OXFORD BULLETIN OF ECONOMICS AND STATISTICS, 0305-9049
doi: 10.1111/obes.12647

Testing for Threshold Effects in the Presence of Heteroskedasticity and Measurement Error With an Application to Italian Strikes

FRANCESCO ANGELINI,† MASSIMILIANO CASTELLANI,†
SIMONE GIANNERINI‡ and GRETA GORACCI§

†Department of Statistical Sciences, University of Bologna, Italy
(e-mail: francesco.angelini7@unibo.it; m.castellani@unibo.it)
‡Department of Economics and Statistics, University of Udine, Italy
(e-mail: simone.giannerini@uniud.it)
§Faculty of Economics and Management, Free University of Bozen-Bolzano, Italy
(e-mail: greta.goracci@unibz.it)

Abstract

We address the issue of testing for threshold nonlinearity in the conditional mean in the presence of conditional heteroskedasticity. We propose a supremum Lagrange multiplier approach to test a linear ARMA-GARCH model versus a TARMA-GARCH model. We derive the asymptotic null distribution of the test statistic, and this requires novel results due to nuisance parameters, absent under the null hypothesis, combined with the nonlinear moving average and GARCH-type innovations. We show that tests that do not account for heteroskedasticity fail to achieve the correct size even for large sample sizes. Moreover, the TARMA specification naturally accounts for the ubiquitous presence of measurement error that affects macroeconomic data. We apply the results to analyse the time series of Italian strikes, and we show that the TARMA-GARCH specification is consistent with the relevant macroeconomic theory while capturing the main features of the Italian strikes dynamics, such as asymmetric cycles and regime-switching.

I. Introduction

Strike activity of labour unions is linked to the business cycle and to workers' organizational and political power. Hence, understanding strikes' dynamics can shed light on related social and political issues and support the economic analysis of industrial conflicts in several countries. It is acknowledged that the dynamics of strikes presents complex features, but to the best of our knowledge, both the economic and the econometric literature appear to lack a comprehensive analysis that considers most of these aspects in a unique framework (Paldam and Pedersen, 1982; Hundley and Koreisha, 1987;

JEL Classification numbers: C12, C22, J52.

Franzosi, 1989; Paldam, 2021). The use of strikes as a tool of political and organizational power differs from country to country depending on their specific institutional context (Franzosi, 1989; Castellani, Fanelli, and Savioli, 2013). In some countries, unions use strikes to influence public policies and the reasons behind them are mostly political, while in other unionized economies, such as Italy, strikes are also used as a bargaining tool in the labour market. Due to its historical and institutional background, the Italian strikes time series is well-suited to investigate the dynamics of this economic and political variable (Lange, Irvin, and Tarrow, 1990; Corneo and Lucifora, 1997; Quaranta, 2012).

Strike activity has a very long history in Western countries and presents some stylized facts, that is, they appear cyclically and occur in waves. For example, during the 1970s and 1980s, OECD countries experienced a long strike wave, while a sharp decline in strikes occurred in the last decades (Godard, 2011). During the 1990s, most European countries and the United States witnessed a significant decrease in strike activity, together with a strong shift towards the ''tertiarisation of conflict'' (Bordogna and Cella, 2002). More recently, strikes against governments have been increasingly engaged by unions across Western Europe (Hamann, Johnston, and Kelly, 2013). A traditional explanation connects the time series of strikes with the business cycle dynamics. Rees (1952) argued that strikes should increase during booms and decrease during the down-swing phase of the business cycle. Wage claims are another cause of strikes.[1] However, the empirical evidence for OECD countries seems to indicate that strikes are procyclical and wage increases have a negative effect on strikes.

Qualitative analyses suggest that strikes' dynamics is characterized by volatility with clustering effects, primarily due to certain types of strikes that can attract more workers or have prolonged durations (Vandaele, 2016). Together with the observed asymmetric cyclical behaviour, this hints at the presence of a regime-switching mechanism with conditional heteroskedasticity. We advocate the threshold autoregressive-moving average (TARMA) model with GARCH innovations as a novel and appropriate specification to deal with the aforementioned features. Threshold nonlinearity offers a feasible approximation of general complex dynamics while retaining a good interpretability. Threshold autoregressive models (Tong, 1978; Tong and Lim, 1980) have been widely applied in Economics (Hansen, 2011; Chan, Hansen, and Timmermann, 2017) and Finance (Chen, Liu, and So, 2011). Threshold models are particularly suitable to describe the phenomenon of *regulation* which plays a fundamental role in Finance and Economics. For instance, in financial time series, it is common to observe a ''band of inaction'' random walk regime, where arbitrage does not occur, and other regimes where mean reversion takes place so that the model is globally stationary, see e.g. Chan *et al.* (2020). Moreover, Pesaran and Potter (1997) used the TAR model to show that the US GDP is also subject to floor and ceiling effects. Koop and Potter (1999) estimate TAR models for US unemployment rates and Altissimo and Violante (2001) propose a threshold VAR model to study the joint dynamics of US GDP and unemployment rates. Note that, none of

[1] Hicks (1932) pointed out that industrial conflict is costly for both workers and firms. Thus, if agents were rational and fully informed, there would be no reason to strike. Although employees usually strike to gain higher wages or better working conditions, its efficacy remains controversial. Shalev (1980) surveyed the literature on industrial relations theory.

these studies included a moving-average component in their specifications, most probably due to the lack of developments on nonlinear ARMA models.

TARMA models combine the well-known threshold autoregressive (TAR) model and the threshold moving-average (TMA) model (Ling and Tong, 2005). The incorporation of the moving-average component in a nonlinear parametric framework achieves a great approximating capability with few parameters (Goracci, 2020, 2021). They allow us to interpret phenomena that change qualitatively across regimes and react differently to shocks, which is a key aspect in macroeconomic dynamics, as also pointed out in Gonçalves *et al.* (2021). Moreover, as shown in Chan *et al.* (2024), they naturally account for the presence of measurement error. Despite these advantages, the theoretical development of TARMA models has been stuck for many years due to unsolved theoretical problems, mainly due to their non-Markovian nature. The impasse has been overcome by Chan and Goracci (2019), which solved the long-standing open problem regarding the probabilistic structure of the first-order TARMA model. This paved the way for substantial inferential developments and practical applications, see e.g. Goracci *et al.* (2024).

The main aim of this paper is to introduce a test for threshold ARMA effects with GARCH innovations and ascertain whether the dynamics of the Italian strikes can be adequately modelled using a TARMA-GARCH specification. Existing tests for threshold effects on the conditional mean are negatively affected by the presence of heteroscedasticity. A partial solution to this problem is to use threshold tests sequentially on the residuals of a GARCH model but this is likely to affect non-trivially the overall significance level of the tests. Moreover, the presence of measurement error would require a full TARMA specification but existing tests do not account for this. In particular, Wong and Li (1997) test an AR-ARCH versus a TAR-ARCH specification, while Li and Li (2008) compare an MA-GARCH against a TMA-GARCH. We fill the gap and develop a test comparing the ARMA-GARCH against the TARMA-GARCH specification using a sup-Lagrange Multiplier (supLM) statistic tailored to this scope. One of the main advantages of supLM tests over other tests, such as Wald and Likelihood Ratio tests, is that they only require estimating the model under the null hypothesis, achieving superior numerical stability and performance, especially in small samples. The test is an extension of the one introduced in Goracci *et al.* (2023), where i.i.d. innovations are assumed.

We derive the asymptotic null distribution of the test statistic and this requires novel results since the inherent difficulties of working with nuisance parameters absent under the null hypothesis are amplified by the nonlinear moving average setting combined with GARCH-type innovations. We show that, in the presence of heteroskedasticity, tests that assume i.i.d. innovations can be severely biased, whereas our test can be used successfully in all those cases where the aim is testing for a nonlinear (threshold) effect in the conditional mean but the series also presents (conditional) heteroskedasticity.

We use the novel results for the analysis of strikes' dynamics, by using Italian data covering the period between January 1949 and December 2009. We identify a regime-switching dynamics with GARCH innovations. We discuss the shortcomings of the linear approach and showcase the adequacy of our nonlinear specification, which turns out to be consistent both with observed stylized facts and macroeconomic theories, rooted in the Italian labour market history and on more general social dynamics of labour markets.

The remainder of the paper is organized as follows. In Section II we introduce our sup-Lagrange multiplier test for threshold effects with GARCH innovations. We present the asymptotic derivations of the null distribution in Section II and, in Section III, we assess the performance of the test in finite samples by comparing it with the test for threshold nonlinearity that assumes i.i.d. innovations (Goracci *et al.*, 2023), including also its heteroskedasticity-robust extension. The extended simulation study includes the impact of model misspecification (Section III) and measurement error (Section III). In Section IV we study the dynamics of Italian strikes time series. We apply our test to the monthly series of non-worked hours due to strikes and propose and validate the TARMA-GARCH specification. Finally, Section V concludes the paper. All the proofs are reported in the Appendix. The online Supporting Information contains additional Monte Carlo results and further analyses of the Italian strikes series.

II. Test for threshold effects with GARCH innovations

The supLM test statistic

Assume the time series $\{X_t : t = 0, \pm 1, \pm 2, \dots\}$ to follow the threshold autoregressive moving-average model with GARCH errors, TARMA(p, q)-GARCH(u, v), defined by the system of difference equations:

$$\begin{aligned} X_t &= \phi_0 + \sum_{i=1}^{p} \phi_i X_{t-i} - \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t \\ &\quad + \left(\varphi_0 + \sum_{i=1}^{p} \varphi_i X_{t-i} - \sum_{j=1}^{q} \vartheta_j \varepsilon_{t-j} \right) I(X_{t-d} \le r) \\ \varepsilon_t &= \sqrt{h_t} z_t; \quad h_t = a_0 + \sum_{i=1}^{u} a_i \varepsilon_{t-i}^2 + \sum_{j=1}^{v} b_j h_{t-j}. \end{aligned} \tag{1}$$

The process $\{z_t\}$ is a sequence of i.i.d. random variables with zero mean, unit variance and finite fourth moment. p and q are the autoregressive and moving-average orders, respectively; d is the delay parameter; u and v are the ARCH and GARCH orders, respectively. We assume p, q, d, u, v to be known positive integers. Moreover, $I(\cdot)$ is the indicator function and $r \in R$ is the threshold parameter. For notational convenience, we abbreviate $I(X_t \le r)$ by $I_r(X_t)$. We define the following vectors containing the parameters of Model (1): $\boldsymbol{\phi} = \left(\phi_0, \phi_1, \dots, \phi_p\right)^\intercal$, $\boldsymbol{\varphi} = \left(\varphi_0, \varphi_1, \dots, \varphi_p\right)^\intercal$, $\boldsymbol{\theta} = \left(\theta_1, \dots, \theta_q\right)^\intercal$, $\boldsymbol{\vartheta} = \left(\vartheta_1, \dots, \vartheta_q\right)^\intercal$, $\mathbf{a} = (a_0, a_1, \dots, a_u)^\intercal$, $\mathbf{b} = (b_1, \dots, b_v)^\intercal$, with $\boldsymbol{\phi}, \boldsymbol{\varphi} \subseteq R^{p+1}$; $\boldsymbol{\theta}, \boldsymbol{\vartheta} \subseteq R^q$; $\mathbf{a} \subseteq R^{u+1}$ and $\mathbf{b} \subseteq R^v$. Moreover, let

$$\begin{aligned} &\boldsymbol{\Psi}_1 = (\boldsymbol{\phi}^\intercal, \boldsymbol{\theta}^\intercal)^\intercal, \quad \boldsymbol{\Psi}_2 = (\boldsymbol{\varphi}^\intercal, \boldsymbol{\vartheta}^\intercal)^\intercal, \quad \boldsymbol{\Psi} = (\boldsymbol{\Psi}_1^\intercal, \boldsymbol{\Psi}_2^\intercal)^\intercal, \\ &\boldsymbol{\lambda} = \left(\boldsymbol{\Psi}_1^\intercal, \mathbf{a}^\intercal, \mathbf{b}^\intercal\right)^\intercal, \quad \boldsymbol{\eta} = \left(\boldsymbol{\lambda}^\intercal, \boldsymbol{\Psi}_2^\intercal\right)^\intercal. \end{aligned} \tag{2}$$

$\boldsymbol{\Psi}_2$ and $\boldsymbol{\Psi}_1$ contain the ARMA parameters to be tested and those who are not, respectively; $\boldsymbol{\lambda}$ is the parametric vector for the ARMA-GARCH model whereas $\boldsymbol{\eta}$ is the vector of all the parameters (excluding the threshold r) in Model (1). We use $\boldsymbol{\Psi}$, $\boldsymbol{\phi}$, etc. to refer to unknown parameters, whereas the true parameters are obtained by adding the 0 subscript, i.e.: $\mathbf{\Psi}_0 = (\mathbf{\Psi}_{0,1}^\top, \mathbf{\Psi}_{0,2}^\top)^\top$, $\boldsymbol{\lambda}_0^\top$ and $\boldsymbol{\eta}_0^\top$. Also, we assume $\boldsymbol{\eta}_0$ to be an interior point of the parameter space.

We test whether a TARMA(p, q)-GARCH(u, v) model provides a significantly better fit than the linear ARMA(p, q)-GARCH(u, v) model by developing a Lagrange multiplier test statistics. Letting $\mathbf{0}$ be the vector of all zeroes, the system of hypothesis results:

$$\begin{cases} H_0 & : \mathbf{\Psi}_{0,2} = \mathbf{0}, \\ H_1 & : \mathbf{\Psi}_{0,2} \neq \mathbf{0}. \end{cases} \tag{3}$$

Under H_0 the process follows a linear ARMA(p, q)-GARCH(u, v) model:

$$\begin{aligned} X_t &= \phi_{0,0} + \sum_{i=1}^{p} \phi_{0,i} X_{t-i} - \sum_{j=1}^{q} \theta_{0,j} \varepsilon_{t-j} + \varepsilon_t \\ \varepsilon_t &= \sqrt{h_t} z_t; \quad h_t = a_{0,0} + \sum_{i=1}^{u} a_{0,i} \varepsilon_{t-i}^2 + \sum_{j=1}^{v} b_{0,j} h_{t-j}. \end{aligned} \tag{4}$$

Define the polynomials $\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p$, $\theta(z) = 1 - \theta_1 z - \theta_2 z^2 - \cdots - \theta_q z^q$, $a(z) = 1 - a_1 z - a_2 z^2 - \cdots - a_u z^u$, $b(z) = 1 - b_1 z - b_2 z^2 - \cdots - b_v z^v$, $\varphi(z) = 1 - \varphi_1 z - \varphi_2 z^2 - \cdots - \varphi_p z^p$, $\vartheta(z) = 1 - \vartheta_1 z - \vartheta_2 z^2 - \cdots - \vartheta_q z^q$. We assume the following:

- A.1 $\phi(z) \neq 0$ and $\theta(z) \neq 0$ for all $z \in \mathbb{C}$ such that $|z| \leq 1$ and they do not share common roots. $\varphi(z)$ and $\vartheta(z)$ are also coprime.
- A.2 $a_0 > 0$, $a_i \geq 0$, $i = 1, \ldots, u$, with at least one strictly positive a_i; $b_j \geq 0$, $j = 1, \ldots, v$; $\left| \sum_{i=1}^{u} a_i + \sum_{j=1}^{v} b_j \right| < 1$; $a(z)$ and $b(z)$ are coprime.
- A.3 $\{z_t\}$ is a sequence of i.i.d. random variables with $E[z_t] = 0$, $E[z_t^2] = 1$ and $E[z_t^4] < \infty$. Moreover, z_t has a continuous and positive density function, $f_z(x)$.
- A.4 The sequence $\{\varepsilon_t\}$ is strictly stationary and ergodic with finite fourth moments.
- A.5 The process $\{X_t\}$ is ergodic and invertible under H_0.

