JOURNAL OF TIME SERIES ANALYSIS
J. Time Ser. Anal. **46**: 458–490 (2025)
Published online 29 August 2024 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1111/jtsa.12770

ORIGINAL ARTICLE

IMPROVED ESTIMATION OF DYNAMIC MODELS OF CONDITIONAL MEANS AND VARIANCES

WEINING WANG[a] JEFFREY M. WOOLDRIDGE[b] AND MENGSHAN XU[c]
[a]*Faculty of Economics and Business, University of Groningen, Groningen, The Netherlands*
[b]*Department of Economics, Michigan State University, East Lansing, MI, USA*
[c]*Department of Economics, University of Mannheim, Mannheim, Germany*

Using 'working' assumptions on conditional third and fourth moments of errors, we propose a method of moments estimator that can have improved efficiency over the popular Gaussian quasi-maximum likelihood estimator (GQMLE). Higher-order moment assumptions are not needed for consistency – we only require the first two conditional moments to be correctly specified – but the optimal instruments are derived under these assumptions. The working assumptions allow both asymmetry in the distribution of the standardized errors as well as fourth moments that can be smaller or larger than that of the Gaussian distribution. The approach is related to the generalized estimation equations (GEE) approach – which seeks the improvement of estimators of the conditional mean parameters by making working assumptions on the conditional second moments. We derive the asymptotic distribution of the new estimator and show that it does not depend on the estimators of the third and fourth moments. A simulation study shows that the efficiency gains over the GQMLE can be non-trivial.

Received 3 January 2022; Accepted 2 August 2024

Keywords: Dynamic models; GEE; QMLE; GARCH; optimal instrument; efficiency.

JEL. C13; C22; C32; G00.

MSC subject classification: 91B84.

1. INTRODUCTION

Nonlinear dynamic models of means, variances and covariances are routinely estimated in financial economics, macroeconomics and various other disciplines. A leading example is the set of models coming from the GARCH (generalized autoregressive conditional heteroskedasticity) and multi-variate GARCH class of models; see Bollerslev (1986). In an influential paper, Bollerslev and Wooldridge (1992) show that the GQMLE has a critical robustness property in the general multi-variate case where means, variances, and covariances can all depend on the conditioning set: provided the first two conditional moments are correctly specified, the GQMLE consistently estimates the parameters indexing the means, variances and covariances under weak regularity conditions. Moreover, Bollerslev and Wooldridge (1992) show how to estimate the asymptotic variance of the GQMLE allowing for arbitrary departures from normality (subject to the existence of enough finite moments). Many empirical papers have computed parameter estimates of the first two conditional moments using the GQMLE with fully robust SEs.

Naturally, as noted by Bollerslev and Wooldridge (1992), the GQMLE is asymptotically inefficient compared to the maximum likelihood estimator (MLE) from a correctly specified model. Of course, this requires a researcher to model the entire distribution of the outcome variables conditional on the observed covariates. A leading example in the univariate case is Bollerslev (1987), who proposes replacing the normal distribution with a *t*-distribution

*Correspondence to: Mengshan Xu, Department of Economics, University of Mannheim, L7, 3-5, 68161 Mannheim, Germany.
E-mail: mengshan.xu@uni-mannheim.de

with unknown degrees of freedom. The degrees-of-freedom parameter is estimated along with the mean and variance parameters. There are some shortcomings to this approach, the most important being that the estimators of the mean and variance parameters are generally inconsistent if the t-distribution is misspecified – which happens if the distribution is asymmetric, or if the conditional fourth moment is not proportional to the square of the variance. In the univariate setting, Newey and Steigerwald (1997) characterize the class of quasi-maximum likelihood estimators that identify parameters in correctly specified first two conditional moments. Symmetry plays an important role, as does the assumption that the standardized innovations are actually i.i.d., rather than a general martingale difference sequence (MDS) with unit conditional variances. As a practical matter, specifying a non-normal distribution is difficult with multi-variate outcomes. One possibility is to combine a marginal distribution, such as a t -distribution, with copulas, as in Patton (2006) and Fan *et al.* (2014). However, one is still making a full parametric joint distributional assumption, and the resulting MLEs are inconsistent if either the marginal distribution or the copula is misspecified.

Rather than assuming a particular distribution for the standardized innovations, one can take a semi-parametric approach to allow flexibility in the distribution. Examples in the univariate case include Engle and Gonzalez-Rivera (1991), Drost *et al.* (1997), Hafner and Rombouts (2007) and Di and Gangopadhyay (2011). However, such approaches still impose a non-trivial restriction on the distribution of the innovations: the innovations are assumed to be i.i.d. This means that the third and fourth moments are restricted to very specific functions of the conditional variance. If those restrictions fail – for example, if the asymmetry or kurtosis changes in general ways with the conditioning variables – the semi-parametric (or adaptive) estimators will be inconsistent. By contrast, the GQMLE does not require i.i.d. innovations for consistency or asymptotic normality, and neither does fully robust inference. In other words, in terms of consistency, semi-parametric methods are less robust than the GQMLE. There is a practically important difference between the standardized innovations being an MDS – which is implied by the correct specification of the first two moments – and assuming they are actually i.i.d. Relaxing the i.i.d. condition on the standardized residuals is not easy; for example, see the discussion in Komunjer and Vuong (2010). Of course, things would be even more difficult in a multi-variate setting. As a practical matter, semi-parametric approaches with multi-variate outcomes can be very difficult computationally – even if one believes that the (matrix) standardized innovations are i.i.d. In addition, proper choices of tuning parameters are crucial to the performance of these methods.

Our purpose in this article is to improve the GQMLE estimator without imposing additional (substantive) assumptions. As just summarized, any non-Gaussian MLE or semi-parametric MLE imposes assumptions beyond those used by Bollerslev and Wooldridge (1992) for consistency and asymptotic normality. In the conclusion of Bollerslev and Wooldridge (1992), the inefficiency of the GQMLE is noted, with the possibility of using the method of moments to improve efficiency suggested as a future research topic. Building on this observation, we propose an estimator that potentially has a smaller asymptotic variance than the GQMLE if some *working* assumptions hold. As in the GEE literature for estimating parameters in conditional mean functions, efficiency gains often arise even when the working assumptions fail. A key point is that the consistency of the 'optimal' instrumental variables estimator does not require these working assumptions.

We extend the GEE idea for estimating conditional means (Liang and Zeger, 1986) to the case where the parameters of the first two conditional moments are of interest and these moments are assumed to be correctly specified. To obtain optimal instrumental variables, we restrict the class of submodels. In particular, as in the GEE literature, we specify a 'working' optimal instrument matrix. This involves a 'working' conditional variance–covariance matrix for the residual function that defines the first two conditional moments. Namely, we consider an estimator that, like the GQMLE, requires only the first two moments to be correctly specified for consistency. Within this class, we want to find an estimator that is motivated by the optimal instrumental variables estimator (OPIV) but requires weaker assumptions. This novel estimator integrates essential characteristics from both OPIV and GEE approaches, and we refer to it as 'working OPIV' (WOPIV). It has the potential to be more efficient than the GQMLE when only the first two moments are correctly specified. As demonstrated through simulations in various univariate and multi-variate models, the WOPIV outperforms the GQMLE when the underlying distribution of the

DOI: 10.1111/jtsa.12770 *Journal of Time Series Analysis* published by John Wiley & Sons Ltd.

innovation term is skew normal, while both methods have similar performances when the underlying distribution is standard normal.

