TI 2021-098/III
Tinbergen Institute Discussion Paper

Conditional Score Residuals and Diagnostic Analysis of Serial Dependence in Time Series Models

F. Blasques[1]
P. Gorgi[1]
S.J. Koopman[2]

[1] Vrije Universiteit Amsterdam, Tinbergen Institute
[2] Vrije Universiteit Amsterdam, Tinbergen Institute, CREATES Aarhus University

Conditional score residuals and diagnostic analysis of serial dependence in time series models*

F. Blasques[a,b], P. Gorgi[a,b], and S. J. Koopman[a,b,c]

[a]*Vrije Universiteit Amsterdam, The Netherlands*

[b]*Tinbergen Institute, The Netherlands*

[c]*CREATES, Aarhus University, Denmark*

November 16, 2021

Abstract

We introduce conditional score residuals and provide a general framework for the diagnostic analysis of time series models. A key feature of conditional score residuals is that they account for the shape of the conditional distribution. These residuals offer reliable and powerful diagnostic tools for testing residual autocorrelation. Furthermore, they can be employed in models of which it is not clear how to define residuals. The asymptotic properties of the empirical autocorrelation function for conditional score residuals are formally derived. The results yield a unified theory for the diagnostic analysis of a wide class of time series models. The practical relevance of the proposed framework is illustrated for heavy-tailed GARCH models. Monte Carlo and empirical results support the finding that conditional score residuals are more reliable in testing residual autocorrelation, when compared to squared GARCH residuals. We finally show how a diagnostic analysis can be designed for dynamic copula models.

Some keywords: conditional score residuals, diagnostic analysis, residual autocorrelation, time series models

JEL codes: C12, C22, C58

*Corresponding author: S.J. Koopman, Department of Econometrics, School of Business and Economics, Vrije Universiteit Amsterdam, De Boelelaan 1105, 1081HV Amsterdam, Netherlands. Email: s.j.koopman@vu.nl.
F. Blasques is thankful to the Dutch Research Council (NWO) grant VIDI.195.099 for financial support. S. J. Koopman acknowledges support from CREATES, Aarhus University, Denmark, funded by Danish National Research Foundation, (DNRF78).

1 Introduction

Model specification, estimation and diagnostic analysis are three of the main building blocks of statistical modeling. The latter plays a key role in evaluating the appropriateness of the model specification. A standard approach to detect misspecification is to analyze the residuals of the model. In time series analysis, residuals are widely used to assess the dynamic specification of a model by means of the empirical autocorrelation function. After parameter estimation, patterns of serial dependence that are not properly explained by the model, such as neglected lags and seasonal components, are typically detected by residual autocorrelations. Early developments on residuals analysis of regression and time series models date back to the seminal work of Durbin and Watson (1950, 1951), Durbin (1970), Box and Pierce (1970) and Ljung and Box (1978). There is much literature on testing residual autocorrelation, providing diagnostic tools for different econometric models. The definition of residuals varies with the class of time series models. In case of autoregressive moving average (ARMA) models, the residuals are defined as the predictive estimates of the error terms. The properties of their empirical autocorrelations have been extensively studied (McLeod, 1978; Francq et al., 2005). In case of generalized autoregressive conditional heteroskedasticity (GARCH) models, residual analysis is based on the so-called squared residuals, which are a quadratic transformation of the estimate of the standardized error term (Li and Mak, 1994; Horváth et al., 2001; Berkes et al., 2003). More generally, in models with non-continuous response variables, diagnostic analysis typically relies on Pearson residuals. For example, Pearson residuals are a standard diagnostic tool for integer-valued time series models, such as Poisson autoregressive models (Jung et al., 2006; Davis et al., 2016).

In this study, we introduce a general framework for diagnostic analysis of time series models based on conditional score residuals. We define conditional score residuals through the score of the conditional probability distribution of the time series process, re-scaled by the square root of the conditional Fisher information. We illustrate how conditional score residuals encompass standard definitions of residuals. It follows that ARMA residuals, squared residuals and Pearson residuals are special cases of conditional score residuals when the conditional distribution of the model belongs to the exponential family. More generally, conditional score residuals provide an alternative definition of residuals when the distribution does not belong to the exponential family. The key advantage of conditional score residuals is that they account for the form of the conditional distribution. For instance, conditional score residuals are robust to extreme observations in heavy-tailed models and they produce more reliable and powerful diagnostic tests. Moreover, conditional score residuals can be employed to test residual autocorrelation for classes of models where otherwise it may not even be clear how the residuals should be defined.

We formally derive the asymptotic properties of empirical autocorrelations of conditional score residuals for a general class of parametric time series models, which includes ARMA, GARCH and integer-valued auto-regressive models as special cases. We propose a consistent estimator of the asymptotic covariance matrix that provides positive definite estimates in small samples under mild conditions. The results deliver a unified theory for diagnostic analysis of observation-driven time series models. We consider two examples as illustrations to validate the finite sample properties of the proposed diagnostic tools and to show the practical relevance of the proposed framework. The first study features GARCH models with Student's t errors. It shows the robustness of conditional score residuals against extreme observations on the basis of a Monte Carlo experiment and an empirical illustration to the modeling of daily returns of the S&P500 index. The overall conclusion is that our framework provides more powerful and reliable diagnostic tests when compared to those based on squared residuals, which are highly unreliable in the presence of extreme observations. The second study considers a bivariate copula model with a dynamic correlation coefficient. It shows the flexibility of the proposed approach and the reliability of the asymptotic results in finite samples on the basis of a Monte Carlo experiment and an empirical application to the modeling of the dependence between oil prices and exchange rates.

The diagnostic analysis of conditional score residuals differs from the specification analysis based on the score of the log-likelihood function. In the latter approach, specification tests rely on the gradient of the log-likelihood function of the model; see White (1994) and Ling and Tong (2011). By contrast, conditional score residuals are defined as the score of the conditional probability density function and they provide a general definition of residuals that includes existing types of residuals as special cases. In the time series literature, conditional scores have been employed for purposes other than diagnostic analysis. Creal et al. (2013) and Harvey (2013) introduce the class of score-driven models where the dynamics of the process is driven by conditional score functions. Harvey and

Thiele (2016) and Calvori et al. (2017) propose the use of score-driven models to test for the presence of time-varying parameters using Lagrange multiplier tests. In the current study, we consider conditional scores only to provide a definition of model-based residuals and we develop a general theoretical framework for testing residual autocorrelation.

The remainder of this paper is structured as follows. Section 2 introduces conditional score residuals and discusses their relationship with Pearson residuals. Section 3 studies the asymptotic properties of empirical autocorrelation vectors of conditional score residuals and derives diagnostic tests for residual autocorrelation. Section 4 presents a Monte Carlo study and an empirical application for the GARCH model with Student's t errors. Section 5 provides a Monte Carlo study and an empirical application for the dynamic bivariate Gaussian copula model. Section 6 concludes.

2 Conditional score residuals

2.1 Definition of conditional score residuals

Let $\{y_t\}_{t\in\mathbb{Z}}$ be a time series process with elements taking values in a sample space Y. Assume that the process has the following conditional distribution

$$y_t|\mathcal{F}_{t-1} \sim p(y_t|f_t;\boldsymbol{\lambda}), \tag{1}$$

where $\mathcal{F}_t$ is the sigma field generated by $\{y_t, y_{t-1}, y_{t-2}, \dots\}$, $p(\cdot|f_t;\boldsymbol{\lambda})$ is a conditional probability density function, $\boldsymbol{\lambda} \in \Lambda \subseteq \mathbb{R}^r$ is a parameter vector, and f_t is a scalar time-varying parameter that takes values in $\mathrm{F} \subseteq \mathbb{R}$. The time-varying parameter f_t is specified through the following observation-driven stochastic recurrence equation

$$f_t = g_{\boldsymbol{\theta}}(f_{t-1}, \dots, f_{t-q}, y_{t-1}, \dots, y_{t-p}), \quad t \in \mathbb{Z}, \tag{2}$$

where $g_{\boldsymbol{\theta}}$ is a parametric updating function that maps $\mathrm{F}^q \times \mathrm{Y}^p$ into F. The updating function $g_{\boldsymbol{\theta}}$ is indexed by the parameter vector $\boldsymbol{\theta} = (\boldsymbol{\xi}^\top, \boldsymbol{\lambda}^\top)^\top$, where $\boldsymbol{\xi} \in \Xi \subseteq \mathbb{R}^s$ is a vector of parameters that are specific to the updating function $g_{\boldsymbol{\theta}}$. The size of the parameter vector of the model $\boldsymbol{\theta} \in \Theta = \Xi \times \Lambda$ is $n = s + r$. We note that this general formulation allows the updating function $g_{\boldsymbol{\theta}}$ to depend on the parameter vector of the conditional density $\boldsymbol{\lambda}$. Models where the updating function does not depend on $\boldsymbol{\lambda}$ are special cases where the function $\boldsymbol{\lambda}$ is constant with respect to $\boldsymbol{\lambda}$. The class of parametric observation-driven models specified in

(1) and (2) is very general and it covers a wide range of time series models. For instance, it includes linear and nonlinear ARMA models (Box and Pierce, 1970), GARCH models (Engle, 1982; Bollerslev, 1986), score-driven models (Creal et al., 2013; Harvey, 2013), autoregressive conditional duration models (Engle and Russell, 1998), autoregressive Poisson models (Davis et al., 2003; Fokianos et al., 2009), and autoregressive conditional copula models (Patton, 2006).

Before introducing conditional score residuals, we define the conditional score errors $\{s_t\}_{t\in\mathbb{Z}}$ of the observation-driven model in (1) and (2) as

$$s_t = \frac{u_t}{\sqrt{I_t}}, \tag{3}$$

with

$$u_t = \frac{\partial \log p(y_t|f;\boldsymbol{\lambda})}{\partial f}\Big|_{f=f_t}, \quad \text{and} \quad I_t = \mathbb{E}\left(u_t^2|\mathcal{F}_{t-1}\right) = \mathcal{I}(f_t,\boldsymbol{\lambda}),$$

where $\mathcal{I}(\cdot,\cdot)$ is a known function. The error u_t is the first derivative of the conditional log-density with respect to the time varying parameter f_t and I_t is the conditional Fisher information, which can be expressed as a function of f_t and the parameter vector $\boldsymbol{\lambda}$, i.e. $I_t = \mathcal{I}(f_t,\boldsymbol{\lambda})$. Under standard regularity conditions on the conditional density function, the sequence of conditional score errors $\{s_t\}_{t\in\mathbb{Z}}$ is a martingale difference sequence, $\mathbb{E}(s_t|\mathcal{F}_{t-1}) = 0$, with conditional variance equal to one, $\mathbb{V}ar(s_t|\mathcal{F}_{t-1}) = 1$. In practice, the parameter vector of the model $\boldsymbol{\theta}$ is unknown and it needs to be estimated from the observed data. Therefore, the conditional score errors defined in (3) are not observable since they depend on $\boldsymbol{\theta}$.

We define conditional score residuals as plug-in estimates of the conditional score errors. More specifically, assume we observe a realized path of size T, $\{y_t\}_{t=1}^{T}$, from the time series process in (1) and (2) with true parameter value $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. Based on the observed data, the parameter vector $\boldsymbol{\theta}$ is estimated by the method of Maximum Likelihood (ML), which is a standard approach for the estimation of parametric time series models. The ML estimator $\hat{\boldsymbol{\theta}}_T = (\hat{\boldsymbol{\xi}}_T^\top, \hat{\boldsymbol{\lambda}}_T^\top)^\top$ is defined as the maximizer of the log-likelihood function

$$\hat{\boldsymbol{\theta}}_T = \arg\max_{\boldsymbol{\theta}\in\Theta} \sum_{t=1}^{T} l_t(\boldsymbol{\theta}), \qquad l_t(\boldsymbol{\theta}) = \log p(y_t|f_t(\boldsymbol{\theta});\boldsymbol{\lambda}),$$

where $f_t(\boldsymbol{\theta})$ is the filtered time-varying parameter obtained form equation (2) using the observed data. From the ML estimate $\hat{\boldsymbol{\theta}}_T$, the plug-in estimate of the time-varying parameter

$\hat{f}_t = f_t(\hat{\boldsymbol{\theta}}_T)$ is obtained. Conditional score residuals are defined as

$$\hat{s}_t = \frac{\hat{u}_t}{\sqrt{\hat{I}_t}}, \tag{4}$$

where

$$\hat{u}_t = \frac{\partial \log p(y_t|f, \hat{\boldsymbol{\lambda}}_T)}{\partial f}\Big|_{f=\hat{f}_t}, \quad \text{and} \quad \hat{I}_t = \mathcal{I}(\hat{f}_t, \hat{\boldsymbol{\lambda}}_T).$$

Throughout the paper, for ease of notation, we consider the convention that any function of $\boldsymbol{\theta}$ is evaluated at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ when there is no explicit dependence on $\boldsymbol{\theta}$ and it is evaluated at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}_T$ when the hat symbol is present on the function. For example, $s_t(\boldsymbol{\theta})$ denotes the score evaluated at a general $\boldsymbol{\theta} \in \Theta$, s_t denotes the score evaluated at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, i.e. $s_t = s_t(\boldsymbol{\theta}_0)$, and $\hat{s}_t$ denotes the score evaluated at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}_T$, i.e. $\hat{s}_t = s_t(\hat{\boldsymbol{\theta}}_T)$. Furthermore, for simplicity of exposition, we shall sometimes refer to conditional score errors as conditional score residuals. A clear distinction between errors and residuals shall be made when needed.