These assumptions are common in deriving the asymptotic behaviour of test for threshold nonlinearity. See, inter alia, Li and Li (2008), Goracci *et al.* (2023), Chan (1990). They also imply that the process $\{h_t\}$ is strictly stationary and ergodic with $E[h_t^2] < \infty$ and it is bounded away from zero with probability 1, see Li and Li (2008, 2011) and Li, Li, and Ling (2011) for further details. In particular, Assumptions A.1 and A.2 allow us to estimate and identify uniquely the parameters of the ARMA and GARCH parts, respectively.

Suppose we observe the time series $X_1, \ldots, X_n$. Omitting a negative constant, the quasi-Gaussian log-likelihood conditional on the initial values $X_0, X_{-1}, \ldots$ is:

$$\ell_n(\boldsymbol{\eta}, r) = -\frac{1}{2} \sum_{t=1}^{n} \frac{\varepsilon_t^2(\boldsymbol{\eta}, r)}{h_t(\boldsymbol{\eta}, r)} + \ln(h_t(\boldsymbol{\eta}, r)), \tag{5}$$

$$
\begin{aligned}
\text{where } \varepsilon_t(\boldsymbol{\eta}, r) &= X_t - \left\{ \phi_0 + \sum_{i=1}^{p} \phi_i X_{t-i} - \sum_{j=1}^{q} \theta_j \varepsilon_{t-j}(\boldsymbol{\eta}, r) \right\} \\
&\quad - \left\{ \varphi_0 + \sum_{i=1}^{p} \varphi_i X_{t-i} - \sum_{j=1}^{q} \vartheta_j \varepsilon_{t-j}(\boldsymbol{\eta}, r) \right\} I_r\left(X_{t-d}\right)
\end{aligned} \tag{6}
$$

$$
h_t(\boldsymbol{\eta}, r) = a_0 + \sum_{i=1}^{u} a_i \varepsilon_{t-i}^2(\boldsymbol{\eta}, r) + \sum_{j=1}^{v} b_j h_{t-j}(\boldsymbol{\eta}, r). \tag{7}
$$

Also:

$$
\varepsilon_t(\boldsymbol{\lambda}) = \varepsilon_t(\boldsymbol{\eta}, -\infty) = X_t - \left\{ \phi_0 + \sum_{i=1}^{p} \phi_i X_{t-i} - \sum_{j=1}^{q} \theta_j \varepsilon_{t-j}(\boldsymbol{\lambda}) \right\}; \tag{8}
$$

$$
h_t(\boldsymbol{\lambda}) = h_t(\boldsymbol{\eta}, -\infty) = a_0 + \sum_{i=1}^{u} a_i \varepsilon_{t-i}^2(\boldsymbol{\lambda}) + \sum_{j=1}^{v} b_j h_{t-j}(\boldsymbol{\lambda}), \tag{9}
$$

and, under the null hypothesis, $\varepsilon_t(\boldsymbol{\eta}_0, r) = \varepsilon_t$ and $h_t(\boldsymbol{\eta}_0, r) = h_t$.

The derivation of the Lagrange multipliers test is based upon the first and second partial derivatives of $\ell_n(\boldsymbol{\eta}, r)$ with respect to $\boldsymbol{\Psi}$. Let

$$
\begin{aligned}
\frac{\partial \ell_n(\boldsymbol{\eta}, r)}{\partial \boldsymbol{\Psi}} &= \left(\left(\frac{\partial \ell_n(\boldsymbol{\eta}, r)}{\partial \boldsymbol{\Psi}_1} \right)^{\mathsf{T}}, \left(\frac{\partial \ell_n(\boldsymbol{\eta}, r)}{\partial \boldsymbol{\Psi}_2} \right)^{\mathsf{T}} \right)^{\mathsf{T}} \\
&= \sum_{t=1}^{n} \left\{ -\frac{\varepsilon_t(\boldsymbol{\eta}, r)}{h_t(\boldsymbol{\eta}, r)} \frac{\partial \varepsilon_t(\boldsymbol{\eta}, r)}{\partial \boldsymbol{\Psi}} + \frac{1}{2} \left(\frac{\varepsilon_t^2(\boldsymbol{\eta}, r)}{h_t^2(\boldsymbol{\eta}, r)} - \frac{1}{h_t(\boldsymbol{\eta}, r)} \right) \frac{\partial h_t(\boldsymbol{\eta}, r)}{\partial \boldsymbol{\Psi}} \right\}
\end{aligned}
$$

with $\partial \varepsilon_t(\boldsymbol{\Psi}, r)/\partial \boldsymbol{\Psi}$ (respectively $\partial h_t(\boldsymbol{\Psi}, r)/\partial \boldsymbol{\Psi}$) being the partial derivative of $\varepsilon_t(\boldsymbol{\Psi}, r)$ ($h_t(\boldsymbol{\Psi}, r)$) with respect to $\boldsymbol{\Psi}$:

$$
\begin{aligned}
&\frac{\partial \varepsilon_t(\boldsymbol{\eta}, r)}{\partial \boldsymbol{\Psi}} = D_t + \sum_{j=1}^{q} \theta_j \frac{\partial \varepsilon_{t-j}(\boldsymbol{\eta}, r)}{\partial \boldsymbol{\Psi}}, \\
&D_t = \big(-1, -X_{t-1}, \ldots, -X_{t-p}, \varepsilon_{t-1}, \ldots, \varepsilon_{t-q}, \\
&-I_r(X_{t-d}), -X_{t-1} I_r(X_{t-d}), \ldots, -X_{t-p} I_r(X_{t-d}), \varepsilon_{t-1} I_r(X_{t-d}), \ldots, \varepsilon_{t-q} I_r(X_{t-d})\big)^{\mathsf{T}} \\
&\frac{\partial h_t(\boldsymbol{\eta}, r)}{\partial \boldsymbol{\Psi}} = 2 \sum_{i=1}^{u} a_i \varepsilon_{t-i} \frac{\partial \varepsilon_{t-i}(\boldsymbol{\eta}, r)}{\partial \boldsymbol{\Psi}} + \sum_{j=1}^{v} b_j \frac{\partial h_{t-j}(\boldsymbol{\eta}, r)}{\partial \boldsymbol{\Psi}}.
\end{aligned}
$$

Lastly, define the block matrix of second derivatives $\mathcal{I}_n(\boldsymbol{\eta}, r)$:

$$
\mathcal{I}_n(\boldsymbol{\eta}, r) = \begin{pmatrix} \mathcal{I}_{n,11}(\boldsymbol{\eta}) & \mathcal{I}_{n,12}(\boldsymbol{\eta}, r) \\ \mathcal{I}_{n,21}(\boldsymbol{\eta}, r) & \mathcal{I}_{n,22}(\boldsymbol{\eta}, r) \end{pmatrix} = \begin{pmatrix} -\frac{\partial^2 \ell_n(\boldsymbol{\eta}, r)}{\partial \boldsymbol{\Psi}_1 \partial \boldsymbol{\Psi}_1^{\mathsf{T}}} & -\frac{\partial^2 \ell_n(\boldsymbol{\eta}, r)}{\partial \boldsymbol{\Psi}_1 \partial \boldsymbol{\Psi}_2^{\mathsf{T}}} \\ -\frac{\partial^2 \ell_n(\boldsymbol{\eta}, r)}{\partial \boldsymbol{\Psi}_2 \partial \boldsymbol{\Psi}_1^{\mathsf{T}}} & -\frac{\partial^2 \ell_n(\boldsymbol{\eta}, r)}{\partial \boldsymbol{\Psi}_2 \partial \boldsymbol{\Psi}_2^{\mathsf{T}}} \end{pmatrix}. \tag{10}
$$

The Lagrange multiplier approach requires estimating the model under the null hypothesis. Let $\hat{\lambda} = (\hat{\phi}^\top, \hat{\theta}^\top, \hat{\mathbf{a}}^\top, \hat{\mathbf{b}}^\top)^\top = \arg\min_{\lambda} \ell_n(\lambda)$, with $\ell_n(\lambda) = \ell_n(\eta, -\infty)$. Hence, $\hat{\lambda}$ is the Maximum Likelihood Estimator (hereafter MLE) of the ARMA-GARCH coefficients in equation (4) and we define $\hat{\eta} = (\hat{\lambda}^\top, \mathbf{0}^\top)^\top$ to be the so-called *restricted* MLE, i.e., under the null hypothesis. We write $\partial \hat{\ell}_n(r)/\partial \mathbf{\Psi}_2$ and $\hat{\mathcal{I}}_n(r)$ to refer to $\partial \ell_n(\eta, r)/\partial \mathbf{\Psi}_2$ and $\mathcal{I}_n(\eta, r)$ evaluated at the restricted MLE $\hat{\eta}$, i.e.:

$$\frac{\partial \hat{\ell}_n(r)}{\partial \mathbf{\Psi}_2} = \frac{\partial \ell_n(\hat{\eta}, r)}{\partial \mathbf{\Psi}_2}; \qquad \hat{\mathcal{I}}_n(r) = \mathcal{I}_n(\hat{\eta}, r) = \begin{pmatrix} \hat{\mathcal{I}}_{n,11} & \hat{\mathcal{I}}_{n,12}(r) \\ \hat{\mathcal{I}}_{n,21}(r) & \hat{\mathcal{I}}_{n,22}(r) \end{pmatrix}.$$

Under the null hypothesis, the threshold parameter r is absent, thereby the standard asymptotic theory is not applicable. To cope with this issue, we firstly develop the Lagrange multiplier test statistic as a function of r ranging in a data-driven set $\mathcal{R} = [r_L, r_U]$, being r_L and r_U some percentiles of the data. Then, we take the overall test statistic as the supremum on $\mathcal{R}$. This approach has become widely used in the literature of tests involving nuisance parameters. This was first proposed in the seminal work of Chan (1990), and followed by Andrews (1993) and Hansen (1996). Within the threshold setting, the idea was deployed in Wong and Li (1997), Ling and Tong (2005), Li and Li (2008), Li *et al.* (2011), Chan *et al.* (2020), Goracci (2021). Recently, Giannerini, Goracci, and Rahbek (2024) adapted it to prove the validity of a bootstrap scheme in testing threshold nonlinearity.

The test statistic is

$$T_n = \sup_{r \in [r_L, r_U]} T_n(r), \text{ where} \tag{11}$$

$$T_n(r) = \left(\frac{\partial \hat{\ell}_n(r)}{\partial \mathbf{\Psi}_2}\right)^\top \left(\hat{\mathcal{I}}_{n,22}(r) - \hat{\mathcal{I}}_{n,21}(r)\hat{\mathcal{I}}_{n,11}^{-1}\hat{\mathcal{I}}_{n,12}(r)\right)^{-1} \frac{\partial \hat{\ell}_n(r)}{\partial \mathbf{\Psi}_2}. \tag{12}$$

Note that, in equation (11), besides taking the supremum, other functions can be used to derive an overall test statistic, see Tong (1990) and Hansen (1996) for a discussion and Kapetanios and Shin (2006) for an example where other functions are used.

The null distribution

In this section we derive the asymptotic distribution of T_n under the null hypothesis that $\{X_t\}$ follows the ARMA(p, q)-GARCH(u, v) process defined in equation (4). Hereafter, all the expectations are taken under the true probability distribution for which H_0 holds. We use $\|\cdot\|$ to refer to the $\mathcal{L}^2$ matrix norm (the Frobenius' norm, i.e. $\|A\| = \sqrt{\sum_{i=1}^n \sum_{j=1}^m |a_{ij}|^2}$, where A is a $n \times m$ matrix). Also, $o_p(1)$ indicates the convergence in probability to zero as n increases. $\mathcal{D}_R(a, b)$, $a < b$, is the space of functions from (a, b) to R that are right continuous with left-hand limits. We assume $\mathcal{D}_R(a, b)$ to be equipped with the topology of uniform convergence on compact sets. Also, define $\mathcal{D}^k(a, b) = \mathcal{D}_R(a, b) \times \cdots \times \mathcal{D}_R(a, b)$ (k times), equipped with the corresponding product Skorohod topology. Weak convergence on $\mathcal{D}^k(-\infty, +\infty)$ is defined as that on $\mathcal{D}^k(a, b)$ for each $a, b \in R$ as n diverges and is denoted by $\Rightarrow$, see e.g. Billingsley (1968) for further details.

In order to obtain its asymptotic distribution in the main theorem, we derive, under the null hypothesis, an (asymptotic) *uniform* approximation of the test statistic depending on the true parameters $\eta_0 = (\boldsymbol{\lambda}_0^\intercal, \mathbf{0}^\intercal)^\intercal$. In this respect, define the vector $\nabla_n(r) = \left(\nabla_{n,1}^\intercal, \nabla_{n,2}^\intercal(r)\right)^\intercal$, with

$$\nabla_{n,1} = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left\{-\frac{\varepsilon_t}{h_t}\frac{\partial\varepsilon_t(\eta_0,r)}{\partial\boldsymbol{\Psi}_1} + \frac{1}{2}\left(\frac{\varepsilon_t^2}{h_t^2} - \frac{1}{h_t}\right)\frac{\partial h_t(\eta_0,r)}{\partial\boldsymbol{\Psi}_1}\right\},$$

$$\nabla_{n,2}(r) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left\{-\frac{\varepsilon_t}{h_t}\frac{\partial\varepsilon_t(\eta_0,r)}{\partial\boldsymbol{\Psi}_2} + \frac{1}{2}\left(\frac{\varepsilon_t^2}{h_t^2} - \frac{1}{h_t}\right)\frac{\partial h_t(\eta_0,r)}{\partial\boldsymbol{\Psi}_2}\right\}$$

and the matrix

$$\begin{aligned}\Lambda(r) &= \begin{pmatrix}\Lambda_{11} & \Lambda_{12}(r)\\ \Lambda_{21}(r) & \Lambda_{22}(r)\end{pmatrix}\\ &= E\left[\frac{1}{h_t}\left(\frac{\partial\varepsilon_t(\eta_0,r)}{\partial\boldsymbol{\Psi}}\right)\left(\frac{\partial\varepsilon_t(\eta_0,r)}{\partial\boldsymbol{\Psi}}\right)^\intercal + \frac{1}{2h_t^2}\left(\frac{\partial h_t(\eta_0,r)}{\partial\boldsymbol{\Psi}}\right)\left(\frac{\partial h_t(\eta_0,r)}{\partial\boldsymbol{\Psi}}\right)^\intercal\right].\end{aligned}$$

In the following Lemma, under the null hypothesis and Assumption A, we provide a uniform approximation that will allow us to derive the asymptotic null distribution of the test statistic T_n in the main theorem.