Our proposed WOPIV method is related to the literature on estimating function methods of dynamic models. Methodologies in a similar spirit were first developed in statistics by Durbin (1960), Godambe (1985), and Godambe and Heyde (2010); see also Heyde (1997). The theory of estimating functions with the plugged-in estimator of nuisance parameters has been applied to the estimation of ARCH-type models by Chandra and Taniguchi (2001). Compared to the above-referenced work, the WOPIV is for a more general class of dynamic models. There are a few works in the literature considering improving the estimator of dynamic models by exploiting higher-order moment conditions. Prono (2010) proposes a GMM method to obtain an efficient estimator for the GARCH(1,1) model; see also Im and Schmidt (2008) for the i.i.d. case. The highly cited work of Harvey and Siddique (1999) includes the third moment to account for the skewness in the innovation distribution. The method considered in Meddahi and Renault (1998) and Li and Turtle (2000) can be regarded as special cases of our framework in the univariate ARCH-type models. An important advantage of our proposed WOPIV method over those based on GMM is that instead of estimating the entire set of optimal instruments, the WOPIV solely relies on the estimated third and fourth moments to construct the instruments, and inconsistent estimators of these moments will not affect the consistency of the WOPIV. In contrast, the consistent estimation of high-order moments is crucial for the above-mentioned GMM-based methods.

The rest of the article is organized as follows. In Section 2, we show the basic univariate framework of the proposed WOPIV. In Section 3, we present its theoretical properties. In Section 4, we extend our approach to multi-variate models. In Section 5, we discuss the performance of the WOPIV in comparison to other relevant methods. Simulations and applications are in Sections 6 and 7. The proofs and other technical details of the WOPIV, as well as tables and figures, are presented in the Appendices.

2. UNIVARIATE MODELS

We start with the univariate case. Let y_t (for $t = 1, \dots, T$) be a scalar response, and $\mathbf{x}_t$ be a vector of conditioning variables, which is finite-dimensional and can generally include lagged values of y_t. $\mathbf{x}_t$ can also include contemporaneous values of some other series, say $\mathbf{z}_t$, as well as lags of $\mathbf{z}_t$.

Let Θ be the parameter set where θ takes its value, and we use $m_t(\mathbf{x}_t, \boldsymbol{\theta})$ and $v_t(\mathbf{x}_t, \boldsymbol{\theta})$, where $v_t(\mathbf{x}_t, \boldsymbol{\theta}) > 0$ for all $\mathbf{x}_t$ and $\boldsymbol{\theta} \in \Theta$, to denote the conditional mean and variance respectively. We assume that the first two moments are correctly specified: for some $\boldsymbol{\theta}_0 \in \Theta$,

$$\mathrm{E}(y_t|\mathbf{x}_t) = m_t(\mathbf{x}_t, \boldsymbol{\theta}_0), \tag{1}$$

$$\mathrm{Var}(y_t|\mathbf{x}_t) = v_t(\mathbf{x}_t, \boldsymbol{\theta}_0). \tag{2}$$

The setup of this model is general, allowing for variance parameters and mean parameters to be completely separate or to overlap. An example of the latter is the ARCH-in-mean type model. See, for example, Engle *et al.* (1987).

Conditions (1) and (2) are standard regularity conditions for dynamic models. However, it is traditional in the settings of interest to assume that the models are dynamically complete in mean and variance:

$$\mathrm{E}(y_t|\mathbf{x}_t, \mathcal{F}_{t-1}) = \mathrm{E}(y_t|\mathbf{x}_t), \tag{3}$$

$$\mathrm{Var}(y_t|\mathbf{x}_t, \mathcal{F}_{t-1}) = \mathrm{Var}(y_t|\mathbf{x}_t), \tag{4}$$

where

$$\mathcal{F}_{t-1} = \sigma(y_{t-1}, \mathbf{x}_{t-1}, y_{t-2}, \mathbf{x}_{t-2}, \dots, y_1, \mathbf{x}_1, y_0, \mathbf{x}_0, \dots),$$

Journal of Time Series Analysis published by John Wiley & Sons Ltd. DOI: 10.1111/jtsa.12770

is the filtration corresponding to time $t-1$. A dynamic completeness assumption in the first two moments is always assumed in ARCH and GARCH models and their extensions. Following the convention, we assume that the first two conditional moments are correctly specified and dynamically complete. This is the assumption imposed in Bollerslev and Wooldridge (1992).

Example 1. The GARCH(1,1) model:

$$\begin{aligned} \varepsilon_t &= \sigma_t \eta_t, \\ \sigma_t^2 &= \omega_0 + \alpha_0 \varepsilon_{t-1}^2 + \beta_0 \sigma_{t-1}^2, \end{aligned} \tag{5}$$

where η_t's are i.i.d. innovations, and $\alpha_0, \beta_0 > 0$, $\alpha_0 + \beta_0 < 1$.

In this example, $y_t = \varepsilon_t$ and $\mathbf{x}_t = (\varepsilon_{t-1}, \sigma_{t-1})$. We shall note that usually for the GARCH model, a joint Markovian assumption holds: the process $\mathbf{x}_t$ is Markovian, but ε_t is not. In addition, $\mathbf{x}_t = \left(\varepsilon_{t-1}, \sigma_{t-1}\right)$ is only partly observed because σ_{t-1} is unobserved; it depends on θ_0 and $\{\varepsilon_{t-i}, i > 1\}$, i.e., an infinite number of past values of the process. We will discuss the estimation strategy in Section 2.2.

By the dynamic completeness assumption, the optimal instrumental variables derived for the WOPIV depend solely on $\mathbf{x}_t$. As we permit lags of y_t or $\mathbf{z}_t$ to be included in $\mathbf{x}_t$, there is no longer a need to use lags of $\mathbf{x}_t$, such as $\mathbf{x}_{t-1}$, for constructing the instruments at time t. In this way, the WOPIV becomes a direct extension of the GQMLE, which can be viewed as a particular instrumental variables estimator (IV) whose instruments depend only on $\mathbf{x}_t$.

