

Article

Second-Order Least Squares Estimation in Nonlinear Time Series Models with ARCH Errors

Mustafa Salamh ¹  and Liqun Wang ^{2,*} ¹ Department of Statistics, Cairo University, Giza 12613, Egypt; Mustafa.Salamh@feps.edu.eg² Department of Statistics, University of Manitoba, Winnipeg, MB R3T 2N2, Canada

* Correspondence: Liqun.Wang@umanitoba.ca

Abstract: Many financial and economic time series exhibit nonlinear patterns or relationships. However, most statistical methods for time series analysis are developed for mean-stationary processes that require transformation, such as differencing of the data. In this paper, we study a dynamic regression model with nonlinear, time-varying mean function, and autoregressive conditionally heteroscedastic errors. We propose an estimation approach based on the first two conditional moments of the response variable, which does not require specification of error distribution. Strong consistency and asymptotic normality of the proposed estimator is established under strong-mixing condition, so that the results apply to both stationary and mean-nonstationary processes. Moreover, the proposed approach is shown to be superior to the commonly used quasi-likelihood approach and the efficiency gain is significant when the (conditional) error distribution is asymmetric. We demonstrate through a real data example that the proposed method can identify a more accurate model than the quasi-likelihood method.

Keywords: nonlinear dynamic model; ARCH error; mixing process; mean nonstationarity; second order least squares; semiparametric efficiency; econometric modeling; financial time series



Citation: Salamh, Mustafa, and Liqun Wang. 2021. Second-Order Least Squares Estimation in Nonlinear Time Series Models with ARCH Errors. *Econometrics* 9: 41. <https://doi.org/10.3390/econometrics9040041>

Academic Editor: Ryo Okui

Received: 20 September 2021
Accepted: 23 November 2021
Published: 27 November 2021

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



Copyright: © 2021 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

Dynamic models have been widely applied in the analysis of economic and financial data. Most theories and methods are developed for mean-stationary data generating processes, specifically ARMA processes, although ARIMA processes have also been studied (Koul and Ling 2006; Ling 2003; Ling and McAleer 2003; Meitz and Saikkonen 2008). However, many financial and economic variables exhibit nonlinear behaviour or relationships (e.g., Enders 2010; Franses and Van Dijk 2000). As pointed out by Li et al. (2002), consistent estimation of variance parameters may be misleading or impossible if the conditional mean function is not adequately specified. Therefore, more general and flexible models with time varying mean functions are desirable to capture the nonlinear dynamic behaviour and the structural relationships in the real data.

On the other hand, the autoregressive conditional heteroscedasticity (ARCH) model and its various generalizations have been widely used to analyze economic and financial data. These models allow both the conditional means and variances of a process to jointly evolve over time. The mainstream method for estimation and inference in generalized ARCH (GARCH) models is likelihood based (e.g., Engle 1982; Engle and Gonzalez-Rivera 1991; Weiss 1986), although the estimating function approach is also studied (Li and Turtle 2000). So far most research focuses on the quasi-likelihood method for various generalizations of the ARCH error component while keeping the process mean function very simple, e.g., constant or linear. Therefore from both theoretical and practical points of view, it is important to develop methodologies for mean nonstationary ARCH processes.

In this paper, we consider a model with a fairly general time varying nonlinear mean function and ARCH error that covers both stationary and mean-nonstationary processes. In

particular, we propose a so-called second-order least squares (SLS) approach based on the first two conditional moments of the process. We establish the consistency and asymptotic normality for the proposed estimator under general mixing conditions. We demonstrate that this approach is more efficient than the commonly used quasi-likelihood method and the efficiency gain is significant when the conditional error distribution is asymmetric. We also carry out extensive simulations to study finite sample properties of our proposed estimator and compare its performance with other related estimators. Our results show that in most cases the optimal SLS estimator has superior performance over the estimating function estimators based on the same set of moments. Finally, we apply our approach to the empirical example of the U.K. inflation of Engle (1982), which leads to a different model specification than the quasi-likelihood method.

It is worthwhile to note that some researchers have obtained more efficient estimators than the Gaussian quasi-maximum likelihood estimator (QMLE) by assuming various parametric families of error distributions (see a recent survey by Zhu and Li (2015)). However, our approach is based on the first two conditional moments of the process only and does not require any distributional assumptions. The SLS method was first used by Wang (2003, 2004) to estimate the nonlinear measurement error models. Later, it was extended to the nonlinear longitudinal data models by Wang (2007) and to the censored linear models by Abarin and Wang (2009). Wang and Leblanc (2008) showed that under a nonlinear cross-sectional data model, the SLS estimator is asymptotically more efficient than the ordinary least squares estimator when the error term has nonzero third moment, and both estimators are equally efficient otherwise. Further, Kim and Ma (2012) showed that the SLS estimator attains the optimal semiparametric efficiency bound in general. More recently, Rosadi and Filzmoser (2019); Rosadi and Peiris (2014) and Salamh and Wang (2021) used this method in dynamic models. It has also been applied in the optimal design problems by several researchers, e.g., Bose and Mukerjee (2015); Gao and Zhou (2017); Yin and Zhou (2017) and He and Yue (2019).

The paper is organized as follows. In Section 2 we introduce the model and SLS estimator and establish its consistency and asymptotic normality. In Section 3 we derive the optimal SLS estimator and propose a feasible version of it. We also investigate the efficiency gain of the optimal SLS estimator relative to the QMLE and highlight the differences between our approach and that of the estimating functions. In Section 4 we carry out Monte Carlo simulations to study the finite sample behavior of the SLS estimator in the cases of both skewed and leptokurtic conditional error distributions, and compare it with the QMLE and some other related estimators. In Section 5 we apply the SLS approach to an empirical analysis of the U.K. inflation data. Finally, conclusions and discussion are given in Section 6, while regularity assumptions and mathematical proofs are given in the Appendix A.

2. Model and SLS Estimation

Let $\{(y_t, x_t')\}$ be a sequence of random vectors defined on a complete probability space (Ω, \mathcal{F}, P) and denote $v_t = (x_t', y_{t-1}, x_{t-1}', \dots, y_{t-\tau}, x_{t-\tau}')'$ for some nonnegative integer τ . We consider the model

$$y_t = f_t(v_t, \theta_0) + \epsilon_t, \quad t \in \mathbb{Z}, \quad (1)$$

where $f_t: \mathbb{R}^v \times \Theta \rightarrow \mathbb{R}^1$ are known measurable functions on \mathbb{R}^v for each $\theta \in \Theta \subset \mathbb{R}^q$, and continuous on Θ uniformly in t a.s.- P . Further, denote the σ -field $\mathcal{F}_{t-1} = \mathcal{F}\{x_i, y_{i-1}, i \leq t\}$ and assume that $\epsilon_t = \sigma_t \varepsilon_t$ satisfying

$$E(\epsilon_t | \mathcal{F}_{t-1}) = 0 \quad \text{a.s.-}P, \quad E(\epsilon_t^2 | \mathcal{F}_{t-1}) = 1 \quad \text{a.s.-}P, \quad \sigma_t^2 = \phi_{00} + \sum_{i=1}^p \phi_{0i} \epsilon_{t-i}^2, \quad (2)$$

where $\phi_{00}, \phi_{0p} > 0$, and $\phi_{0i} \geq 0$ for $i = 1, 2, \dots, p-1$. It is easy to see that the linear model with ARCH error is a special case of model (1) and (2). Our main goal is to estimate $\gamma_0 = (\theta_0', \phi_0')' \in \Gamma$ which is a compact subset of \mathbb{R}^{q+p+1} .

Given the random sample $(y_T, \mathbf{x}'_T, \dots, y_{1-p-\tau}, \mathbf{x}'_{1-p-\tau})$, let $\epsilon_t(\boldsymbol{\theta}) = y_t - f_t(\mathbf{v}_t, \boldsymbol{\theta})$, $\sigma_t^2(\gamma) = \phi_0 + \sum_{i=1}^p \phi_i \epsilon_{t-i}^2(\boldsymbol{\theta})$ and $\mathbf{h}_t(\gamma) = (\epsilon_t(\boldsymbol{\theta}), y_t^2 - f_t^2(\mathbf{v}_t, \boldsymbol{\theta}) - \sigma_t^2(\gamma))'$. Then the second-order least squares (SLS) estimator is defined as the \mathcal{F} -measurable function satisfying

$$\hat{\gamma}_T = \underset{\gamma \in \Gamma}{\operatorname{argmin}} Q_T(\gamma) \quad \text{a.s.-}P, \tag{3}$$

where

$$Q_T(\gamma) = T^{-1} \sum_{t=1}^T \mathbf{h}'_t(\gamma) \mathbf{W}_t \mathbf{h}_t(\gamma)$$

and \mathbf{W}_t is a nonnegative definite matrix that is measurable, with respect to $\mathcal{F}\{\mathbf{v}_t, \dots, \mathbf{v}_{t-p}\}$.

Next we establish the consistency and asymptotic normality of $\hat{\gamma}_T$ under the regularity conditions given in Appendix A. The consistency of $\hat{\gamma}_T$ follows from the uniform convergence of $Q_T(\gamma)$ on Γ to a non-random sequence $\bar{Q}_T(\gamma)$ which have unique minimizer at γ_0 for sufficiently large T .

Theorem 1. Under Assumptions A1–A3, $\hat{\gamma}_T \xrightarrow{a.s.} \gamma_0$, as $T \rightarrow \infty$.