2.2 Relationship with Pearson residuals and examples

In this section, we illustrate how conditional score residuals encompass standard definitions of residuals that are typically employed in time series analysis. More specifically, conditional score residuals are equivalent to Pearson residuals when the conditional density belongs to the exponential family of distributions and they provide an alternative definition of residuals when the distribution is not a member of the exponential family.

Assume that the conditional density function in (1) belongs to the one-parameter exponential family of distributions with respect to the time-varying parameter f_t

$$p(y_t|f_t; \boldsymbol{\lambda}) = \exp\{\eta(f_t)T(y_t) - A(\eta(f_t))\}h(y_t), \tag{5}$$

where $\eta(\cdot)$, $T(\cdot)$, $A(\cdot)$ and $h(\cdot)$ are known functions, which may depend on the parameter vector $\boldsymbol{\lambda}$. The parameter $\eta(f_t)$ is often referred to as the canonical parameter of the exponential family. The conditional expectation of $T(y_t)$ is $\mathbb{E}(T(y_t)|\mathcal{F}_{t-1}) = A'(\eta(f_t))$ and the conditional variance is $\mathbb{V}ar(T(y_t)|\mathcal{F}_{t-1}) = A''(\eta(f_t))$. The definition of conditional score residuals entails that

$$s_t = \frac{T(y_t) - A'(\eta(f_t))}{\sqrt{A''(\eta(f_t))}}, \tag{6}$$

since

$$u_t = \eta'(f_t)\{T(y_t) - A'(\eta(f_t))\}, \quad \text{and} \quad I_t = \eta'(f_t)^2 A''(\eta(f_t)).$$

Therefore, from equation (6), we can see that conditional score residuals coincide with Pearson residuals when $T(\cdot)$ is the identity and the conditional density function belongs to the exponential family in (5). Furthermore, when $T(\cdot)$ is not the identity, conditional score residuals can be interpreted as Pearson residuals with respect to transformed variable $T(y_t)$ since s_t is simply obtained by standardizing $T(y_t)$ with its conditional mean and variance. For instance, as we shall see below, squared residuals of GARCH models may be interpreted as Pearson residuals based on the transformed variable y_t^2. Some examples for specific classes of models are provided below.

Examples 2.1-2.3 feature three classes of time series models where the conditional distribution is a member of the exponential family and therefore conditional score residuals are equivalent to Pearson residuals. In particular, Example 2.1 considers ARMA models with Gaussian errors, Example 2.2 GARCH models with Gaussian errors, and Example 2.3 Poisson autoregressive models, which are also known as Poisson INGARCH models.

Example 2.1 (Gaussian ARMA models)**.** *Consider Gaussian ARMA-type processes, specified as*

$$y_t = \mu_t + \sigma\varepsilon_t, \qquad \varepsilon_t \sim N(0,1), \qquad t \in \mathbb{Z},$$

where $\mu_t = g_{\boldsymbol{\theta}}(\mu_{t-1}, \ldots, \mu_{t-q}, y_{t-1}, \ldots, y_{t-p})$ *and* $\sigma > 0$*. Note that the linear ARMA model is a special case of this model formulation when* $g_{\boldsymbol{\theta}}$ *is a linear function. Conditional score residuals are given by*

$$s_t = \frac{y_t - \mu_t}{\sigma}.$$

These residuals are equivalent to standard ARMA residuals, which are also equal to Pearson residuals.

Example 2.2 (Gaussian GARCH models)**.** *Consider Gaussian GARCH-type models, specified as*

$$y_t = \sqrt{h_t}\varepsilon_t, \qquad \varepsilon_t \sim N(0,1), \qquad t \in \mathbb{Z},$$

where $h_t = g_{\boldsymbol{\theta}}(h_{t-1}, \ldots, h_{t-q}, y_{t-1}, \ldots, y_{t-p})$*. Note that the standard GARCH model is a special case when* $g_{\boldsymbol{\theta}}$ *is a linear function of lagged* h_t *and lagged* y_t^2*. Conditional score residuals are given by*

$$s_t = \frac{1}{\sqrt{2}}\left(\frac{y_t^2}{h_t} - 1\right).$$

These residuals coincide with squared residuals from GARCH models, which can be interpreted as Pearson residuals with respect to the statistic y_t^2*.*

Example 2.3 (Poisson INGARCH models)**.** *Consider Poisson autoregressive models, specified as*

$$y_t|\mathcal{F}_{t-1} \sim \mathcal{P}(\mu_t), \qquad t \in \mathbb{Z},$$

where $\mu_t = g_{\boldsymbol{\theta}}(\mu_{t-1}, \ldots, \mu_{t-q}, y_{t-1}, \ldots, y_{t-p})$ *and* $\mathcal{P}(\mu_t)$ *denotes a Poisson with mean* μ_t*. Conditional score residuals are*

$$s_t = \frac{y_t}{\sqrt{\mu_t}} - \sqrt{\mu_t}.$$

These residuals coincide with standard Pearson residuals for Poisson models.

Examples 2.4 and 2.5 illustrate how conditional score residuals differ from ARMA residuals and squared residuals of GARCH models when the conditional distribution of the error is not normal. Example 2.4 considers the ARMA model with the Student's t error distribution. The resulting conditional score residuals are robust to extreme observations. Example 2.5 features GARCH models with a Student's t distribution of the error term. Also in this case, conditional score residuals are robust to extreme observations. Example 2.5 shall be discussed in detail in Section 4 through a simulation study and an empirical application. As we shall see, conditional score residuals deliver more powerful tests for residual autocorrelation and they allows us to conduct inference even when the error term has only two finite moments. Instead, squared residuals require four finite moments and they are unreliable in the presence of extreme observations.

Example 2.4 (Student's t-ARMA models)**.** *Consider ARMA models with Student's t errors, specified as*

$$y_t = \mu_t + \sigma\varepsilon_t, \qquad \varepsilon_t \sim t_v(0,1), \qquad t \in \mathbb{Z},$$

where $\mu_t = g_{\boldsymbol{\theta}}(\mu_{t-1}, \ldots, \mu_{t-q}, y_{t-1}, \ldots, y_{t-p})$ *and* $t_v(0,1)$ *is the Student's t distribution with mean* 0*, variance* 1 *and degrees of freedom* $v > 2$*. Conditional score residuals are given by*

$$s_t = \sqrt{\frac{(v+3)(v-2)}{v(v+1)}} \left(\frac{\sigma(v+1)(y_t - \mu_t)}{(v-2)\sigma^2 + (y_t - \mu_t)^2} \right).$$

Example 2.5 (Student's t-GARCH models)**.** *Consider GARCH models with Student's t errors, specified as*

$$y_t = \sqrt{h_t}\varepsilon_t, \qquad \varepsilon_t \sim t_v(0,1), \qquad t \in \mathbb{Z},$$

where $h_t = g_{\boldsymbol{\theta}}(h_{t-1}, \ldots, h_{t-q}, y_{t-1}, \ldots, y_{t-p})$*. Conditional score residuals are given by*

$$s_t = \sqrt{\frac{v+3}{2v}} \left(\frac{(v+1)y_t^2}{(v-2)h_t + y_t^2} - 1 \right).$$

Finally, Examples 2.6 and 2.7 illustrate how conditional score residuals are useful to define residuals in general classes of observation-driven time series models. Example 2.6 features a bivariate Gaussian copula model with dynamic correlation coefficient. This example shall be discussed in more detail in Section 5. Example 2.7 features a dynamic ordered probit model where the observable variable is not numerical but only ordinal. An example of ordinal time series where dynamic ordered probit/logit models are used is given by credit ratings (Creal et al., 2014). Given the non-numerical nature of the observable variable, it may not be clear how to define Pearson-type residuals in this setting. Instead, conditional score residuals provide a natural way of defining the model's residuals and inference on residual autocorrelation can be carried out within the general theoretical framework that shall be discussed in Section 3.

Example 2.6 (Dynamic copula models)**.** *Consider Gaussian copula models for a bivariate vector* $y_t = (y_{1t}, y_{2t})^\top$*, specified as*

$$y_t | \mathcal{F}_{t-1} \sim \mathcal{C}_{\mathcal{G}}(p_t), \qquad t \in \mathbb{Z},$$

where $\mathcal{C}_{\mathcal{G}}$ *is a bivariate Gaussian copula with time-varying correlation parameter* $p_t = g_{\boldsymbol{\theta}}(p_{t-1}, \ldots, p_{t-q}, y_{t-1}, \ldots, y_{t-p})$*. Conditional score residuals are given by*

$$s_t = \frac{p_t + x_{1t}x_{2t} - p_t(x_{1t}^2 + x_{2t}^2) + p_t^2(x_{1t}x_{2t} - p_t)}{\sqrt{1 + p_t^2}(1 - p_t^2)},$$

where $x_{1t} = \Phi^{-1}(y_{1t})$ *and* $x_{2t} = \Phi^{-1}(y_{2t})$*, and* $\Phi(\cdot)$ *denotes the cumulative distribution function of the standard normal distribution.*

Example 2.7 (Dynamic ordered probit)**.** *Consider dynamic ordered probit models, specified as*

$$y_t = \begin{cases} 0 & \text{if } -\infty < y_t^* \leq \lambda_1, \\ 1 & \text{if } \lambda_1 < y_t^* \leq \lambda_2, \\ \vdots & \\ m & \text{if } \lambda_m < y_t^* < \infty, \end{cases} \qquad y_t^* = \mu_t + \varepsilon_t, \qquad \varepsilon_t \sim N(0,1), \qquad t \in \mathbb{Z},$$

where y_t takes values in the set $\{0, 1, \dots, m\}$ and $\mu_t = g_{\boldsymbol{\theta}}(\mu_{t-1}, \dots, \mu_{t-q}, y_{t-1}, \dots, y_{t-p})$. Conditional score residuals are given by

$$s_t = \sum_{j=0}^{m} I_j(y_t) \frac{d_{j,t} - d_{j+1,t}}{g_{j+1,t} - g_{j,t}} \left(\sum_{j=0}^{m} \frac{(d_{j,t} - d_{j+1,t})^2}{g_{j+1,t} - g_{j,t}} \right)^{-\frac{1}{2}},$$

where $g_{j,t} = \Phi(\lambda_j - \mu_t)$, $d_{j,t} = \Phi'(\lambda_j - \mu_t)$, $\lambda_0 = -\infty$, $\lambda_{m+1} = \infty$ and $I_j(y_t)$ is an indicator function such that $I_j(y_t) = 1$, if $y_t = j$, and $I_j(y_t) = 0$, otherwise.

3 Testing residual autocorrelation

3.1 The empirical autocorrelation function

Residuals of time series models are routinely used to detect the presence of residual autocorrelation that remains unexplained by the model. Autocorrelation functions are a useful diagnostic tool to asses whether the specification of the time varying parameter needs to be extended by, for instance, adding more lags of the observable variable or including seasonal components in the dynamic equation of the model. In this section, we discuss the derivation of empirical autocorrelation functions of conditional score residuals and their asymptotic properties.

The kth lag empirical autocorrelations of conditional score errors $\rho_{T,k}$ and of conditional score residuals $\hat{\rho}_{T,k}$ are given by

$$\rho_{T,k} = \frac{\sum_{t=1}^{T} s_t s_{t-k}}{\sum_{t=1}^{T} s_t^2}, \quad \text{and} \quad \hat{\rho}_{T,k} = \frac{\sum_{t=1}^{T} \hat{s}_t \hat{s}_{t-k}}{\sum_{t=1}^{T} \hat{s}_t^2}, \tag{7}$$

for $k \in \{1, 2, 3, \dots\}$. As noted before, conditional score errors s_t are not observable and therefore only the empirical autocorrelation function of conditional score residuals is feasible. The asymptotic properties of the residual autocorrelations $\hat{\rho}_{T,k}$ are different from the asymptotic properties of the autocorrelations of the errors $\rho_{T,k}$. It can be easily shown that, under the null hypothesis of correct specification of the model and hence no residual autocorrelation, the asymptotic distribution of $\sqrt{T}\rho_{T,k}$ is standard normal. However, this is not the case for $\sqrt{T}\hat{\rho}_{T,k}$ as its distribution depends on the asymptotic distribution of the ML estimator $\hat{\boldsymbol{\theta}}_T$. This fact is well-known in the literature; see, for instance, Li and Mak (1994) for a discussion on the distribution of squared residuals of GARCH models. For notational convenience, we stack the first K empirical autocorrelations of conditional score errors and residuals into vectors, which are denoted as $\boldsymbol{\rho}_T = (\rho_{T,1}, \ldots, \rho_{T,K})^\top$ and $\hat{\boldsymbol{\rho}}_T = (\hat{\rho}_{T,1}, \ldots, \hat{\rho}_{T,K})^\top$, respectively. In the next section, we derive the asymptotic properties of the empirical autocorrelation vector $\hat{\boldsymbol{\rho}}_T$.