In an effort to improve the GQMLE method, we proceed to define the error term as well as the standardized error as

$$\begin{aligned} u_t &= y_t - m_t(\mathbf{x}_t, \boldsymbol{\theta}_0), \\ e_t &= \frac{u_t}{\sqrt{v_t(\mathbf{x}_t, \boldsymbol{\theta}_0)}}. \end{aligned}$$

By construction,

$$\begin{aligned} \mathrm{E}\left(u_t|\mathbf{x}_t\right) = 0,\ \mathrm{Var}\left(u_t|\mathbf{x}_t\right) &= v_t(\mathbf{x}_t, \boldsymbol{\theta}_0), \\ \mathrm{E}\left(e_t|\mathbf{x}_t\right) = 0,\ \mathrm{Var}\left(e_t|\mathbf{x}_t\right) &= 1, \end{aligned}$$

and these conditional moments continue to hold conditional on $\mathcal{F}_{t-1}$. It is important to observe that e_t is not guaranteed to be even independent of $\mathbf{x}_t$, let alone its further lags before time t. Treatments such as Newey and Steigerwald (1997) make the strong assumption (which is not required in our article):

$$e_t \text{ is independent of } \left(\mathbf{x}_t, \mathbf{x}_{t-1}, \ldots\right),\ t = 1, 2, \ldots \tag{6}$$

In contrast, conditions (3) and (4) require only that $\{e_t\}_{t=1}^T$ is an MDS. As discussed in Bollerslev and Wooldridge (1992), the correct specification of the first two conditional moments implies that the vector of score functions of the quasi-log likelihood (evaluated at θ_0) forms an MDS. Along with weak dependence requirements, the MDS properties ensure that the GQMLE is $\sqrt{T}$-asymptotically normal. Condition (6) can be used to simplify the verification of regularity conditions, but it has no substantive effect on the asymptotic properties of the GQMLE. See also Wooldridge (1994) for a more general discussion.

To obtain a simple estimator that can be more efficient than the GQMLE, we start with the following two conditions:

$$\mathrm{E}\left(e_t^3|\mathbf{x}_t\right) = \mathrm{E}\left(e_t^3\right) \equiv \kappa_3^0, \tag{7}$$

DOI: 10.1111/jtsa.12770 *Journal of Time Series Analysis* published by John Wiley & Sons Ltd.

$$\mathrm{E}\left(e_t^4|\mathbf{x}_t\right) = \mathrm{E}\left(e_t^4\right) \equiv \kappa_4^0, \tag{8}$$

where κ_3^0 and κ_4^0 are two constants. Written in terms of the errors u_t,

$$\mathrm{E}\left(u_t^3|\mathbf{x}_t\right) = \kappa_3^0\left[v_t(\mathbf{x}_t, \boldsymbol{\theta}_0)\right]^{3/2}, \tag{9}$$

$$\mathrm{E}\left(u_t^4|\mathbf{x}_t\right) = \kappa_4^0\left[v_t(\mathbf{x}_t, \boldsymbol{\theta}_0)\right]^2. \tag{10}$$

It shall be noted that conditions (7) and (8) are not really assumptions imposed on the proposed WOPIV method. Rather, they are introduced to motivate the construction of the working variance–covariance matrix of the conditional moment restrictions shown below. However, it is worth noting that our method implicitly assumes the existence of the third and fourth moments of e_t. Therefore we shall not target applications, where the third and the fourth moments do not exist, for example, the cases considered in Fan *et al.* (2014). Under the assumption of normality, we have $\kappa_3^0 = 0$ and $\kappa_4^0 = 3$. Bollerslev and Wooldridge (1992) show that neither of these restrictions is necessary for the consistency of the GQMLE. In fact, neither is the assumption that these $\mathrm{E}\left(e_t^3|\mathbf{x}_t\right)$ and $\mathrm{E}\left(e_t^4|\mathbf{x}_t\right)$ are constant. For the estimators here, we use (7) and (8) to derive working optimal instruments for estimating $\boldsymbol{\theta}_0$, but these conditions are not required for the consistency of the WOPIV. Later, we will need to estimate κ_3^0 and κ_4^0, but this is easily done given an initial preliminary estimator of $\boldsymbol{\theta}_0$, which is typically the GQMLE. In deriving the asymptotic properties, we will only assume that the estimators converge to some constants without invoking (7) or (8).

Our proposed WOPIV is motivated by finding the working optimal instruments under the assumption that the model is dynamically complete in the first two moments and the auxiliary conditions (7) and (8). For each t and $\mathbf{w}_t \overset{\text{def}}{=} (y_t, \mathbf{x}_t')'$, define the 2×1 matrix of residual functions:

$$\mathbf{r}_t(\mathbf{w}_t, \boldsymbol{\theta}) = \begin{pmatrix} y_t - m_t(\mathbf{x}_t, \boldsymbol{\theta}) \\ \left[y_t - m_t(\mathbf{x}_t, \boldsymbol{\theta})\right]^2 - v_t(\mathbf{x}_t, \boldsymbol{\theta}) \end{pmatrix}.$$

If the model is dynamically complete, then

$$\mathrm{E}\left[\mathbf{r}_t(\mathbf{w}_t, \boldsymbol{\theta}_0)|\mathbf{x}_t\right] = \mathrm{E}\left[\mathbf{r}_t(\mathbf{w}_t, \boldsymbol{\theta}_0)|\mathbf{x}_t, \mathcal{F}_{t-1}\right] = \mathbf{0}.$$

As discussed in Wooldridge (1994), the optimal instrumental variables based on these moment conditions depend on $\mathrm{E}\left[\nabla_{\boldsymbol{\theta}}\mathbf{r}_t(\mathbf{w}_t, \boldsymbol{\theta}_0)|\mathbf{x}_t\right]$ and $\mathrm{Var}\left[\mathbf{r}_t(\mathbf{w}_t, \boldsymbol{\theta}_0)|\mathbf{x}_t\right]$. Let P denote the dimension of θ_0. Under the correct specification of the first and the second moments, we can obtain the $2 \times P$ matrix

$$\mathbf{R}_t(\mathbf{x}_t, \boldsymbol{\theta}_0) \equiv \mathrm{E}\left[\nabla_{\boldsymbol{\theta}}\mathbf{r}_t(\mathbf{w}_t, \boldsymbol{\theta}_0)|\mathbf{x}_t\right] = -\begin{pmatrix} \nabla_{\boldsymbol{\theta}} m_t(\mathbf{x}_t, \boldsymbol{\theta}_0) \\ \nabla_{\boldsymbol{\theta}} v_t(\mathbf{x}_t, \boldsymbol{\theta}_0) \end{pmatrix}.$$

