$  is the  $i$ th fitted observation from the full model applied on the original data and  $E_i$  is the  $i$ th residual in (1).

3. Fit both the full and reduced models and compute  $\hat{\theta}_{boot}^U(1)$  and  $\hat{\theta}_{boot}^R(1)$ , then obtain  $\hat{\theta}_{boot}^{PT}(1)$ ,  $\hat{\theta}_{boot}^S(1)$  and  $\hat{\theta}_{boot}^{S+}(1)$ .
4. Compute the predicted values  $\hat{Y}_i^*$  using the estimated parameters of all estimators

$$\hat{Y}_i^*(1) = X\hat{\theta}_{boot}(1),$$

where  $\hat{\theta}_{boot}^*(1) \in \{\hat{\theta}_{boot}^U(1), \hat{\theta}_{boot}^R(1), \hat{\theta}_{boot}^{PT}(1), \hat{\theta}_{boot}^S(1), \hat{\theta}_{boot}^{S+}(1)\}$ .

5. Compute the Bootstrapping Mean Squared Error (MSEB) of  $\hat{\theta}_{boot}^*(1)$  the estimator  $\hat{\theta}^*$  as follows:

$$MSEB\hat{\theta}_{boot}^*(1) = \sum_{i=1}^n (\hat{\theta}_{boot}^*(1) - \hat{\theta}^*)^2. \tag{34}$$

6. Repeat steps (1)–(5) a number of times  $K$  until stable results are obtained. We found that  $K = 3000$  is an adequate number of iterations.
7. Compute the relative efficiency of the mean squared error (RMSE) as follows,

$$RMSE(\theta^*) = \frac{\text{Average of MSEB for } \hat{\theta}^U}{\text{Average of MSEB for } \hat{\theta}^*}. \tag{35}$$

Results of the RMSEs for our data are reported in Table 1.

**Table 1.** Relative MSE with respect to  $\hat{\theta}^U$  for the S&P500 stock market daily closing prices.

Estimator	$\hat{\theta}^R$	$\hat{\theta}^{PT}$	$\hat{\theta}^S$	$\hat{\theta}^{S+}$
MSE	1.1236	1.0010	1.0236	1.0382

It is clear that  $\hat{\theta}^R$  outperforms all other estimators which indicates that the restriction given by the null hypothesis is correct.  $\hat{\theta}^{S+}$  comes the second and then it is  $\hat{\theta}^S$ .  $\hat{\theta}^{PT}$  performs better than  $\hat{\theta}^U$  even though it was the worst among other estimators. This may be an indication that the AIC/BIC selection criteria worked quite well on this dataset.

### 7. Conclusions

In this article, we investigated the performance of the pretest and James–Stein (shrinkage) estimators to estimate the parameter's vector  $\theta$  of the ARCH model. These estimators were first analytically compared via their asymptotic quadratic risk and asymptotic mean square error matrices and then numerically compared using simulated and real datasets to confirm our analytical results. However, the reduced model in some cases might not be the right choice: analytical and numerical results showed that the pretest and James–Stein estimators represent a safeguard against the high risks associated with the reduced model that we obtain under the set of local alternatives.

Historically, the ARCH model is the simplest version of ARCH family models; however, its drawback is that it requires many parameters to adequately describes the volatility

of such phenomena, and the positive James–Stein estimator should successfully overcome this dilemma by providing a parsimonious submodel (reduced model). To obtain a UPI, we used AIC and BIC selection criteria to select the reduced model.

According to our research findings, it is recommended that the positive James–Stein estimator is used as it outperforms all other estimators regardless of whether the restriction given by the null hypothesis is true. In addition, the proposed estimation strategy can be applied to different ARCH family models.

**Author Contributions:** Conceptualization, A.B.A.D. and M.A.-M.; methodology, A.B.A.D. and M.A.-M.; software, A.B.A.D.; formal analysis, A.B.A.D.; investigation, M.A.-M.; writing—original draft, A.B.A.D.; writing—review and editing, A.B.A.D. and M.A.-M.; supervision, M.A.-M. All authors have read and agreed to the published version of the manuscript.

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

**Data Availability Statement:** The data used in the real example part can be found in fGarch/R- package, as produced by [Wuertz and Chalabi \(2008\)](#).

**Acknowledgments:** The authors thank the Academic Editor and the three reviewers for their valuable comments and suggestions which significantly improved the paper.