3.2 Asymptotic distribution of empirical autocorrelations

We formally derive the asymptotic distribution of empirical autocorrelations of conditional score residuals. We start by introducing a set of regularity assumptions on the class of observation-driven models defined in (1) and (2). The first assumption imposes some regularity conditions on the conditional probability density function $p(y|f;\boldsymbol{\lambda})$. More specifically, Assumption 3.1 ensures that the conditional probability density function is such that its expected score is equal to zero and the Fisher information equality holds. In addition, it imposes some smoothness assumptions on the probability density function.

Assumption 3.1. *The conditional density function $p(y|f;\boldsymbol{\lambda})$ satisfies the following regularity conditions:*

(i) The function $(f,\boldsymbol{\lambda}) \mapsto \log p(y|f;\boldsymbol{\lambda})$ is twice continuously differentiable in $\mathrm{F} \times \Lambda$ for any $y \in \mathrm{Y}$.

(ii) The expected score is zero

$$\int_{\mathrm{Y}} d_1(y,f,\boldsymbol{\lambda}) p(y|f;\boldsymbol{\lambda}) dy = \mathbf{0}_r,$$

and the Fisher information equality holds

$$\boldsymbol{\Omega}(f,\boldsymbol{\lambda}) = \int_{\mathrm{Y}} d_1(y,f,\boldsymbol{\lambda}) d_1(y,f,\boldsymbol{\lambda})^\top p(y|f;\boldsymbol{\lambda}) dy = -\int_{\mathrm{Y}} d_2(y,f,\boldsymbol{\lambda}) p(y|f;\boldsymbol{\lambda}) dy,$$

for any $(f,\boldsymbol{\lambda}) \in \mathrm{F} \times \Lambda$, where

$$d_1(y,f,\boldsymbol{\lambda}) = \frac{\partial \log p(y|f;\boldsymbol{\lambda})}{\partial (f,\boldsymbol{\lambda}^\top)^\top} \quad \text{and} \quad d_2(y,f,\boldsymbol{\lambda}) = \frac{\partial^2 \log p(y|f;\boldsymbol{\lambda})}{\partial (f,\boldsymbol{\lambda}^\top)^\top \partial (f,\boldsymbol{\lambda}^\top)}.$$

(iii) The information matrix $\boldsymbol{\Omega}(f,\boldsymbol{\lambda})$ is positive definite for any $(f,\boldsymbol{\lambda}) \in \mathrm{F} \times \Lambda$, and the function $(f,\boldsymbol{\lambda}) \mapsto \boldsymbol{\Omega}(f,\boldsymbol{\lambda})$ is continuously differentiable.

The conditions in Assumption 3.1 are standard regularity conditions which are satisfied for most parametric probability density functions of practical interest. Assumption 3.2 imposes some regularity conditions on the time series process and the filtered time-varying parameter $f_t(\boldsymbol{\theta})$.

Assumption 3.2. *The sequence* $\{(y_t, f_t(\boldsymbol{\theta}))\}_{t\in\mathbb{Z}}$ *is stationary and ergodic and* $f_t(\boldsymbol{\theta})$ *is* $\mathcal{F}_{t-1}$*-measurable for any* $\boldsymbol{\theta} \in \Theta$*. Furthermore,* $\boldsymbol{\theta} \mapsto f_t(\boldsymbol{\theta})$ *is twice continuously differentiable in* Θ *with probability one. The parameter set* Θ *is compact.*

The stationarity and ergodicity of the time series process is a standard condition. The $\mathcal{F}_{t-1}$-measurability and stationarity and ergodicity of the filtered parameter $f_t(\boldsymbol{\theta})$ follows from the so-called invertibility of the model (Straumann and Mikosch, 2006; Blasques et al., 2018). These conditions are typically required for time-varying parameter models in order to ensure consistent estimation of the parameters.

Assumption 3.3 requires some uniform moment conditions on the conditional score errors. The norm $\|\cdot\|$ denotes the $L1$ norm when applied to a vector and the operator norm induced by the $L1$ norm when applied to a matrix. Furthermore, $\|\cdot\|_\Theta$ denotes the supremum norm. Given a function $g : \Theta \mapsto \mathbb{R}^{a\times b}$, $a, b \in \mathbb{N}$, the supremum norm is $\|g\|_\Theta = \sup_{\theta\in\Theta} \|g(\theta)\|$.

Assumption 3.3. *The following uniform moment conditions are satisfied*

$$\mathbb{E}\left\|\frac{\partial s_t}{\partial \boldsymbol{\theta}}\frac{\partial s_t}{\partial \boldsymbol{\theta}^\top}\right\|_\Theta < \infty, \qquad \textit{and} \qquad \mathbb{E}\|s_t^2\|_\Theta < \infty.$$

Assumption 3.4 imposes conditions on the asymptotic behavior of the ML estimator $\hat{\boldsymbol{\theta}}_T$.

Assumption 3.4. *The ML estimator* $\hat{\boldsymbol{\theta}}_T$ *satisfies*

$$\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) = \boldsymbol{\Sigma}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\partial l_t}{\partial \boldsymbol{\theta}} + o_p(1), \tag{8}$$

where

$$\boldsymbol{\Sigma} = -\mathbb{E}\left[\frac{\partial^2 l_t}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^\top}\right]^{-1} = \mathbb{E}\left[\frac{\partial l_t}{\partial \boldsymbol{\theta}}\frac{\partial l_t}{\partial \boldsymbol{\theta}^\top}\right]^{-1} \tag{9}$$

is positive definite.

The conditions in (8) and (9) follow under standard regularity assumptions that are typically needed to ensure the asymptotic normality of the ML estimator. In fact, Assumption 3.4 ensures that the ML estimator is asymptotically normal with asymptotic covariance matrix given by the inverse of the Fisher information.

Assumption 3.5 imposes a linear independence condition to ensure that the asymptotic covariance matrix of the autocorellation vector $\hat{\boldsymbol{\rho}}_T$ is positive definite. We denote with $\boldsymbol{s}_{K,t}$ the vector that contains the first K lags of the conditional score errors, i.e. $\boldsymbol{s}_{K,t} = (s_{t-1}, \ldots, s_{t-K})^\top$.

Assumption 3.5. *The following inequality holds with positive probability:*

$$\mathbf{z}^\top \boldsymbol{s}_{K,t} + \mathbf{x}^\top \sqrt{I_t} \frac{\partial f_t}{\partial \boldsymbol{\xi}} \neq 0,$$

for any $\mathbf{x} \in \mathbb{R}^s$, $\mathbf{z} \in \mathbb{R}^K$, $\mathbf{z} \neq \mathbf{0}_K$.

Theorem 3.1 delivers the main result on the asymptotic distribution of the empirical autocorrelation vector $\hat{\boldsymbol{\rho}}_T$ for the general class of time series models in (1) and (2) under Assumptions 3.1-3.5.

Theorem 3.1. *Let Assumptions 3.1-3.5 hold. Then, the empirical autocorrelation vector* $\hat{\boldsymbol{\rho}}_T$ *has the following asymptotic distribution*

$$\sqrt{T}\hat{\boldsymbol{\rho}}_T \xrightarrow{d} N(\mathbf{0}_K, \boldsymbol{V}), \qquad \boldsymbol{V} = \boldsymbol{I}_K - \boldsymbol{\Phi}\boldsymbol{\Sigma}\boldsymbol{\Phi}^\top,$$

where $\boldsymbol{\Sigma}$ *is defined in (9) and*

$$\boldsymbol{\Phi} = \mathbb{E}(\boldsymbol{s}_{K,t}\boldsymbol{n}_t^\top), \quad \textit{with} \quad \boldsymbol{n}_t = \sqrt{I_t}\frac{\partial f_t}{\partial \boldsymbol{\theta}} + \frac{1}{\sqrt{I_t}}\boldsymbol{k}_t, \tag{10}$$

$$\boldsymbol{k}_t = \begin{bmatrix} \mathbf{0}_s \\ \boldsymbol{j}_{\boldsymbol{\lambda} f,t} \end{bmatrix}, \qquad \boldsymbol{j}_{\boldsymbol{\lambda} f,t} = \mathbb{E}(u_t \boldsymbol{u}_{\boldsymbol{\lambda},t} | \mathcal{F}_{t-1}), \qquad \boldsymbol{u}_{\boldsymbol{\lambda},t} = \frac{\partial \log p(y_t | f; \boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}}\Big|_{f=f_t}.$$

Furthermore, the covariance matrix $\boldsymbol{V}$ *is positive definite.*

The proof of the theorem is given in Appendix A.1. Theorem 3.1 provides the form of the asymptotic covariance matrix $\boldsymbol{V}$ of empirical autocorrelations of conditional score residuals. This poses the basis for statistical inference on residual autocorrelations. Some special cases of this general result can be found in existing literature. For instance, for the ARMA model with a Gaussian error term, conditional score residuals are equivalent to standard ARMA residuals as discussed in Example 2.1 and the asymptotic distribution in Theorem 3.1 coincides with the result in McLeod (1978). Similarly, for the GARCH model with Gaussian errors, conditional score residuals are equivalent to squared residuals as discussed in Example 2.2 and the asymptotic distribution in Theorem 3.1 coincides with the result in Li and Mak (1994). In the next section, we propose a reliable estimator for $\boldsymbol{V}$ and formally discuss its properties.

3.3 Estimation of the asymptotic covariance matrix

In order to conduct inference on residual autocorrelations, we need a consistent estimate of the asymptotic covariance matrix $\boldsymbol{V}$. Under some regularity conditions, the covariance matrix $\boldsymbol{V}$ may be consistently estimated by means of a plug-in estimator, where the matrices $\boldsymbol{\Sigma}$ and $\boldsymbol{\Phi}$ are replaced by their sample equivalent with the ML estimate $\hat{\boldsymbol{\theta}}_T$ plugged-in in place of the true parameter value. However, the use of a plug-in estimator does not ensure that the estimated covariance matrix is positive definite in finite samples. This is a relevant problem in practical applications since a positive definite covariance is needed to derive confidence intervals and to test for residual autocorrelation. We propose a consistent estimator that ensures positive definiteness in small samples under mild conditions. Consider the covariance matrix estimate $\hat{\boldsymbol{V}}$ given by

$$\hat{\boldsymbol{V}} = \frac{1}{T}\sum_{t=1}^{T}\hat{\boldsymbol{s}}_{K,t}\hat{\boldsymbol{s}}_{K,t}^{\top} - \frac{1}{T}\sum_{t=1}^{T}\hat{\boldsymbol{s}}_{K,t}\hat{\boldsymbol{n}}_t^{\top}\left(\frac{1}{T}\sum_{t=1}^{T}(\hat{\boldsymbol{n}}_t\hat{\boldsymbol{n}}_t^{\top} + \hat{\boldsymbol{H}}_t)\right)^{-1}\frac{1}{T}\sum_{t=1}^{T}\hat{\boldsymbol{n}}_t\hat{\boldsymbol{s}}_{K,t}^{\top}, \tag{11}$$

with

$$\hat{\boldsymbol{H}}_t = \hat{\boldsymbol{J}}_t - \frac{1}{\hat{I}_t}\hat{\boldsymbol{k}}_t\hat{\boldsymbol{k}}_t^{\top},$$

where $\hat{\boldsymbol{s}}_{K,t}$, $\hat{\boldsymbol{n}}_t$ and $\hat{\boldsymbol{k}}_t$ are plug-in estimates of $\boldsymbol{s}_{K,t}$, $\boldsymbol{n}_t$ and $\boldsymbol{k}_t$, respectively, and $\hat{\boldsymbol{J}}_t$ is the plug-in estimate of $\boldsymbol{J}_t$, which is defined as

$$\boldsymbol{J}_t = \begin{bmatrix} \boldsymbol{0}_{s\times s} & \boldsymbol{0}_{s\times r} \\ \boldsymbol{0}_{r\times s} & \boldsymbol{j}_{\boldsymbol{\lambda\lambda},t} \end{bmatrix}, \qquad \boldsymbol{j}_{\boldsymbol{\lambda\lambda},t} = \mathbb{E}(\boldsymbol{u}_{\boldsymbol{\lambda},t}\boldsymbol{u}_{\boldsymbol{\lambda},t}^{\top}|\mathcal{F}_{t-1}).$$

From the expression of $\hat{\boldsymbol{V}}$, we can see that if $\sum_{t=1}^{T}\hat{\boldsymbol{H}}_t$ is equal to zero, then $\hat{\boldsymbol{V}}$ is the sample covariance matrix of the residuals of the regression of $\hat{\boldsymbol{s}}_{K,t}$ on $\hat{\boldsymbol{n}}_t$. Furthermore, the matrix $\sum_{t=1}^{T}\hat{\boldsymbol{H}}_t$ is positive semidefinite since $\hat{I}_t\,\hat{\boldsymbol{j}}_{\boldsymbol{\lambda\lambda},t} - \hat{\boldsymbol{j}}_{\boldsymbol{\lambda}f,t}\hat{\boldsymbol{j}}_{\boldsymbol{\lambda}f,t}^{\top}$ is positive definite for any t under the positive definiteness of the conditional Fisher information in Assumption 3.1. Therefore, it is immediate to see that $\hat{\boldsymbol{V}}$ is positive definite if the columns of the matrices $[\hat{\boldsymbol{s}}_{K,1},\dots,\hat{\boldsymbol{s}}_{K,T}]^\top$ and $[\hat{\boldsymbol{n}}_1,\dots,\hat{\boldsymbol{n}}_T]^\top$ are linearly independent. Furthermore, given the positive definiteness of $\hat{I}_t\,\hat{\boldsymbol{j}}_{\boldsymbol{\lambda}\boldsymbol{\lambda},t}-\hat{\boldsymbol{j}}_{\boldsymbol{\lambda}f,t}\hat{\boldsymbol{j}}_{\boldsymbol{\lambda}f,t}^\top$, it can be shown that the linear independence between the columns of $[\hat{\boldsymbol{s}}_{K,1},\dots,\hat{\boldsymbol{s}}_{K,T}]^\top$ and the first s columns of $[\hat{\boldsymbol{n}}_1,\dots,\hat{\boldsymbol{n}}_T]^\top$ is actually sufficient for the positive definiteness of $\hat{\boldsymbol{V}}$. This weaker condition is, in fact, the sample equivalent of Assumption 3.5.