Next, $\mathrm{Var}\left[\mathbf{r}_t(\mathbf{w}_t, \boldsymbol{\theta}_0)|\mathbf{x}_t\right]$ can generally be any positive semi-definite matrix function of $\mathbf{x}_t$, making it difficult to implement an always efficient IV estimator. Our key innovation is to impose a working version of $\mathrm{Var}\left[\mathbf{r}_t(\mathbf{w}_t, \boldsymbol{\theta}_0)|\mathbf{x}_t\right]$, where we borrow the term 'working' from the GEE literature (for example, Zeger and Liang, 1986). In particular, if we impose (7) and (8), then

$$\mathbf{D}_t(\mathbf{x}_t, \boldsymbol{\theta}_0, \boldsymbol{\kappa}_0) \equiv \mathrm{Var}\left[\mathbf{r}_t(\mathbf{w}_t, \boldsymbol{\theta}_0)|\mathbf{x}_t\right] = \begin{pmatrix} v_t(\mathbf{x}_t, \boldsymbol{\theta}_0) & \kappa_3^0\left[v_t(\mathbf{x}_t, \boldsymbol{\theta}_0)\right]^{3/2} \\ \kappa_3^0\left[v_t(\mathbf{x}_t, \boldsymbol{\theta}_0)\right]^{3/2} & \left(\kappa_4^0 - 1\right)\left[v_t(\mathbf{x}_t, \boldsymbol{\theta}_0)\right]^2 \end{pmatrix},$$

Journal of Time Series Analysis published by John Wiley & Sons Ltd. DOI: 10.1111/jtsa.12770

where $\kappa_0 \overset{\text{def}}{=} (\kappa_3^0, \kappa_4^0)'$. Rather than being unrestricted, $\mathbf{D}_t(\mathbf{x}_t, \boldsymbol{\theta}_0, \boldsymbol{\kappa}_0)$ has a relatively simple form and depends on only two additional parameters. Under normality or other symmetric distributions, this structure holds with $\kappa_3^0 = 0$. Specific distributions also imply values for κ_4^0, or in some cases (such as the skew normal distribution or the t-distribution) we treat it as a parameter to be estimated using an MLE approach. We can use this structure to obtain an estimator that has the potential to be more efficient than the GQMLE. As discussed in the GEE literature, we expect efficiency will carry over even if we drop (7) and (8).

As in Wooldridge (1994) (Page 61, equation (7.32)), the working optimal instruments, obtained only from the moments conditional on $\mathbf{x}_t$, are

$$\mathbf{Z}_t \overset{\text{def}}{=} \mathbf{D}_t(\mathbf{x}_t, \boldsymbol{\theta}_0, \boldsymbol{\kappa}_0)^{-1}\mathbf{R}_t(\mathbf{x}_t, \boldsymbol{\theta}_0).$$

In practice, obtaining this working optimal instrument matrix for some models may be infeasible since $\mathbf{x}_t$ might depend on the unknown parameter θ_0 or unobserved historical values, such as y_t with $t \leq 0$. Thus, we shall distinguish estimation strategies for fully observed $\mathbf{x}_t$ and for partially unobserved $\mathbf{x}_t$. The detailed estimation steps in these two cases are outlined in the following two subsections.

2.1. $\mathbf{x}_t$ is fully observed

To implement the WOPIV, we need initial estimators of $\boldsymbol{\theta}_0$ and $\boldsymbol{\kappa}_0$, which are denoted by $\check{\boldsymbol{\theta}}$ and $\check{\kappa}$. For $\boldsymbol{\theta}_0$, an obvious choice is the GQMLE. For κ_0, we employ a method-of-moments estimator to obtain the standardized residuals:

$$\check{e}_t = \frac{\check{u}_t}{\sqrt{v_t(\mathbf{x}_t, \check{\boldsymbol{\theta}})}},$$

where $\check{u}_t = y_t - m_t(\mathbf{x}_t, \check{\boldsymbol{\theta}})$. Then, we can obtain

$$\check{\kappa}_3 \overset{\text{def}}{=} \frac{1}{T}\sum_{t=1}^{T}\check{e}_t^3 = \frac{1}{T}\sum_{t=1}^{T}\left[\frac{\check{u}_t}{\sqrt{v_t(\mathbf{x}_t, \check{\boldsymbol{\theta}})}}\right]^3. \tag{11}$$

Next, define $\tau_4^0 = \kappa_4^0 - 1$, such that $\tau_4^0 = \mathrm{E}\left[\left(e_t^2 - 1\right)^2\right]$. Then, a method-of-moments estimator that ensures non-negativity is

$$\check{\tau}_4 \overset{\text{def}}{=} \frac{1}{T}\sum_{t=1}^{T}\left(\check{e}_t^2 - 1\right)^2 = \frac{1}{T}\sum_{t=1}^{T}\left[\frac{\check{u}_t^2}{v_t(\mathbf{x}_t, \check{\theta})} - 1\right]^2. \tag{12}$$

Given the preliminary estimators $\check{\theta}$ and $\check{\kappa} \overset{\text{def}}{=} (\check{\kappa}_3, \check{\kappa}_4) = (\check{\kappa}_3, \check{\tau}_4 + 1)$, we can obtain estimators of the working optimal instruments under the correct model specification, the dynamic completeness, and (7) and (8):

$$\check{\mathbf{Z}}_t = \check{\mathbf{D}}_t^{-1}\check{\mathbf{R}}_t, \tag{13}$$

where $\check{\mathbf{D}}_t = \mathbf{D}_t(\mathbf{x}_t, \check{\theta}, \check{\kappa})$ and $\check{\mathbf{R}}_t = \mathbf{R}_t(\mathbf{x}_t, \check{\theta})$. The WOPIV is obtained by solving θ from

$$\sum_{t=1}^{T}\check{\mathbf{Z}}_t'\mathbf{r}_t(\mathbf{w}_t, \theta) = \mathbf{0}, \tag{14}$$

which is a set of P nonlinear equations with P unknowns in $\boldsymbol{\theta}$. We use $\hat{\theta}$ to denote its solution. It is worth noting that the existence of a solution to (14) is not a trivial issue. See, e.g., Definition 2.1 and the accompanying remarks in Jacod and Sørensen (2018) for a further discussion.

2.2. $\mathbf{x}_t$ is partially observed

If $\mathbf{x}_t$ is not fully observed, it can be denoted by $\mathbf{x}_t \equiv \mathbf{x}_t(\theta_0) = \left(\mathbf{x}_{1t}', \mathbf{x}_{2t}(\theta_0)'\right)'$, where $\mathbf{x}_{1t}$ is fully observed, and $\mathbf{x}_{2t}(\theta_0)$ depends on the unknown θ_0 and potentially infinite number of observation lags. However, in practice, only historical realizations $(y_t, \mathbf{x}_t)$ for $t \geq 1$ are available. Thus, we need to replace the unobserved part $\mathbf{x}_{2t}(\theta_0)$ with a proxy containing finite lags of available observations and some initial values. Let $\tilde{\mathbf{x}}_{2t}(\theta)$ be a feasible version of $\mathbf{x}_{2t}(\theta_0)$ that is generated with a given θ and an initial value $\mathbf{w}_0$ (recall that $\mathbf{w}_t \overset{\text{def}}{=} (y_t, \mathbf{x}_t')'$). We define $\tilde{\mathbf{x}}_t(\theta) = \left(\mathbf{x}_{1t}', \tilde{\mathbf{x}}_{2t}(\theta)'\right)'$ and $\tilde{\mathbf{w}}_t(\theta) = \left(y_t, \tilde{\mathbf{x}}_t(\theta)'\right)'$. Next, we define

$$\mathbf{r}_t(\tilde{\mathbf{w}}_t(\theta), \boldsymbol{\theta}) = \begin{pmatrix} y_t - m_t(\tilde{\mathbf{x}}_t(\theta), \boldsymbol{\theta}) \\ \left[y_t - m_t(\tilde{\mathbf{x}}_t(\theta), \boldsymbol{\theta})\right]^2 - v_t(\tilde{\mathbf{x}}_t(\theta), \boldsymbol{\theta}) \end{pmatrix}.$$