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

#### Appendix A. Proof of Theorem 4

- (1) The proof follows from ([Francq and Zakoian 2004](#); [Francq and Zakoian 2012](#); [Weiss 1986](#)).
- (2)

$$\begin{aligned}
 T_n^{(2)} &= \sqrt{n}(\hat{\theta}^R - \theta) = \sqrt{n}\{\hat{\theta}^U + C^{-1}R'[RC^{-1}R']^{-1}(r - R\hat{\theta}^U) - \theta\} \\
 &= \sqrt{n}(\hat{\theta}^U - \theta) + \sqrt{n}\{C^{-1}R'[RC^{-1}R']^{-1}(r - R\hat{\theta}^U)\} \\
 &= \sqrt{n}(\hat{\theta}^U - \theta) - \sqrt{n}\{C^{-1}R'[RC^{-1}R']^{-1}(R(\hat{\theta}^U - \theta) \\
 &\quad + R\theta - r)\} \\
 &= \sqrt{n}(\hat{\theta}^U - \theta) - C^{-1}R'[RC^{-1}R']^{-1}R\sqrt{n}(\hat{\theta}^U - \theta) + \\
 &\quad C^{-1}R'[RC^{-1}R']^{-1}\sqrt{n}(R\theta - r) \\
 &= T_n^{(1)} + C^{-1}R'[RC^{-1}R']^{-1}RT_n^{(1)} - C^{-1}R'[RC^{-1}R']^{-1}\sqrt{n}(R\theta - r) \\
 &= \left[ I_q - C^{-1}R'[RC^{-1}R']^{-1}R \right] T_n^{(1)} - \sqrt{n}\{C^{-1}R'[RC^{-1}R']^{-1}(R\theta - r)\},
 \end{aligned}$$

where **I** is the identity matrix.

$T_n^{(2)}$  is a linear combination in  $T_n^{(1)}$  that can be represented in a matrix format as  $T_n^{(2)} = A_2T_n^{(1)} - \mathbf{B}_2$  where  $A_2$  and  $\mathbf{B}_2$  are given as follows

$$A_2 = \left[ I_q - C^{-1}R'[RC^{-1}R']^{-1}R \right]_{q \times q}, \quad \mathbf{B}_2 = \left[ C^{-1}R'[RC^{-1}R']^{-1}(R\theta - r) \right]_{q \times 1}.$$

From Theorem 4 part (1), as  $n \rightarrow \infty$ ,  $T_n^{(2)} \xrightarrow{L} T^{(2)}$  and by Slutsky's Theorem,

$$T_n^{(2)} \xrightarrow{L} T^{(2)} \sim \mathcal{N}_q\left(\boldsymbol{\mu}^{(2)}, \boldsymbol{\Sigma}^{(2)}\right), \tag{A1}$$

with  $\boldsymbol{\mu}^{(2)}$  and  $\boldsymbol{\Sigma}^{(2)}$  which are given by

$$\begin{aligned}
 \boldsymbol{\mu}^{(2)} &= -C^{-1}R'[RC^{-1}R']^{-1}(R\theta - r) \\
 &= -\boldsymbol{\delta},
 \end{aligned}$$

and

$$\begin{aligned} \Sigma^{(2)} &= \sigma^2 C^{-1} - 2\sigma^2 A + C^{-1} R' [RC^{-1} R']^{-1} R \sigma^2 C^{-1} [RC^{-1} R']^{-1} RC^{-1} \\ &= \sigma^2 [C^{-1} - 2A + A] \\ &= \sigma^2 [C^{-1} - A]. \end{aligned}$$

Similarly, we can prove Formulas (1)–(6).

### Appendix B. Proof of Theorem 5

The proof of Formulas (1) and (2) are straightforward.

(3)

$$\begin{aligned} \sqrt{n}(\hat{\theta}^{PT} - \theta) &= \sqrt{n} \left( \hat{\theta}^U - (\hat{\theta}^U - \hat{\theta}^R) I(\mathcal{L} \leq \mathcal{L}_\alpha) - \theta \right) \\ &= \sqrt{n}(\hat{\theta}^U - \theta) - \sqrt{n}(\hat{\theta}^U - \hat{\theta}^R) I(\mathcal{L} \leq \mathcal{L}_\alpha) \\ &= T_n^{(1)} + \left( \sqrt{n}(X'X)^{-1} R' [RC^{-1} R']^{-1} \right. \\ &\quad \left. (r - R\theta) I(\mathcal{L} \leq \mathcal{L}_\alpha) \right), \end{aligned}$$

As  $n \rightarrow \infty$ , with Slutsky's theorem, we have  $\mathcal{L}_n \xrightarrow{L} \mathcal{L} \sim \chi_m^2$  and  $\mathcal{L}_{n,\alpha} \xrightarrow{L} \mathcal{L} \sim \chi_m^2(\alpha)$ , then,

$$\begin{aligned} \sqrt{n}(\hat{\theta}^{PT} - \theta) &= C^{-1} R' [RC^{-1} R']^{-1} (r - R\theta) G_{m+2}(\chi_m^2(\alpha); \Delta^2) \\ &= -\delta G_{m+2}(\chi_m^2(\alpha); \Delta^2) \\ B_3(\hat{\theta}^{PT}) &= \Delta^2 [G_{m+2}(\chi_m^2(\alpha); \Delta^2)]^2. \end{aligned}$$

Similarly, we can prove Formulas (4) and (5).