Theorem 2. Under Assumptions A1–A9, $\mathbf{V}_T^{-1/2} \bar{\mathbf{A}}_T(\gamma_0) \sqrt{T}(\hat{\gamma}_T - \gamma_0) \xrightarrow{d} N(\mathbf{0}, \mathbf{I}_{q+p+1})$, as $T \rightarrow \infty$, where

$$\bar{\mathbf{A}}_T(\gamma_0) = 2T^{-1} \sum_{t=1}^T E\{\nabla_{\gamma} \mathbf{h}'_t(\gamma_0) \mathbf{W}_t \nabla_{\gamma'} \mathbf{h}_t(\gamma_0)\} \tag{4}$$

and

$$\mathbf{V}_T = 4T^{-1} \sum_{t=1}^T E\{\nabla_{\gamma} \mathbf{h}'_t(\gamma_0) \mathbf{W}_t \mathbf{h}_t(\gamma_0) \mathbf{h}'_t(\gamma_0) \mathbf{W}_t \nabla_{\gamma'} \mathbf{h}_t(\gamma_0)\}. \tag{5}$$

The proofs are given in the Appendix A.

3. Optimal SLS Estimator

From Theorem 2 the asymptotic covariance (acov) of $\sqrt{T}(\hat{\gamma}_T - \gamma_0)$ is given by $\bar{\mathbf{A}}_T^{-1}(\gamma_0) \mathbf{V}_T \bar{\mathbf{A}}_T^{-1}(\gamma_0)$ which depends on the weights \mathbf{W}_t , $t = 1, 2, \dots, T$. Therefore it is of interest to find the (asymptotically) optimal estimator, say $\hat{\gamma}_T^o$, which has the smallest asymptotic variance in the class of estimators defined by (3), i.e., for any estimator $\hat{\gamma}_T$ satisfying (3), $\operatorname{acov} \sqrt{T}(\hat{\gamma}_T - \gamma_0) - \operatorname{acov} \sqrt{T}(\hat{\gamma}_T^o - \gamma_0)$ is nonnegative definite. The following theorem gives the optimal choice of \mathbf{W}_t to achieve this goal.

Theorem 3. Suppose $\mathbf{U}_t = E\{\mathbf{h}_t(\gamma_0) \mathbf{h}'_t(\gamma_0) | \mathbf{v}_t, \dots, \mathbf{v}_{t-p}\}$ is nonsingular a.s.- P , and Assumptions A2, A3, A6–A9 hold with $\mathbf{W}_t = \mathbf{U}_t^{-1}$. Then, the asymptotically optimal SLS (OSLS) estimator $\hat{\gamma}_T^o$ is obtained by using $\mathbf{W}_t = \mathbf{U}_t^{-1}$, $t = 1, 2, \dots, T$. Further, the corresponding (inverse) optimal covariance matrix is given by

$$\operatorname{acov}^{-1} \sqrt{T}(\hat{\gamma}_T^o - \gamma_0) = T^{-1} \sum_{t=1}^T E\{\nabla_{\gamma} \mathbf{h}'_t(\gamma_0) \mathbf{U}_t^{-1} \nabla_{\gamma'} \mathbf{h}_t(\gamma_0)\}, \tag{6}$$

$$= T^{-1} \sum_{t=1}^T E\{\mathbf{B}'_t \boldsymbol{\Omega}_t^{-1} \mathbf{B}_t\}, \tag{7}$$

where

$$\mathbf{B}'_t = \begin{pmatrix} \nabla_{\boldsymbol{\theta}} f_t(\mathbf{v}_t, \boldsymbol{\theta}_0) & \nabla_{\boldsymbol{\theta}} \sigma_t^2(\gamma_0) \\ \mathbf{0} & \nabla_{\boldsymbol{\phi}} \sigma_t^2(\gamma_0) \end{pmatrix}, \tag{8}$$

and

$$\boldsymbol{\Omega}_t = \sigma_t^2(\gamma_0) \begin{pmatrix} 1 & \sigma_t(\gamma_0) E(\epsilon_t^3 | \mathbf{v}_t, \dots, \mathbf{v}_{t-p}) \\ \cdot & \sigma_t^2(\gamma_0) [E(\epsilon_t^4 | \mathbf{v}_t, \dots, \mathbf{v}_{t-p}) - 1] \end{pmatrix}. \tag{9}$$

However, the OLS estimator $\hat{\gamma}_T^0$ is infeasible because \mathbf{U}_t depends on γ_0 and $E(\epsilon_t^j | \mathbf{v}_t, \dots, \mathbf{v}_{t-p})$, $j = 3, 4$. In practice a two-step procedure can be used as follows. First, a first-step consistent estimator of γ_0 is calculated, such as the QMLE or simply the SLS $\hat{\gamma}_T$ using the identity weight matrix. Second, the residuals $\hat{\epsilon}_t$ are calculated and suitable autoregressive models are fitted to $\hat{\epsilon}_t^3$ and $\hat{\epsilon}_t^4$ respectively. Finally, these fitted values are substituted in \mathbf{U}_t and the second-step estimator is calculated using the estimated optimal weights $\mathbf{W}_t = \hat{\mathbf{U}}_t^{-1}$. Under some general conditions this two-step estimator is consistent and, moreover, it has the same asymptotic variance as in (7) if $\hat{\mathbf{U}}_t$ is consistent for \mathbf{U}_t . Henceforth, this two-step estimator will be called the feasible optimal SLS estimator (FSLS). Note that fitting the autoregressive models is useful if the errors $\{\epsilon_t\}$ are not i.i.d., otherwise the sample means of $\hat{\epsilon}_t^3$ and $\hat{\epsilon}_t^4$ can be used. Alternatively, if the conditioning set $\{\mathbf{v}_t, \dots, \mathbf{v}_{t-p}\}$ is reasonably small, one can use nonparametric estimators of the conditional skewness and kurtosis to obtain $\hat{\mathbf{U}}_t$. More details about the two-step estimators can be found in White (1996, Section 6.3).

In the rest of this section we investigate the efficiency gain of the OLS estimator compared to the Gaussian QMLE, which is one of the most popular methods of estimation in GARCH models. The asymptotic properties of the QMLE are studied by Weiss (1986) and Bollerslev and Wooldridge (1992). Specifically for model (1) and (2), the Gaussian QMLE is defined as

$$\hat{\gamma}_T^Q = \underset{\gamma \in \Gamma}{\operatorname{argmin}} T^{-1} \sum_{t=1}^T \log \sigma_t^2(\gamma) + \frac{\epsilon_t^2(\theta)}{\sigma_t^2(\gamma)} \quad \text{a.s.-}P. \tag{10}$$

Under similar conditions as Assumptions A2–A9 and similar to the proofs of Theorems 1 and 2, we can show that $\hat{\gamma}_T^Q$ is \sqrt{T} -consistent with $\operatorname{acov} \sqrt{T} (\hat{\gamma}_T^Q - \gamma_0)$ given by

$$T \left(\sum_{t=1}^T E \{ \mathbf{B}'_t \boldsymbol{\Sigma}_t^{-1} \mathbf{B}_t \} \right)^{-1} \left(\sum_{t=1}^T E \{ \mathbf{B}'_t \boldsymbol{\Sigma}_t^{-1} \boldsymbol{\Omega}_t \boldsymbol{\Sigma}_t^{-1} \mathbf{B}_t \} \right) \left(\sum_{t=1}^T E \{ \mathbf{B}'_t \boldsymbol{\Sigma}_t^{-1} \mathbf{B}_t \} \right)^{-1}, \tag{11}$$

where \mathbf{B}_t and $\boldsymbol{\Omega}_t$ are defined in (8) and (9) respectively and

$$\boldsymbol{\Sigma}_t = \begin{pmatrix} \sigma_t^2(\gamma_0) & 0 \\ 0 & 2\sigma_t^4(\gamma_0) \end{pmatrix}.$$

Further, similar to the proof of Theorem 3, we can show that

$$\operatorname{acov} \sqrt{T} \mathbf{a}' (\hat{\gamma}_T^0 - \gamma_0) \leq \operatorname{acov} \sqrt{T} \mathbf{a}' (\hat{\gamma}_T^Q - \gamma_0)$$

for any $\mathbf{a} \in \mathbb{R}^{q+p+1}$, and the equality holds if and only if for $t = 1, 2, \dots, T$,

$$\boldsymbol{\Omega}_t \boldsymbol{\Sigma}_t^{-1} \mathbf{B}_t \mathbf{a} = \mathbf{B}_t \left(\sum_{t=1}^T E \{ \mathbf{B}'_t \boldsymbol{\Omega}_t^{-1} \mathbf{B}_t \} \right)^{-1} \left(\sum_{t=1}^T E \{ \mathbf{B}'_t \boldsymbol{\Sigma}_t^{-1} \mathbf{B}_t \} \right) \mathbf{a} \quad \text{a.s.-}P. \tag{12}$$

The above general condition can be simplified under specific settings. For example, if the process $\{(y_t, \mathbf{x}'_t, \sigma_t, \epsilon_t)\}$ is stationary with $E(\epsilon_t^3 | \mathbf{v}_t, \dots, \mathbf{v}_{t-p}) = 0$, and $E(\epsilon_t^4 | \mathbf{v}_t, \dots, \mathbf{v}_{t-p}) = \mu_4$, then it can be shown that Equation (12) is equivalent to

$$\mathbf{a}'_1 (\mathbf{I}_q - \mathbf{C}) \nabla_{\theta} f_t(\mathbf{v}_t, \boldsymbol{\theta}_0) = 0, \quad \mathbf{a}'_1 \left(\frac{\mu_4 - 1}{2} \mathbf{I}_q - \mathbf{C} \right) \nabla_{\theta} \sigma_t^2(\gamma_0) = 0, \tag{13}$$

where $\mathbf{C} = \left(\mathbf{C}_1 + \frac{1}{2} \mathbf{C}_2 \right) \left(\mathbf{C}_1 + \frac{1}{\mu_4 - 1} \mathbf{C}_2 \right)^{-1}$, $\mathbf{C}_1 = E \left\{ \sigma_t^{-2}(\gamma_0) \nabla_{\theta} f_t(\mathbf{v}_t, \boldsymbol{\theta}_0) \nabla_{\theta'} f_t(\mathbf{v}_t, \boldsymbol{\theta}_0) \right\}$, $\mathbf{C}_2 = E \left\{ \sigma_t^{-4}(\gamma_0) \nabla_{\theta} \sigma_t^2(\gamma_0) \nabla_{\theta'} \sigma_t^2(\gamma_0) \right\}$ and \mathbf{a}_1 is the subvector of the first q elements of \mathbf{a} .