Lemma 1. Under Assumptions A.1–A.5 and under H_0, the following quantities are $o_p(1)$:

$$(i)\ \sup_{r\in[r_L,r_U]}\left\|\left(\frac{\hat{\mathcal{I}}_{n,22}(r)}{n} - \frac{\hat{\mathcal{I}}_{n,21}(r)}{n}\left(\frac{\hat{\mathcal{I}}_{n,11}}{n}\right)^{-1}\frac{\hat{\mathcal{I}}_{n,12}(r)}{n}\right)^{-1} - \left(\Lambda_{22}(r) - \Lambda_{21}(r)\Lambda_{11}^{-1}\Lambda_{12}(r)\right)^{-1}\right\|,$$

$$(ii)\ \sup_{r\in[r_L,r_U]}\left\|\frac{1}{\sqrt{n}}\frac{\partial\hat{\ell}_n(r)}{\partial\boldsymbol{\Psi}_2} - \left(\nabla_{n,2}(r) - \Lambda_{21}(r)\Lambda_{11}^{-1}\nabla_{n,1}\right)\right\|.$$

Define the process: $\{Q(r), r\in R\}$, with $Q(r) = \left(\nabla_{n,2}(r) - \Lambda_{21}(r)\Lambda_{11}^{-1}\nabla_{n,1}\right)$. Note that $\{Q(r)\}$ is a marked empirical process with infinitely many markers. In the next theorem we derive a novel Functional Central Limit Theorem (hereafter FCLT) for $\{Q(r)\}$.

Theorem 2. Let $\{\xi(r),\ r\in R\}$ be a centred Gaussian vector process of dimension $(p+q+1)$, with covariance kernel $\Sigma(r,s) = \Lambda_{22}(r\wedge s) - \Lambda_{21}(r)\Lambda_{11}^{-1}\Lambda_{12}(s)$, where $r\wedge s$ indicate min (r,s). Under Assumptions A.1–A.5 and H_0, it holds that

$$\begin{aligned}&(i)\ Q(r) \Rightarrow \xi(r) \text{ in } \mathcal{D}^{p+q+1}(-\infty,+\infty).\\ &(ii)\ \text{The Lagrange multiplier test statistic } T_n \text{ has the same distribution of}\\ &\qquad \sup_{r\in[r_L,r_U]}\xi(r)^\intercal\Sigma(r,r)^{-1}\xi(r).\end{aligned} \tag{13}$$

As also shown in Goracci *et al.* (2023), the asymptotic null distribution of supLM tests is pivotal and coincides with that of Andrews (1993) and Chan (1991). The limiting process in equation (13) is a standardized quadratic functional of a Gaussian process whose covariance kernel depends upon (all) the estimated parameters. The main difference with respect to the test of Goracci *et al.* (2023) is that the quasi-Gaussian log-likelihood contains extra terms depending upon the GARCH specification. These enter both the score vector and the Fisher information matrix and this brings in technical difficulties, especially proving the tightness of the score vector, the key step of proving the FCLT. Note that, the null distribution depends only upon the number of tested parameters, and further estimated parameters, such as those of the GARCH part, do not affect the number of degrees of freedom but only enter the standardization term.

III. Finite sample performance

In this section, we study the finite sample performance of our test (Eq. 11), which we denote by sLMg, and compare it with the sLM test of Goracci *et al.* (2023), which assumes i.i.d. innovations, we call it sLMi. We also include its heteroskedasticity-robust version (see e.g., Hansen, 1996), which we denote with sLMh. It takes the following form:

$$\begin{aligned} \text{sLMh} &= \sup_{r\in[r_L,r_U]} [\Xi_n(r)]^{\intercal} \, [V_n(r)]^{-1} \, \Xi_n(r), \quad \text{where} \\ \Xi_n(r) &= \left(\frac{\partial \hat{\ell}_n(r)}{\partial \boldsymbol{\Psi}_2} - \hat{\mathcal{I}}_{n,21}(r)\hat{\mathcal{I}}_{n,11}^{-1} \frac{\partial \hat{\ell}_n}{\partial \boldsymbol{\Psi}_1} \right) \quad \text{and} \quad V_n(r) = \Xi_n(r) \, [\Xi_n(r)]^{\intercal}. \end{aligned} \tag{14}$$

The length of the series is $n = 100, 200, 500$ and z_t, $t = 1, \ldots, n$, is generated from a standard Gaussian white noise. The nominal size of the tests is $\alpha = 5\%$ and the number of Monte Carlo replications is 10000. Furthermore, we use the tabulated critical values of Andrews (2003) and the threshold is searched from percentile 25th to 75th of the sample. Note that the tabulated critical values of Andrews are derived under symmetric threshold ranges, i.e. $r_L = \pi_0$ and $r_U = 1 - \pi_0$.

Size

We study the empirical size by simulating from the following ARMA(1, 1)-GARCH(1, 1) data generating process (DGP):

$$\begin{aligned} X_t &= \phi_1 X_{t-1} + \theta_1 \varepsilon_{t-1} + \varepsilon_t. \\ \varepsilon_t &= \sqrt{h_t} z_t; \quad h_t = a_0 + a_1 \varepsilon_{t-1}^2 + b_1 h_{t-1}. \end{aligned} \tag{15}$$

where $\phi_1 = (-0.9, -0.6, -0.3, 0.0, 0.3, 0.6, 0.9)$ and $\theta_1 = (-0.8, -0.4, 0.0, 0.4, 0.8)$. The associated coefficients of the Wold representation are reported in Table S1 of the Supporting Information. We combine these with the following parameters for the GARCH specification: $(a_0, a_1, b_1) = (1, 0.04, 0.95)$ (case A), $(1, 0.3, 0.0)$ (case B), $(1, 0.4, 0.4)$ (case C), so as to obtain 35 different parameter settings for each case. Notice that case B corresponds to an ARCH(1) process. Also, all three cases fulfil the condition of finite fourth moments. The results are presented in Figure 1, where the boxplots group together the 35 configurations for each of the three cases. The full results are reported in Tables S2 to S4 of the Supporting Information. Clearly, the standard sLMi test is oversized and,

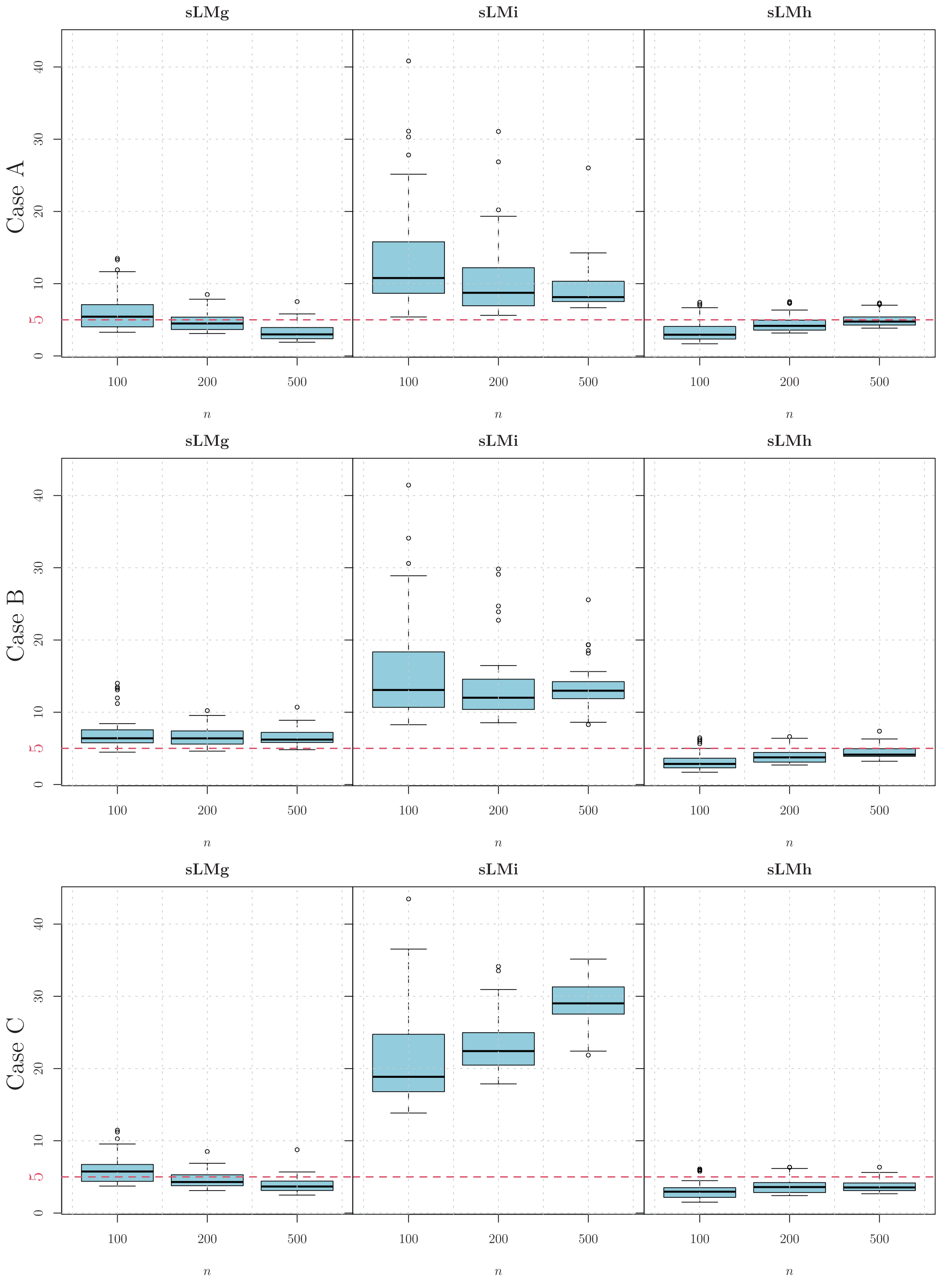

Figure 1. Empirical size (percent) of the sLMg, sLMi and sLMh tests, at nominal level $\alpha = 5\%$ for the ARMA(1,1)-GARCH(1,1) process of equation (15). Cases A,B,C [Colour figure can be viewed at *wileyonlinelibrary.com*]

especially for case C, the bias increases with the sample size and can be severe. On the contrary, both the TARMA-GARCH sLMg and the sLMh tests have correct size in almost every setting. Note that this holds also for corner cases, such as near integrated DGPs or when near cancellation of the AR and MA polynomials occurs. For smaller sample sizes ($n = 100$) the sLMh test is slightly undersized and is more conservative than the sLMg test. As the sample size increases, the empirical size of the sLMg test remains slightly above the nominal 5% for case B, whereas it seems to settle around lower values for cases A and C. As we will show in the next section, this slight undersize does not impinge negatively upon the power and results in a conservative test, which is generally appealing in practical applications. The size for the three cases grouped together is reported in Figure S1 of the Supporting Information. Finally, the size of the sLMg and sLMh tests is not affected if the GARCH part has infinite fourth moments. This is shown in Section S1.2 of the Supporting Information and implies a good practical applicability of the two tests.

Size in presence of misspecification

In this section, we assess the impact upon the size of the tests of different levels of misspecification in the order of the tested model. We assume that the DGP is the following ARMA(2,1) – GARCH(1,1)

$$\begin{aligned} X_t &= \phi_1 X_{t-1} + \phi_2 X_{t-2} + \theta_1 \varepsilon_{t-1} + \varepsilon_t. \\ \varepsilon_t &= \sqrt{h_t} z_t \\ h_t &= a_0 + a_1 \varepsilon_{t-1}^2 + b_1 h_{t-1}. \end{aligned} \tag{16}$$

where $\theta_1 = (-0.8, -0.4, 0.0, 0.4, 0.8)$ and the AR and GARCH parameters are in Table 1. Overall, we obtain 60 different parameters' combinations. To each of these, we apply the following 8 tests that entail different levels of misspecification:

1. sLMg: ARMA(2,1) – GARCH(1,1) (no misspecification)
2. sLMg: ARMA(1,1) – GARCH(1,1) (ARMA part misspecified)
3. sLMg: ARMA(2,1) – ARCH(1) (GARCH part misspecified)
4. sLMg: ARMA(1,1) – ARCH(1) (both ARMA and GARCH part misspecified)
5. sLMi: ARMA(2,1) – (no ARMA misspecification, i.i.d. innovations assumed)
6. sLMi: ARMA(1,1) – (ARMA part misspecified, i.i.d. innovations assumed)
7. sLMh: ARMA(2,1) – (no ARMA misspecification)
8. sLMh: ARMA(1,1) – (ARMA part misspecified)

The results are shown in Figure 2, which reports the boxplots (60 parameters settings) of the empirical size (percent) at nominal 5% for each of the eight tests and sample size $n = 100, 200, 500$.