### Appendix C. Proof of Theorem 6

(1) Part (1) is straightforward.

(2) Note that  $n(\hat{\theta}^R - \theta)(\hat{\theta}^R - \theta)' = T_n^{(2)} T_n^{(2)'}.$  Then, by Theorem 4 part (2), we have

$$\begin{aligned} M_2(\hat{\theta}^R) &= E[n(\hat{\theta}^R - \theta)(\hat{\theta}^R - \theta)'] \\ &= E\{n\{(\hat{\theta}^U - \theta) - C^{-1} R' [RC^{-1} R']^{-1} (R\theta - r)\} \\ &\quad \{(\hat{\theta}^U - \theta) - C^{-1} R' [RC^{-1} R']^{-1} (R\theta - r)\}'\} \\ &= \sigma^2 C^{-1} + C^{-1} R' [RC^{-1} R']^{-1} E\{(R\hat{\theta}^U - r)(R\hat{\theta}^U - r)'\} \times \\ &\quad [RC^{-1} R']^{-1} RC^{-1} - 2C^{-1} R' [RC^{-1} R']^{-1} \times E\{(R\hat{\theta}^U - r) \\ &\quad (R\hat{\theta}^U - r)'\} \\ &= \sigma^2 C^{-1} + C^{-1} R' [RC^{-1} R']^{-1} \{\sigma^2 (RC^{-1} R') + (R\theta - r)(R\theta - r)'\} \times \\ &\quad [RC^{-1} R']^{-1} RC^{-1} - 2\sigma^2 C^{-1} R' [RC^{-1} R']^{-1} RC^{-1} \\ &= \sigma^2 C^{-1} - \sigma^2 C^{-1} R' [RC^{-1} R']^{-1} RC^{-1} + \delta \delta \\ &= \sigma^2 [C^{-1} - A] + \delta \delta. \\ R_2(\hat{\theta}^R, W) &= \sigma^2 tr(W(C^{-1} - A)) + \delta W \delta. \end{aligned}$$

Similarly, Formulas (3)–(5) can be proven.