Next, we focus on the consistency of the covariance matrix estimator $\hat{\boldsymbol{V}}$. We impose some additional uniform moment conditions to ensure the convergence of the sample averages in (11) to the corresponding population quantities.

Assumption 3.6. *The following uniform moment conditions are satisfied*

$$\mathbb{E}\,\|\boldsymbol{n}_t\|_\Theta^2<\infty,\quad \textit{and}\quad \mathbb{E}\,\|\boldsymbol{H}_t\|_\Theta<\infty.$$

Theorem 3.2 delivers the consistency of the covariance matrix estimator $\hat{\boldsymbol{V}}$.

Theorem 3.2. *Let Assumptions 3.1-3.6 hold. Then, the estimator $\hat{\boldsymbol{V}}$ defined in (11) is consistent,*

$$\hat{\boldsymbol{V}}\xrightarrow{p}\boldsymbol{V}\quad \textit{as}\quad T\to\infty.$$

The asymptotic results provided by Theorems 3.1 and 3.2 can be employed to test the null hypothesis of no residual autocorrelation. In particular, under the null hypothesis of correct specification, the autocorrellation of conditional score errors is equal to zero, i.e. $\rho_k=0$ for any $k\geq 1$ where $\rho_k=\mathbb{E}(s_t s_{t-k})/\mathbb{E}(s_t^2)$. An element-wise test for $\rho_k=0$ can be implemented based on the standardized empirical autocorrelation

$$\hat{r}_k=\sqrt{T}\frac{\hat{\rho}_{T,k}}{\sqrt{\hat{v}_{kk}}},$$

where $\hat{v}_{kk}$ denotes the kth diagonal element of $\hat{\boldsymbol{V}}$. From Theorems 3.1 and 3.2, it follows immediately that $\hat{r}_k\xrightarrow{d}N(0,1)$. This result can be used to test for residual autocorrelation at a given lag k and also to derive a graphical representation of the autocorrelation function together with 95%-level confidence intervals, which is a standard diagnostic tool that is typically implemented in statistical software packages.

A Portmanteau test for the null hypothesis that the first K autocorrelations are equal to zero $\rho_k=0$, $k=1,\dots,K$, can be constructed based on the test statistic

$$Q_K=T\hat{\boldsymbol{\rho}}_T^\top\hat{\boldsymbol{V}}^{-1}\hat{\boldsymbol{\rho}}_T.$$

Theorems 3.1 and 3.2 entail that, under the null hypothesis, Q_K has an asymptotic chi-squared distribution with K degrees of freedom, that is $Q_K \xrightarrow{d} \chi^2_K$. Therefore, the test rejects the hypothesis of no residual autocorellation for large values of the test statistic according to the critical values of χ^2_K. Sections 5 and 4 provide two examples that validate the use of the asymptotic results in small samples.

4 Residuals for GARCH models with heavy-tailed errors

4.1 Conditional score residuals for t-GARCH models

Time series datasets with extreme observations are often encountered in empirical applications. GARCH models are typically embedded with heavy-tailed distributions to describe extreme events in conditionally heteroskedastic time series data. The Student's t distribution is widely used for this purpose and statistical software packages have default implementations of such models. For instance, heavy-tailed GARCH models are employed to model price changes of financial assets, which are well-known to display volatility clustering as well as extreme observations. In this section, we illustrate the appealing properties of conditional score residuals for diagnostic analysis of GARCH models with Student's t errors, which are typically referred to as t-GARCH models in the literature. Consider the general class of t-GARCH models as given by

$$y_t = \sqrt{h_t}\varepsilon_t, \quad h_t = g_{\boldsymbol{\theta}}(h_{t-1}, \dots, h_{t-q}, y_{t-1}, \dots, y_{t-p}), \tag{12}$$

where ε_t has a standardized Student's t distribution, $\varepsilon_t \sim t_v(0,1)$, with zero mean, unit variance, and degrees of freedom parameter v. The specification of the conditional variance is such that $h_t > 0$ with probability one. We refer the reader to Straumann (2005) for a formal discussion on the stochastic properties and the ML estimation theory of the general class of t-GARCH models in (12).

For this class of models, conditional score residuals are given by

$$s_t = \sqrt{\frac{v+3}{2v}} \left(\frac{(v+1)y_t^2}{(v-2)h_t + y_t^2} - 1 \right).$$

From the expression of s_t, we notice that conditional score residuals converge to standard squared residuals as $v \to \infty$. This result follows intuitively from the fact that the Student's t distribution approaches the Gaussian distribution as $v \to \infty$. A key advantage of conditional score residuals compared to squared residuals is that they are robust to extreme observations. In particular, $|s_t|$ is bounded by a constant with probability one and therefore s_t has finite moments of any order irrespective of the moments of the error term ε_t. On the other hand, squared residuals require a finite fourth moment for the error term, $\mathbb{E}(\varepsilon_t^4) < \infty$, in order to ensure that squared residuals have a finite second moment. We note that a finite second moment for the residuals is necessary for the autocorrelation function to be well defined. Furthermore, as we shall illustrate below, squared residuals perform poorly for diagnostic analysis in time series datasets with extreme observations even when the fourth moment is finite. Instead, conditional score residuals deliver more reliable and powerful diagnostic tools.

4.2 Monte Carlo study

We perform a simulation study to evaluate the size and power of diagnostic tests based on conditional score residuals in comparison with squared residuals. We start by focusing on the size of tests for the null hypothesis of no residual autocorrelation. For this purpose, we consider the following t-GARCH(1,1) model

$$y_t = \sqrt{h_t}\varepsilon_t, \quad h_t = \alpha_0 + \beta_1 h_{t-1} + \alpha_1 y_{t-1}^2, \tag{13}$$

where $\varepsilon_t \sim t_v(0,1)$. For this model, the unknown parameter vector $\boldsymbol{\theta} = (\alpha_0, \beta_1, \alpha_1, v)^\top$ is estimated by ML. We consider the diagnostic tests presented in Section 3 for the first five lags of the autocorrelation function. More specifically, we test the null hypothesizes of no residual autocorrelation based on the test statistics, $\hat{r}_k$ for $k = 1, \ldots, 5$, and Q_K for $K = 5$.

Table 1 reports the empirical size of the tests for conditional score residuals and squared residuals. Different sample sizes and different values of the degrees of freedom parameter v are considered. The results show that tests based on squared residuals are highly oversized when $v = 5$. As the sample size T increases, the empirical size seems to approach the nominal level, however, the tests are still significantly oversized even for relatively large sample sizes ($T = 2500$). This indicates that tests based on squared residuals are not reliable when extreme observations are present. For $v = 10$, we can see that the performance of squared residuals improves as the tests are oversized in small samples ($T = 500$) but they are properly sized in larger samples. When looking at conditional score residuals, the results indicate that the empirical size is close to the nominal one irrespective of the value of the tail parameter v. The results are consistent across the different sample sizes considered in the study. This indicates that the presence of extreme observations in the data have no relevant effect on the empirical size of tests based on conditional score residuals and the asymptotic distribution of the test statistics can be regarded as an accurate approximation to the finite sample distribution even for relatively small sample sizes.

Table 1: *Empirical size of autocorrelation tests for conditional score residuals. The results are obtained from* 5000 *Monte Carlo replications. The nominal sizes considered are* $\alpha = 0.05$ *and* $\alpha = 0.10$*. The true parameter vector is* $\boldsymbol{\theta} = (1.0, 0.60, 0.20, v)^\top$ *for* $v = 5$ *and* $v = 10$.

			score residuals			squared residuals		
			$T = 500$	$T = 1000$	$T = 2500$	$T = 500$	$T = 1000$	$T = 2500$
		$\hat{r}_1$	0.053	0.050	0.047	0.083	0.075	0.067
		$\hat{r}_2$	0.053	0.049	0.048	0.109	0.095	0.080
	$v = 5$	$\hat{r}_3$	0.051	0.047	0.045	0.104	0.093	0.081
		$\hat{r}_4$	0.053	0.049	0.049	0.111	0.098	0.080
		$\hat{r}_5$	0.056	0.052	0.049	0.120	0.104	0.081
$\alpha = 0.05$		Q_5	0.056	0.050	0.053	0.209	0.187	0.157
		$\hat{r}_1$	0.059	0.054	0.052	0.046	0.045	0.046
		$\hat{r}_2$	0.053	0.051	0.048	0.055	0.050	0.049
	$v = 10$	$\hat{r}_3$	0.046	0.051	0.046	0.052	0.058	0.049
		$\hat{r}_4$	0.052	0.049	0.049	0.059	0.053	0.052
		$\hat{r}_5$	0.058	0.058	0.049	0.060	0.055	0.047
		Q_5	0.057	0.052	0.052	0.075	0.064	0.056
		$\hat{r}_1$	0.105	0.105	0.101	0.118	0.114	0.111
		$\hat{r}_2$	0.102	0.100	0.094	0.152	0.139	0.118
	$v = 5$	$\hat{r}_3$	0.103	0.096	0.095	0.159	0.144	0.130
		$\hat{r}_4$	0.101	0.105	0.101	0.159	0.144	0.130
		$\hat{r}_5$	0.108	0.095	0.098	0.163	0.154	0.122
$\alpha = 0.10$		Q_5	0.113	0.102	0.105	0.256	0.238	0.207
		$\hat{r}_1$	0.111	0.105	0.104	0.086	0.083	0.091
		$\hat{r}_2$	0.107	0.102	0.100	0.101	0.090	0.094
	$v = 10$	$\hat{r}_3$	0.100	0.102	0.093	0.095	0.102	0.090
		$\hat{r}_4$	0.099	0.096	0.097	0.102	0.094	0.093
		$\hat{r}_5$	0.117	0.107	0.103	0.111	0.103	0.096
		Q_5	0.108	0.095	0.097	0.123	0.104	0.103

We investigate next the power of the diagnostic autocorrelation tests based on the score residuals in comparison with the power of the same tests based on squared residuals. In our simulation study, we consider the following t-GARCH(1,2) model as the data generating process

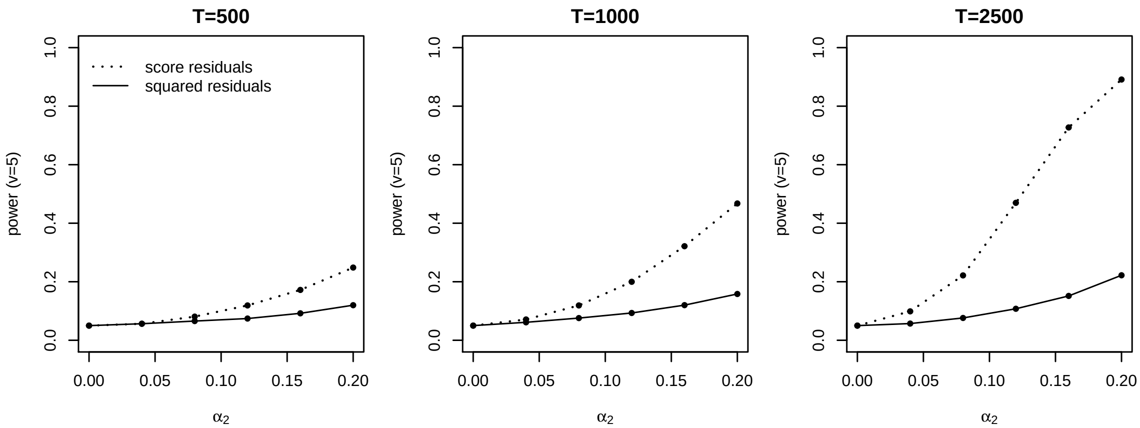
$$h_t = 1.0 + 0.6h_{t-1} + 0.15y_{t-1}^2 + \alpha_2 y_{t-2}^2.$$

On the other hand, the misspecified t-GARCH(1,1) model in (13) is considered for estimation and analysis, that is $\alpha_2 = 0$. A grid of values of α_2, ranging from 0 to 0.20, is considered. This provides different levels of misspecification where $\alpha_2 = 0$ represents the case of no misspecifcation and $\alpha_2 = 0.20$ can be regarded as the more severe misspecification. For all considered values of α_2, the simulation study is conducted for different sample sizes ($T = 500,\ 1000,\ 2500$) and for two different values of the degrees of freedom parameter ($v = 5,\ 10$).