Since it is difficult to quantify the difference between the asymptotic covariance matrices in (7) and (11) in general, in the following we calculate some examples under a simple AR(1) model with ARCH(1) error

$$y_t = \theta_0 y_{t-1} + \epsilon_t, \quad \sigma_t^2 = 1 - \phi_0 + \phi_0 \epsilon_{t-1}^2, \tag{14}$$

where the innovations $\{\epsilon_t = \epsilon_t/\sigma_t\}$ are i.i.d. with zero mean and unit variance. We consider both symmetric and skewed distributions for ϵ_t . In particular, we choose Gamma distribution with various shape parameters and Student's t distribution with various degrees of freedom to reflect the different degrees of skewness and kurtosis, respectively.

The asymptotic variances of the OLS and QMLE for various parameter values are given in Table 1, where the asymptotic variance of the true maximum likelihood estimator (MLE) is also given as a benchmark. The results show clearly that the efficiency gain of the OLS over the QMLE is significant in the case of highly skewed distributions such as Gamma, while this is less so in the case of symmetric distributions such as Student's t . Note that the QMLE and OLS of ϕ_0 have the same asymptotic variances under the Student's t distribution, which is consistent with the theoretical result (13).

Table 1. Asymptotic variances of OLS, QMLE, and MLE under AR(1)-ARCH(1) model.

		$v(\hat{\theta}_0)$	$v(\hat{\phi}_0)$	$v(\hat{\theta}_0)$	$v(\hat{\phi}_0)$	$v(\hat{\theta}_0)$	$v(\hat{\phi}_0)$	$v(\hat{\theta}_0)$	$v(\hat{\phi}_0)$
		$\theta_0 = 0.2$	$\theta_0 = 0.2$	$\theta_0 = 0.2$	$\theta_0 = 0.2$	$\theta_0 = 0.8$	$\theta_0 = 0.8$	$\theta_0 = 0.8$	$\theta_0 = 0.8$
		$\phi_0 = 0.2$	$\phi_0 = 0.6$						
Gamma (2)	OSLS	0.83	2.89	0.81	1.23	0.24	2.77	0.19	1.23
	QML	1.35	4.48	1.63	2.02	0.42	4.48	0.38	2.03
	ML	0.15	0.25	0.08	0.14	0.04	0.06	0.04	0.13
Gamma (8)	OSLS	0.97	2.08	0.88	0.94	0.31	2.06	0.22	0.94
	QML	1.18	2.52	1.09	1.15	0.38	2.51	0.28	1.14
	ML	0.87	1.44	0.69	0.64	0.27	1.35	0.19	0.62
Gamma (12)	OSLS	1.02	2.00	0.90	0.90	0.32	1.99	0.23	0.90
	QML	1.17	2.30	1.05	1.04	0.37	2.29	0.27	1.04
	ML	0.97	1.58	0.79	0.69	0.30	1.50	0.21	0.72
Gamma (20)	OSLS	1.06	1.93	0.91	0.88	0.34	1.93	0.24	0.88
	QML	1.15	2.11	1.00	0.96	0.37	2.11	0.26	0.96
	ML	1.03	1.67	0.86	0.73	0.33	1.62	0.23	0.75
t (5)	OSLS	1.34	6.32	1.51	2.86	0.37	6.26	0.29	2.86
	QML	1.56	6.32	2.34	2.86	0.41	6.26	0.41	2.86
	ML	1.05	2.44	1.04	1.11	0.29	2.41	0.21	1.11
t (7)	OSLS	1.26	3.30	1.26	1.51	0.37	3.32	0.29	1.51
	QML	1.30	3.30	1.41	1.51	0.38	3.32	0.32	1.51
	ML	1.10	2.31	1.05	1.05	0.32	2.34	0.25	1.05
t (13)	OSLS	1.20	2.31	1.08	1.06	0.37	2.33	0.26	1.06
	QML	1.20	2.31	1.10	1.06	0.37	2.33	0.26	1.06
	ML	1.15	2.12	1.03	0.97	0.36	2.17	0.24	0.97

In order to see how much loss of efficiency in the QMLE is recovered by the OLS estimator, next we calculate the relative reduction in the QMLE efficiency-loss (inefficiency)

$$RIEL(a'\gamma_0) = 100 \times \frac{\text{acov}\sqrt{T} a'(\hat{\gamma}_T^Q - \gamma_0) - \text{acov}\sqrt{T} a'(\hat{\gamma}_T^o - \gamma_0)}{\text{acov}\sqrt{T} a'(\hat{\gamma}_T^Q - \gamma_0) - \text{acov}\sqrt{T} a'(\hat{\gamma}_T^M - \gamma_0)},$$

where $\hat{\gamma}_T^M$ is the true MLE for γ_0 . This measure also indicates which estimator approaches the asymptotic variance lower bound faster.

Figure 1 contains a sample of the numerical outputs. First, Figure 1a shows that 45–60% of the inefficiency of $\hat{\theta}_T^Q$ is recovered by $\hat{\theta}_T^o$ when the innovations $\{\varepsilon_t\}$ have a heavy tail Student's t distribution. Further, $RIEL(\theta_0)$ declines sharply as the degrees of freedom increases, indicating that as the distribution of $\{\varepsilon_t\}$ gets close to the Gaussian, the QMLE improves quickly and it gets close to the OLS estimator, and both of them approach the variance lower bound. However, the situation in Figure 1b,c is opposite, where $RIEL(\theta_0)$ and $RIEL(\phi_0)$ are increasing with the shape parameter of the Gamma distribution. This indicates that the OLS improves significantly faster than the QMLE as the skewed distribution gets closer to the Gaussian. In other words, the efficiency loss of the QMLE is persistent and therefore it is not desirable under asymmetric conditional error distribution.

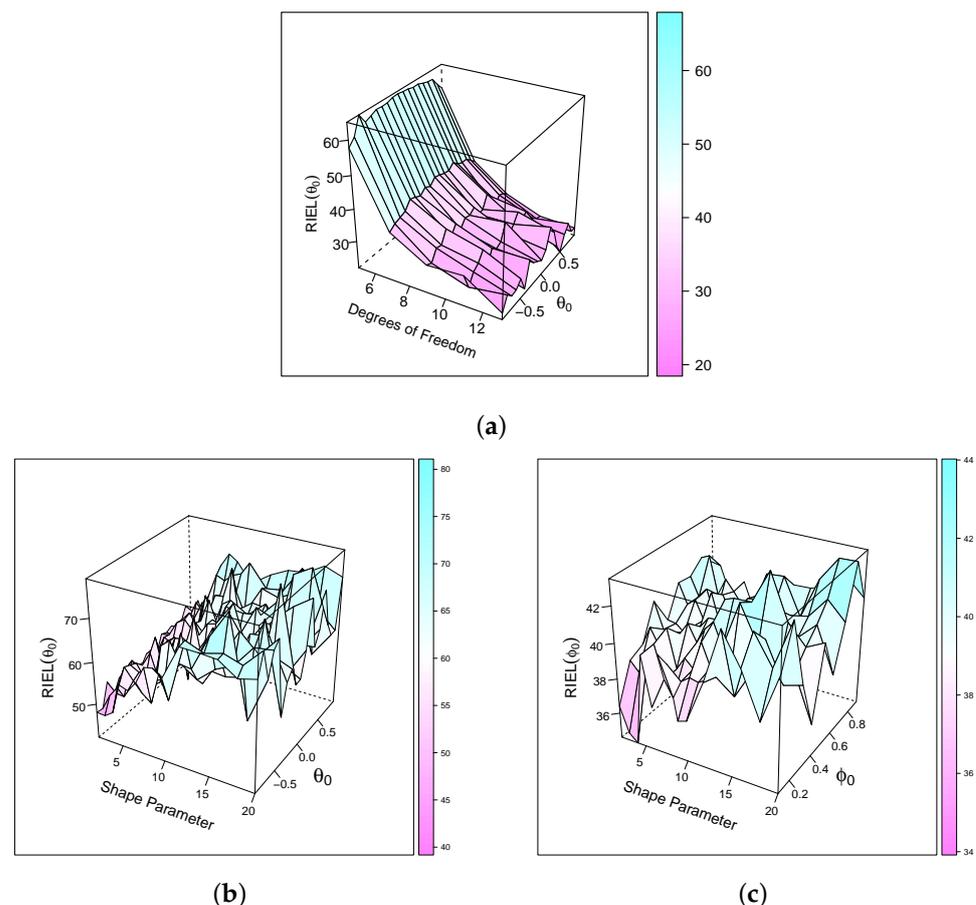


Figure 1. Reduction (%) in the QMLE efficiency-loss. (a) $\varepsilon_t \sim$ Student's t , $\phi_0 = 0.5$; (b) $\varepsilon_t \sim$ Gamma, $\phi_0 = 0.5$; (c) $\varepsilon_t \sim$ Gamma, $\theta_0 = 0.5$.

4. Simulation Studies

In this section we carry out Monte Carlo simulations to study the finite sample behaviour of the feasible optimal SLS estimator (FSLs) $\hat{\gamma}_T^o$ and compare it with some other related, commonly used estimators.