Tests 1 (sLMg) and 7 (sLMh) involve no misspecification and their size fluctuates around the nominal 5%. Consistent with the results of Section III, the sLMh test is slightly undersized. In tests 2 (sLMg) and 8 (sLMh), the ARMA part is misspecified but this has a small effect on the size: it produces some undersize for the sLMg test and some increase in variability for both the sLMg and the sLMh test. When the GARCH part is misspecified

TABLE 1
AR and GARCH parameters for the ARMA(2,1)-GARCH(1,1) DGP of equation (16)

ϕ_1		ϕ_2	
−0.65		0.25	
−0.35		−0.45	
0.45		0.25	
0.45		−0.55	
	a_1	b_1	a_0
A	0.04	0.95	1
B	0.30	0.00	1

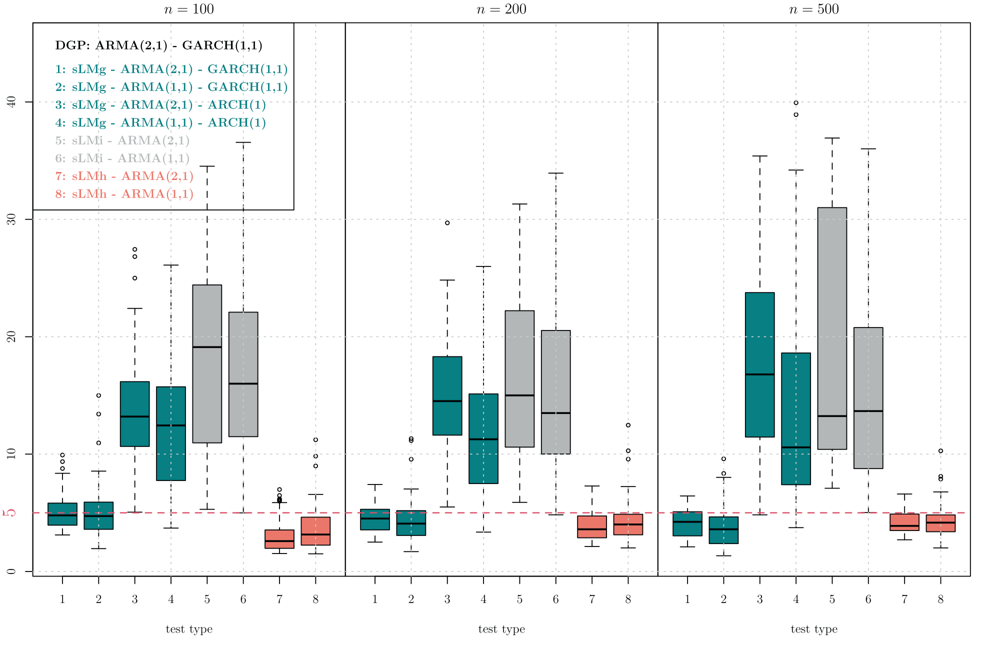

Figure 2. Empirical size (percent) at nominal level $\alpha = 5\%$ under the ARMA(2,1) – GARCH(1,1) DGP of equation (16) and different, possibly misspecified, sLM tests: sLMg, turquoise, tests 1-4; sLMi, grey, tests 5,6; sLMh, salmon, tests 7,8 [Colour figure can be viewed at *wileyonlinelibrary.com*]

the sLMg test is oversized as expected (tests 3 and 4). Note that misspecifiying the ARMA part partly compensates the oversize (test 4). The same holds for the sLMi tests, (tests 4 and 5). The general conclusion that can be drawn is that misspecifiying the ARMA part does not affect appreciably the performance of the tests, and this is consistent with the results of Goracci *et al.* (2023). On the contrary, it is important to account properly for the heteroskedasticity, possibly by using a high-order GARCH in the sLMg test and/or use it in combination with the sLMh test.

Power

In order to study the power of the tests we simulate from the following TARMA(1, 1)-GARCH(1, 1) DGP:

$$\begin{aligned} X_t &= 0.5 + 0.5X_{t-1} + 0.5\varepsilon_{t-1} + (\varphi_0 + \varphi_1 X_{t-1} + \vartheta_1 \varepsilon_{t-1})\, I(X_{t-1} \leq 0) + \varepsilon_t. \\ \varepsilon_t &= \sqrt{h_t} z_t; \quad h_t = a_0 + a_1 \varepsilon_{t-1}^2 + b_1 h_{t-1}. \end{aligned} \tag{17}$$

where $\varphi_0 = \varphi_1 = \vartheta_1 = 0.5 - \Psi$, and $\Psi = (0.00, 0.15, 0.30, 0.45, 0.60, 0.75, 0.90, 1.05)$. Hence, the system of hypotheses of equation (3) becomes

$$\begin{cases} H_0 & : \boldsymbol{\Psi} = \mathbf{0}, \\ H_1 & : \boldsymbol{\Psi} \neq \mathbf{0}, \end{cases}$$

where $\boldsymbol{\Psi} = (\Psi, \Psi, \Psi)$, so that the parameter Ψ represents the departure from the null hypothesis. As above, we combine these with the following parameters for the GARCH specification: $(a_0, a_1, b_1) = (1, 0.1, 0.8)$ (case A), $(1, 0.4, 0.4)$ (case B), $(1, 0.8, 0.1)$ (case C). The size-corrected power of the tests (in percentage) is presented in Table 2. The behaviour of the three tests is very similar, especially when the distance from the null hypothesis is large. However, for small to moderate departures from H_0 ($\Psi \leq 0.45$) the sLMg test is always superior to both the sLMi and the sLMh tests and, in some instances, is markedly superior, see case C and $n = 500$. The power loss incurred by estimating the GARCH parameters in the sLMg test is very small and is overwhelmingly compensated by the correct size in the presence of heteroskedasticity. The raw power is reported in the Supporting Information, Table S5. The general conclusion that can be drawn is that, even in the absence of information regarding the presence of heteroskedasticity, it is safe to use tests that account for it as it will lower the risk of a false rejection while retaining a good discriminating power.

Measurement error

We assess the effect of measurement error on the size of the sLMg test. We simulate X_t from the following AR(1)-GARCH(1,1) DGP:

$$\begin{aligned} X_t &= \phi_1 X_{t-1} + \varepsilon_t. \\ \varepsilon_t &= \sqrt{h_t} z_t; \quad h_t = a_0 + a_1 \varepsilon_{t-1}^2 + b_1 h_{t-1}, \end{aligned} \tag{18}$$

where, as above, $\phi_1 = (-0.9, -0.6, -0.3, 0.0, 0.3, 0.6, 0.9)$ and $(a_0, a_1, b_1) = (1, 0.04, 0.95)$ (case A), $(1, 0.3, 0.0)$ (case B), $(1, 0.4, 0.4)$ (case C). We add measurement noise as follows: $Y_t = X_t + \eta_t$, where the measurement error $\eta_t \sim N(0, \sigma_\eta^2)$ is such that the signal-to-noise ratio $\text{SNR} = \sigma_X^2 / \sigma_\eta^2$ is equal to $\{\infty, 50, 10, 5\}$. Here, σ_X^2 is the unconditional variance of X_t computed by means of simulation. The case without noise ($\text{SNR} = \infty$) is taken as the benchmark. The empirical size (rejection percentages) for the three sample sizes is presented in Table 3. Clearly, the size of the sLMg test is either minimally affected or not affected at all by the presence of measurement error, even for high levels of noise (signal-to-noise ratio $= 5$).

TABLE 2
Size-corrected power (percent) of the sLMg, sLMi and sLMh tests, at nominal level α = 5% for the TARMA(1,1)-GARCH(1,1) process of equation (17)

		n = 100			n = 200			n = 500		
	Ψ	sLMg	sLMi	sLMh	sLMg	sLMi	sLMh	sLMg	sLMi	sLMh
A	0.00	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0
	0.15	9.3	10.0	9.7	16.9	17.2	16.0	44.9	42.7	39.6
	0.30	25.8	27.4	25.6	53.7	55.3	50.7	96.7	95.5	94.6
	0.45	50.0	53.3	47.8	86.9	88.4	84.9	100.0	100.0	99.9
	0.60	72.5	77.2	69.4	98.1	98.7	97.6	100.0	100.0	100.0
	0.75	87.0	91.5	84.2	99.5	99.8	99.6	100.0	100.0	100.0
	0.90	92.3	96.5	90.7	99.6	99.9	99.8	100.0	100.0	100.0
	1.05	93.0	98.1	93.6	99.4	99.9	99.8	100.0	100.0	100.0
B	0.00	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0
	0.15	8.0	7.3	6.9	11.5	11.5	10.6	25.6	25.1	23.5
	0.30	20.0	18.0	18.1	40.5	41.1	39.2	82.5	83.6	83.1
	0.45	46.2	44.2	45.4	79.3	81.1	81.4	99.5	99.7	99.7
	0.60	71.7	72.4	73.9	96.0	97.3	97.4	100.0	100.0	100.0
	0.75	87.3	89.8	90.3	98.1	99.1	99.2	99.9	100.0	100.0
	0.90	92.0	95.9	95.4	98.3	99.5	99.5	99.7	99.9	99.9
	1.05	94.2	97.7	97.7	98.1	99.6	99.6	99.7	99.9	100.0
C	0.00	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0
	0.15	8.5	7.9	7.9	13.1	10.7	11.3	31.8	16.7	21.7
	0.30	18.0	17.6	18.4	39.0	31.0	35.3	84.0	62.1	73.7
	0.45	36.6	35.9	39.8	71.1	65.3	71.9	99.1	95.9	96.6
	0.60	58.3	60.2	65.0	91.2	89.5	92.6	100.0	99.8	99.6
	0.75	74.9	79.9	82.7	97.0	97.4	97.6	100.0	100.0	99.8
	0.90	83.1	91.2	91.1	97.7	98.9	99.0	99.9	99.9	99.9
	1.05	85.9	95.3	94.1	97.7	99.1	99.2	99.9	100.0	100.0

As for the power of the test, the presence of high levels of measurement noise impinges negatively upon it, so larger sample sizes are required to compensate for this. The study is reported in Section S1.4 of the Supporting Information.

IV. Testing and modelling the Italian strikes time series

In this section, we study the dynamics of the monthly series of hours not worked due to strikes in Italy between January 1949 and December 2009 ($n = 732$). The data were obtained from the Italian Statistical Institute (ISTAT). Despite their importance, to the best of our knowledge, this is the first time that these data are used within a macroeconometric approach. Note that, also for Italy, ISTAT has suspended the collection of monthly data from 2009 onwards, so more recent data are not available. The time plot is characterized by cyclical oscillations and a structural change in variance starting from the mid-1980s, see Figure 3 (left). In order to identify the change point, we have applied to the squared series the moving-sum test developed in Cho and Kirch (2022) and implemented in the `mosum` R package (Meier, Kirch, and Cho, 2021). The method allows for heavy tails and dependence. We have run the test on a grid of bandwidths and the results indicate robustly that the change point occurred in June 1983, see the red dashed vertical line in Figure 3 (left). This can be explained by a series of historical occurrences such as the increase in

TABLE 3

Empirical size of the sLMg test at nominal level $\alpha = 5\%$ for the AR(1)-GARCH(1,1) process of equation (18) with three levels of measurement error

		$n = 100$				$n = 200$				$n = 500$			
	ϕ_1	∞	50	10	5	∞	50	10	5	∞	50	10	5
A	−0.9	3.7	2.6	2.8	4.2	3.2	2.7	3.8	4.5	2.7	2.8	4.6	5.1
	−0.6	3.5	4.0	3.7	4.5	3.0	3.2	3.6	3.7	2.5	2.1	2.1	3.1
	−0.3	4.7	5.4	5.4	6.8	3.0	2.3	3.0	4.1	2.0	2.7	2.1	3.1
	0.0	7.7	8.4	9.3	8.7	7.5	7.9	7.8	8.0	4.2	2.4	3.8	4.1
	0.3	4.0	4.8	3.8	4.6	2.8	2.5	3.1	3.4	1.3	1.4	1.7	2.2
	0.6	3.0	2.8	3.4	3.4	2.9	3.7	3.2	3.4	2.0	2.8	2.5	3.6
	0.9	3.3	3.7	4.2	5.0	3.4	4.1	4.3	4.7	2.8	2.3	2.6	3.5
B	−0.9	3.7	3.3	3.0	3.6	4.8	4.3	4.3	4.9	3.5	5.3	4.9	4.8
	−0.6	4.4	3.9	4.1	3.9	4.5	4.5	6.1	5.3	6.6	5.9	6.2	6.8
	−0.3	4.8	5.6	7.0	5.0	4.5	5.5	6.3	5.7	5.5	4.1	5.7	6.3
	0.0	9.4	9.6	7.8	9.8	6.4	6.4	7.1	7.1	7.6	8.1	7.2	9.2
	0.3	4.9	4.2	5.0	5.1	3.1	4.1	5.5	5.2	5.3	4.9	4.3	6.1
	0.6	4.0	4.4	4.1	4.4	4.4	4.3	4.7	5.0	6.3	5.3	5.6	6.5
	0.9	4.3	4.8	4.6	5.1	4.5	4.3	3.9	3.5	5.2	5.4	5.5	5.8
C	−0.9	3.3	2.9	3.8	5.7	2.5	3.6	4.6	4.1	3.7	5.6	5.9	6.3
	−0.6	2.4	2.1	3.3	4.0	2.8	3.3	3.8	4.5	2.2	2.4	3.5	3.4
	−0.3	4.5	4.9	3.3	6.8	2.3	2.5	3.7	4.6	2.6	3.5	2.0	3.2
	0.0	6.0	5.8	7.5	8.5	3.9	5.2	6.0	7.1	4.0	4.6	4.5	6.1
	0.3	3.4	3.9	2.7	4.8	1.9	3.1	3.7	4.0	2.4	1.8	3.1	3.8
	0.6	3.7	3.4	3.1	4.3	2.5	3.2	3.9	5.6	2.8	2.8	4.9	6.0
	0.9	4.5	4.8	5.4	4.8	3.8	4.7	7.1	5.0	3.0	3.3	5.6	10.4

the workers' opportunity cost associated with the decision to strike, which reduced the incentive to strike for economic reasons, e.g. wage and/or contractual claims.

We consider the logarithm (in base 10) as a variance-stabilizing transformation and show the result in Figure 3 (right). The month plot is shown in Figure S4 of the Supporting Information. The series has seasonal oscillations and August is the month where the strikes undergo a consistent drop, followed by September, January and December, where the phenomenon is less pronounced. August is a vacation month in Italy, and September is a postvacation month, while January and December are characterized by less working days due to Christmas holidays. The series also presents some seasonality in the minima and the maxima, due to the strikes being more likely to be observed in those months when they are more effective. As strikes are costly for both firms and workers, the seasonality of economic activity, as well as the economic cycle, make the strike strategy more effective during the high seasons or expansionary phases while discouraging workers during low seasons or recessionary phases.

The autocorrelation of the series decays slowly hinting at the presence of a trend and a strong seasonal component.[2] The spectral density function (smoothed periodogram with a Daniell window) of Figure 4 shows the main periodicities of the series. Besides the yearly periodicity, the 7-year peak together with its harmonics (3.5, 1.7, and 0.85 years) is related to the business cycle. The 2.1-year periodicity could be the first indication

[2] The correlograms are reported in Supporting Information, Section S2.1.