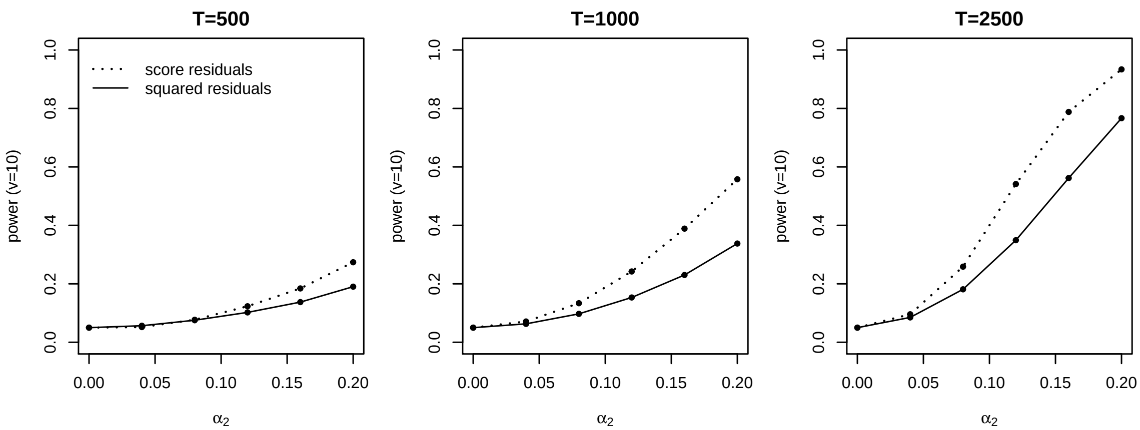
Figure 1 displays the size-adjusted power of the test statistics Q_K, $K = 5$, for conditional score residuals and for squared residuals. The results show that conditional score residuals outperform squared residuals in terms of power in all configurations of the simulations. For small values of the tail parameter, $v = 5$, squared residuals have low power in detecting the misspecification of the conditional variance h_t. Even for the largest sample size, $T = 2500$, and highest degree of misspecification, $\alpha_2 = 0.20$, the power remains below 0.3. On the other hand, conditional score residuals have a much higher power and they do not suffer from the presence of extreme observations. In particular, we find that the test power from conditional score residuals does not decline for the lower value of v. Overall, the results indicate that squared residuals are not well suited for diagnostic analysis of time series with extreme observations. On the contrary, conditional score residuals provide robust diagnostic tools to detect residual autocorrelation. The low values of the degrees of freedom parameter v considered in this simulation study are commonly encountered in empirical applications. For example, financial time series typically exhibit extreme observations that require such low degree of freedom values v for their treatment.

4.3 Empirical illustration for US stock returns

In this section, we present an empirical application that illustrates the robustness of conditional score residuals compared to squared residuals. The dataset consists of 20 years of daily log-returns of the S&P500 stock index from January 2000 to December 2019.



(a)



(b)

Figure 1: *Size-adjusted power of the test statistic* Q_5*, using* 5000 *Monte Carlo replications. The dashed and solid lines represent the power of the test for conditional score residuals and squared residuals, respectively. Panels (a),(b) show the power for* $v = 5, 10$.

The sample size is 5030 observations. The idea is to estimate the t-GARCH(1,1) model in (13) and perform diagnostic analysis based on both conditional score residuals and squared residuals. Table 2 reports the ML estimates of the t-GARCH(1,1) as well as the t-GARCH(1,2). The tail parameter v is estimated to be around 6.5. This indicates the presence of heavy tails in the error term of the model. The information criteria suggest that the t-GARCH(1,2) fits the data better than the t-GARCH(1,1), which can be noted as the t-GARCH(1,2) has lower AIC and BIC. Therefore, we may expect the presence of residual autocorrelation when analyzing the residuals of the t-GARCH(1,1) .

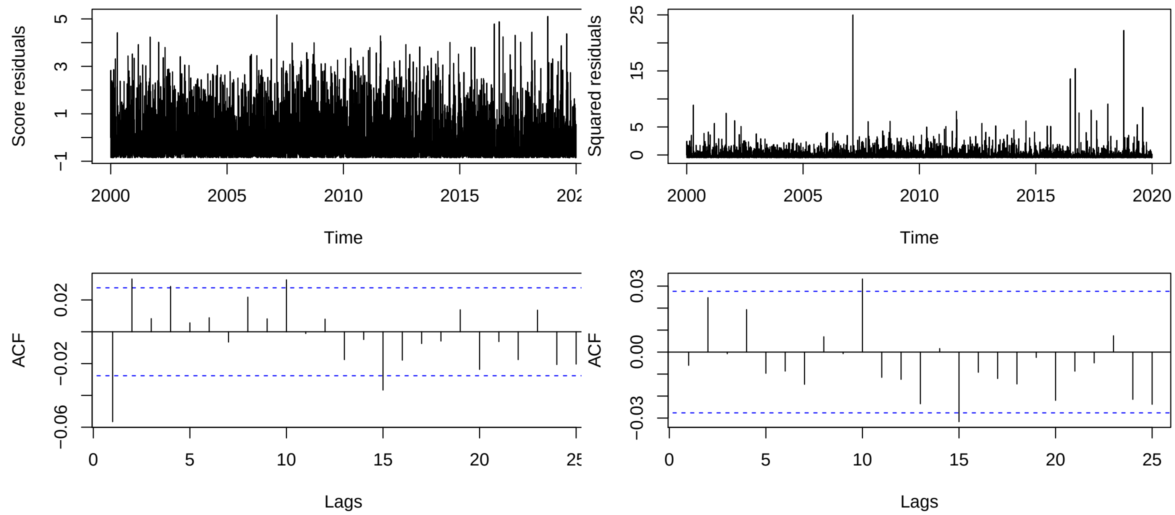
Table 2: *Maximum likelihood estimates of t-GARCH(1,1) and t-GARCH(1,2) models, the standard errors are in brackets. The last three columns contain the maximized log-likelihood value (log-lik), the Akaike information criterion (AIC) and the Bayesian information criterion (BIC).*

	α_0	β_1	α_1	α_2	v	log-lik	AIC	BIC
GARCH(1,1)	2.723	0.891	0.105	-	6.713	-6730.8	13469.6	13495.7
	(0.996)	(0.010)	(0.010)		(0.590)			
GARCH(1,2)	3.029	0.859	0.048	0.087	6.470	-6723.3	13456.6	13489.2
	(1.253)	(0.015)	(0.015)	(0.021)	(0.547)			

Figure 2 displays conditional score residuals and squared residuals together with their standardized empirical autocorrelation function for the t-GARCH(1,1) model. The standardized empirical autocorrelations are plotted together with 95%-level asymptotic confidence intervals under the null hypothesis of no residual autocorrelation, as described in Section 3.3. Form the plots of the residuals, we can see the robustness of conditional score residuals to extreme observations. Instead, squared residuals present several extreme values. Conditional score residuals show significant autocorrelation in the first and second lags, providing evidence of misspecification of the t-GARCH(1,1) model. In particular, the first lag shows significant negative autocorrelation and the second lag positive autocorrelation. This autocorreation structure in the residuals is coherent with the estimation results that suggest a better fit of the t-GARCH(1,2) compared to the t-GARCH(1,1). On the contrary, squared residuals show no evidence of significant residual autocorrelation. This may be explained by the low power of tests based on squared residuals in the presence of extreme observations as illustrated in the simulation study.

Table 3: *Test statistic Q_K and corresponding p-values for conditional score residuals and squared residuals of the t-GARCH(1,1) model.*

		$K = 5$	$K = 10$	$K = 15$	$K = 20$
Conditional score residuals	Q_K	21.49	32.95	40.30	46.21
	p-value	< 0.001	< 0.001	< 0.001	< 0.001
Squared residuals	Q_K	5.45	12.8	21.07	25.86
	p-value	0.363	0.235	0.135	0.170



(a) (b)

Figure 2: *Panel (a) displays the conditional score residuals of the GARCH(1,1) and their standardized empirical autocorrelation function with 95%-level confidence intervals under the null hypothesis of no residual autocorrelation. Panel (b) displays squared residuals of the GARCH(1,1) and their standardized empirical autocorrelation function with 95%-level confidence intervals.*

Finally, Table 3 reports the test statistic Q_K for conditional score residuals and squared residuals for different values of K. The results show that the null hypothesis of no residual autocorrelation is strongly rejected for conditional score residuals. Instead, the null hypothesis is not rejected for squared residuals. These findings are coherent with the empirical autocorellation functions reported in Figure 2.

5 Residuals for dynamic Gaussian copula models

5.1 Conditional score residuals for a bivariate Gaussian copula model

Dynamic copula models are widely used in financial econometrics to model the dependence between time series processes in settings where the multivariate Gaussian density is not appropriate; see the discussions in Patton (2006) and Salvatierra and Patton (2015). Although the literature on the dynamic modeling with copulas has been growing over the years, see Patton (2012) for an overview, a formally defined residual for testing residual dependence in dynamic copula models is not available. Conditional score residuals can be employed for this purpose. Here, we define conditional score residual for a specific Gaussian copula model, however, our framework can be applied to all other forms of copula functions that are typically used in the literature. Consider the observable bivariate vector $y_t = (y_{1t}, y_{2t})^\top$ and specify the bivariate Gaussian copula model as

$$y_t|\mathcal{F}_{t-1} \sim \mathcal{C}_{\mathcal{G}}(p_t), \quad p_t = g_{\boldsymbol{\theta}}(p_{t-1}, \ldots, p_{t-q}, y_{t-1}, \ldots, y_{t-p}), \tag{14}$$

where $C_G(p_t)$ denotes a bivariate Gaussian copula distribution with dynamic correlation parameter p_t. The updating function $g_{\boldsymbol{\theta}}$ is a parametric function such that $p_t \in [-1, 1]$ with probability one. For this class of dynamic copula models, conditional score residuals are given by

$$s_t = \frac{p_t + x_{1t}x_{2t} - p_t(x_{1t}^2 + x_{2t}^2) + p_t^2(x_{1t}x_{2t} - p_t)}{\sqrt{1 + p_t^2(1 - p_t^2)}},$$

where $x_{1t} = \Phi^{-1}(y_{1t})$ and $x_{2t} = \Phi^{-1}(y_{2t})$. Next, we present a simulation study and an empirical illustration on testing residual autrocorrelation for a dynamic copula model using conditional score residuals. In this case, there is not an alternative definition of residuals that can be used as a benchmark.

5.2 Monte Carlo study

We study the small sample properties of the autocorrelation tests presented in Section 3 for the conditional score residuals in terms of size and power. We start by investigating the size of the tests. The design of the Monte Carlo experiment is equivalent to the one in Section 4. For the specification of the conditional correlation coefficient p_t in (14), we follow Patton (2006) and consider the following link function and dynamic equation,

$$p_t = \frac{\exp(g_t) - 1}{\exp(g_t) + 1}, \qquad g_{t+1} = \alpha_0 + \beta_1 g_t + \alpha_1 \Phi^{-1}(y_{1t})\Phi^{-1}(y_{2t}). \tag{15}$$

Table 4 reports the empirical size of the tests based on the statistics $\hat{r}_k$, for $k = 1, \ldots, 5$, and Q_K, with $K = 5$, for different sample sizes. The reported results show that the empirical size is close to its corresponding nominal value for all sample sizes. We therefore conclude that the asymptotic results can be used as a reliable approximation for the distribution of the test statistics in relatively small samples.

Table 4: *Empirical size of autocorrelation tests for conditional score residuals. The results are obtained from* 5000 *Monte Carlo replications. The nominal sizes considered are* $\alpha = 0.05$ *and* $\alpha = 0.10$*. The true parameter vector is* $\boldsymbol{\theta} = (0.2, 0.60, 0.20)^\top$.

	$\alpha = 0.10$			$\alpha = 0.05$		
	$T = 500$	$T = 1000$	$T = 2500$	$T = 500$	$T = 1000$	$T = 2500$
$\hat{r}_1$	0.103	0.100	0.097	0.051	0.052	0.048
$\hat{r}_2$	0.113	0.096	0.106	0.057	0.051	0.049
$\hat{r}_3$	0.101	0.097	0.105	0.050	0.048	0.056
$\hat{r}_4$	0.102	0.098	0.099	0.047	0.047	0.050
$\hat{r}_5$	0.107	0.096	0.098	0.052	0.046	0.054
Q_5	0.107	0.103	0.105	0.058	0.056	0.049

Next, our focus is on assessing the power of the autocorrelation test based on the statistic Q_K, with $K = 5$. The data generating process in our simulation study is based on model (14) with the dynamic correlation coefficient p_t given by equation (15) and the dynamic process for g_t given by

$$g_{t+1} = 0.2 + 0.6g_t + 0.15\Phi^{-1}(y_{1t})\Phi^{-1}(y_{2t}) + \alpha_2\Phi^{-1}(y_{1t-1})\Phi^{-1}(y_{2t-1}).$$

The estimation of the model is for the misspecified dynamic copula model in (15). The model is correctly specified when $\alpha_2 = 0$. We consider different values for α_2, ranging from 0 to 0.2. Figure 3 displays the power of the test as a function of α_2, for different sample sizes T. We see that the power increases as α_2 increases. This first finding indicates that the test rejects the null hypothesis with higher likelihood when a more severe form of misspecification is present. Furthermore, the power function of the test increases as the sample size increases. This second finding suggests that the test is consistent in detecting the dynamic misspecification of the model.