And for

$$\tilde{u}_t = y_t - m_t(\tilde{\mathbf{x}}_t(\check{\boldsymbol{\theta}}), \check{\boldsymbol{\theta}}), \tilde{e}_t = \frac{\tilde{u}_t}{\sqrt{v_t(\tilde{\mathbf{x}}_t(\check{\boldsymbol{\theta}}), \check{\boldsymbol{\theta}})}},$$

where $\check{\boldsymbol{\theta}}$ is a consistent initial estimator of θ_0. Then, we have

$$\tilde{\kappa}_3 = \frac{1}{T}\sum_{t=1}^{T}\left[\frac{\tilde{u}_t}{\sqrt{v_t(\tilde{\mathbf{x}}_t(\check{\boldsymbol{\theta}}), \check{\boldsymbol{\theta}})}}\right]^3, \tilde{\tau}_4 = \frac{1}{T}\sum_{t=1}^{T}\left[\frac{\tilde{u}_t^2}{v_t(\tilde{\mathbf{x}}_t(\check{\boldsymbol{\theta}}), \check{\boldsymbol{\theta}})} - 1\right]^2.$$

Now, a feasible sample score function is

$$\sum_{t=1}^{T} \tilde{\mathbf{Z}}_t' \mathbf{r}_t(\tilde{\mathbf{w}}_t(\theta), \boldsymbol{\theta}) = \mathbf{0}, \tag{15}$$

where $\tilde{\mathbf{Z}}_t \overset{\text{def}}{=} \tilde{\mathbf{D}}_t^{-1}\tilde{\mathbf{R}}_t$, $\tilde{\mathbf{D}}_t \overset{\text{def}}{=} \mathbf{D}_t(\tilde{\mathbf{x}}_t(\check{\boldsymbol{\theta}}), \check{\boldsymbol{\theta}}, \tilde{\kappa})$, $\tilde{\mathbf{R}}_t \overset{\text{def}}{=} \mathbf{R}_t(\tilde{\mathbf{x}}_t(\check{\boldsymbol{\theta}}), \check{\boldsymbol{\theta}})$, and $\tilde{\kappa} \overset{\text{def}}{=} (\tilde{\kappa}_3, \tilde{\kappa}_4) = (\tilde{\kappa}_3, \tilde{\tau}_4 + 1)$.

To show the asymptotic properties of the solution to (15), we can follow a strategy that is commonly applied in the GARCH literature (e.g., p. 143 of Francq and Zakoian (2010)). For a given θ, let $\mathbf{x}_t(\theta)$ (and $\mathbf{w}_t(\theta)$) be series that depend on θ and possibly unobserved historical values prior to the time point $t = 1$. With some abuse of notations, let $\check{\kappa}$ be defined similarly to (11) and (12) except that in those formulae, $\mathbf{x}_t$'s are replaced by $\mathbf{x}_t(\check{\boldsymbol{\theta}})$'s. Similarly, we can define $\check{\mathbf{D}}_t = \mathbf{D}_t(\mathbf{x}_t(\check{\boldsymbol{\theta}}), \check{\boldsymbol{\theta}}, \check{\kappa})$, $\check{\mathbf{R}}_t = \mathbf{R}_t(\mathbf{x}_t(\check{\boldsymbol{\theta}}), \check{\boldsymbol{\theta}})$, and $\check{\mathbf{Z}}_t = \check{\mathbf{D}}_t^{-1}\check{\mathbf{R}}_t$. In Section 3, we assume that the estimator from (15) has the same first-order asymptotic behavior as the one from $\sum_{t=1}^{T}\check{\mathbf{Z}}_t'\mathbf{r}_t(\mathbf{w}_t(\theta), \theta) = \mathbf{0}$, and then we show the asymptotic properties of the latter. In Appendix B.3, we show that for the case of GARCH(1,1), this assumption is valid, i.e., the impact caused by initial values vanishes as T goes to infinity.

Example 1 (continued)**.** For the GARCH(1,1) model, we have

$$\mathbf{r}_t(\mathbf{w}_t, \theta_0) = \begin{pmatrix} \varepsilon_t \\ \varepsilon_t^2 - (\omega_0 + \alpha_0\varepsilon_{t-1}^2 + \beta_0\sigma_{t-1}^2) \end{pmatrix}, \mathbf{R}_t = -\begin{pmatrix} 0 & 0 & 0 \\ 1 & \varepsilon_{t-1}^2 & \sigma_{t-1}^2 \end{pmatrix},$$

Journal of Time Series Analysis published by John Wiley & Sons Ltd. DOI: 10.1111/jtsa.12770

and

$$\mathbf{D}_t(\mathbf{x}_t,\theta_0,\kappa_0) \equiv \mathrm{Var}\left[\mathbf{r}_t(\mathbf{w}_t,\theta_0)|\mathbf{x}_t\right] = \begin{pmatrix} v_t(\mathbf{x}_t,\theta_0) & \kappa_3^0[v_t(\mathbf{x}_t,\theta_0)]^{3/2} \\ \kappa_3^0[v_t(\mathbf{x}_t,\theta_0)]^{3/2} & (\kappa_4^0-1)[v_t(\mathbf{x}_t,\theta_0)]^2 \end{pmatrix}$$
$$= \begin{pmatrix} \sigma_t^2 & \kappa_3^0\sigma_t^3 \\ \kappa_3^0\sigma_t^3 & (\kappa_4^0-1)\sigma_t^4 \end{pmatrix}.$$