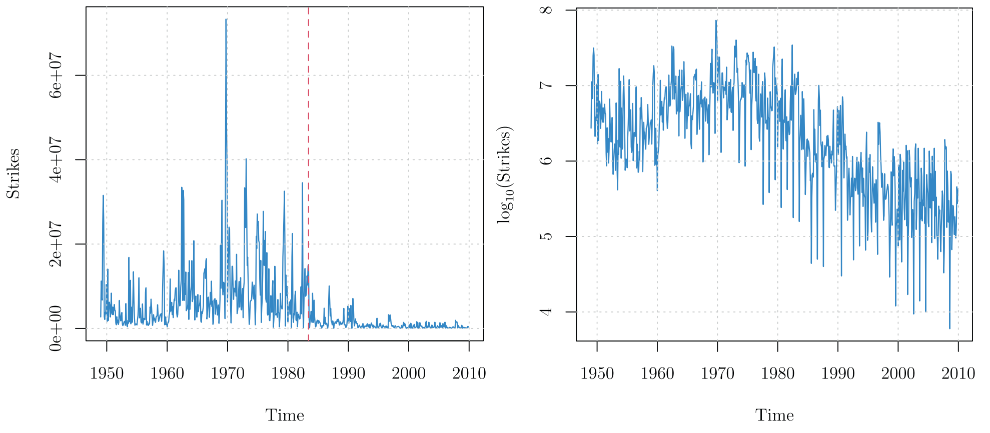

Figure 3. Time series of the Italian Strikes from January 1949 to December 2009. (Left) raw time series with the change point in variance identified by the MOSUM test (red dashed vertical line). (Right) $\log_{10}$-transformed series [Colour figure can be viewed at *wileyonlinelibrary.com*]

of a nonlinear dynamics emerging as a resonance (non-trivial combination of the main frequencies). This periodicity is also in line with the existence of a cycle in Italian strikes recalled by Franzosi (1980) based on the spectral analysis of this same series, using data predating the 1980s. Indeed, strike activity concentrates at the expiration of the contracts, whose average duration is 2–3 years (Myers and Saretto, 2016). Labour strikes are a tool for bargaining better working conditions, both for those who strike and for those who do not, and this poses a threat to firms (Barrett and Pattanaik, 1989). Strikes have an upper bound in the number of hours they can be used, though, since above that level the tool can backfire and damage the workers themselves. This can be due, for example, to the firm being so damaged by the strikes that it has to reduce its production and hence lay off some workers. This implies that the series cannot present a unit root, be it regular or seasonal. The application of seasonal unit root tests seems to rule out the possibility of seasonal unit roots (see Section S2.2 of the Supporting Information). However, as also discussed in the following, since it has been proven that unit root tests can be severely oversized in the presence of moving average terms, the results have to be taken with caution. Moreover, cointegrating relations are not expected. We assess this conjecture by applying a battery of unit root tests to the series of the Italian strikes and to the following covariates: *salary*: Monthly index of salaries of industrial workers; *price*: Monthly index of consumer prices.[3] Following the labour supply models, we have chosen these two economic variables because they represent a measure of the opportunity cost of strikes and one of the main economic motivations behind strikes for wage demands. The results are shown in Table 4. The first row presents the test $\overline{MZ}_{\alpha}^{\text{GLS}}$ as proposed by Perron and Qu (2007), which is essentially the same as the MZ_{α}^{GLS} test of Ng and Perron (2001) but the lag of the ADF regression is selected on OLS detrended data. The second row shows the results for the $\overline{MP}_{t}^{\text{GLS}}$, the modified feasible point optimal test (see also eq. (9) in Ng and Perron (2001)). The third row lists the GLS detrended version of the ADF

[3] All the series have been obtained from the Italian National Institute of Statistics (ISTAT).

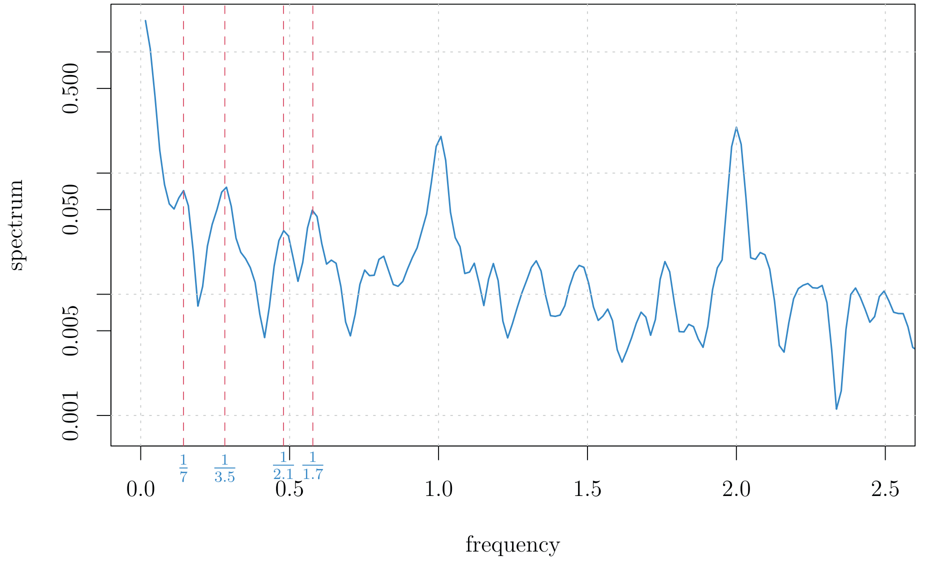

Figure 4. Spectral density function, up to a frequency of 2.5 (cycles per year) of the Italian Strikes (in log_{10}) from January 1949 to December 2009. The periods corresponding to the dominant peaks are added in blue [Colour figure can be viewed at *wileyonlinelibrary.com*]

TABLE 4

Values of three unit root test statistics applied to the series of the Italian strikes plus the covariates

	Strikes	*Salary*	*Price*	*5% critical value*
$\overline{M}Z_{\alpha}^{GLS}$	−1.71	−1.23	−6.14	−21.30
$\overline{M}P_{t}^{GLS}$	52.92	70.04	14.83	5.48
ADF^{GLS}	−1.08	−1.52	−1.64	−2.91

Note: The last column reports the 5% critical values of the null distribution.

test (denoted by ADF^{GLS}). Following Chan *et al.* (2020, 2024), we have chosen them since they are the best performers among all those proposed in Ng and Perron (2001) and Perron and Qu (2007).[4] The last column contains the critical values of the null asymptotic distribution at the 5% level. Clearly, none of the three tests manages to reject the null hypothesis of a unit root in any of the series. For this reason, we perform a pairwise cointegration analysis. We regress the series of strikes on the three covariates and apply the above tests to the residuals of the fitted models. The results are presented in Table 5. Again, none of the tests is able to reject the null hypothesis and this renders the whole analysis inconclusive in that all the series appear to be integrated but no cointegrating relationships can be found.

One possible reason for the above results is the lack of power of unit root tests against a nonlinear alternative. This could be due to them not explicitly encompassing

[4]We have ported to R the original Gauss routines of Ng and Perron (2001). The results for the remaining tests are similar. All the material is available upon request.

TABLE 5
Cointegration tests: application of the unit root tests to the residuals of the regression of the Italian strikes on the covariates

	Salary	*Price*	*5% critical value*
$\overline{M}Z_{\alpha}^{\mathrm{GLS}}$	−1.81	−2.92	−21.30
$\overline{M}P_{t}^{\mathrm{GLS}}$	50.21	31.19	5.48
$\mathrm{ADF}^{\mathrm{GLS}}$	−1.21	−1.83	−2.91

Note: The last column reports the 5% critical values of the null distribution.

the nonlinearity within their specification. On the other hand, tests for a unit root against a threshold autoregressive alternative are severely biased in the presence of MA components. This is discussed in Chan *et al.* (2020, 2024), which solves the problem by proposing a unit root test where the null hypothesis entails an integrated MA against the alternative of a stationary threshold ARMA model, possessing a unit root regime. The application of such a test to the time series of Italian strikes produces a test statistic equal to 13.95, with an associated (heteroskedastic robust) wild bootstrap p-value of 0.012, and this points to an alternative explanation of the strikes' dynamics based upon a stationary regime-switching mechanism.

In Section S2.3 of the Supporting Information, we adopt a linear modelling approach based on a two-step ARIMA-GARCH fit. The results show that both regular and seasonal differences are needed and this is not consistent with stylized economic facts and difficult to interpret. Moreover, the residual analysis hints at the presence of unaccounted nonlinear dependence (see Figure S10 of the Supporting Information). Given the inadequacy of the non-stationary linear modelling approach, and also in view of the results of the unit root test of Chan *et al.* (2024), we are encouraged to seek for a different explanation for the strikes' dynamics, based on a stationary regime-switching model with conditional heteroskedasticity. Hence, we test the ARMA-GARCH model against the TARMA-GARCH specification by using our novel sup-Lagrange Multiplier sLMg test. We use the consistent Hannan−Rissanen criterion to select the order of the ARMA model to be tested (Hannan and Rissanen, 1982). The maximum order tested is the ARMA(12,12) and the procedure used selects the ARMA(1,1) model.[5] As for the order of the GARCH model specification we select the GARCH(1,2) on the basis of the previous investigation shown in Section S2.3 of the Supporting Information. Hence, we test the ARMA(1,1)-GARCH(1,2) against the TARMA(1,1)-GARCH(1,2) specification with $d = 1$. We obtained a value of the sLMg statistic equal to 24.71 (threshold $= 6.27$). Given that the critical value at 1% level results 16.65 (from table I of Andrews (2003) with $\pi_0 = 0.30$), our asymptotic sLMg test rejects and points to a significant threshold effect either in the intercept and/or at lag 1. Also, the robustified sLMh test statistic results are significant at 1% level: (sLMh $= 26.16$, threshold $= 6.28$). Thus, the results of the tests corroborate the hypothesis that the dynamic of strikes is governed by a regime-switching dynamics with

[5]The Hannan−Rissanen ARMA model selection is a multistage procedure. First, the series is prefiltered using an AR(p) model, using a sufficiently large p. Second, the residuals of the AR fit are used as predictors in a regression that includes them (mimicking the MA part), plus the lagged series (AR part). Such regression is repeated on a grid of AR and MA orders and the order that minimizes the BIC is selected. See also Choi (1992) for more details.

conditional heteroskedasticity. We propose the following two-stage TARMA-GARCH model:

$$\begin{aligned} X_t &= \begin{cases} \phi_{1,0} + \phi_{1,1}X_{t-1} + \phi_{1,12}X_{t-12} & \\ \quad + \theta_{1,3}\varepsilon_{t-3} + \theta_{1,12}\varepsilon_{t-12} + \theta_{1,13}\varepsilon_{t-13} + \varepsilon_t & \text{if } X_{t-1} \le r \\ \phi_{2,0} + \phi_{2,1}X_{t-1} + \phi_{2,12}X_{t-12} + \theta_{2,1}\varepsilon_{t-1} + \theta_{2,3}\varepsilon_{t-3} + \varepsilon_t, & \text{if } X_{t-1} > r. \end{cases} \\ \varepsilon_t &= \sqrt{h_t} z_t; \quad h_t = a_0 + a_1\varepsilon_{t-1}^2 + b_1 h_{t-1}. \end{aligned} \tag{19}$$

We adopt a two-stage estimation approach since the existing results for TARMA models include least squares estimators for which consistency and asymptotic normality have been established (Li *et al.*, 2011; Giannerini and Goracci, 2021). In this way, we can adopt a maximum likelihood approach for estimating the GARCH part, exploiting mature implementations such as those of the R package `rugarch` (Ghalanos, 2022). Equation (20) reports the estimated coefficients of the model, with standard errors in parentheses below each coefficient. The threshold is estimated to be 6.23, equivalent to roughly 1.7 million non-worked hours due to strikes in a month.

$$X_t = \begin{cases} \underset{(0.17)}{0.42} + \underset{(0.04)}{0.07}X_{t-1} + \underset{(0.05)}{0.85}X_{t-12} + & \\ \underset{(0.05)}{0.09}\varepsilon_{t-3} - \underset{(0.06)}{0.51}\varepsilon_{t-12} - \underset{(0.07)}{0.19}\varepsilon_{t-13} + \varepsilon_t, & \text{if } X_{t-1} \le 6.23 \\ \underset{(0.47)}{0.76} + \underset{(0.09)}{0.41}X_{t-1} + \underset{(0.06)}{0.47}X_{t-12} + \underset{(0.09)}{0.25}\varepsilon_{t-1} + \underset{(0.05)}{0.05}\varepsilon_{t-3} + \varepsilon_t, & \text{if } X_{t-1} > 6.23 \end{cases} \tag{20}$$

$$\varepsilon_t = \sqrt{h_t} z_t; \quad h_t = \underset{(0.09)}{0.20} + \underset{(0.04)}{0.10}\varepsilon_{t-1}^2 + \underset{(0.10)}{0.70}h_{t-1}. \tag{21}$$

The estimated TARMA model fulfils the sufficient conditions of ergodicity and invertibility as discussed in Chan and Goracci (2019). Figure 5 shows the series (light blue line) together with the fitted values from the TARMA model (blue line), the estimated conditional standard deviation $\hat{h}_t^{1/2}$ from the GARCH(1,1) fit (green line), and the estimated threshold ($\hat{r} = 6.23$, dashed red line). The TARMA specification for the conditional mean has a lower regime which is characterized by a persistent seasonality with a clear threshold effect at lag 12. The MA(3) parameters are not significant but play a role in providing a solution that mimics the observed dynamics. The lags in the TARMA model have been selected by combining information criteria, the ability of the fit to reproduce the observed nonlinear features of the data (i.e. asymmetric distribution, autocorrelation properties, spectral peaks), and on the basis of the residual analysis. Note that having some lags only in one specific regime is key to explain the asymmetric behaviour of the series. In particular, the upper regime contains no seasonal MA lags and this is a strong indication of the presence of a threshold effect in the moving average part, something that conventional TAR models are not able to reproduce or, at best, can only reproduce partially by incorporating higher order lags. Clearly, the dynamics reacts differently to shocks depending upon the regime. The threshold identifies the time corresponding to the structural break in the variance of the series and splits it into two periods. Indeed, from the mid-80s onwards, the series belongs to the lower regime where the dynamics is

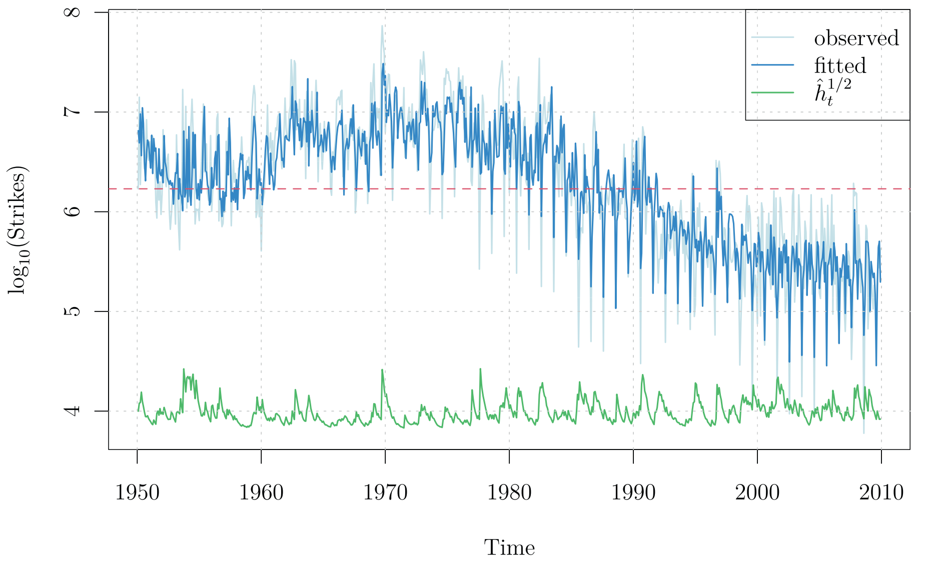

Figure 5. Time series of the Italian Strikes (light blue) with the fitted series from the TARMA-GARCH model (blue) and the estimated conditional standard deviations $\hat{h}_t^{1/2}$ (green). The estimated threshold $\hat{r} = 6.23$ is indicated as a dashed horizontal red line [Colour figure can be viewed at *wileyonlinelibrary.com*]

much more characterized by seasonality. The series of the conditional standard deviation (in green) shows both clusters of volatility and asymmetric peaks, and appears to change starting from the late 1990s.