4.1. Comparison with Quasi-MLE

We first compare the FSLs with the quasi-maximum likelihood estimator (QMLE). Specifically, we generate the data from the AR(1)-ARCH(1) model in (14) with innovations $\{\varepsilon_t\}$ drawn from the standardized distributions of different levels of skewness and kurtosis. We consider various sample sizes including $T = 10,000$ to approximate the asymptotic results. In each simulation, we vary the values of the parameters (θ_0, ϕ_0) to represent different levels of persistence in the mean and variance components. For each estimator, we calcu-

late the mean estimates $(\hat{\theta}_0, \hat{\phi}_0)$ and the root mean squared errors $(RMSE(\hat{\theta}_0), RMSE(\hat{\phi}_0))$ based on 3000 independent replications over 4 different pairs of true values (θ_0, ϕ_0) .

Table 2 contains the summary results for Gamma (2,1) innovations, which show clearly that the FSLs outperforms the QMLE for all sample sizes and panels, while both estimators have the same degree of bias. We have also done the simulations with Student’s $t(5)$ innovations and the results show that the FSLs has moderate gain of efficiency over the QMLE, which performs fairly well in small samples.

Table 2. Simulation results for FSLs and QMLE under AR(1)-ARCH(1) model with $\varepsilon_t \sim \text{Gamma}(2,1)$.

T		$\hat{\theta}_0$	$RMSE(\hat{\theta}_0)$	$\hat{\phi}_0$	$RMSE(\hat{\phi}_0)$	$\hat{\theta}_0$	$RMSE(\hat{\theta}_0)$	$\hat{\phi}_0$	$RMSE(\hat{\phi}_0)$
		(a): $\theta_0 = \phi_0 = 0.2$				(b): $\theta_0 = 0.2, \phi_0 = 0.6$			
60	QMLE	0.21	0.149	0.30	0.213	0.19	0.152	0.58	0.163
	FSLs	0.21	0.119	0.27	0.172	0.19	0.121	0.59	0.136
100	QMLE	0.20	0.116	0.25	0.169	0.20	0.121	0.58	0.134
	FSLs	0.20	0.091	0.23	0.138	0.20	0.094	0.59	0.108
1000	QMLE	0.20	0.038	0.20	0.063	0.20	0.040	0.60	0.037
	FSLs	0.20	0.029	0.20	0.052	0.20	0.028	0.60	0.030
10,000	QMLE	0.20	0.012	0.20	0.020	0.20	0.013	0.60	0.012
	OSLS	0.20	0.009	0.20	0.016	0.20	0.009	0.60	0.010
		(c): $\theta_0 = 0.8, \phi_0 = 0.2$				(d): $\theta_0 = 0.8, \phi_0 = 0.6$			
60	QMLE	0.77	0.098	0.29	0.215	0.77	0.097	0.59	0.161
	FSLs	0.78	0.073	0.27	0.178	0.78	0.077	0.59	0.139
100	QMLE	0.78	0.074	0.25	0.172	0.78	0.071	0.58	0.128
	FSLs	0.79	0.054	0.23	0.140	0.79	0.054	0.59	0.106
1000	QMLE	0.80	0.021	0.20	0.062	0.80	0.021	0.60	0.038
	FSLs	0.80	0.016	0.19	0.050	0.80	0.015	0.60	0.030
10,000	QMLE	0.80	0.007	0.20	0.020	0.80	0.007	0.60	0.012
	OSLS	0.80	0.005	0.20	0.016	0.80	0.005	0.60	0.010

To understand how the values of (θ_0, ϕ_0) , shape parameter, and sample size T affect the relative RMSE of the FSLs compared to QMLE, we use the numerical results of the simulations to fit two regression equations with $RRMSE(\hat{\theta}_T^o) = RMSE(\hat{\theta}_T^o)/RMSE(\hat{\theta}_T^Q)$ and $RRMSE(\hat{\phi}_T^o)$ as the response variables, respectively. The results in Table 3 show that the shape parameter has a positive effect on both $RRMSE(\hat{\theta}_T^o)$ and $RRMSE(\hat{\phi}_T^o)$, while T is negatively associated with both RRMSEs, indicating that the outperformance of the FSLs in large samples is more evident than in small samples if the innovations distribution is skewed. Moreover, the negative sign of θ_0 in the $RRMSE(\hat{\theta}_T^o)$ equation indicates that the performance of the QMLE improves quickly as the value of θ_0 gets larger.

Table 3. Effect of the shape, sample size, and parameter values on the RRMSEs under Gamma distribution. All coefficients are significant at 0.0001 level.

	Coefficients					R^2	Error df
	Const	Shape	T	θ_0	ϕ_0		
$RRMSE(\hat{\theta}_T^o)$	0.76546	0.02887	−0.00007	−0.02783	0.02525	0.78692	4785
$RRMSE(\hat{\phi}_T^o)$	0.79952	0.02078	−0.00003	0.01013	0.00962	0.70135	4785

4.2. Comparison with Estimating Function Estimators

Our approach is related to the estimating function (EF) approach. Following Durairajan (1992), it can be shown that under models (1) and (2) the EF estimator $\hat{\gamma}_T^{EF}$ is obtained by solving the estimating equation

$$\sum_{t=1}^T \mathbf{B}'_t(\gamma) \Omega_t^{-1}(\gamma) \begin{pmatrix} \epsilon_t(\gamma) \\ \epsilon_t^2(\gamma) - \sigma_t^2(\gamma) \end{pmatrix} = \mathbf{0}, \quad (15)$$

where $\mathbf{B}'_t(\gamma)$ and $\Omega_t(\gamma)$ are given in (8) and (9) with θ_0, γ_0 and $E(\epsilon_t^j | \mathbf{v}_t, \dots, \mathbf{v}_{t-p}), j = 3, 4$ replaced by θ, γ and $E_\gamma(\epsilon_t^j(\gamma) | \mathbf{v}_t, \dots, \mathbf{v}_{t-p}), j = 3, 4$ respectively. The EF in (15) is optimal with respect to the so-called Godambe's information criterion. Moreover, under some regularity conditions similar to Assumptions A2–A9, the EF estimator can be shown to be \sqrt{T} -consistent with $\text{acov}\sqrt{T}(\hat{\gamma}_T^{EF} - \gamma_0)$ given by Equation (7). However, although the FLS and EF estimators have the same asymptotic variance, they are distinct in the following aspects. First, the FLS is an extremum estimator while the EF estimator represents a solution of the optimal estimating Equation (15). Second, if $E_\gamma(\epsilon_t^j(\gamma) | \mathbf{v}_t, \dots, \mathbf{v}_{t-p}), j = 3, 4$ are known functions of γ , then the EF estimator can be calculated in one step, while the FLS remains to be a two-step estimator due the dependence of \mathbf{W}_t on γ_0 . Third, the two estimators may behave differently in finite-sample situations because they have different estimating equations. This can be seen by comparing Equation (15) with the first order condition for the FLS which can be written as

$$\sum_{t=1}^T \mathbf{B}'_t(\gamma) \mathbf{H}_t(\theta) \Omega_t^{-1} \mathbf{H}'_t(\theta) \begin{pmatrix} \epsilon_t(\gamma) \\ \epsilon_t^2(\gamma) - \sigma_t^2(\gamma) \end{pmatrix} = \mathbf{0},$$

where

$$\mathbf{H}_t(\theta) = \begin{pmatrix} 1 & 2f_t(\mathbf{v}_t, \theta) - 2f_t(\mathbf{v}_t, \theta_0) \\ 0 & 1 \end{pmatrix}.$$

Next we calculate some numerical examples to compare the FLS with four different versions of the EF estimators that are commonly used in practice. First, since ϵ_t are i.i.d., Equation (15) can be written as

$$\sum_{t=1}^T \mathbf{B}'_t(\gamma_1) \Omega_t^{-1}(\gamma_2, \mu_3, \mu_4) \begin{pmatrix} \epsilon_t(\gamma) \\ \epsilon_t^2(\gamma) - \sigma_t^2(\gamma) \end{pmatrix} = \mathbf{0}. \quad (16)$$

Then we calculate four variants of the EF estimator as follows: the estimator EF0 is obtained by taking $\gamma_1 = \gamma_2 = \hat{\gamma}_T^Q, \mu_3 = 1/T \sum_{t=1}^T \epsilon_t^3(\hat{\gamma}_T^Q)$ and $\mu_4 = 1/T \sum_{t=1}^T \epsilon_t^4(\hat{\gamma}_T^Q)$; EF1 is the same as EF0 except that $\gamma_1 = \gamma$; EF is the same as EF0 except that $\gamma_1 = \gamma_2 = \gamma$; and EF2 is obtained by letting $\gamma_1 = \gamma_2 = \gamma, \mu_3 = 1/T \sum_{t=1}^T \epsilon_t^3(\gamma)$ and $\mu_4 = 1/T \sum_{t=1}^T \epsilon_t^4(\gamma)$. The four variants have the same asymptotic covariance matrix.

Figure 2 shows the relative RMSE of the FLS over the RMSE of the QMLE and EF estimators, respectively. Figure 2a is based on 4790 simulations with 3000 independent replications where $\{\epsilon_t\}$ are generated from the standardized Gamma distribution with shape parameters 2, 3, 4, 5, 6, and 7 respectively. Similarly, Figure 2b is based on 4050 simulations with $\{\epsilon_t\}$ generated from the standardized Student's t distribution with df 5, 6, 7, 8, and 9, respectively. In all cases, sample size T varies over a range (30, 40, 50, 60, 70, 80, 90, 100, 500, and 1000) and the RMSE of the estimators are calculated on the parameters grid $\{(0.1, 0.1), (0.1, 0.2), \dots, (0.1, 0.9), (0.2, 0.1), \dots, (0.9, 0.9)\}$.