Let $\mathbf{D}_t \overset{\text{def}}{=} \mathbf{D}_t(\mathbf{x}_t,\theta_0,\kappa_0)$, we have

$$\mathbf{D}_t^{-1} = \frac{1}{(\kappa_4^0-1)\sigma_t^6-(\kappa_3^0)^2\sigma_t^6}\begin{pmatrix} (\kappa_4^0-1)\sigma_t^4 & -\kappa_3^0\sigma_t^3 \\ -\kappa_3^0\sigma_t^3 & \sigma_t^2 \end{pmatrix}$$
$$= \frac{1}{c_\kappa^0}\begin{pmatrix} (\kappa_4^0-1)\sigma_t^{-2} & -\kappa_3^0\sigma_t^{-3} \\ -\kappa_3^0\sigma_t^{-3} & \sigma_t^{-4} \end{pmatrix},$$

where $c_\kappa^0 = \kappa_4^0 - 1 - (\kappa_3^0)^2$. Then

$$\mathbf{D}_t^{-1}\mathbf{R}_t = -\frac{1}{c_\kappa^0}\begin{pmatrix} (\kappa_4^0-1)\sigma_t^{-2} & -\kappa_3^0\sigma_t^{-3} \\ -\kappa_3^0\sigma_t^{-3} & \sigma_t^{-4} \end{pmatrix}\begin{pmatrix} 0 & 0 & 0 \\ 1 & \varepsilon_{t-1}^2 & \sigma_{t-1}^2 \end{pmatrix}$$
$$= -\frac{1}{c_\kappa^0}\begin{pmatrix} -\kappa_3^0\sigma_t^{-3} & -\kappa_3^0\sigma_t^{-3}\varepsilon_{t-1}^2 & -\kappa_3^0\sigma_t^{-3}\sigma_{t-1}^2 \\ \sigma_t^{-4} & \sigma_t^{-4}\varepsilon_{t-1}^2 & \sigma_t^{-4}\sigma_{t-1}^2 \end{pmatrix}. \tag{16}$$

Recall that in Example 1, we have that $\mathbf{x}_t = \left(\mathbf{x}_{1t}', \mathbf{x}_{2t}(\theta_0)'\right)' = \left(\varepsilon_{t-1}, \sigma_{t-1}\right)$ is only partially observed since $\sigma_{t-1}^2 = \omega_0 + \alpha_0\varepsilon_{t-2}^2 + \beta_0\sigma_{t-2}^2 = \frac{\omega_0}{1-\beta_0} + \alpha_0\sum_{k=0}^{\infty}\beta_0^k\varepsilon_{t-2-k}^2$ – the last infinite sum depends on the unknown parameter θ_0 and involves unobservable historical values. To construct the working optimal instrument matrix, we have to apply a feasible version of σ_t^2. A possible choice is to generate $\tilde{\sigma}_t^2(\theta)$: for $t = 1, \ldots, T$, we have

$$\mathbf{x}_{1t} = \varepsilon_{t-1},$$
$$\tilde{\mathbf{x}}_{2t}(\theta) = \tilde{\sigma}_{t-1}(\theta) = \left(\frac{\omega}{1-\beta} + \alpha\sum_{k=0}^{t-1}\beta^k\varepsilon_{t-2-k}^2\right)^{1/2}, \tag{17}$$

for some initial values $(\varepsilon_0^2, \varepsilon_{-1}^2)$ and a given $\theta = (\omega,\alpha,\beta)'$ satisfying $\alpha > 0$, $\beta > 0$ and $\alpha + \beta < 1$. In general, the impact of initial values is asymptotically vanishing, as discussed in Appendix B.3.

Based on (17), we can write down $\tilde{\mathbf{x}}_t(\theta)$ and $\tilde{\mathbf{w}}_t(\theta)$. Then, we have

$$\mathbf{r}_t(\tilde{\mathbf{w}}_t(\theta),\theta) = \begin{pmatrix} \varepsilon_t \\ \varepsilon_t^2 - (\omega + \alpha\varepsilon_{t-1}^2 + \beta\tilde{\sigma}_{t-1}^2(\theta)) \end{pmatrix},$$
$$\tilde{\mathbf{Z}}_t = \tilde{\mathbf{D}}_t^{-1}\tilde{\mathbf{R}}_t = -\frac{1}{\tilde{c}_\kappa}\begin{pmatrix} -\tilde{\kappa}_3\tilde{\sigma}_t^{-3}(\check{\theta}) & -\tilde{\kappa}_3\tilde{\sigma}_t^{-3}(\check{\theta})\varepsilon_{t-1}^2 & -\tilde{\kappa}_3\tilde{\sigma}_t^{-3}(\check{\theta})\tilde{\sigma}_{t-1}^2(\check{\theta}) \\ \tilde{\sigma}_t^{-4}(\check{\theta}) & \tilde{\sigma}_t^{-4}(\check{\theta})\varepsilon_{t-1}^2 & \tilde{\sigma}_t^{-4}(\check{\theta})\tilde{\sigma}_{t-1}^2(\check{\theta}) \end{pmatrix},$$

DOI: 10.1111/jtsa.12770 *Journal of Time Series Analysis* published by John Wiley & Sons Ltd.

where $\tilde{c}_\kappa = \tilde{\kappa}_4 - 1 - \tilde{\kappa}_3^2$. Finally, the WOPIV can be obtained by solving $\theta = (\omega, \alpha, \beta)'$ from

$$\sum_{t=1}^{T} \tilde{\mathbf{Z}}_t' \mathbf{r}_t(\tilde{\mathbf{w}}_t(\theta), \theta) = -\frac{1}{\tilde{c}_\kappa} \sum_{t=1}^{T} \begin{pmatrix} \frac{\varepsilon_t^2 - (\omega + \alpha\varepsilon_{t-1}^2 + \beta\tilde{\sigma}_{t-1}^2(\theta))}{\tilde{\sigma}_t^4(\check{\theta})} - \tilde{\kappa}_3 \tilde{\sigma}_t^{-3}(\check{\theta})\varepsilon_t \\ \frac{\varepsilon_t^2 - (\omega + \alpha\varepsilon_{t-1}^2 + \beta\tilde{\sigma}_{t-1}^2(\theta))}{\tilde{\sigma}_t^4(\check{\theta})}\varepsilon_{t-1}^2 - \tilde{\kappa}_3 \tilde{\sigma}_t^{-3}(\check{\theta})\varepsilon_{t-1}^2\varepsilon_t \\ \frac{\varepsilon_t^2 - (\omega + \alpha\varepsilon_{t-1}^2 + \beta\tilde{\sigma}_{t-1}^2(\theta))}{\tilde{\sigma}_t^4(\check{\theta})}\tilde{\sigma}_{t-1}^2(\check{\theta}) - \tilde{\kappa}_3 \tilde{\sigma}_t^{-3}(\check{\theta})\tilde{\sigma}_{t-1}^2(\check{\theta})\varepsilon_t \end{pmatrix} = \mathbf{0}. \tag{18}$$

2.3. Some remarks

It is true that, under the correct specification of the first two moments (dynamic completeness is not needed for this), the preliminary estimation of $\boldsymbol{\theta}_0$ and $\boldsymbol{\kappa}_0$ does not affect the consistency of the WOPIV. In fact, in the formal statement of our result, we drop (7) and (8), and simply assume that $\check{\theta}$ and $\check{\kappa}$ converge in probability to some constant vectors; see the remark following Theorem 1 in Section 3.1. The proposed estimator will not be the OPIV without (7) and (8), but it can still be asymptotically more efficient than the GQMLE. Intuitively, the WOPIV plugs in sample averages of the third and the fourth sample moments of residuals that are normalized by the square root of the conditional second moment, while the GQMLE implicitly imposes $\kappa_3^0 = 0$ and $\kappa_4^0 = 3$. Therefore, we expect the WOPIV to outperform the GQMLE in many cases. We illustrate in Section 5.1 the scenarios in which the WOPIV could potentially be more efficient than the GQMLE.