The parameter estimates for the GARCH(1,1) part of the model are shown in equation (21). The estimates indicate a non-negligible heteroskedastic structure in the innovations. Also, in this case, the estimated model results are stationary and ergodic and its adequacy is assessed in the Supporting Information S2.4, which also reports model diagnostics. The fit is able to reproduce the strong asymmetric behaviour of the series: the histogram of the data is reported in Figure 6 (left), where we superimposed a kernel density estimate over a simulated trajectory of 100k observations from the fitted model (blue line). This is also witnessed by the Kolmogorov−Smirnov two-sample test, computed between the distribution of the observed data and that of the simulated series (p-value = 0.08). Moreover, we test the residuals of the model for the presence of nonlinear serial dependence, up to lag 12, with the entropy metric S_ρ, (Giannerini, Maasoumi, and Bee Dagum, 2015) see Figure 6 (right). Since no lags exceed the bootstrap rejection band we can be confident that the model managed to capture the underlying nonlinear features of the series. Finally, in Figure S15 of the Supporting Information, we show that there is no unaccounted cross-dependence between the residuals of the fitted model and the covariates *salary* and *price*, introduced at the beginning of the section.

The existence of a threshold-type dynamics is consistent with the underlying theory, see e.g. Granovetter (1978), where individuals make choices that are influenced by the behaviour of other individuals. In particular, Granovetter's model posits that individuals' decisions to participate in collective action (such as strikes) depend on the number of other participants. This may create a nonlinear, threshold-activated dynamics where small

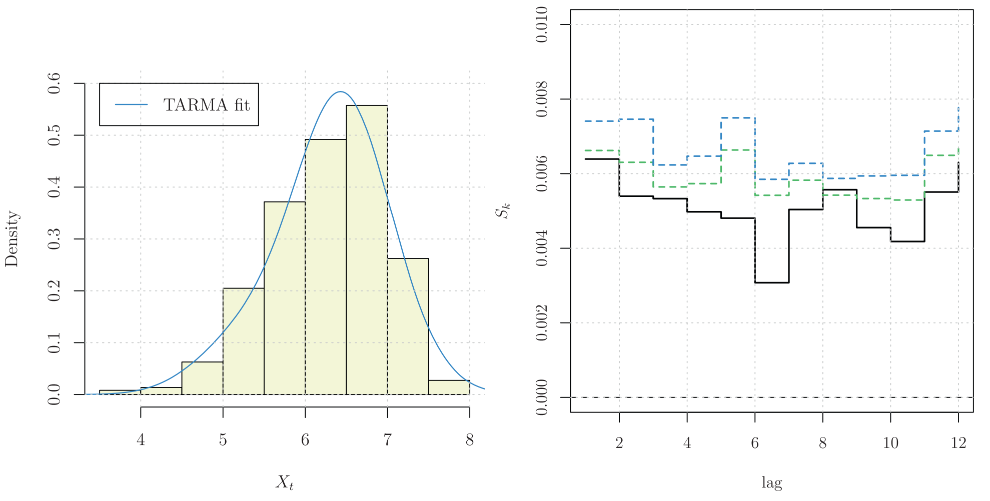

Figure 6. (left) Histogram of the time series of the Italian strikes. The smooth blue curve is the density estimate based on 100 k data simulated from the fitted TARMA model. (right) Entropy-based metric up to lag 12, computed on the residuals of the TARMA-GARCH model of equation (19). The dashed lines correspond to bootstrap rejection bands at levels 95% (green) and 99% (blue) under the null hypothesis of linear serial dependence [Colour figure can be viewed at *wileyonlinelibrary.com*]

changes in participation can lead to large-scale collective action if thresholds are crossed. The theory explains the clustering effect observed in strike activities. When a critical mass of workers engages in a strike, it lowers the perceived risk and increases the potential benefits for others to join, leading to a cascading effect. This behaviour aligns with the observed volatility and clustering, as initial strikes by a few workers can quickly grow into larger movements. The behavioural implications of such threshold dynamics, compatible with the empirical observation we presented, are significant. They suggest that strike activities are not only responses to economic conditions but are also heavily influenced by social dynamics and peer behaviour. Also, Bursztyn *et al.* (2021) stress the importance of the activity of people within their social network when choosing to take part in a protest. The existence of strong ties between individuals is important for the persistence of the protests and the clustering effect.[6] As for the structural change in the variance, we note that the strengthening of the government and the exclusion of unions from the process of reforming policies has downplayed collective strikes as a political and social bargaining tool (Hamann *et al.*, 2013). Godard (2011) argues that the change in the dynamics of strikes observed since the 1980s can be due to four different reasons: (i) they have been diverted into alternative forms of conflict; (ii) the capitalist system managed to reduce the number of citizens who disapprove of the economic and political system itself, or at least it managed to reduce their will to act upon it; (iii) the conflict underwent a transformation

[6]Nevertheless, social and political instability, strongly linked with economic growth, also depends on fiscal policy, which can be influenced by strikes (Alcántar-Toledo and Venieris, 2014). In particular, political instability negatively affects economic growth (Jong-A-Pin, 2009), but also weak ties and the behaviour of the whole population play a role in this issue (see e.g. Passarelli and Tabellini (2017) and Cantoni *et al.* (2019)). Therefore, policymakers must consider these complex dynamics when addressing labour disputes and designing economic policies to mitigate the adverse effects of strikes while supporting workers' rights and economic growth.

and has become more deeply embedded and linked to general behaviour within and outside the workplace, such as cynicism and escapism, but also depression; (iv) the conflict has become dormant. While some of these reasons appear more convincing than others, their multifactorial interaction could have played a role in explaining the observable shift. In addition, other possible reasons, partially overlapping the aforementioned, are the reduction of the union density, the birth of autonomous and non-political trade unions, and the creation of local agreements, three events that characterized Italy after the 1970s and that are particularly relevant in the 1990s and 2000s, see also Giangrande (2021).

V. Conclusions

Our proposals allow us to test for nonlinearity in the conditional mean, without being affected by heteroskedasticity, making it possible to assess the two types of nonlinearity separately. The test sLMh is robust against an unspecified heteroskedasticity so that there is no need to model it directly as in the sLMg test. On the other hand, the sLMh test is generally undersized and this produces a moderate power loss with respect to the sLMg test, especially in small samples. Since threshold ARMA models can accommodate a wide range of complex dynamics, we expect the tests to have power against most types of nonlinearity in the conditional mean, besides threshold models. For this reason, they can be used as omnibus explorative nonlinearity tests with good properties of robustness against heteroskedasticity. This aspect will be further explored in future investigations.

The results point to a significant threshold effect in the series of Italian strikes, which appears to be governed by a regime-switching dynamics with conditional heteroskedasticity. The analysis of the strikes offers some evidence that TARMA models provide a flexible and parsimonious specification that describes some of its empirical nonlinear stylized facts. Our findings are coherent with both features of the history of the Italian labour market, such as the decrease in the social turmoil after the first half of the 1980s (Godard, 2011), and the social aspects of unrest and labour markets in general, based on the threshold model by Granovetter (1978) and also in line with more recent studies of the impact of social ties and the existence of social networks able to trigger political engagement and participation to protest (Bursztyn *et al.*, 2021).

Acknowledgments

We thank the Editor Anindya Banerjee and three anonymous referees for the comments that helped improving the paper. We also thank Valentina Corradi and Wai Keung Li for useful comments and suggestions.

CONFLICTS OF INTEREST

The authors declare no competing interest.

Final Manuscript Received: March 2024

References

Alcántar-Toledo, J. and Venieris, Y. (2014). 'Fiscal policy, growth, income distribution and sociopolitical instability', *European Journal of Political Economy*, Vol. 34, pp. 315–331.

Altissimo, F. and Violante, G. (2001). 'The non-linear dynamics of output and unemployment in the US', *Journal of Applied Econometrics*, Vol. 16, pp. 461–486.

Andrews, D. (1993). 'Tests for parameter instability and structural change with unknown change point', *Econometrica*, Vol. 61, pp. 821–856. http://www.jstor.org/stable/2951764.

Andrews, D. (2003). 'Tests for parameter instability and structural change with unknown change point: A corrigendum', *Econometrica*, Vol. 71, pp. 395–397. https://doi.org/10.1111/1468-0262.00405.

Barrett, C. and Pattanaik, P. (1989). 'Bargaining over employment and wages, with threats of strikes and lockouts: An extension of the nash solution', *European Journal of Political Economy*, Vol. 5, pp. 333–345.

Billingsley, P. (1968). *Convergence of probability measures*, Wiley, New York.

Bordogna, L. and Cella, G. (2002). 'Decline or transformation? change in industrial conflict and its challenges', *Transfer: European Review of Labour and Research*, Vol. 8, pp. 585–607.

Bursztyn, L., Cantoni, D., Yang, D., Yuchtman, N. and Zhang, Y. (2021). 'Persistent political engagement: Social interactions and the dynamics of protest movements', *American Economic Review: Insights*, Vol. 3, pp. 233–250.

Cantoni, D., Yang, D., Yuchtman, N. and Zhang, Y. (2019). 'Protests as strategic games: experimental evidence from Hong Kong's democracy movement', *Quarterly Journal of Economics*, Vol. 134, pp. 1021–1077.

Castellani, M., Fanelli, L. and Savioli, M. (2013). *Government Fiscal Efforts vs. Labour Union Strikes: An Estimated Cournot-Nash Policy Game*. Technical report, Mimeo, Bologna, Italy: Collana Quaderni-Working Paper DSE, University of Bologna.

Chan, K. S. (1991). 'Percentage points of likelihood ratio tests for threshold autoregression', *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, Vol. 53, pp. 691–696. http://www.jstor.org/stable/2345598.

K.-S. Chan. Testing for threshold autoregression. *Annals of Statistics*, 18(4):1886–1894, 12 1990. https://doi.org/10.1214/aos/1176347886.

Chan, K.-S., Giannerini, S., Goracci, G. and Tong, H. (2020). Testing for threshold regulation in presence of measurement error with an application to the PPP hypothesis. https://arxiv.org/abs/2002.09968.

Chan, K.-S., Giannerini, S., Goracci, G. and Tong, H. (2024). 'Testing for threshold regulation in presence of measurement error', *Statistica Sinica*, Vol. 34, pp. 1413–1434. https://doi.org/10.5705/ss.202022.0125.

Chan, K.-S. and Goracci, G. (2019). 'On the ergodicity of first-order threshold autoregressive moving-average processes', *Journal of Time Series Analysis*, Vol. 40, pp. 256–264.

Chan, K.-S., Hansen, B. and Timmermann, A. (2017). 'Guest editors' introduction: Regime switching and threshold models', *Journal of Business & Economic Statistics*, Vol. 35, pp. 159–161. https://doi.org/10.1080/07350015.2017.1236521.

Chen, C., Liu, F.-C. and So, M. K. (2011). 'A review of threshold time series models in finance', *Statistics and its Interface*, Vol. 4, pp. 167–181.

Cho, H. and Kirch, C. (2022). 'Two-stage data segmentation permitting multiscale change points, heavy tails and dependence', *Annals of the Institute of Statistical Mathematics*, Vol. 74, pp. 653–684. https://doi.org/10.1007/s10463-021-00811-5.

Choi, B. (1992). *ARMA Model Identification*, Springer, New York, NY. https://doi.org/10.1007/978-1-4613-9745-8.1.

Corneo, G. and Lucifora, C. (1997). 'Wage formation under union threat effects: theory and empirical evidence', *Labour Economics*, Vol. 4, pp. 265–292.

Franzosi, R. (1980). 'Gli scioperi in Italia: Analisi esplorativa dei dati', *Rivista di Politica Economica*, Vol. 70, pp. 1275–1322.

Franzosi, R. (1989). 'One hundred years of strike statistics: Methodological and theoretical issues in quantitative strike research', *Industrial and Labor Relations Review*, Vol. 42, pp. 348–362.

Ghalanos, A. (2022). rugarch: Univariate GARCH models. R Package Version 1.4-9.

Giangrande, N. (2021). *Political Economy of Contemporary Italy: The Economic Crisis and State Intervention*, Routledge, London.

Giannerini, S. and Goracci, G. (2021). *Estimating and Forecasting with Threshold ARMA Models*. Technical report, Bologna, Italy: Department of Statistical Sciences, University of Bologna (Italy).

Giannerini, S., Goracci, G. and Rahbek, A. (2024). 'The validity of bootstrap testing in the threshold framework', *Journal of Econometrics*, Vol. 239, 105379. https://www.sciencedirect.com/science/article/pii/S0304407623000040 Climate Econometrics.

Giannerini, S., Maasoumi, E. and Bee Dagum, E. (2015). 'Entropy testing for nonlinear serial dependence in time series', *Biometrika*, Vol. 102, pp. 661–675. http://biomet.oxfordjournals.org/content/102/3/661.abstract.

Godard, J. (2011). 'What has happened to strikes?', *British Journal of Industrial Relations*, Vol. 49, pp. 282–305.

Gonçalves, S., Herrera, A. M., Kilian, L. and Pesavento, E. (2021). 'Impulse response analysis for structural dynamic models with nonlinear regressors', *Journal of Econometrics*, Vol. 225, pp. 107–130. https://doi.org/10.1016/j.jeconom.2021.06.009.

Goracci, G. (2020). 'Revisiting the canadian lynx time series analysis through TARMA models', *Statistica*, Vol. 80, pp. 357–394. https://doi.org/10.6092/issn.1973-2201/11478.

Goracci, G. (2021). 'An empirical study on the parsimony and descriptive power of TARMA models', *Statistical Methods & Applications*, Vol. 30, pp. 109–137. https://doi.org/10.1007/s10260-020-00516-8.