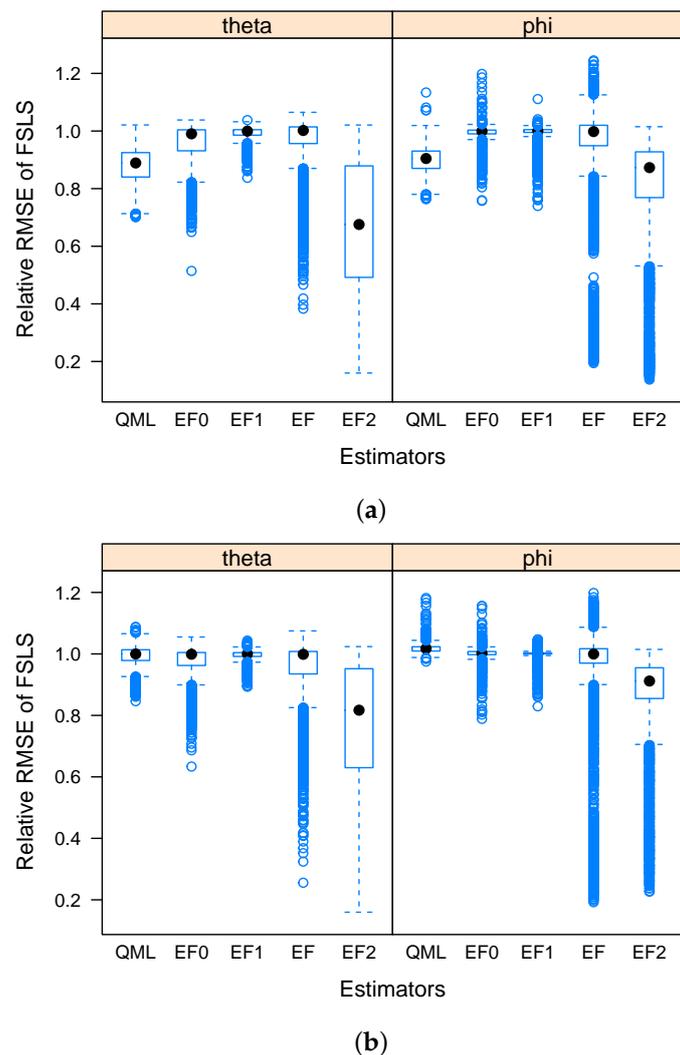


Figure 2. Ratio of the RMSE of FLS to the RMSE of QML and EF estimators respectively. (a) $\varepsilon_t \sim$ Gamma with shape parameter (2, 3, 4, 5, 6, 7). (b) $\varepsilon_t \sim$ Student's t with df (5, 6, 7, 8, 9).

The results show clearly that the FLS outperform the EF estimators for θ_0 in almost all cases and for ϕ_0 in the majority of the cases. In particular, results of EF2 show that replacing the nuisance parameters with highly nonlinear functions of the estimated parameter makes the performance worse. Therefore in practice, a two-step EF estimator such as the EF or EF1 should be recommended. Finally, the QMLE performs reasonably well in the case of symmetric error distributions, while not so well in the case of skewed error distributions.

5. Application

In this section we apply our method to an empirical example of Engle (1982) (see also Enders 2010), who used an AR model with ARCH error to study the wage/price spiral in the U.K. over the period 1958Q2–1977Q2. Specifically, let p_t denote the log of the consumer price index and w_t denote the log of the index of the nominal wage rates. Then $y_t = p_t - p_{t-1}$ and $r_t = w_t - p_t$ are the rate of inflation and real wage, respectively. Engle (1982) first fitted the following equation using the least squares (LS) method

$$y_t = \begin{matrix} 0.0257 & + & 0.334 & y_{t-1} & + & 0.408 & y_{t-4} & - & 0.404 & y_{t-5} & + & 0.0559 & r_{t-1} & + & \varepsilon_t, \\ (0.006) & & (0.103) & & (0.110) & & (0.114) & & (0.014) & & & & & & \end{matrix} \quad (17)$$

$$\hat{\sigma}_t^2 = 8.9 \times 10^{-5},$$

where the standard errors are in parentheses. Since the Lagrange multiplier (LM) test for ARCH(1) error was not significant, but for ARCH(4) process was instead, the following conditional variance equation was specified

$$\sigma_t^2 = \phi_0 + \phi_1(0.4\epsilon_{t-1}^2 + 0.3\epsilon_{t-2}^2 + 0.2\epsilon_{t-3}^2 + 0.1\epsilon_{t-4}^2), \quad (18)$$

where the two-parameter variance function with declining weights was chosen to satisfy the nonnegativity and stationarity constraints. Further, Engle (1982) fitted Equations (17) and (18) jointly using the Gaussian quasi-likelihood method, where all coefficients (except the first lag in the inflation rate) were significant at level 0.05.

Since we could not obtain the wage rates before 1963, we use the data from 1963Q1 through 1982Q1 to compensate for the 19 missing quarters. The data were obtained from the OECD website <http://dx.doi.org/10.1787/data-00052-en> (accessed on 15 January 2014, see OECD, Main Economic Indicators—complete database). To see if there is major structural difference in our data compared to the data used by Engle (1982), we have done an initial investigation and found no evidence of structural change by replacing the 19 quarters. So our results are to some extent comparable with those in Engle (1982).

We start by fitting the following regression model for inflation using the LS method

$$y_t = \theta_0 + \theta_1 y_{t-1} + \theta_2 y_{t-2} + \theta_3 y_{t-3} + \theta_4 y_{t-4} + \theta_5 y_{t-5} + \theta_6 r_{t-1} + \epsilon_t \quad (19)$$

The results are shown in the first part of Table 4 under Model-I (LS), where the White's correction for the standard errors are reported in parentheses.

We calculate the Ljung-Box statistic Q for ϵ_t (denoted by Q1) and $\hat{\epsilon}_t^2$ (denoted by Q2) at lags 5, 10, 15, and 20. They are all insignificant at level 0.1 except for Q2(5), which agrees with Engle (1982) in including four lags in the variance equation. Further, we use the squared residuals from this regression to fit an ARCH(4) model for the conditional variance

$$\sigma_t^2 = \phi_0 + \phi_1 \epsilon_{t-1}^2 + \phi_2 \epsilon_{t-2}^2 + \phi_3 \epsilon_{t-3}^2 + \phi_4 \epsilon_{t-4}^2 \quad (20)$$

The results are shown in the second part of Table 4 under Model-I (LS). Again, the ARCH(4) model is confirmed by the LM test at significance level 0.05. We report only the LS estimates of the variance function without standard errors because those estimates are only used as starting values to compute the QMLE. The Q1 and Q2 statistics from Model-I (LS) (in the third part of Table 4) indicate that the mean and variance equations are fairly well specified since none of these diagnostics are significant at level 0.1. Therefore we fit the Model-I again using the QMLE, which is more efficient than the LS procedure. Although the diagnostics of the standardized innovations from Model-I (QMLE) do not show serial correlation of the first or second order, all coefficients in the variance function are insignificant except for the constant term. This contradicts the ARCH(4) that we found before to be correctly specified. However, this can be explained by the lack of efficiency in the QMLE due to the moderate level of skewness in the corresponding residuals.

On the other hand, our FLS estimation yields a significant fourth lag in the variance function in addition to the correct specification as indicated by Q1 and Q2. Accordingly, we use the model fitted by FLS in stepwise regression algorithm to obtain a reduced model (Model-II in Table 4)

$$y_t = \theta_0 + \theta_1 y_{t-1} + \theta_4 y_{t-4} + \theta_5 y_{t-5} + \theta_6 r_{t-1} + \epsilon_t \quad (21)$$

$$\sigma_t^2 = \phi_0 + \phi_4 \epsilon_{t-4}^2 \quad (22)$$

Note that while the mean equation is identical to that in Engle (1982), only the fourth lag is significant in the variance equation. The above ARCH structure can only be detected by using the full model that is more flexible than the two-parameter variance function in Equation (18). Moreover, the more efficient FLS estimation yields the ARCH(4) structure, while the QMLE would conclude with a misspecified homoscedastic model.

Table 4. Fitted Model-I (19) and (20) and Model-II (21) and (22), where standard errors are in parentheses. The superscript ^a indicates statistical significance at 5% level. $Q1(n)$ ($Q2(n)$) is the Ljung-Box statistic for the (squared) standardized innovations and the corresponding p-values are in parentheses. JB is the standard Jarque-Bera test.

Coef.	Model-I			Model-II	
	LS	QMLE	FSLs	QMLE	FSLs
Conditional Mean Equation					
θ_0	0.071 ^a (0.016)	0.079 ^a (0.015)	0.049 ^a (0.014)	0.067 ^a (0.015)	0.063 ^a (0.013)
θ_1	0.417 ^a (0.141)	0.339 ^a (0.1)	0.280 ^a (0.093)	0.323 ^a (0.096)	0.257 ^a (0.086)
θ_2	−0.039 (0.095)	0.004 (0.086)	0.050 (0.081)	–	–
θ_3	−0.180 (0.148)	−0.235 ^a (0.101)	−0.158 (0.094)	–	–
θ_4	0.436 ^a (0.184)	0.481 ^a (0.108)	0.563 ^a (0.101)	0.328 ^a (0.116)	0.339 ^a (0.104)
θ_5	−0.350 ^a (0.101)	−0.310 ^a (0.094)	−0.294 ^a (0.088)	−0.246 ^a (0.094)	−0.234 ^a (0.084)
θ_6	0.076 ^a (0.018)	0.086 ^a (0.017)	0.051 ^a (0.016)	0.073 ^a (0.017)	0.067 ^a (0.015)
Conditional Variance Equation					
ϕ_0	0.0001	0.000 ^a (0.000)	0.000 (0.000)	0.000 ^a (0.000)	0.000 (0.000)
ϕ_1	0.1064	0.093 (0.134)	0.021 (0.126)	–	–
ϕ_2	0.0000	0.000 (0.077)	0.000 (0.072)	–	–
ϕ_3	0.0806	0.100 (0.146)	0.102 (0.137)	–	–
ϕ_4	0.3364	0.389 (0.252)	0.479 ^a (0.235)	0.553 (0.308)	0.556 ^a (0.281)
Diagnostic Statistics of the Standardized Innovations					
Q1(5)	0.9 (0.97)	1.0 (0.96)	2.1 (0.83)	2.0 (0.85)	1.9 (0.86)
Q1(10)	4.4 (0.93)	7.4 (0.68)	6.4 (0.78)	8.4 (0.59)	8.5 (0.58)
Q1(15)	12.8 (0.62)	16.1 (0.38)	14.1 (0.52)	19.6 (0.19)	18.7 (0.23)
Q2(5)	3.7 (0.59)	4.4 (0.50)	0.4 (0.99)	6.5 (0.26)	6.4 (0.27)
Q2(10)	5.7 (0.84)	6.9 (0.74)	1.3 (0.99)	8.2 (0.61)	8.1 (0.62)
Q2(15)	9.2 (0.87)	9.8 (0.83)	2.7 (0.99)	9.1 (0.87)	9.4 (0.85)
Skewness	0.78	0.61	1.67	0.73	0.99
Kurtosis	4.07	3.49	8.48	3.73	4.27
JB	11.0 ^a	5.35	115 ^a	8.12 ^a	16.93 ^a