To avoid solving the nonlinear first-order condition, we can consider a one-step estimation. The idea of a one-step estimator is to consider a linear approximation to the moment condition in (14) or (15). Provided that we have a well-chosen initial estimator $\check{\theta}$ (for example, a consistent GQMLE), we can improve on the estimator $\check{\theta}$ using a one-step procedure. Taking (14) as an example, the corresponding one-step estimator can be obtained by

$$\overline{\theta} = \check{\theta} - \left[\sum_{t=1}^{T} \check{\mathbf{Z}}_t' \nabla \mathbf{r}_t(\mathbf{w}_t, \check{\theta}) \right]^{-1} \sum_{t=1}^{T} \check{\mathbf{Z}}_t' \mathbf{r}_t(\mathbf{w}_t, \check{\theta}). \tag{19}$$

Suppose that $\check{\theta}$ is a consistent and asymptotically normal estimator, and $\hat{\theta}$ is the solution to (14), an 'ideal' consistent estimator for θ_0. Then, $\overline{\theta}$ is asymptotically first-order equivalent to $\hat{\theta}$.

3. THEOREMS

Here, we outline the assumptions and theoretical properties of the proposed WOPIV.

3.1. Consistency

Our general estimation equation is defined on a compact parameter space $\Theta \times \Gamma$. In this space, $\theta \in \Theta$ is the parameter of interest, and $\kappa = (\kappa_3, \kappa_4) \in \Gamma$ is a set of nuisance parameters. Recall the notations defined in Section 2.2: For a given θ, we use $\mathbf{x}_t(\theta)$ (and $\mathbf{w}_t(\theta)$) to denote series that depend on θ and possibly unobserved historical values prior to the time point $t = 1$.

Define the score function

$$Q_T(\check{\theta}, \check{\kappa}, \theta) = \frac{1}{T} \sum_{t=1}^{T} \mathbf{R}_t(\mathbf{x}_t(\check{\theta}), \check{\theta})' \mathbf{D}_t(\mathbf{x}_t(\check{\theta}), \check{\theta}, \check{\boldsymbol{\kappa}})^{-1} \mathbf{r}_t(\mathbf{w}_t(\theta), \theta), \tag{20}$$

Journal of Time Series Analysis published by John Wiley & Sons Ltd. DOI: 10.1111/jtsa.12770

where $\check{\theta}$ is a consistent initial estimator of θ_0; $\check{\kappa}$ is defined similarly to (11) and (12) except that in those formulae, $\mathbf{x}_t$'s are replaced by $\mathbf{x}_t(\check{\boldsymbol{\theta}})$'s. Furthermore, we define

$$Q_\infty(\theta_0, \kappa_0, \theta) = \mathrm{E}[\mathbf{R}_t(\theta_0)' \mathbf{D}_t(\mathbf{x}_t, \boldsymbol{\theta}_0, \boldsymbol{\kappa}_0)^{-1} \mathbf{r}_t(\mathbf{w}_t(\theta), \theta)].$$

The true parameter of interest should satisfy $\theta_0 = \mathrm{argzero}_{\theta\in\Theta} Q_\infty(\theta_0, \kappa_0, \theta)$, where $\mathrm{argzero}_{\theta\in\Theta} f(\theta)$ denotes the root of $f(\theta) = 0$. Our proposed estimator is

$$\hat{\theta} = \mathrm{argzero}_{\theta\in\Theta} Q_T(\check{\theta}, \check{\kappa}, \theta), \tag{21}$$

with plug-in $\check{\theta}$ and $\check{\kappa}$.

Here, we show the consistency, $\hat{\theta} \to_p \theta_0$. When $\mathbf{x}_t$ is not fully observed, we shall estimate θ_0 from an alternative score function, $\tilde{Q}_T(\check{\theta}, \tilde{\kappa}, \theta) = \frac{1}{T}\sum_{t=1}^T \tilde{\mathbf{Z}}_t' \mathbf{r}_t(\tilde{\mathbf{w}}_t(\theta), \boldsymbol{\theta}) = 0$; see equation (15) in Section 2.2. This score function involves $\{\mathbf{y}_t, \mathbf{x}_{t1}, \tilde{\mathbf{x}}_{t2}(\theta)\}_t$, where $\tilde{\mathbf{x}}_{t2}(\theta)$ is generated from a given θ and some initial values. We hope that the feasible $\tilde{Q}_T(\check{\theta}, \tilde{\kappa}, \theta)$ is first-order equivalent to $Q_T(\check{\theta}, \check{\kappa}, \theta)$, such that we can focus on the latter in the entire Section 3. To this end, we assume that:

A.0 Asymptotically, the choice of initial values does not matter. Namely, $\tilde{Q}_T(\check{\theta}, \tilde{\kappa}, \theta)$ is close to $Q_T(\check{\theta}, \check{\kappa}, \theta)$, in the sense that $\sup_{\dot{\theta},\kappa,\theta} |\nabla_\theta^l \tilde{Q}_T(\dot{\theta}, \kappa, \theta) - \nabla_\theta^l Q_T(\dot{\theta}, \kappa, \theta)|_2 = O_p(T^{-1})$ for $l = 0, 1$. Here, ∇_θ^l represents the lth-order derivative with respect to θ (assuming that both $\tilde{Q}_T(\dot{\theta}, \kappa, \theta)$ and $Q_T(\dot{\theta}, \kappa, \theta)$ are l-times differentiable with respect to θ).

We will verify this assumption in Appendix B.3 for the GARCH(1,1) model. When $\mathbf{x}_t$ is fully observed, the score function $Q_T(\check{\theta}, \check{\kappa}, \theta)$ can be simplified to (14), and all the subsequent analysis remains valid. Therefore, in the rest of Section 3, we will work with $Q_T(\check{\theta}, \check{\kappa}, \theta)$ that is presented in (20).

To proceed, we define the $|v|_2$ as the l_2-norm of a vector v, and $|\mathbf{A}|_2$ as the spectral norm of a matrix $\mathbf{A}$. Additionally, for a random variable X, we define $\|X\|_p$ as its L_p-norm given a positive number p. Based on these definitions, we impose the following assumptions:

A.1 (Initial estimators) For the initial estimators $\check{\theta}$ and $\check{\kappa}$, it holds that $\check{\theta} \to_p \theta_0$ and $\check{\kappa} \equiv (\check{\kappa}_3, \check{\kappa}_4) \to_p \kappa_0$, where κ_0 is a constant vector; it holds that $\sup_{\theta\in\Theta} |Q_T(\check{\theta}, \check{\kappa}, \theta) - Q_T(\theta_0, \kappa_0, \theta)|_2 = o_p(1)$.
A.2 (Uniform consistency) $\sup_{\theta\in\Theta} |Q_T(\theta_0, \kappa_0, \theta) - Q_\infty(\theta_0, \kappa_0, \theta)|_2 \to_p 0$.
A.3 (Identification) For any constant $\varepsilon > 0$, we have $\inf_{|\theta-\theta_0|_2\geq\varepsilon} |Q_\infty(\theta_0, \kappa_0, \theta)|_2 > 0 = |Q_\infty(\theta_0, \kappa_0, \theta_0)|_2$.
A.4 (Existence of an estimate) The root of the estimating equation $Q_T(\check{\theta}, \check{\kappa}, \hat{\theta}) = 0$ exists.