Goracci, G., Ferrari, D., Giannerini, S. and Ravazzolo, F. (2024). 'Robust estimation for threshold autoregressive moving-average models', *Journal of Business & Economic Statistics*, Vol. 0, pp. 1–32. https://doi.org/10.1080/07350015.2024.2412011.

Goracci, G., Giannerini, S., Chan, K.-S. and Tong, H. (2023). 'Testing for threshold effects in the TARMA framework', *Statistica Sinica*, Vol. 33, pp. 1879–1901. https://doi.org/10.5705/ss.202021.0120.

Granovetter, M. (1978). 'Threshold models of collective behavior', *The American Journal of Sociology*, Vol. 83, pp. 1420–1443.

Hamann, K., Johnston, A. and Kelly, J. (2013). 'Unions against governments: Explaining general strikes in Western Europe, 1980–2006', *Comparative Political Studies*, Vol. 46, pp. 1030–1057.

Hannan, E. and Rissanen, J. (1982). 'Recursive estimation of mixed autoregressive-moving average order', *Biometrika*, Vol. 69, pp. 81–94. http://www.jstor.org/stable/2335856.

Hansen, B. (1996). 'Inference when a nuisance parameter is not identified under the null hypothesis', *Econometrica*, Vol. 64, pp. 413–430. http://www.jstor.org/stable/2171789.

Hansen, B. (2011). 'Threshold autoregression in economics', *Statistics and its Interface*, Vol. 4, pp. 123–127.

Hicks, J. (1932). *The Theory of Wages*, MacMillan & Co., London.

Hundley, G. and Koreisha, S. (1987). 'The specification of econometric strike models: A VARMA approach', *Applied Economics*, Vol. 19, pp. 511–530.

Jong-A-Pin, R. (2009). 'On the measurement of political instability and its impact on economic growth', *European Journal of Political Economy*, Vol. 25, pp. 15–29.

Kapetanios, G. and Shin, Y. (2006). 'Unit root tests in three-regime SETAR models', *The Econometrics Journal*, Vol. 9, pp. 252–278.

Koop, G. and Potter, S. (1999). 'Dynamic asymmetries in US unemployment', *Journal of Business & Economic Statistics*, Vol. 17, pp. 298–312.

Lange, P., Irvin, C. and Tarrow, S. (1990). 'Mobilization, social movements and party recruitment: The Italian Communist party since the 1960s', *British Journal of Political Science*, Vol. 20, pp. 15–42.

Li, D., Li, W. and Ling, S. (2011). 'On the least squares estimation of threshold autoregressive and moving-average models', *Statistics and Its Interface*, Vol. 4, pp. 183–196. https://doi.org/10.4310/SII.2011.v4.n2.a13.

G. Li and W. Li. Testing for threshold moving average with conditional heteroscedasticity. *Statistica Sinica*, 18:647–665, 04 2008.

Li, G. and Li, W. (2011). 'Testing a linear time series model against its threshold extension', *Biometrika*, Vol. 98, pp. 243–250. https://doi.org/10.1093/biomet/asq074.

Ling, S. and Tong, H. (2005). 'Testing for a linear MA model against threshold MA models', *Ann. Statist.*, Vol. 33, pp. 2529–2552. https://doi.org/10.1214/009053605000000598.

Meier, A., Kirch, C. and Cho, H. (2021). 'mosum: A package for moving sums in change-point analysis', *Journal of Statistical Software*, Vol. 97, pp. 1–42. https://doi.org/10.18637/jss.v097.i08.
Myers, B. and Saretto, A. (2016). 'Does capital structure affect the behavior of nonfinancial stakeholders? an empirical investigation into leverage and union strikes', *Management Science*, Vol. 62, pp. 3235–3253.
Ng, S. and Perron, P. (2001). 'Lag length selection and the construction of unit root tests with good size and power', *Econometrica*, Vol. 69, pp. 1519–1554. https://doi.org/10.1111/1468-0262.00256.
Paldam, M. (2021). 'The political economy of strikes', in Zimmermann K. (ed), *Handbook of Labor* Human Resources and Population Economics, Springer, Cham, pp. 1–21.
Paldam, M. and Pedersen, P. (1982). 'The macroeconomic strike model: a study of seventeen countries, 1948–1975', *Industrial and Labor Relations Review*, Vol. 35, pp. 504–521.
Passarelli, F. and Tabellini, G. (2017). 'Emotions and political unrest', *Journal of Political Economy*, Vol. 125, pp. 903–946.
Perron, P. and Qu, Z. (2007). 'A simple modification to improve the finite sample properties of Ng and Perron's unit root tests', *Economics Letters*, Vol. 94, pp. 12–19.
Pesaran, M. and Potter, S. (1997). 'A floor and ceiling model of US output', *Journal of Economic Dynamics and Control*, Vol. 21, pp. 661–695.
Quaranta, M. (2012). *The rise of unconventional political participation in Italy: Measurement equivalence and trend over time, 1976-2009*, Irvine, CA: University of California. https://escholarship.org/uc/item/50x011vk.
Rees, A. (1952). 'Industrial conflict and business fluctuations', *The Journal of Political Economy*, Vol. 60, pp. 371–382.
Shalev, M. (1980). 'Industrial relations theory and the comparative study of industrial relations and industrial conflict', *British Journal of Industrial Relations*, Vol. 18, pp. 26–43.
Tong, H. (1978). 'On a threshold model', in Chen C. (ed), *Pattern Recognition and Signal Processing* NATO ASI Series E: Applied ScVol. 29, Sijthoff & Noordhoff, Amsterdam, pp. 575–586.
Tong, H. (1990). *Non-linear Time Series: A Dynamical System Approach*, Oxford, UK: Oxford University Press.
Tong, H. and Lim, K. (1980). 'Threshold autoregression, limit cycles and cyclical data', *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, Vol. 42, pp. 245–268.
Vandaele, K. (2016). 'Interpreting strike activity in western europe in the past 20 years: the labour repertoire under pressure', *Transfer: European Review of Labour and Research*, Vol. 22, pp. 277–294.
Wong, C. and Li, W. (1997). 'Testing for threshold autoregression with conditional heteroscedasticity', *Biometrika*, Vol. 84, pp. 407–418. http://www.jstor.org/stable/2337466.

Supporting Information

Additional Supporting Information may be found in the online Appendix:

Data S1. Supporting Information.

Data replication package: the data replication package is available at https://doi.org/10.3886/E188961

Appendix A: PROOFS

Proof of Lemma 1

Part (i): By deploying the same argument in Goracci *et al.* (2023), it is possible to prove that $n^{-1}\hat{\mathcal{I}}_n(r) = n^{-1}\mathcal{I}_n(\eta_0, r) + o_p(1)$. The ergodicity of $\{X_t\}$ implies that, for each r, $n^{-1}\mathcal{I}_n(\eta_0, r)$ converges in probability to $E[\partial^2 \ell_n(\eta_0, r)/\partial \boldsymbol{\Psi} \partial \boldsymbol{\Psi}^{\intercal}]$. On the other hand, it holds that:

$$\frac{\partial^2 \ell_n(\eta_0, r)}{\partial \boldsymbol{\Psi} \partial \boldsymbol{\Psi}^\intercal}$$
$$= \frac{1}{h_t}\left(\frac{\partial \varepsilon_t(\eta_0, r)}{\partial \boldsymbol{\Psi}}\right)\left(\frac{\partial \varepsilon_t(\eta_0, r)}{\partial \boldsymbol{\Psi}}\right)^\intercal - \frac{\varepsilon_t}{h_t^2}\left(\frac{\partial h_t(\eta_0, r)}{\partial \boldsymbol{\Psi}}\right)\left(\frac{\partial \varepsilon_t(\eta_0, r)}{\partial \boldsymbol{\Psi}}\right)^\intercal$$
$$+ \frac{\varepsilon_t}{h_t}\frac{\partial^2 \varepsilon_t(\eta_0, r)}{\partial \boldsymbol{\Psi} \partial \boldsymbol{\Psi}^\intercal} - \frac{\varepsilon_t}{h_t^2}\left(\frac{\partial \varepsilon_t(\eta_0, r)}{\partial \boldsymbol{\Psi}}\right)\left(\frac{\partial h_t(\eta_0, r)}{\partial \boldsymbol{\Psi}}\right)^\intercal$$
$$+ \frac{\varepsilon_t^2}{h_t^3}\left(\frac{\partial h_t(\eta_0, r)}{\partial \boldsymbol{\Psi}}\right)\left(\frac{\partial h_t(\eta_0, r)}{\partial \boldsymbol{\Psi}}\right)^\intercal - \frac{1}{2}\frac{1}{h_t^2}\left(\frac{\partial h_t(\eta_0, r)}{\partial \boldsymbol{\Psi}}\right)\left(\frac{\partial h_t(\eta_0, r)}{\partial \boldsymbol{\Psi}}\right)^\intercal$$
$$+ \frac{1}{2}\left(\frac{\varepsilon_t^2}{h_t^2} - \frac{1}{h_t}\right)\frac{\partial^2 h_t(\eta_0, r)}{\partial \boldsymbol{\Psi} \partial \boldsymbol{\Psi}^\intercal}.$$

By combining the law of iterated expectations with $E[\varepsilon_t|\mathcal{F}_{t-1}] = 0$ and $E[\varepsilon_t^2|\mathcal{F}_{t-1}] = h_t$ we have that, for each r,

$$\|M_n(r)\| = o_p(1), \qquad \text{where } M_n = \frac{\mathcal{I}_n(\eta_0, r)}{n} - \Lambda(r).$$

In the following, we prove that the result holds uniformly, i.e.:

$$\sup_{r \in [r_l, r_u]} \|M_n(r)\| = o_p(1).$$

To this aim fix $\kappa_1 < \kappa_2$ and consider a grid $\kappa_1 = r_0 < r_1 < \ldots < r_m = \kappa_2$ with equal mesh size, i.e. $r_i - r_{i-1} \equiv c$, for some $c > 0$. It holds that $\sup_{r \in [r_{i-1}, r_i]} \|M_n(r) - M_n(r_{i-1})\| \leq C_n$, for all i. Moreover, $E(C_n) \to 0$ as $c \to 0$. Because for any $r \in [\kappa_1, \kappa_2]$, there exists an i such that $r_{i-1} \leq r \leq r_i$ and hence $M_n(r) = M_n(r) - M_n(r_{i-1}) + M_n(r_{i-1})$ and $\sup_{r \in [\kappa_1, \kappa_2]} \|M_n(r)\| \leq \max_{i=0, \ldots, m} M_n(r_i) + C_n$. The proof is complete since for fixed m,

$$\max_{i=0, \ldots, m} M_n(r_i) \to 0 \text{ in probability}$$

and $E(C_n) \to 0$ as $c \to 0$ in probability.

Part (ii): Within this proof, all the $o_p(1)$ terms hold uniformly on $r \in [r_L, r_U]$. We need to prove that:

$$\sup_{r \in [r_L, r_U]} \left\| \frac{1}{\sqrt{n}} \frac{\partial \hat{\ell}_n(r)}{\partial \boldsymbol{\Psi}_2} - \left(\nabla_{n,2}(r) - \Lambda_{21}(r)\Lambda_{11}^{-1}\nabla_{n,1}\right) \right\| = o_p(1).$$

Since $\sqrt{n}(\hat{\boldsymbol{\Psi}}_1 - \boldsymbol{\Psi}_{0,1}) = \Lambda_{11}^{-1} n^{-1/2} \partial \ell_n / \partial \boldsymbol{\Psi}_1 + o_p(1)$ and $\Lambda_{21}(r) = O_p(1)$ uniformly in $r \in [r_L, r_U]$, then it is sufficient to prove that

$$\sup_{r \in [r_L, b_U]} \left\| \frac{1}{\sqrt{n}} \frac{\partial \hat{\ell}_n(r)}{\partial \boldsymbol{\Psi}_2} - \frac{1}{\sqrt{n}} \frac{\partial \ell_n(r)}{\partial \boldsymbol{\Psi}_2} + \Lambda_{21}(r)\sqrt{n}(\hat{\boldsymbol{\Psi}}_1 - \boldsymbol{\Psi}_{0,1}) \right\| = o_p(1).$$

which holds by using the same arguments developed in Theorem 2.1 in Li and Li (2008).

Proof of Theorem 2

To prove the FCLT of point (i) we follow two steps: (1) we verify that $Q_n(r)$ converges to $\xi_n(r)$ in terms of finite distributions and (2) we show the asymptotic equicontinuity of $Q_n(r)$. The first step is readily implied by combining the central limit theorem and the Cramer-Wold device ergo the theorem will be proved upon showing the tightness of the score vector $\nabla_n(r)$ componentwise. Note that the proof is different from that in Goracci *et al.* (2023), because the score vector contains extra terms that depend upon the GARCH specification. We rely upon the ARCH(∞) representation and use a combination of Jensen's inequality and the law of iterated expectations. For the sake of presentation and without loss of generality, we detail the case $p = q = u = v = 1$ and prove the tightness of the following process:

$$
\begin{aligned}
G_n(r) &= \frac{1}{\sqrt{n}} \frac{\partial \ell_n(r)}{\partial \varphi_1} \\
&= -\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{\varepsilon_t}{h_t} \frac{\partial \varepsilon_t(\eta_0, r)}{\partial \varphi_1} + \frac{1}{2} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{1}{h_t} \left(\frac{\varepsilon_t^2}{h_t} - 1 \right) \frac{\partial h_t(\eta_0, r)}{\partial \varphi_1} \\
&= \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{\varepsilon_t}{h_t} \sum_{\iota=0}^{t-1} \theta_1^{\iota} X_{t-1-\iota} I(X_{t-1-\iota} \le r) \\
&\quad + \frac{a_1}{\sqrt{n}} \sum_{t=1}^{n} \frac{1}{h_t} \left(1 - \frac{\varepsilon_t^2}{h_t} \right) \sum_{\tau=0}^{t-1} b_1^{\tau} \varepsilon_{t-1-\tau} \sum_{\iota=0}^{t-2-\tau} \theta_1^{\iota} X_{t-2-\tau-\iota} I(X_{t-2-\tau-\iota} \le r).
\end{aligned}
$$

For each $r, s \in R$, define

$$
\begin{aligned}
g_t^{(1)}(r, s) &= \frac{\varepsilon_t}{h_t} \sum_{\iota=0}^{t-1} \theta_1^{\iota} X_{t-1-\iota} I(s < X_{t-1-\iota} \le r), \\
g_t^{(2)}(r, s) &= \frac{a_1}{h_t} \left(1 - \frac{\varepsilon_t^2}{h_t} \right) \sum_{\tau=0}^{t-1} b_1^{\tau} \varepsilon_{t-1-\tau} \sum_{\iota=0}^{t-2-\tau} \theta_1^{\iota} X_{t-2-\tau-\iota} I(s < X_{t-2-\tau-\iota} \le r).
\end{aligned}
$$