6. Conclusions and Discussion

Although the ARCH-type models have been extensively studied for decades, most theories and methods are developed for stationary data processes whose conditional mean functions are either ARMA or simple linear functions of covariates. Moreover, recent research has been focusing on generalization of the error component while leaving the mean function to be simple linear form. However, many economic and financial time series are nonlinear and/or nonstationary in the mean, therefore, data transformation is required in order to apply standard methodologies in the analysis.

In this paper, we proposed the second-order least squares (SLS) approach to estimate a flexible and general model with nonlinear and time-varying conditional mean and ARCH conditional variance function. This approach is applicable to both stationary and mean-nonstationary processes and therefore can be used to analyze the transformed as well as the

original data. Another advantage of this approach is that it does not require specification of the underlying distribution for the errors. Moreover, the feasible optimal SLS estimator (FSLs) is more efficient than the commonly used QMLE, and the efficiency gain is significant when the conditional error distribution is asymmetric. We have demonstrated through a real data example that the efficiency gain of the proposed approach leads to a more accurate model than the QMLE. The third and fourth conditional moments of the innovation provide useful information that is utilized by the FSLs (through the weight matrix) to gain efficiency over the QMLE. This information is even more important in the case of skewed and/or leptokurtic error distributions. Our simulation studies also show that the SLS approach has better finite sample properties than the estimating function approach based on the same set of conditional moments.

There are some issues remaining to be studied in the future. Some assumptions for the asymptotic theories are sufficient but not necessary. They are adopted here mainly due to the way of the proofs we used. As is common in statistics and econometrics, there are usually different ways to prove an asymptotic theory, and each of them requires a specific set of assumptions. Therefore it is interesting to explore the possibilities of establishing the asymptotic properties for the FSLs estimator under the similar conditions for the QMLE. Indeed, our Monte Carlo simulation studies have shown that the FSLs estimator performs well in the finite sample situations even if the innovation does not have the finite moments of order higher than four. From the application point of view, it is also important to extend the method of this paper to models with more general GARCH errors. This is possible by modifying some assumptions to adapt for the mixing process.

Author Contributions: Conceptualization, M.S. and L.W.; methodology, M.S. and L.W.; software, M.S.; validation, M.S. and L.W.; formal analysis, M.S.; investigation, M.S.; resources, L.W.; data curation, M.S.; writing—original draft preparation, M.S.; writing—review and editing, L.W.; visualization, M.S.; supervision, L.W.; project administration, L.W.; funding acquisition, L.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Natural Sciences and Engineering Research Council of Canada (NSERC) grant number 546719.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The dataset in the Application section was accessible at OECD website <http://dx.doi.org/10.1787/data-00052-en> (accessed on 15 January 2014).

Acknowledgments: We are grateful to the Editor and the three anonymous referees for their comments and suggestions, which were helpful in improving the previous version of this paper. The research was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC). The first author also gratefully acknowledges financial support by the University of Manitoba Graduate Fellowship and Manitoba Graduate Scholarship.

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

Appendix A

In this Appendix we first list the technical assumptions that are sufficient for the asymptotic properties of the SLS estimator $\hat{\gamma}_T$, followed by the mathematical proofs for the theorems. To simplify the notation, we denote $f_t(\cdot, \theta) = f_t(v_t, \theta)$.

Appendix A.1. Regularity Conditions

We make the following assumptions for the consistency of the SLS estimator $\hat{\gamma}_T$.

Assumption A1. The process $\{(y_t, \mathbf{x}'_t)\}$ is strong mixing of size $-a$, for some $a > 1$. That is, there exists $\delta > 0$ such that $\alpha(m) = O(m^{-a-\delta})$, where

$$\alpha(m) = \sup_n \sup_{\{F \in \mathcal{F}_{-\infty}^n, G \in \mathcal{F}_{n+m}^\infty\}} |P(F \cap G) - P(F)P(G)|,$$

$$\mathcal{F}_{-\infty}^n = \mathcal{F}\{(y_t, \mathbf{x}'_t), t \leq n\} \text{ and } \mathcal{F}_{n+m}^\infty = \mathcal{F}\{(y_t, \mathbf{x}'_t), t \geq n+m\}.$$

This is a high-level assumption which allows for considerable dependence and heterogeneity in the underlying process. As noted by White and Domowitz (1984), it preserves the asymptotic independence of the observed process even under further transformations. This assumption can be justified on a case by case basis. For example, if $\sum_{i=1}^p \phi_{0i} < 1$, then $\{\epsilon_t\}$ is strong mixing with geometric rate if the innovation noise sequence $\{\epsilon_t\}$ is i.i.d., has finite second moment, and Lebesgue density being strictly positive in a neighbourhood of zero (Lindner 2009). The geometric memory decay implies that a can be set to an arbitrarily large number. It can also be shown that finite order Gaussian ARMA processes are strong mixing (Ibragimov and Linnik 1971, pp. 312–13).

Assumption A2. Let $\|\cdot\|$ be the Euclidean norm. Then for some $r > \frac{a}{a-1}$,

$$\sup_{t \in \mathbb{N}} E \left\{ \|\mathbf{W}_t\| \left(1 + \sum_{i=0}^p \epsilon_{t-i}^4 + \sup_{\Theta} f_{t-i}^4(\cdot, \theta) \right) \right\}^r < \infty.$$

Assumption A3. For any open neighbourhood $\mathcal{N} \subsetneq \Gamma$ of γ_0 , there exists $T_0(\mathcal{N})$ such that

$$\inf_{T \geq T_0} T^{-1} \sum_{t=1}^T \min_{\gamma \in \mathcal{N}^c \cap \Gamma} E \left\{ (\mathbf{h}_t(\gamma) - \mathbf{h}_t(\gamma_0))' \mathbf{W}_t (\mathbf{h}_t(\gamma) - \mathbf{h}_t(\gamma_0)) \right\} > 0.$$

Note that Assumption A2 ensures the uniform convergence of $Q_T(\gamma)$ and Assumption A3 is sufficient for parameter identification. If the process (y_t, \mathbf{x}'_t) is stationary, $f_t = f : \mathbb{R}^v \times \Theta \rightarrow \mathbb{R}^1$ and $\mathbf{W}_t = \mathbf{W}(v_t, \dots, v_{t-p})$ is positive definite a.s.-P, then Assumption A3 is equivalent to that $f(\cdot, \theta) = f(\cdot, \theta_0)$ a.s.-P only if $\theta = \theta_0$.

Further, we make the following additional assumptions for the asymptotic normality of $\hat{\gamma}_T$.

Assumption A4. The true value γ_0 is an interior point of Γ .

Assumption A5. The random functions $f_t(\cdot, \theta)$ are twice continuously differentiable on Γ uniformly in t a.s.-P.

Assumption A6. For some $r > \frac{a}{a-1}$, it holds

$$\sup_{t \in \mathbb{N}} E \left\{ \|\mathbf{W}_t\| \sup_{\Theta} \left(\left\| \nabla_{\theta}^2 f_t(\cdot, \theta) \right\|^2 + \sum_{i=0}^p \|\nabla_{\theta} f_{t-i}(\cdot, \theta)\|^4 + \sum_{i=1}^p \epsilon_{t-i}^2 \left\| \nabla_{\theta}^2 f_{t-i}(\cdot, \theta) \right\|^2 + \sum_{i=0}^p f_{t-i}^2(\cdot, \theta) \left\| \nabla_{\theta}^2 f_{t-i}(\cdot, \theta) \right\|^2 \right) \right\}^r < \infty.$$

Assumption A7. The sequence $\bar{A}_T(\gamma_0) = 2T^{-1} \sum_{t=1}^T E \{ \nabla_{\gamma} \mathbf{h}'_t(\gamma_0) \mathbf{W}_t \nabla_{\gamma} \mathbf{h}_t(\gamma_0) \}$ is bounded and $\liminf_{T \rightarrow \infty} |\bar{A}_T(\gamma_0)| > 0$.

Assumption A8. For some $r > \frac{a}{a-1}$,

$$\sup_{t \in \mathbb{N}} E \left\{ \|W_t\|^2 \left(1 + f_t^2(\cdot, \theta_0) \|\nabla_{\theta} f_t(\cdot, \theta_0)\|^2 + \sum_{i=1}^p \epsilon_{t-i}^2 \|\nabla_{\theta} f_{t-i}(\cdot, \theta_0)\|^2 + \|\nabla_{\theta} f_t(\cdot, \theta_0)\|^2 + \sum_{i=1}^p \epsilon_{t-i}^4 \right) \left(1 + \sum_{i=0}^p \epsilon_{t-i}^4 + \epsilon_t^2 f_t^2(\cdot, \theta_0) \right) \right\}^r < \infty.$$

Assumption A9. The sequence $V_T = 4T^{-1} \sum_{t=1}^T E \{ \nabla_{\gamma} h_t'(\gamma_0) W_t h_t(\gamma_0) h_t'(\gamma_0) W_t \nabla_{\gamma'} h_t(\gamma_0) \}$ is bounded and $\liminf_{T \rightarrow \infty} |V_T| > 0$.