## References

- Ahmed, S. Ejaz, Abdulkadir Hussein, and Marwan Al-Momani. 2015. Efficient estimation for the conditional autoregressive model. *Journal of Statistical Computation and Simulation* 85: 2569–81. [\[CrossRef\]](#)
- Ahmed, S. Ejaz, and Bahadır Yüzbaşı. 2016. Big data analytics: Integrating penalty strategies. *International Journal of Management Science and Engineering Management* 11: 105–15. [\[CrossRef\]](#)
- Ahmed, S. Ejaz, and Bahadır Yüzbaşı. 2017. High dimensional data analysis: Integrating submodels. In *Big and Complex Data Analysis*. Cham: Springer, pp. 285–304.
- Ahmed, S. Ejaz, and Enayetur Raheem. 2012. Shrinkage and absolute penalty estimation in linear regression models. *Wiley Interdisciplinary Reviews: Computational Statistics* 4: 541–53. [\[CrossRef\]](#)
- Ahmed, S. Ejaz. 2014. *Penalty, Shrinkage and Pretest Strategies: Variable Selection and Estimation*. Berlin and Heidelberg: Springer.
- Akaike, Hirotugu. 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19: 716–23. [\[CrossRef\]](#)
- Al-Momani, Marwan, Abdulkadir Hussein, and S. Ejaz. Ahmed. 2016. Penalty and related estimation strategies in the spatial error model. *Statistica Neerlandica* 71: 4–30. [\[CrossRef\]](#)
- Al-Momani, Marwan, Syed Ejaz Ahmed, and Abdul A. Hussein. 2019. Efficient Estimation Strategies for Spatial Moving Average Model. In *Biometrics Proceedings of the Thirteenth International Conference on Management Science and Engineering Management*. Cham: Springer, pp. 520–43.
- Bancroft, Theodore Alfonso. 1944. On biases in estimation due to the use of preliminary tests of significance. *The Annals of Mathematical Statistics* 15: 190–204. [\[CrossRef\]](#)
- Brockwell, Peter J., and Richard A. Davis. 2016. *Introduction to Time Series and Forecasting*. New York: Springer.
- Dawod, Abdaljbbar, Marwan Al-Momani, and Saddam Akber Abbasi. 2018. On efficient estimation strategies in monitoring of linear profiles. *The International Journal of Advanced Manufacturing Technology* 96: 3977–91. [\[CrossRef\]](#)
- Ding, Zhuangxin, Clive W. J. Granger, and Robert F. Engle. 1993. A long memory property of stock market returns and a new model. *Journal of Empirical Finance* 1: 83–106. [\[CrossRef\]](#)
- Emmert-Streib, Frank, and Matthias Dehmer. 2019. High-Dimensional LASSO-Based Computational Regression Models: Regularization, Shrinkage, and Selection. *Machine Learning and Knowledge Extraction* 1: 359–83. [\[CrossRef\]](#)
- Engle, Robert F. 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the Econometric Society* 50: 987–1007. [\[CrossRef\]](#)
- Franco, Christian, and Jean-Michel Zakoïan. 2004. Maximum likelihood estimation of pure garch and arma-garch processes. *Bernoulli* 10: 605–37. [\[CrossRef\]](#)
- Franco, Christian, and Jean-Michel Zakoïan. 2010. *Garch Models: Structure, Statistical Inference and Financial Applications*. Hoboken: John Wiley & Sons.
- Franco, Christian, and Jean-Michel Zakoïan. 2012. Strict stationarity testing and estimation of explosive and stationary generalized autoregressive conditional heteroscedasticity models. *Econometrica* 80: 821–61.
- Freedman, David A. 1981. Bootstrapping regression models. *The Annals of Statistics* 9: 1218–28. [\[CrossRef\]](#)
- Grublytė, Ieva, Donatas Surgailis, and Andrius Škarnulis. 2017. Qmle for quadratic arch model with long memory. *Journal of Time Series Analysis* 38: 535–51. [\[CrossRef\]](#)
- Hansen, Lars Peter. 1982. Large sample properties of generalized method of moments estimators. *Econometrica: Journal of the Econometric Society* 50: 1029–54. [\[CrossRef\]](#)
- Neter, John, Michael H. Kutner, Christopher J. Nachtsheim, and William Wasserman. 1996. *Applied Linear Statistical Models*. Chicago: Irwin Chicago, vol. 4.
- Peiris, Ushan, and T. S. G. Peiris. 2011. Measuring stock market volatility in an emerging economy: Empirical evidence from the colombo stock exchange (CSE). In *International Research Conference*. Durgapur: University of Kelaniya-Sri Lanka.
- Rathnayaka, R. M., D. M. Seneviratne, and S. C. Nagahawatta. 2013. Statistical techniques approach for evaluating the market fluctuations; the case study in colombo stock exchange. In *International Conference on Business & Information*. Durgapur: University of Kelaniya-Sri Lanka.
- Rich, Robert W., Jennie Raymond, and John Scott Butler. 1991. Generalized instrumental variables estimation of autoregressive conditional heteroskedastic models. *Economics Letters* 35: 179–85. [\[CrossRef\]](#)
- Saleh, A. K. Md Ehsanes. 2006. *Theory of Preliminary Test and Stein-Type Estimation with Applications*. Hoboken: John Wiley & Sons, vol. 517.
- Schwarz, Gideon. 1978. Estimating the dimension of a model. *The Annals of Statistics* 6: 461–64. [\[CrossRef\]](#)
- Stein, Charles. 1966. An approach to the recovery of interblock information in balanced incomplete block designs. In *Research Paper in statistics: Festschrift for J. Neyman*. London: Wiley, pp. 351–66.
- Stein, Charles. 1956. Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *Volume 1 Contribution to the Theory of Statistics, Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*; Berkeley: University of California Press, vol. 1, pp. 197–206.
- Teräsvirta, Timo. 2009. An introduction to univariate garch models. In *Handbook of Financial Time Series*. Berlin and Heidelberg: Springer, pp. 17–42.
- Thomson, T., S. Hossain, and M. Ghahramani. 2016. Efficient estimation for time series following generalized linear models. *Australian & New Zealand Journal of Statistics* 58: 493–513.

- Tsay, Ruey S. 2005. *Analysis of Financial Time Series*. Hoboken: John Wiley & Sons, vol. 48. [[CrossRef](#)]
- Wang, Changlin. 2021. Different garch models analysis of returns and volatility in bitcoin. *Data Science in Finance and Economics* 1: 37–59. [[CrossRef](#)]
- Weiss, Andrew A. 1986. Asymptotic theory for arch models: Estimation and testing. *Econometric Theory* 2: 107–31. [[CrossRef](#)]
- Wuertz, Diethelm, and Yohan Chalabi. 2008. fgarch: Rmetrics-Autoregressive Conditional Heteroskedastic Modelling, r Package Version 290.76. Available online: <https://www.rmetrics.org> (accessed on 3 April 2022).
- Yüzbaşı, Bahadır, and S. Ejaz Ahmed. 2020. Ridge Type Shrinkage Estimation of Seemingly Unrelated Regressions Furthermore, Analytics of Economic and Financial Data from “Fragile Five” Countries. *Journal of Risk and Financial Management* 13: 131. [[CrossRef](#)]