Clearly,

$$
G_n(r) - G_n(s) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left\{ g_t^{(1)}(r, s) + g_t^{(2)}(r, s) \right\}.
$$

Using the same approach of Wong and Li (1997) it suffices to prove that there exists a constant C such that

$$
E\left[\sup_{|r-s|<\delta} \left| g_t^{(1)}(r, s) + g_t^{(2)}(r, s) \right|^2 \right] \le C\delta, \tag{A1}
$$

which follows upon proving that there exist two constants, say C_1 and C_2, such that

$$E\left[\left|g_t^{(1)}(r,s)\right|^2\right] \leq C_1(r-s), \tag{A2}$$

$$E\left[\left|g_t^{(2)}(r,s)\right|^2\right] \leq C_2(r-s). \tag{A3}$$

In the following we use $\mathcal{C}$ to refer to a generic constant that can change across lines. In order to prove (A2) note that:

$$E\left[\frac{\varepsilon_t^2}{h_t^2}\middle| \mathcal{F}_{t-1}\right] = \frac{1}{h_t} \leq \frac{1}{a_0}.$$

Hence (A2) is verified by using the law of iterated expectations, Jensen's inequality and the fact that $|\theta_1| < 1$:

$$\begin{aligned} E\left[\left|g_t^{(1)}(r,s)\right|^2\right] &= E\left[\left\{\sum_{\iota=0}^{t-1}\theta_1^{\iota}X_{t-1-\iota}I(s < X_{t-1-\iota} \leq r)\right\}^2 E\left[\frac{\varepsilon_t^2}{h_t^2}\middle|\mathcal{F}_{t-1}\right]\right] \\ &\leq \frac{1}{a_0(1-|\theta_1|)}\sum_{\iota=0}^{t-1}|\theta_1|^{\iota}E\left[X_{t-1-\iota}^2 I(s < X_{t-1-\iota} \leq r)\right] \leq C_1(r-s). \end{aligned}$$

As concerns (A3) note that:

$$E\left[\left|1-\frac{\varepsilon_t^2}{h_t}\right|^2\middle|\mathcal{F}_{t-1}\right] \leq E\left[4\left(1+\frac{\varepsilon_t^4}{h_t^2}\right)\middle|\mathcal{F}_{t-1}\right] = 4E\left[1+\frac{z_t^4h_t^2}{h_t^2}\middle|\mathcal{F}_{t-1}\right] = 4(1+E[z_t^4]) < \infty.$$

The law of iterated expectations and Jensen's inequality imply that

$$\begin{aligned} E\left[\left|g_t^{(2)}(r,s)\right|^2\right] &= E\left[\left|\frac{a_1}{h_t}\left(1-\frac{\varepsilon_t^2}{h_t}\right)\sum_{\tau=0}^{t-1}b_1^{\tau}\varepsilon_{t-1-\tau}\sum_{\iota=0}^{t-2-\tau}\theta_1^{\iota}X_{t-2-\tau-\iota}I(s < X_{t-2-\tau-\iota} \leq r)\right|^2\right] \\ &= a_1^2E\left[\frac{1}{h_t^2}\left|\sum_{\tau=0}^{t-1}b_1^{\tau}\varepsilon_{t-1-\tau}\sum_{\iota=0}^{t-2-\tau}\theta_1^{\iota}X_{t-2-\tau-\iota}I(s < X_{t-2-\tau-\iota} \leq r)\right|^2 E\left[\left|1-\frac{\varepsilon_t^2}{h_t}\right|^2\middle|\mathcal{F}_{t-1}\right]\right] \\ &\leq \mathcal{C}E\left[\frac{1}{h_t^2}\left|\sum_{\tau=0}^{t-1}b_1^{\tau}\varepsilon_{t-1-\tau}\sum_{\iota=0}^{t-2-\tau}\theta_1^{\iota}X_{t-2-\tau-\iota}I(s < X_{t-2-\tau-\iota} \leq r)\right|^2\right] \\ &\leq \mathcal{C}\sum_{\tau=0}^{t-1}|b_1|^{\tau}\sum_{\iota=0}^{t-2-\tau}|\theta_1|^{\iota}E\left[\varepsilon_{t-1-\tau}^2X_{t-2-\tau-\iota}^2I(s < X_{t-2-\tau-\iota} \leq r)\right] \\ &= \mathcal{C}\sum_{\tau=0}^{t-1}|b_1|^{\tau}\sum_{\iota=0}^{t-2-\tau}|\theta_1|^{\iota}E\left[X_{t-2-\tau-\iota}^2I(s < X_{t-2-\tau-\iota} \leq r)E\left[\varepsilon_{t-1-\tau}^2\middle|\mathcal{F}_{t-2-\tau}\right]\right] \\ &= \mathcal{C}\sum_{\tau=0}^{t-1}|b_1|^{\tau}\sum_{\iota=0}^{t-2-\tau}|\theta_1|^{\iota}E\left[h_{t-1-\tau}X_{t-2-\tau-\iota}^2I(s < X_{t-2-\tau-\iota} \leq r)\right]. \end{aligned} \tag{A4}$$

For each $k \in N, k \neq 0$, consider $E[h_{t-1-\tau}X^2_{t-k}I(s < X_{t-k} \leq r)]$. Since, with the convention that $\prod_{i=1}^{0} \cdot = 1$,

$$h_t = a_0 \sum_{j=0}^{k-1} \prod_{i=1}^{j} \left(a_1 z^2_{t-i} + b_1\right) + \prod_{i=1}^{k} \left(a_1 z^2_{t-i} + b_1\right) h_{t-k}$$

it holds that

$$\begin{aligned} &E[h_{t-1-\tau}X^2_{t-k}I(s < X_{t-k} \leq r)] \\ &= E\left[\left\{ a_0 \sum_{j=0}^{k-1} \prod_{i=1}^{j} \left(a_1 z^2_{t-i} + b_1\right) + \prod_{i=1}^{k} \left(a_1 z^2_{t-i} + b_1\right) h_{t-k} \right\} X^2_{t-k} I(s < X_{t-k} \leq r)\right] \\ &= E\left[a_0 \sum_{j=0}^{k-1} \prod_{i=1}^{j} \left(a_1 z^2_{t-i} + b_1\right) X^2_{t-k} I(s < X_{t-k} \leq r)\right] \end{aligned} \tag{A5}$$

$$+ E\left[\prod_{i=1}^{k} \left(a_1 z^2_{t-i} + b_1\right) h_{t-k} X^2_{t-k} I(s < X_{t-k} \leq r)\right]. \tag{A6}$$

We prove separately that (A5) and (A6) are bounded by $\mathcal{C}(r - s)$. Since $E[z^2_{t_1} X^2_{t_2}] = E[z^2_{t_1}]E[X^2_{t_2}]$ for each $t_1 > t_2$, and $E[z^2_t] = 1$ for each t (A5) equals

$$\begin{aligned} &E\left[a_0 \sum_{j=0}^{k-1} \prod_{i=1}^{j} \left(a_1 z^2_{t-i} + b_1\right) X^2_{t-k} I(s < X_{t-k} \leq r)\right] \\ &\quad = a_0 \sum_{j=0}^{k-1} \prod_{i=1}^{j} (a_1 + b_1)\, E\left[X^2_{t-k} I(s < X_{t-k} \leq r)\right] \leq \mathcal{C}(r - s). \end{aligned}$$

Analogously, (A6) results

$$\begin{aligned} &E\left[\prod_{i=1}^{k} \left(a_1 z^2_{t-i} + b_1\right) h_{t-k} X^2_{t-k} I(s < X_{t-k} \leq r)\right] \\ &\quad = (a_1 + b_1)^k E\left[h_{t-k} X^2_{t-k} I(s < X_{t-k} \leq r)\right] \end{aligned}$$

We claim that

$$E\left[h_{t-k} X^2_{t-k} I(s < X_{t-k} \leq r)\right] \leq \mathcal{C}(r - s). \tag{A7}$$

Therefore (A6) is bounded by $(a_1 + b_1)^k \mathcal{C}(r - s)$, whereas (A4) is bounded by

$$\mathcal{C} \sum_{\tau=0}^{t-1} |b_1|^\tau \sum_{\iota=0}^{t-2-\tau} |\theta_1|^\iota \left\{(r - s) + (a_1 + b_1)^{\iota+1}(r - s)\right\} \leq \mathrm{C}_2(r - s).$$

Hence (A1) holds and the tightness of the process follows by using the same argument of Wong and Li (1997). It remains to verify the Claim A7. To this aim, define

$$\Upsilon_t^{(1)} = \phi_0 + \phi_1 X_t - \theta_1 \varepsilon_t \quad \text{and} \quad \Upsilon_t^{(2)} = \phi_0^2 + \phi_1^2 X_t^2 + \theta_1^2 \varepsilon_t^2.$$

Since $X_{t-k} = \Upsilon_{t-k-1}^{(1)} + \varepsilon_{t-k}$ and $X_{t-k}^2 \leq \mathcal{C}\left(\Upsilon_{t-k-1}^{(2)} + \varepsilon_{t-k}^2\right)$, we have that

$$\begin{aligned}
E\left[h_{t-k}X_{t-k}^2 I(s < X_{t-k} \leq r)\right] &\leq E\left[h_{t-k}\left(\Upsilon_{t-k-1}^{(2)} + \varepsilon_{t-k}^2\right) I(s < \Upsilon_{t-k-1}^{(1)} + \varepsilon_{t-k} \leq r)\right] \\
&= \mathcal{C}E\left[h_{t-k}\Upsilon_{t-k-1}^{(2)} I(s < \Upsilon_{t-k-1}^{(1)} + \varepsilon_{t-k} \leq r)\right] \\
&\quad + \mathcal{C}E\left[h_{t-k}\varepsilon_{t-k}^2 I(s < \Upsilon_{t-k-1}^{(1)} + \varepsilon_{t-k} \leq r)\right] \\
&= \mathcal{C}E\left[h_{t-k}\phi_0^2 I(s < \Upsilon_{t-k-1}^{(1)} + \varepsilon_{t-k} \leq r)\right]
\end{aligned} \tag{A8}$$

$$+ \mathcal{C}E\left[h_{t-k}\phi_1^2 X_{t-k-1}^2 I(s < \Upsilon_{t-k-1}^{(1)} + \varepsilon_{t-k} \leq r)\right] \tag{A9}$$

$$+ \mathcal{C}E\left[h_{t-k}\theta_1^2 \varepsilon_{t-k-1}^2 I(s < \Upsilon_{t-k-1}^{(1)} + \varepsilon_{t-k} \leq r)\right] \tag{A10}$$

$$+ \mathcal{C}E\left[h_{t-k}\varepsilon_{t-k}^2 I(s < \Upsilon_{t-k-1}^{(1)} + \varepsilon_{t-k} \leq r)\right] \tag{A11}$$

We prove that (A8) to (A11) are bounded by $\mathcal{C}(r-s)$. In this respect, we show that

$$\Pr\left[s < \Upsilon_{t-k-1}^{(1)} + \varepsilon_{t-k} \leq r \,\middle|\, \mathcal{F}_{t-k-1}\right] \leq \frac{\mathcal{C}}{\sqrt{a_0}}(r-s), \tag{A12}$$

Indeed:

$$\begin{aligned}
\Pr\left[s < \Upsilon_{t-k-1}^{(1)} + \varepsilon_{t-k} \leq r \,\middle|\, \mathcal{F}_{t-k-1}\right] &= \Pr\left[\frac{s - \Upsilon_{t-k-1}^{(1)}}{h_{t-k}^{1/2}} < z_{t-k} \leq \frac{r - \Upsilon_{t-k-1}^{(1)}}{h_{t-k}^{1/2}} \,\middle|\, \mathcal{F}_{t-k-1}\right] \\
&= \int_{h_{t-k}^{-1/2}(s-\Upsilon_{t-k-1}^{(1)})}^{h_{t-k}^{-1/2}(r-\Upsilon_{t-k-1}^{(1)})} f_z(x)dx \leq \mathcal{C}\frac{1}{h_{t-k}^{1/2}}(r-s) \leq \frac{\mathcal{C}}{\sqrt{a_0}}(r-s).
\end{aligned}$$

Routine algebra and (A12) imply that

$$(A8) \leq \frac{\mathcal{C}^2}{\sigma_\varepsilon}\sqrt{a_0}(r-s); \qquad (A9) \leq \frac{\mathcal{C}^2}{\sigma_\varepsilon}\sqrt{a_0}(r-s)E[h_{t-k}X_{t-k-1}^2]$$

$$(A10) \leq \frac{\mathcal{C}^2}{\sigma_\varepsilon}\sqrt{a_0}(r-s)E[h_{t-k}\varepsilon_{t-k-1}^2]; \qquad (A11) \leq \frac{\mathcal{C}^2}{\sigma_\varepsilon}\sqrt{a_0}(r-s)$$

with $\sigma_\varepsilon^2 = E[h_t]$ being the (unconditional) variance of ε_t which is finite by Assumption A.3 Moreover it is not hard to prove that $E[h_{t-k}X_{t-k-1}^2]$ and $E[h_{t-k}\varepsilon_{t-k-1}^2]$ are finite. For completeness we provide a sketch of the proof that $E[h_{t-k}X_{t-k-1}^2] < \infty$. Without loss of generality consider the case $k = 1$. Combining the MA(∞)-representation of the ARMA process, the ARCH(∞)-representation of the GARCH process with Jensen's inequality, it holds that $E[h_t X_{t-1}^2]$ is bounded by

$$\mathcal{C}E\left[\left\{\frac{a_0}{1+b_1}+a_1\sum_{j=0}^{t-1}b_1^j\varepsilon_{t-1-j}^2\right\}\right.$$
$$\left.\times\left\{\frac{\phi_0^2}{(1-|\theta_1|)^2}+\theta_1^2\sum_{i=0}^{t-2}|\phi_1|^i\varepsilon_{t-2-i}^2+\sum_{i=0}^{t-2}|\phi_1|^i\varepsilon_{t-1-i}^2\right\}\right]$$
$$\leq\mathcal{C}\left\{\frac{a_0\phi_0^2}{(1-b_1)(1-|\phi_1|)^2}+\frac{a_0(1+\theta_1^2)}{(1-b_1)(1-|\phi_1|)}\right.$$
$$\left.+\frac{a_0\phi_0^2\sigma_\varepsilon^2}{(1-b_1)(1-|\phi_1|)^2}+\frac{|\phi_1|}{(1-b_1|\phi_1|)(1-|\phi_1|(a_1+b_1))}\right\}$$

which is finite and this completes the proof of point (i). Point (ii) follows by applying the continuous mapping theorem and this completes the whole proof.