5.3 Empirical illustration

We employ the bivariate dynamic copula model in (15) to model dependence between oil prices and exchange rates. The dataset consists of daily log-differences of WTI oil prices and EU/USD exchange rates from January 2000 to December 2019. The sample size is 4975 observations. The series are obtained form FRED dataset and days where the price of one of the two time series is not available are removed form the sample. We

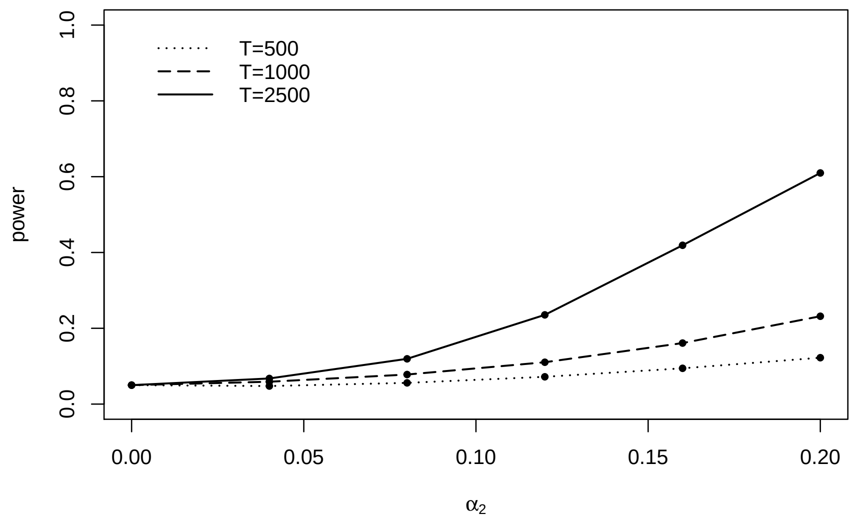


Figure 3: *Empirical power of the test statistic* Q_K, $K = 5$, *for different sample sizes. The results are based on 5000 Monte Carlo replications.*

model the marginal distribution of each series using the t-GARCH(1,1) model and obtain the probability integral transform from the original data. We estimate the dynamic copula model in (15) of order (1,1) and of order (0,0). The latter corresponds to a static dynamic copula model where the correlation parameter p_t is assumed to be constant over time. We then evaluate residual dependence using conditional score residuals.

Table 5 reports the estimation results of the two models. The AIC and BIC indicate that the model of order (1,1) fits the data better than the model of order (0,0). Figure 4 displays the conditional score residuals and their standardized autocorrelaion function with 95%-level confidence intervals. The autocorrelation function of the model of order (0,0) suggests that there is significant autocorrelation in conditional score residuals, indicating that a dynamic specification of the correlation parameter p_t is needed to properly describe the relationship between oil prices and exchange rates. The autocorelation function of the model of order (1,1) suggests that there is no evidence of residual autocorrelation and therefore the bivariate dynamic copula of order (1,1) seems to be able to properly describe the time-variation in the dependence parameter between the two series. These findings are confirmed by the results in Table 6, which reports the results of the test Q_K for different values of K. There is strong evidence of residual autocorrelation for the copula model of order (0,0) while there is none for the copula model of order (1,1).

Table 5: *Maximum likelihood estimates of dynamic copula models of order (0,0) and (1,1). Standard errors are in brackets. The last three columns contain the maximized log-likelihood, the AIC and the BIC.*

	α_0	β_1	α_1	log-lik	AIC	BIC
Copula of order (0,0)	0.321 (0.028)	-	-	63.8	-125.6	-119.1
Copula of order (1,1)	0.060 (0.039)	0.990 (0.003)	0.016 (0.004)	120.2	-234.4	-214.8

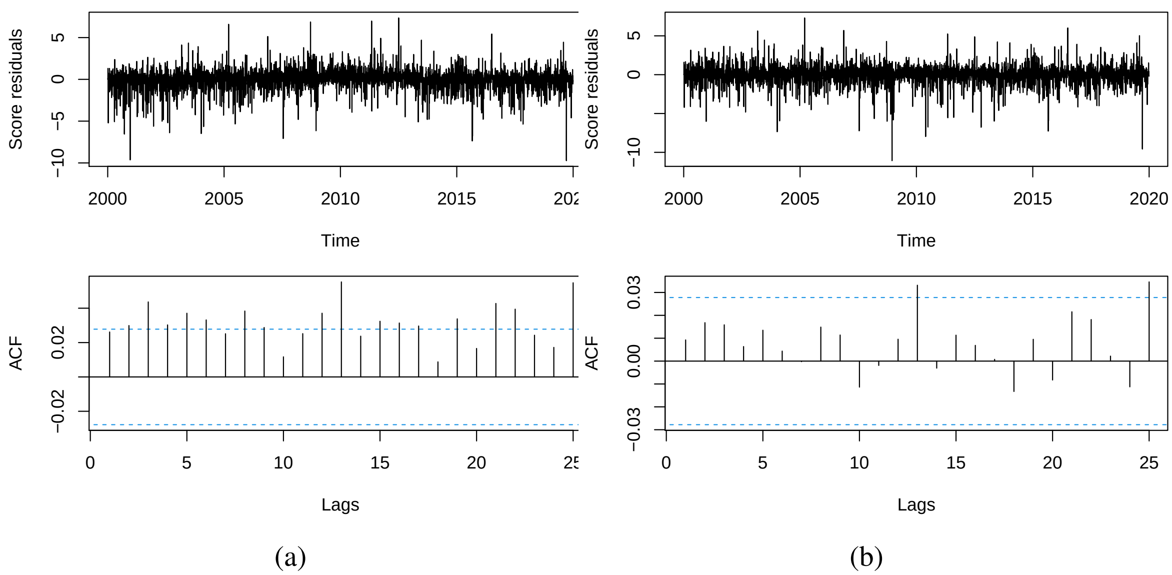


Figure 4: *Panel (a) displays the conditional score residuals of the dynamic copula model of order (0,0) and their standardized empirical autocorrelation function with 95%-level confidence intervals. Panel (b) displays the conditional score residuals of the dynamic copula model of order (1,1) and their standardized empirical autocorrelation function with 95%-level confidence intervals.*

6 Conclusion

We have introduced a general framework for the diagnostic analysis of parametric time series models based on conditional score residuals. The asymptotic properties of empirical autocorrelations of conditional score residuals are derived under general conditions and a consistent and positive definite estimator of the asymptotic covariance matrix is proposed.

Table 6: *Test statistic Q_K and corresponding p-values for conditional score residuals of dynamic coupla models of order (0,0) and (1,1).*

		$K = 5$	$K = 10$	$K = 15$	$K = 20$
Copula of order (0,0)	Q_K	49.47	32.95	82.70	99.41
	p-value	< 0.001	< 0.001	< 0.001	< 0.001
Copula of order (1,1)	Q_K	3.96	6.53	13.18	15.48
	p-value	0.555	0.769	0.588	0.748

These results provide a unified theory for testing residual autocorrelation in time series models. The practical relevance of the proposed framework is discussed through several examples and illustrated in two empirical applications and Monte Carlo studies. Future research may concern the use of the conditional score residuals for Student's t GARCH models as quasi residuals. Conditional score residuals are robust to extreme observations and they have clear benefits compared to squared residuals as illustrated in the paper. Therefore, they may also be interpreted as quasi residuals and employed in semiparametric model settings where the conditional variance is estimated via quasi ML.

A Appendix

A.1 Proofs of the results

Proof of Theorem 3.1. Assumptions 3.1 and 3.2 imply that $\boldsymbol{\rho}_T(\boldsymbol{\theta})$ is continuously differentiable in Θ. An application of the mean value theorem about $\boldsymbol{\theta}_0$ yields

$$\sqrt{T}\hat{\boldsymbol{\rho}}_T = \sqrt{T}\boldsymbol{\rho}_T + \sqrt{T}\frac{\partial \boldsymbol{\rho}_T(\tilde{\boldsymbol{\theta}}_T)}{\partial \boldsymbol{\theta}^\top}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0), \tag{16}$$

where $\tilde{\boldsymbol{\theta}}_T$ is a point between $\boldsymbol{\theta}_0$ and $\hat{\boldsymbol{\theta}}_T$. Next, we notice that $\{\partial l_t / \partial \boldsymbol{\theta}\}$ is a stationary and ergodic martingale difference sequence by Assumptions 3.1 and 3.2, and $\mathbb{E}\|\frac{\partial l_t}{\partial \boldsymbol{\theta}}\|^2 < \infty$ by Assumption 3.4. An application of the central limit theorem for stationary and ergodic martingale difference sequences, as in Billingsley (2013), yields

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\partial l_t}{\partial \boldsymbol{\theta}} \xrightarrow{d} N(\mathbf{0}_n, \boldsymbol{\Sigma}^{-1}).$$

This, together with (8) in Assumption 3.4, immediately implies that $\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) \xrightarrow{d} N(\mathbf{0}_n, \boldsymbol{\Sigma})$. As a result, $\tilde{\boldsymbol{\theta}}_T \xrightarrow{p} \boldsymbol{\theta}_0$ since $\hat{\boldsymbol{\theta}}_T \xrightarrow{p} \boldsymbol{\theta}_0$ and hence we obtain that $\frac{\partial \boldsymbol{\rho}_T(\tilde{\boldsymbol{\theta}}_T)}{\partial \boldsymbol{\theta}^\top} \xrightarrow{p} -\boldsymbol{\Phi}$ by an application of Lemma A.1. Therefore, an application Slutsky's theorem together with equation (16) yields

$$\sqrt{T}\hat{\boldsymbol{\rho}}_T = \sqrt{T}\boldsymbol{\rho}_T - \boldsymbol{\Phi}\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) + o_p(1). \tag{17}$$

Furthermore, from the expression of $\boldsymbol{\rho}_T$, we obtain that

$$\sqrt{T}\boldsymbol{\rho}_T = \frac{1}{\sqrt{T}} \sum_{t=k+1}^{T} s_t \boldsymbol{s}_{K,t} + o_p(1) \tag{18}$$

since $\frac{1}{T}\sum_{t=1}^{T} s_t^2 \xrightarrow{p} 1$ and $\frac{1}{\sqrt{T}}\sum_{t=1}^{T} s_t \boldsymbol{s}_{K,t}$ converges in distribution to a normal, as $\{s_t \boldsymbol{s}_{K,t}\}$ is a stationary and ergodic martingale difference sequence with a finite second moment by Assumptions 3.1 and 3.2. Therefore, combining (17) with (18) and (8), we obtain that

$$\sqrt{T}\hat{\boldsymbol{\rho}}_T = \frac{1}{\sqrt{T}} \sum_{t=k+1}^{T} \boldsymbol{g}_t + o_p(1), \tag{19}$$

where

$$\boldsymbol{g}_t = s_t \boldsymbol{s}_{K,t} - \boldsymbol{\Phi}\boldsymbol{\Sigma}\frac{\partial l_t}{\partial \boldsymbol{\theta}}.$$

Next, we notice that $\{\boldsymbol{g}_t\}$ is a stationary and ergodic martingale difference sequence with a finite second moment by Assumptions 3.1, 3.2 and 3.4. An application of the central limit theorem yields

$$\frac{1}{\sqrt{T}} \sum_{t=k+1}^{T} \boldsymbol{g}_t \xrightarrow{d} N\left(\mathbf{0}_K, \boldsymbol{V}\right), \tag{20}$$

where $\boldsymbol{V} = \mathbb{E}(\boldsymbol{g}_t \boldsymbol{g}_t^\top)$. The matrix $\boldsymbol{V}$ is positive definite by Lemma A.2 as it ensures that the elements of the vector $\boldsymbol{g}_t$ are linearly independent random variables. Finally, we show that $\boldsymbol{V} = \boldsymbol{I}_K - \boldsymbol{\Phi}\boldsymbol{\Sigma}\boldsymbol{\Phi}^\top$. The matrix $\boldsymbol{V}$ can be expressed as

$$\boldsymbol{V} = \boldsymbol{I}_K + (\boldsymbol{\Gamma} - \boldsymbol{\Phi})\boldsymbol{\Sigma}(\boldsymbol{\Gamma}^\top - \boldsymbol{\Phi}^\top) - \boldsymbol{\Gamma}\boldsymbol{\Sigma}\boldsymbol{\Gamma}^\top,$$

where

$$\boldsymbol{\Gamma} = \mathbb{E}\left(s_t \boldsymbol{s}_{K,t} \frac{\partial l_t}{\partial \boldsymbol{\theta}^\top}\right),$$

since $\mathbb{E}(s_t^2 \boldsymbol{s}_{K,t} \boldsymbol{s}_{K,t}^\top) = \boldsymbol{I}_K$ and $\mathbb{E}\left(\frac{\partial l_t}{\partial \boldsymbol{\theta}} \frac{\partial l_t}{\partial \boldsymbol{\theta}^\top}\right) = \boldsymbol{\Sigma}^{-1}$. Lemma A.3 ensures that $\boldsymbol{\Gamma} = \boldsymbol{\Phi}$. Therefore, we immediately obtain $\boldsymbol{V} = \boldsymbol{I}_K - \boldsymbol{\Phi}\boldsymbol{\Sigma}\boldsymbol{\Phi}^\top$. The final result $\sqrt{T}\hat{\boldsymbol{\rho}}_T \xrightarrow{d} N\left(\mathbf{0}_K, \boldsymbol{V}\right)$

follows from the convergence results in equations (19) and (20) by an application of Slutsky's theorem. □