Assumptions A2, A6, and A8 are for general cases and can be simplified for specific choice of W_t . For example, for the optimal weight $W_t = U_t^{-1}$ in Section 3, these assumptions can be simplified to the following assumptions, respectively.

Assumption A10 (Assumption A2). For $k = 0, 1, \dots, p$, and some $r > \frac{a}{a-1}$,

$$\sup_{t \in \mathbb{N}} E \left\{ \epsilon_t^4 + \sigma_t^{-4} \sup_{\Theta} f_{t-k}^4(\cdot, \theta) \right\}^r < \infty.$$

Assumption A11 (Assumption A6). For $s = 1, 2$ and $k = 0, 1, \dots, p$, and some $r > \frac{a}{a-1}$,

$$\sup_{t \in \mathbb{N}} E \left\{ \sigma_t^{-4} \sup_{\Theta} \|\nabla_{\theta}^s f_{t-k}(\cdot, \theta)\|^4 \right\}^r < \infty.$$

Assumption A12 (Assumption A8). For $k = 0, 1, \dots, p$, and some $r > \frac{a}{a-1}$,

$$\sup_{t \in \mathbb{N}} E \left\{ \epsilon_t^8 + \sigma_t^{-8} f_t^8(\cdot, \theta_0) + \sigma_t^{-8} \|\nabla_{\theta} f_{t-k}(\cdot, \theta_0)\|^8 \right\}^r < \infty.$$

Appendix A.2. Proof of Theorem 1

First, by using Hölder’s inequality and C_r inequality, we can easily verify that the sequence $\{h_t'(\gamma) W_t h_t(\gamma)\}$ is dominated by uniformly L_r -bounded variables (i.e., $\sup_t E |h_t'(\gamma) W_t h_t(\gamma)|^r < \infty$). Therefore, $\bar{Q}_T(\gamma) = T^{-1} \sum_{t=1}^T E \{h_t'(\gamma) W_t h_t(\gamma)\}$ is well defined and is continuous on Γ uniformly in T . Then by the uniform law of large numbers (ULLN) (White and Domowitz 1984, Theorem 2.3), we have $\sup_{\gamma \in \Gamma} |Q_T(\gamma) - \bar{Q}_T(\gamma)| \xrightarrow{a.s.} 0$ as $T \rightarrow \infty$. Further, since $\{h_t(\gamma_0), \mathcal{F}_t\}$ is a martingale difference sequence and W_t is measurable- \mathcal{F}_{t-1} , we have

$$E \{h_t'(\gamma) W_t h_t(\gamma)\} = E \left\{ (h_t(\gamma) - h_t(\gamma_0))' W_t (h_t(\gamma) - h_t(\gamma_0)) \right\} + E \{h_t'(\gamma_0) W_t h_t(\gamma_0)\}.$$

Since W_t is non-negative definite a.s.- P , Assumption A3 ensures the uniqueness of the minimum of $\bar{Q}_T(\gamma)$ for sufficiently large T . Thus the result follows from Theorem 3.4 of White (1996). □

Appendix A.3. Proof of Theorem 2

The proof consists of the following four steps.

(i) First we apply the mean value theorem for random functions to the first order condition for a minimum of $Q_T(\gamma)$. Since $\hat{\gamma}_T \xrightarrow{a.s.} \gamma_0$ and γ_0 is interior to Γ , there is a neighbourhood $\mathcal{N} \subset \Gamma$ of γ_0 such that $\hat{\gamma}_T \in \mathcal{N}$ a.s. for sufficiently large T . Further, since $f_t(\theta)$ is twice

continuously differentiable on Θ uniformly in t , by Jennrich (1969, Lemma 3), for sufficiently large T ,

$$\nabla_{\gamma}^2 Q_T(\tilde{\gamma}_T)(\hat{\gamma}_T(\omega) - \gamma_0) = -\nabla_{\gamma} Q_T(\gamma_0), \tag{A1}$$

where $\nabla_{\gamma}^2 Q_T(\gamma)$ is the Hessian matrix of $Q_T(\gamma)$ and $\|\tilde{\gamma}_T - \gamma_0\| \leq \|\hat{\gamma}_T - \gamma_0\|$.

(ii) Let $\bar{A}_T(\gamma) = 2T^{-1} \sum_{t=1}^T E\{\nabla_{\gamma} h'_t(\gamma) \mathbf{W}_t \nabla_{\gamma'} h_t(\gamma)\}$ and we show that

$$\left\| \nabla_{\gamma}^2 Q_T(\tilde{\gamma}_T) - \bar{A}_T(\gamma_0) \right\| \xrightarrow{a.s.} 0 \text{ as } T \rightarrow \infty. \tag{A2}$$

Using Hölder's, triangle and C_r inequalities we can verify that Assumptions A2 and A6 imply that $\|A_t(\gamma)\|$ are dominated by uniformly L_r -bounded functions, where

$$A_t(\gamma) = 2\nabla_{\gamma} h'_t(\gamma) \mathbf{W}_t \nabla_{\gamma'} h_t(\gamma) + 2\left(h'_t(\gamma) \mathbf{W}_t \otimes \mathbf{I}_{q+p+1}\right) \nabla_{\gamma'} \text{vec}(\nabla_{\gamma} h'_t(\gamma)).$$

Hence by the ULLN we have

$$\sup_{\gamma \in \Gamma} \left\| \nabla_{\gamma}^2 Q_T(\gamma) - T^{-1} \sum_{t=1}^T EA_t(\gamma) \right\| \xrightarrow{a.s.} 0 \text{ as } T \rightarrow \infty.$$

Moreover, by the triangle inequality we have

$$\begin{aligned} \left\| \nabla_{\gamma}^2 Q_T(\tilde{\gamma}_T) - T^{-1} \sum_{t=1}^T EA_t(\gamma_0) \right\| &\leq \sup_{\gamma \in \Gamma} \left\| \nabla_{\gamma}^2 Q_T(\gamma) - T^{-1} \sum_{t=1}^T EA_t(\gamma) \right\| \\ &\quad + \sup_{K \in \mathbb{N}} \left\| K^{-1} \sum_{t=1}^K EA_t(\tilde{\gamma}_T) - K^{-1} \sum_{t=1}^K EA_t(\gamma_0) \right\| \quad a.s. \end{aligned}$$

Since $EA_t(\gamma)$ is continuous on Γ uniformly in t and $EA_t(\gamma_0) = 2E\{\nabla_{\gamma} h'_t(\gamma_0) \mathbf{W}_t \nabla_{\gamma'} h_t(\gamma_0)\}$, equation (A2) follows by letting $T \rightarrow \infty$ in the last inequality. Further, since $\bar{A}_T(\gamma_0)$ and \mathbf{V}_T are uniformly nonsingular by Assumptions A7 and A9 respectively, for sufficiently large T , we have

$$\begin{aligned} \mathbf{V}_T^{-1/2} \bar{A}_T(\gamma_0) \sqrt{T}(\hat{\gamma}_T - \gamma_0) &= -\mathbf{V}_T^{-1/2} \sqrt{T} \nabla_{\gamma} Q_T(\gamma_0) \\ &\quad + \mathbf{V}_T^{-1/2} \bar{A}_T(\gamma_0) \left(\bar{A}_T^{-1}(\gamma_0) - \nabla_{\gamma}^2 Q_T^{-1}(\tilde{\gamma}_T) \right) \mathbf{V}_T^{1/2} \mathbf{V}_T^{-1/2} \sqrt{T} \nabla_{\gamma} Q_T(\gamma_0). \end{aligned} \tag{A3}$$

(iii) Now we use Cramér-Wold device (Rao 1973, p. 123) to show that

$$\mathbf{V}_T^{-1/2} \sqrt{T} \nabla_{\gamma} Q_T(\gamma_0) \xrightarrow{d} N(\mathbf{0}, \mathbf{I}_{q+p+1}) \text{ as } T \rightarrow \infty.$$

Let $\lambda \in \mathbb{R}^{q+p+1}$ with $\|\lambda\| = 1$. Then it is sufficient to show that $T^{-1/2} \sum_{t=1}^T \lambda' \mathbf{V}_T^{-1/2} \mathbf{S}_t(\gamma_0) \xrightarrow{d} N(0,1)$ as $T \rightarrow \infty$, where $\mathbf{S}_t(\gamma_0) = 2\nabla_{\gamma} h'_t(\gamma_0) \mathbf{W}_t h_t(\gamma_0)$. By Assumption A9 we have $\mathbf{V}_T^{-1/2} = O(1)$ and, therefore, by Assumption A8 and applying again the Hölder's and C_r inequality the double array $\{m_{Tt} = \lambda' \mathbf{V}_T^{-1/2} \mathbf{S}_t(\gamma_0)\}$ is uniformly L_r -bounded for all T sufficiently large. Further, since $\{h_t(\gamma_0), \mathcal{F}_t\}$ is a martingale difference, we have $E(m_{Tt}) = 0$ and $\text{var}\left(T^{-1/2} \sum_{t=1}^T m_{Tt}\right) = 1$ for all T sufficiently large. It follows from Theorem 14.1 of Davidson (1994) and Assumption A1 that $\{m_{Tt}\}$ to be strong mixing of size $-a$. Hence by Theorem 5.20 of White (2001) we have $T^{-1/2} \sum_{t=1}^T m_{Tt} \xrightarrow{d} N(0,1)$ as $T \rightarrow \infty$.