Proof of Theorem 3.2. The consistency result $\hat{\boldsymbol{V}} \xrightarrow{p} \boldsymbol{V}$ follows immediately by showing that

(a) $\frac{1}{T}\sum_{t=1}^{T} \hat{\boldsymbol{s}}_{K,t}\hat{\boldsymbol{s}}_{K,t}^{\top} \xrightarrow{p} \boldsymbol{I}_K$;

(b) $\frac{1}{T}\sum_{t=1}^{T} (\hat{\boldsymbol{n}}_t\hat{\boldsymbol{n}}_t^{\top} + \hat{\boldsymbol{H}}_t) \xrightarrow{p} \boldsymbol{\Sigma}^{-1}$;

(c) $\frac{1}{T}\sum_{t=1}^{T} \hat{\boldsymbol{s}}_{K,t}\hat{\boldsymbol{n}}_t^{\top} \xrightarrow{p} \boldsymbol{\Phi}$.

(a) First, we obtain that $\frac{1}{T}\sum_{t=1}^{T} \boldsymbol{s}_{K,t}(\boldsymbol{\theta})\boldsymbol{s}_{K,t}(\boldsymbol{\theta})^{\top}$ converges almost surely and uniformly in Θ to the limit $\mathbb{E}(\boldsymbol{s}_{K,t}(\boldsymbol{\theta})\boldsymbol{s}_{K,t}(\boldsymbol{\theta})^{\top})$. This result follows by an application of the ergodic theorem of Rao (1962) for stationary and ergodic sequences of continuous functions on a compact domain with a finite uniform moment. In particular, we have that $\{\boldsymbol{s}_{K,t}(\boldsymbol{\theta})\boldsymbol{s}_{K,t}(\boldsymbol{\theta})^{\top}\}$ is a stationary and ergodic sequence of continuous functions, Θ is a compact set and $\mathbb{E}\|\boldsymbol{s}_{K,t}\|_{\Theta}^{2} < \infty$ holds by Assumption 3.3. Therefore, the conditions of the ergodic theorem of Rao (1962) are satisfied. Finally, given the uniform convergence result and the consistency of the ML estimator $\hat{\boldsymbol{\theta}}_T$, an application of the continuous mapping theorem yields

$$\frac{1}{T}\sum_{t=1}^{T} \hat{\boldsymbol{s}}_{K,t}\hat{\boldsymbol{s}}_{K,t}^{\top} \xrightarrow{p} \mathbb{E}(\boldsymbol{s}_{K,t}\boldsymbol{s}_{K,t}^{\top}) = \boldsymbol{I}_K.$$

(b) As discussed in (a), we obtain that $\frac{1}{T}\sum_{t=1}^{T} \boldsymbol{n}_t(\boldsymbol{\theta})\boldsymbol{n}_t(\boldsymbol{\theta})^{\top} + \boldsymbol{H}_t(\boldsymbol{\theta})$ converges almost surely and uniformly in Θ to $\mathbb{E}(\boldsymbol{n}_t(\boldsymbol{\theta})\boldsymbol{n}_t(\boldsymbol{\theta})^{\top} + \boldsymbol{H}_t(\boldsymbol{\theta}))$ by the ergodic theorem of Rao (1962) since $\{\boldsymbol{n}_t(\boldsymbol{\theta})\boldsymbol{n}_t(\boldsymbol{\theta})^{\top} + \boldsymbol{H}_t(\boldsymbol{\theta})\}$ is a stationary and ergodic sequence of continuous functions, and $\mathbb{E}\|\boldsymbol{n}_t\|_{\Theta}^{2} < \infty$ and $\mathbb{E}\|\boldsymbol{H}_t\|_{\Theta} < \infty$ hold by Assumption 3.6. Therefore, from the consistency of the ML estimator, together with an application of the continuous mapping theorem, we obtain that

$$\frac{1}{T}\sum_{t=1}^{T} (\hat{\boldsymbol{n}}_t\hat{\boldsymbol{n}}_t^{\top} + \hat{\boldsymbol{H}}_t) \xrightarrow{p} \mathbb{E}(\boldsymbol{n}_t\boldsymbol{n}_t^{\top} + \boldsymbol{H}_t)$$

Finally, we notice that

$$\mathbb{E}\left(\frac{\partial l_t}{\partial \boldsymbol{\theta}}\frac{\partial l_t}{\partial \boldsymbol{\theta}^{\top}}\middle|\mathcal{F}_{t-1}\right) = \boldsymbol{n}_t\boldsymbol{n}_t^{\top} + \boldsymbol{H}_t,$$

and therefore $\mathbb{E}(\boldsymbol{n}_t\boldsymbol{n}_t^{\top} + \boldsymbol{H}_t) = \boldsymbol{\Sigma}^{-1}$ follows by the law of total expectation.

(c) Once again, we obtain that $\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{s}_{K,t}(\boldsymbol{\theta})\boldsymbol{n}_t(\boldsymbol{\theta})^\top$ converges almost surely and uniformly in Θ to $\mathbb{E}(\boldsymbol{s}_{K,t}(\boldsymbol{\theta})\boldsymbol{n}_t(\boldsymbol{\theta})^\top)$ by an application of the ergodic theorem of Rao (1962). The conditions of the theorem are satisfied since $\{\boldsymbol{s}_{K,t}(\boldsymbol{\theta})\boldsymbol{n}_t(\boldsymbol{\theta})^\top\}$ is a stationary and ergodic sequence of continuous functions with a finite uniform moment. In particular, $\mathbb{E}(\|\boldsymbol{s}_{K,t}\|_\Theta\|\boldsymbol{n}_t\|_\Theta) < \infty$ follows by the Cauchy Schwarz inequality together with Assumptions 3.3 and 3.6, which ensure that $\|\boldsymbol{s}_{K,t}\|_\Theta$ and $\|\boldsymbol{n}_t\|_\Theta$ have a finite second moment. Finally, the consistency of the ML estimator and an application of the continuous mapping theorem yield

$$\frac{1}{T}\sum_{t=1}^{T}\hat{\boldsymbol{s}}_{K,t}\hat{\boldsymbol{n}}_t^\top \xrightarrow{p} \mathbb{E}(\boldsymbol{s}_{K,t}\boldsymbol{n}_t^\top) = \boldsymbol{\Phi}.$$

□

A.2 Lemmas

Lemma A.1. *Assume that $\tilde{\boldsymbol{\theta}}_T = \boldsymbol{\theta}_0 + o_p(1)$ and let the assumptions of Theorem 3.1 hold. Then,*

$$\frac{\partial \boldsymbol{\rho}_T(\tilde{\boldsymbol{\theta}}_T)}{\partial \boldsymbol{\theta}^\top} \xrightarrow{p} -\boldsymbol{\Phi}, \qquad T \to \infty.$$

Proof of Lemma A.1. First, we note that

$$\frac{\partial \boldsymbol{\rho}_T(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top} = \begin{bmatrix} \frac{\partial \rho_{1,T}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top} \\ \vdots \\ \frac{\partial \rho_{K,T}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top} \end{bmatrix}, \quad \text{where}$$

$$\begin{aligned} \frac{\partial \rho_{k,T}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = & \frac{\frac{1}{T}\sum_{t=1}^{T}\frac{\partial s_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}s_{t-k}(\boldsymbol{\theta}) + s_t(\boldsymbol{\theta})\frac{\partial s_{t-k}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}}{\frac{1}{T}\sum_{t=1}^{T}s_t^2(\boldsymbol{\theta})} \\ & - 2\frac{\left(\frac{1}{T}\sum_{t=1}^{T}s_t(\boldsymbol{\theta})s_{t-k}(\boldsymbol{\theta})\right)\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\partial s_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)}{\left(\frac{1}{T}\sum_{t=1}^{T}s_t^2(\boldsymbol{\theta})\right)^2}, \end{aligned} \tag{21}$$

for $k = 1, \ldots, K$. The uniform moment conditions in Assumption 3.3 ensure that each of the averages in (21) converges almost surely and uniformly in Θ to a deterministic continuous function. This follows by an application of the ergodic theorem of Rao (1962) since

$\{s_t(\boldsymbol{\theta})\}$ and $\{\partial s_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\}$ are stationary and ergodic sequences of continuous functions and all the terms of the averages in (21) have finite uniform moments. Therefore, given the continuity of the limit function of each of the sample averages in (21) together with $\tilde{\boldsymbol{\theta}}_T \xrightarrow{p} \boldsymbol{\theta}_0$, we can apply the continuous mapping theorem and obtain

$$\frac{\partial \rho_{k,T}(\tilde{\boldsymbol{\theta}}_T)}{\partial \boldsymbol{\theta}} \xrightarrow{p} \frac{\mathbb{E}\left(\frac{\partial s_t}{\partial \boldsymbol{\theta}} s_{t-k} + s_t \frac{\partial s_{t-k}}{\partial \boldsymbol{\theta}}\right)}{\mathbb{E}(s_t^2)} - 2\frac{\mathbb{E}(s_t s_{t-k})\mathbb{E}\left(\frac{\partial s_t}{\partial \boldsymbol{\theta}}\right)}{\mathbb{E}(s_t^2)^2}.$$

From the above expression, we notice that $\mathbb{E}(s_t^2) = 1$, $\mathbb{E}(s_t s_{t-k}) = 0$ and $\mathbb{E}\left(\frac{\partial s_t}{\partial \boldsymbol{\theta}}\right) < \infty$. Therefore, we have that

$$\frac{\partial \rho_T(\tilde{\boldsymbol{\theta}}_T)}{\partial \boldsymbol{\theta}^\top} \xrightarrow{p} \mathbb{E}\left(\boldsymbol{s}_{K,t}\frac{\partial s_t}{\partial \boldsymbol{\theta}^\top} + s_t \frac{\partial \boldsymbol{s}_{K,t}}{\partial \boldsymbol{\theta}^\top}\right). \tag{22}$$

We conclude the proof by showing that the expectation in (22) is equal to $-\boldsymbol{\Phi}$. Applying basic calculus together with the fact that $\mathbb{E}(s_t|\mathcal{F}_{t-1}) = 0$, we obtain

$$\begin{aligned}
\mathbb{E}\left(\boldsymbol{s}_{K,t}\frac{\partial s_t}{\partial \boldsymbol{\theta}^\top} + s_t \frac{\partial \boldsymbol{s}_{K,t}}{\partial \boldsymbol{\theta}^\top}\right) &= \mathbb{E}\left(\boldsymbol{s}_{K,t}\frac{\partial s_t}{\partial \boldsymbol{\theta}^\top}\right) + \mathbb{E}\left(\mathbb{E}\left(s_t|\mathcal{F}_{t-1}\right)\frac{\partial \boldsymbol{s}_{K,t}}{\partial \boldsymbol{\theta}^\top}\right) \\
&= \mathbb{E}\left(-\boldsymbol{s}_{K,t}\frac{\partial I_t}{\partial \boldsymbol{\theta}^\top}\frac{s_t}{2I_t} + \frac{1}{\sqrt{I_t}}\boldsymbol{s}_{K,t}\frac{\partial u_t}{\partial \boldsymbol{\theta}^\top}\right) \\
&= \mathbb{E}\left(-\frac{\partial I_t}{\partial \boldsymbol{\theta}^\top}\boldsymbol{s}_{K,t}\frac{1}{2I_t}\mathbb{E}(s_t|\mathcal{F}_{t-1})\right) + \mathbb{E}\left(\frac{1}{\sqrt{I_t}}\boldsymbol{s}_{K,t}\frac{\partial u_t}{\partial \boldsymbol{\theta}^\top}\right) \\
&= \mathbb{E}\left(\frac{1}{\sqrt{I_t}}\boldsymbol{s}_{K,t}\frac{\partial u_t}{\partial f_t}\frac{\partial f_t}{\partial \boldsymbol{\theta}^\top}\right) + \mathbb{E}\left(\frac{1}{\sqrt{I_t}}\boldsymbol{s}_{K,t}\boldsymbol{v}_t^\top\right),
\end{aligned} \tag{23}$$

where

$$\boldsymbol{v}_t = \begin{bmatrix} \mathbf{0}_s \\ \boldsymbol{u}_{\boldsymbol{\lambda}f,t} \end{bmatrix}, \qquad \boldsymbol{u}_{\boldsymbol{\lambda}f,t} = \frac{\partial^2 \log p(y_t|f;\boldsymbol{\lambda})}{\partial f \partial \boldsymbol{\lambda}}\Big|_{f=f_t}.$$

The first expectation in (23) is equal to

$$\begin{aligned}
\mathbb{E}\left(\frac{1}{\sqrt{I_t}}\boldsymbol{s}_{K,t}\frac{\partial u_t}{\partial f_t}\frac{\partial f_t}{\partial \boldsymbol{\theta}^\top}\right) &= \mathbb{E}\left(\frac{1}{\sqrt{I_t}}\boldsymbol{s}_{K,t}\frac{\partial f_t}{\partial \boldsymbol{\theta}^\top}\mathbb{E}\left(\frac{\partial u_t}{\partial f_t}\Big|\mathcal{F}_{t-1}\right)\right) \\
&= -\mathbb{E}\left(\sqrt{I_t}\boldsymbol{s}_{K,t}\frac{\partial f_t}{\partial \boldsymbol{\theta}^\top}\right),
\end{aligned}$$

where the first equality follows by the law of total expectation and the second equality follows by (ii) in Assumption 3.1, which entails $\mathbb{E}(\frac{\partial u_t}{\partial f_t}|\mathcal{F}_{t-1}) = -\mathbb{E}(u_t^2|\mathcal{F}_{t-1}) = -I_t$.

Next, we obtain that the second expectation in (23) is equal to

$$\mathbb{E}\left(\frac{1}{\sqrt{I_t}}\boldsymbol{s}_{K,t}\boldsymbol{v}_t^\top\right) = \mathbb{E}\left(\frac{1}{\sqrt{I_t}}\boldsymbol{s}_{K,t}\mathbb{E}\left(\boldsymbol{v}_t^\top \middle| \mathcal{F}_{t-1}\right)\right)$$
$$= -\mathbb{E}\left(\frac{1}{\sqrt{I_t}}\boldsymbol{s}_{K,t}\boldsymbol{k}_t^\top\right),$$

where the first equality follows by the law of total expectation, the second equality follows by (ii) in Assumption 3.1, which implies $\mathbb{E}(\boldsymbol{u}_{\boldsymbol{\lambda}f,t}|\mathcal{F}_{t-1}) = -\mathbb{E}(u_t\boldsymbol{u}_{\boldsymbol{\lambda},t}|\mathcal{F}_{t-1}) = -\boldsymbol{j}_{\boldsymbol{\lambda}f,t}$ and hence $\mathbb{E}(\boldsymbol{v}_t|\mathcal{F}_{t-1}) = -\boldsymbol{k}_t$. Therefore, we conclude that $\frac{\partial \rho_T(\tilde{\boldsymbol{\theta}}_T)}{\partial \boldsymbol{\theta}^\top} \xrightarrow{p} -\boldsymbol{\Phi}$. □

Lemma A.2. *Let the assumptions of Theorem 3.1 hold. Then, the elements of the vector*

$$\boldsymbol{g}_t = s_t\boldsymbol{s}_{K,t} - \boldsymbol{\Phi}\boldsymbol{\Sigma}\frac{\partial l_t}{\partial \boldsymbol{\theta}}$$

are linearly independent random variables.

Proof. First, we note that $\frac{\partial l_t}{\partial \boldsymbol{\theta}}$ can be expressed as

$$\frac{\partial l_t}{\partial \boldsymbol{\theta}} = u_t\begin{bmatrix}\frac{\partial f_t}{\partial \boldsymbol{\xi}}\\ \frac{\partial f_t}{\partial \boldsymbol{\lambda}}\end{bmatrix} + \begin{bmatrix}\mathbf{0}_s\\ \boldsymbol{u}_{\boldsymbol{\lambda},t}\end{bmatrix}. \tag{24}$$

Next, for any $\mathbf{z} \in \mathbb{R}^K$, we obtain

$$\mathbf{z}^\top\boldsymbol{g}_t = u_t\mathbf{z}^\top\left(\frac{1}{\sqrt{I_t}}\boldsymbol{s}_{K,t} - \boldsymbol{A}_1\frac{\partial f_t}{\partial \boldsymbol{\xi}} - \boldsymbol{A}_2\frac{\partial f_t}{\partial \boldsymbol{\lambda}}\right) - \mathbf{z}^\top\boldsymbol{A}_2\boldsymbol{u}_{\boldsymbol{\lambda},t}, \tag{25}$$

where $\boldsymbol{A}_1$ is a $K \times s$ matrix that contains the first s columns of the matrix $\boldsymbol{\Phi}\boldsymbol{\Sigma}$ and $\boldsymbol{A}_2$ is a $K \times r$ matrix that contains the last r columns of the matrix $\boldsymbol{\Phi}\boldsymbol{\Sigma}$, such that $\boldsymbol{\Phi}\boldsymbol{\Sigma} = [\boldsymbol{A}_1, \boldsymbol{A}_2]$. Assumption 3.1 ensures that the Fisher information matrix of the conditional density function $p(y|f;\boldsymbol{\lambda})$ is positive definite for any $f \in \mathrm{F}$. Therefore, the elements of the vector $(u_t, \boldsymbol{u}_{\boldsymbol{\lambda},t}^\top)^\top$ are linearly independent random variables conditional on $\mathcal{F}_{t-1}$. This implies that $\mathbf{z}^\top\boldsymbol{g}_t = 0$ with probability 1 only if $\mathbf{z}^\top\boldsymbol{A}_2 = \mathbf{0}_r$. As a result, from equation (25), we have that $\mathbf{z}^\top\boldsymbol{g}_t = 0$ with probability 1 only if

$$\mathbf{z}^\top\left(\frac{1}{\sqrt{I_t}}\boldsymbol{s}_{K,t} - \boldsymbol{A}_1\frac{\partial f_t}{\partial \boldsymbol{\xi}}\right) = 0 \quad \text{with probability 1.}$$

However, this can be true only if $\mathbf{z} = \mathbf{0}_K$ by Assumption 3.5. This concludes the proof of the lemma. □

Lemma A.3. *Let the assumptions of Theorem 3.1 hold. Then,*

$$\mathbb{E}\left(s_t \boldsymbol{s}_{K,t} \frac{\partial l_t}{\partial \boldsymbol{\theta}^\top}\right) = \boldsymbol{\Phi}.$$

Proof of Lemma A.3. From the expression of $\partial l_t / \partial \boldsymbol{\theta}$ in (24), we obtain

$$\begin{aligned}
\mathbb{E}\left(s_t \boldsymbol{s}_{K,t} \frac{\partial l_t}{\partial \boldsymbol{\theta}^\top}\right) &= \mathbb{E}\left(s_t^2 \sqrt{I_t} \boldsymbol{s}_{K,t} \frac{\partial f_t}{\partial \boldsymbol{\theta}^\top}\right) + \mathbb{E}\left(s_t \boldsymbol{s}_{K,t} \begin{bmatrix}\mathbf{0}_s^\top & \boldsymbol{u}_{\boldsymbol{\lambda},t}^\top\end{bmatrix}\right) \\
&= \mathbb{E}\left(\sqrt{I_t} \boldsymbol{s}_{K,t} \frac{\partial f_t}{\partial \boldsymbol{\theta}^\top} \mathbb{E}(s_t^2 | \mathcal{F}_{t-1})\right) + \mathbb{E}\left(\frac{1}{\sqrt{I_t}} \boldsymbol{s}_{K,t} \begin{bmatrix}\mathbf{0}_s^\top & \mathbb{E}(u_t \boldsymbol{u}_{\boldsymbol{\lambda},t}^\top | \mathcal{F}_{t-1})\end{bmatrix}\right) \\
&= \mathbb{E}\left(\sqrt{I_t} \boldsymbol{s}_{K,t} \frac{\partial f_t}{\partial \boldsymbol{\theta}^\top} + \frac{1}{\sqrt{I_t}} \boldsymbol{s}_{K,t} \boldsymbol{k}_t^\top\right) = \boldsymbol{\Phi},
\end{aligned}$$

where the second equality follows by the law of total expectation since $\boldsymbol{s}_{K,t}$, I_t and $\frac{\partial f_t}{\partial \boldsymbol{\theta}}$ are $\mathcal{F}_{t-1}$-measurable, and the third equality follows since $\mathbb{E}(s_t^2 | \mathcal{F}_{t-1}) = 1$ and $\mathbb{E}(u_t \boldsymbol{u}_{\boldsymbol{\lambda},t} | \mathcal{F}_{t-1}) = \boldsymbol{j}_{\boldsymbol{\lambda} f,t}$. □

References

Berkes, I., Horváth, L., and Kokoszka, P. (2003). Asymptotics for GARCH squared residual correlations. *Econometric Theory*, pages 515–540.

Billingsley, P. (2013). *Convergence of probability measures*. John Wiley & Sons.

Blasques, F., Gorgi, P., Koopman, S. J., and Wintenberger, O. (2018). Feasible invertibility conditions and maximum likelihood estimation for observation-driven models. *Electronic Journal of Statistics*, 12(1):1019–1052.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327.

Box, G. E. P. and Pierce, D. A. (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association*, 65(332):1509–1526.

Calvori, F., Creal, D., Koopman, S. J., and Lucas, A. (2017). Testing for parameter instability across different modeling frameworks. *Journal of Financial Econometrics*, 15(2):223–246.

Creal, D., Koopman, S. J., and Lucas, A. (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics*, 28(5):777–795.

Creal, D., Schwaab, B., Koopman, S. J., and Lucas, A. (2014). Observation-driven mixed-measurement dynamic factor models with an application to credit risk. *Review of Economics and Statistics*, 96(5):898–915.

Davis, R. A., Dunsmuir, W. T. M., and Streett, S. B. (2003). Observation-driven models for Poisson counts. *Biometrika*, 90(4):777–790.

Davis, R. A., Holan, S. H., Lund, R., and Ravishanker, N. (2016). *Handbook of discrete-valued time series*. CRC Press.

Durbin, J. (1970). Testing for serial correlation in least-squares regression when some of the regressors are lagged dependent variables. *Econometrica*, 38(4):410–421.

Durbin, J. and Watson, G. (1950). Testing for serial correlation in least-squares regression, I. *Biometrika*, 37(4):409–428.

Durbin, J. and Watson, G. (1951). Testing for serial correlation in least-squares regression, II. *Biometrika*, 38(2):159–178.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4):987–1007.

Engle, R. F. and Russell, J. R. (1998). Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data. *Econometrica*, 66(5):1127–1162.

Fokianos, K., Rahbek, A., and Tjøstheim, D. (2009). Poisson autoregression. *Journal of the American Statistical Association*, 104(488):1430–1439.

Francq, C., Roy, R., and Zakoïan, J.-M. (2005). Diagnostic checking in ARMA models with uncorrelated errors. *Journal of the American Statistical Association*, 100(470):532–544.

Harvey, A. (2013). *Dynamic Models for Volatility and Heavy Tails: With Applications to Financial and Economic Time Series*. New York: Cambridge University Press.

Harvey, A. and Thiele, S. (2016). Testing against changing correlation. *Journal of Empirical Finance*, 38:575–589.

Horváth, L., Kokoszka, P., and Teyssiere, G. (2001). Empirical process of the squared residuals of an ARCH sequence. *Annals of Statistics*, 29(2):445–469.

Jung, R. C., Kukuk, M., and Liesenfeld, R. (2006). Time series of count data: modeling, estimation and diagnostics. *Computational Statistics & Data Analysis*, 51(4):2350–2364.

Li, W. and Mak, T. (1994). On the squared residual autocorrelations in non-linear time series with conditional heteroskedasticity. *Journal of Time Series Analysis*, 15(6):627–636.

Ling, S. and Tong, H. (2011). Score based goodness-of-fit tests for time series. *Statistica Sinica*, 21:1807–1829.

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

McLeod, A. I. (1978). On the distribution of residual autocorrelations in Box-Jenkins models. *Journal of the Royal Statistical Society: Series B (Methodological)*, 40(3):296–302.

Patton, A. J. (2006). Modelling asymmetric exchange rate dependence. *International Economic Review*, 47(2):527–556.

Patton, A. J. (2012). A review of copula models for economic time series. *Journal of Multivariate Analysis*, 110:4–18.

Rao, R. R. (1962). Relations between weak and uniform convergence of measures with applications. *The Annals of Mathematical Statistics*, 33(2):659–680.

Salvatierra, I. D. L. and Patton, A. J. (2015). Dynamic copula models and high frequency data. *Journal of Empirical Finance*, 30:120–135.

Straumann, D. (2005). Estimation in Conditionally Heteroschedastic Time Series Models. *Springer, New York*, 181.

Straumann, D. and Mikosch, T. (2006). Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: A stochastic recurrence equations approach. *Annals of Statistics*, 34(5):2449–2495.

White, H. (1994). *Estimation, Inference and Specification Analysis*. Econometric Society Monographs. Cambridge University Press.