(iv) By (A2), Assumption A7 and Theorem 2.16 of White (2001), we have $\nabla_{\gamma}^2 Q_T^{-1}(\tilde{\gamma}_T) - \bar{A}_T^{-1}(\gamma_0) = o_p(1)$. Since $\mathbf{V}_T^{-1/2} \sqrt{T} \nabla_{\gamma} Q_T(\gamma_0) = O_p(1)$ from (iii), it follows that

$$\mathbf{V}_T^{-1/2} \bar{A}_T(\gamma_0) \left(\bar{A}_T^{-1}(\gamma_0) - \nabla_{\gamma}^2 Q_T^{-1}(\tilde{\gamma}_T) \right) \sqrt{T} \nabla_{\gamma} Q_T(\gamma_0) = o_p(1).$$

By the method of subsequences (Davidson 1994, Theorem 18.6), there exists a subsequence $\{T'\}$ such that

$$\mathbf{V}_{T'}^{-1/2} \bar{\mathbf{A}}_{T'}(\gamma_0) \left(\bar{\mathbf{A}}_{T'}^{-1}(\gamma_0) - \nabla_{\gamma}^2 \mathbf{Q}_{T'}^{-1}(\hat{\gamma}_{T'}) \right) \sqrt{T'} \nabla_{\gamma} \mathbf{Q}_{T'}(\gamma_0) \xrightarrow{a.s.} 0 \text{ as } T' \rightarrow \infty,$$

which implies

$$\mathbf{V}_{T'}^{-1/2} \sqrt{T'} (\bar{\mathbf{A}}_{T'}(\gamma_0) (\hat{\gamma}_{T'} - \gamma_0) + \nabla_{\gamma} \mathbf{Q}_{T'}(\gamma_0)) \xrightarrow{a.s.} 0 \text{ as } T' \rightarrow \infty.$$

Finally since $\{T'\}$ is arbitrary, we have

$$\mathbf{V}_T^{-1/2} \bar{\mathbf{A}}_T(\gamma_0) \sqrt{T} (\hat{\gamma}_T - \gamma_0) + \mathbf{V}_T^{-1/2} \sqrt{T} \nabla_{\gamma} \mathbf{Q}_T(\gamma_0) \xrightarrow{P} 0 \text{ as } T \rightarrow \infty,$$

and the proof is completed by applying the result (2c.4.12) of Rao (1973). \square

Appendix A.4. Proof of Theorem 3

Let $\mathbf{R}' = \frac{1}{\sqrt{T}}(\mathbf{R}'_1, \mathbf{R}'_2, \dots, \mathbf{R}'_T)$, $\mathbf{M}' = \frac{1}{\sqrt{T}}(\mathbf{M}'_1, \mathbf{M}'_2, \dots, \mathbf{M}'_T)$, $\mathbf{R}'_t = \nabla_{\gamma} \mathbf{h}'_t(\gamma_0) \mathbf{W}_t \mathbf{U}_t^{1/2}$ and $\mathbf{M}'_t = \nabla_{\gamma} \mathbf{h}'_t(\gamma_0) \mathbf{U}_t^{-1/2}$. Then the proof follows by noting that

$$E \left\{ \left(\mathbf{R} - \mathbf{M} E^{-1} \{ \mathbf{M}' \mathbf{M} \} E \{ \mathbf{M}' \mathbf{R} \} \right)' \left(\mathbf{R} - \mathbf{M} E^{-1} \{ \mathbf{M}' \mathbf{M} \} E \{ \mathbf{M}' \mathbf{R} \} \right) \right\} \quad (\text{A4})$$

is nonnegative definite matrix. Moreover the equality in (A4) holds if $\mathbf{W}_t = \mathbf{U}_t^{-1}$, $t = 1, 2, \dots, T$, which justifies Equation (6). The equivalence between Equations (6) and (7) follows from substituting $\mathbf{\Omega}_t^{-1}$ and \mathbf{B}_t in Equation (7). \square

References

- Abarin, Taraneh, and Liqun Wang. 2009. Second-order least squares estimation of censored regression models. *Journal of Statistical Planning and Inference* 139: 125–35. [\[CrossRef\]](#)
- Bollerslev, Tim, and Jeffrey M. Wooldridge. 1992. Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric Reviews* 11: 143–72. [\[CrossRef\]](#)
- Bose, Mausumi, and Rahul Mukerjee. 2015. Optimal design measures under asymmetric errors, with application to binary design points. *Journal of Statistical Planning and Inference* 159: 28–36. [\[CrossRef\]](#)
- Davidson, James. 1994. *Stochastic Limit Theory: An Introduction for Econometricians*. Oxford: Oxford University Press.
- Durairajan, T. M. 1992. Optimal estimating function for non-orthogonal model. *Journal of Statistical Planning and Inference* 33: 381–84. [\[CrossRef\]](#)
- Enders, Walter. 2010. *Applied Econometric Time Series*. New York: Wiley.
- Engle, Robert F. 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica* 50: 987–1007. [\[CrossRef\]](#)
- Engle, Robert F., and Gloria Gonzalez-Rivera. 1991. Semiparametric ARCH models. *Journal of Business & Economic Statistics* 9: 345–59.
- Franses, Philip Hans, and Dick Van Dijk. 2000. *Non-Linear Time Series Models in Empirical Finance*. Cambridge: Cambridge University Press.
- Gao, Lucy L., and Julie Zhou. 2017. D-optimal designs based on the second-order least squares estimator. *Statistical Papers* 58: 77–94. [\[CrossRef\]](#)
- He, Lei, and Rong-Xian Yue. 2019. R-optimality criterion for regression models with asymmetric errors. *Journal of Statistical Planning and Inference* 199: 318–26. [\[CrossRef\]](#)
- Ibragimov, I. A., and Yu V. Linnik. 1971. *Independent and Stationary Sequences of Random Variables*. Gröningen: Wolters-Noordhoff.
- Jennrich, Robert I. 1969. Asymptotic properties of non-linear least squares estimators. *The Annals of Mathematical Statistics* 40: 633–43. [\[CrossRef\]](#)
- Kim, Mijeong, and Yanyuan Ma. 2012. The efficiency of the second-order nonlinear least squares estimator and its extension. *Annals of the Institute of Statistical Mathematics* 64: 751–64. [\[CrossRef\]](#)
- Koul, Hira L., and Shiqing Ling. 2006. Fitting an error distribution in some heteroscedastic time series models. *The Annals of Statistics* 34: 994–1012. [\[CrossRef\]](#)
- Li, David X., and Harry J. Turtle. 2000. Semiparametric ARCH models: An estimating function approach. *Journal of Business & Economic Statistics* 18: 174–86.

- Li, Wai Keung, Shiqing Ling, and Michael McAleer. 2002. Recent theoretical results for time series models with GARCH errors. *Journal of Economic Surveys* 16: 245–69. [[CrossRef](#)]
- Lindner, Alexander M. 2009. *Stationarity, Mixing, Distributional Properties and Moments of GARCH (p, q)-Processes*. Handbook of Financial Time Series. Berlin: Springer, pp. 43–69.
- Ling, Shiqing. 2003. Adaptive estimators and tests of stationary and nonstationary short-and long-memory ARFIMA–GARCH models. *Journal of the American Statistical Association* 98: 955–67. [[CrossRef](#)]
- Ling, Shiqing, and Michael McAleer. 2003. On adaptive estimation in nonstationary ARMA models with GARCH errors. *The Annals of Statistics* 31: 642–74. [[CrossRef](#)]
- Meitz, Mika, and Pentti Saikkonen. 2008. Stability of nonlinear AR-GARCH models. *Journal of Time Series Analysis* 29: 453–75. [[CrossRef](#)]
- Rao, Calyampudi Radhakrishna. 1973. *Linear Statistical Inference and Its Applications*. New York: Wiley.
- Rosadi, D, and Peter Filzmoser. 2019. Robust second-order least-squares estimation for regression models with autoregressive errors. *Statistical Papers* 60: 105–22. [[CrossRef](#)]
- Rosadi, Dedi, and Shelton Peiris. 2014. Second-order least-squares estimation for regression models with autocorrelated errors. *Computational Statistics* 29: 931–43. [[CrossRef](#)]
- Salamh, Mustafa, and Liqun Wang. 2021. Second-order least squares method for dynamic panel data models with application. *Journal of Risk and Financial Management* 14: 410. [[CrossRef](#)]
- Wang, Liqun. 2003. Estimation of nonlinear Berkson-type measurement error models. *Statistica Sinica* 13: 1201–10.
- Wang, Liqun. 2004. Estimation of nonlinear models with Berkson measurement errors. *The Annals of Statistics* 32: 2559–79. [[CrossRef](#)]
- Wang, Liqun. 2007. A unified approach to estimation of nonlinear mixed effects and Berkson measurement error models. *Canadian Journal of Statistics* 35: 233–48. [[CrossRef](#)]
- Wang, Liqun, and Alexandre Leblanc. 2008. Second-order nonlinear least squares estimation. *Annals of the Institute of Statistical Mathematics* 60: 883–900. [[CrossRef](#)]
- Weiss, Andrew A. 1986. Asymptotic theory for ARCH models: estimation and testing. *Econometric Theory* 2: 107–31. [[CrossRef](#)]
- White, Halbert. 1996. *Estimation, Inference and Specification Analysis*. Cambridge, UK: Cambridge University Press.
- White, Halbert. 2001. *Asymptotic Theory for Econometricians*. New York: Academic Press.
- White, Halbert, and Ian Domowitz. 1984. Nonlinear regression with dependent observations. *Econometrica* 52: 143–61. [[CrossRef](#)]
- Yin, Yue, and Julie Zhou. 2017. Optimal designs for regression models using the second-order least squares estimator. *Statistica Sinica* 27: 1841–56. [[CrossRef](#)]
- Zhu, Ke, and Wai Keung Li. 2015. A new Pearson-type QMLE for conditionally heteroskedastic models. *Journal of Business and Economic Statistics* 33: 552–65. [[CrossRef](#)]