Assumption A.1 ensures that the initial estimators will not impact the consistency of the second-stage estimation. Assumptions A.2 and A.3 are adapted from the general consistency theorem for the Z-estimator (see, e.g., Theorem 5.9 of Van der Vaart, 2000). Assumption A.4 ensures that a solution exists for (21). We acknowledge that the existence of such a solution is not a trivial issue; however, throughout this section, we assume that a solution exists. We refer to, e.g., Jacod and Sørensen (2018), for further discussions on this issue. In addition, Appendix B.1 provides a detailed discussion of Assumptions A.1 to A.4 in the context of the GARCH(1,1) model.

Theorem 1. Under Assumptions A.1, A.2, A.3, and A.4, it holds that $\hat{\theta} \to_p \theta_0$.

The proof of Theorem 1 can be found in Appendix A.1.1.

Remark. Note that in Assumption A.1, κ_0 is only labeled as a constant vector, and we do not impose conditions (7) and (8) presented in Section 2. This is because the convergence of the nuisance parameters to the true values is not necessary for the consistency of $\hat{\theta}$. In fact, in the case that $\check{\theta}$ and $\check{\kappa}$ converge to some other θ^* and κ^*, Theorem 1 still holds if all the $Q_T(\theta_0, \kappa_0, \cdot)$ and $Q_\infty(\theta_0, \kappa_0, \cdot)$ in Assumptions A.1–A.4 are replaced with $Q_T(\theta^*, \kappa^*, \cdot)$ and $Q_\infty(\theta^*, \kappa^*, \cdot)$.

DOI: 10.1111/jtsa.12770

We shall note that Assumption A.2, which pertains to uniform convergence, is comparatively strong. However, we have the option to adopt an alternative set of assumptions. For instance,

A.2' (i) (Semi-continuity) For every $\theta \in \Theta$, $\liminf_{T\to\infty} |Q_T(\theta_0, \kappa_0, \theta)|_2 \geq \lim_{T\to\infty} |\mathrm{E}Q_T(\theta_0, \kappa_0, \theta)|_2$, almost surely; (ii) (Compactness and boundedness) Θ is compact, and $\mathrm{E}\left(\sup_{\theta\in\Theta} |Q_T(\theta_0, \kappa_0, \theta)|_2\right) < \infty$; (iii) (Pointwise convergence) $|Q_T(\theta_0, \kappa_0, \theta) - Q_\infty(\theta_0, \kappa_0, \theta)| \to_p 0$ holds for every $\theta \in \Theta$.

See Appendix B.2 for the verification of Assumption A.2'.

Theorem 2. Under Assumptions A.1, A.2', A.3, and A.4, it holds that $\hat{\theta} \to_p \theta_0$.

The proof of Theorem 2 can be found in Appendix A.1.2.

3.2. Asymptotic normality

Here, we show the asymptotic normality of the WOPIV. We first introduce some definitions and abbreviations. Recall that $\mathbf{r}_t(\mathbf{w}_t, \theta_0) = \mathbf{r}_t(\mathbf{w}_t(\theta_0), \theta_0)$; for any $\theta, \tilde{\theta} \in \Theta$ and initial estimators $(\check{\theta}, \check{\kappa})$, we define

$$\begin{aligned}
&\mathbf{r}_t(\theta) = \mathbf{r}_t(\mathbf{w}_t(\theta), \theta),\\
&\mathbf{R}_t(\tilde{\theta}) = \nabla \mathbf{r}_t(\tilde{\theta}) = \partial \mathbf{r}_t(\theta)/\partial\theta'|_{\theta=\tilde{\theta}},\\
&\check{\mathbf{R}}_t = \mathbf{R}_t(\check{\theta}), \mathbf{R}_t = \mathbf{R}_t(\theta_0), \mathbf{D}_t = \mathbf{D}_t(\mathbf{x}_t, \boldsymbol{\theta}_0, \boldsymbol{\kappa}_0).
\end{aligned}$$

Furthermore, we define

$$\begin{aligned}
\check{\mathbf{A}}_T(\theta) &= \frac{1}{T}\sum_{t=1}^{T} \check{\mathbf{R}}_t' \mathbf{D}_t(\mathbf{x}_t(\check{\theta}), \check{\boldsymbol{\theta}}, \check{\boldsymbol{\kappa}})^{-1} \mathbf{r}_t(\theta),\\
\mathbf{A}_T(\theta) &= \frac{1}{T}\sum_{t=1}^{T} \mathbf{R}_t(\theta_0)' \mathbf{D}_t(\mathbf{x}_t, \boldsymbol{\theta}_0, \boldsymbol{\kappa}_0)^{-1} \mathbf{r}_t(\theta),
\end{aligned}$$

and

$$\begin{aligned}
\check{\mathbf{B}}_T(\theta) &= \frac{1}{T}\sum_{t=1}^{T} \check{\mathbf{R}}_t' \mathbf{D}_t(\mathbf{x}_t(\check{\theta}), \check{\boldsymbol{\theta}}, \check{\boldsymbol{\kappa}})^{-1} \mathbf{R}_t(\theta),\\
\check{\mathbf{B}}_0(\theta) &= \mathrm{E}(\check{\mathbf{R}}_t' \mathbf{D}_t(\mathbf{x}_t(\check{\theta}), \check{\boldsymbol{\theta}}, \check{\boldsymbol{\kappa}})^{-1} \mathbf{R}_t(\theta)),\\
\mathbf{B}_0(\theta) &= \mathrm{E}(\mathbf{R}_t(\theta_0)' \mathbf{D}_t(\mathbf{x}_t, \boldsymbol{\theta}_0, \boldsymbol{\kappa}_0)^{-1} \mathbf{R}_t(\theta)),\\
\mathbf{B}_0 &= \mathbf{B}_0(\theta_0),\\
\mathbf{C}_0 &= \mathrm{E}(\mathbf{R}_t(\theta_0)' \mathbf{D}_t(\mathbf{x}_t, \boldsymbol{\theta}_0, \boldsymbol{\kappa}_0)^{-1} \mathbf{r}_t(\theta_0) \mathbf{r}_t(\theta_0)' \mathbf{D}_t(\mathbf{x}_t, \boldsymbol{\theta}_0, \boldsymbol{\kappa}_0)^{-1} \mathbf{R}_t(\theta_0)).
\end{aligned}$$

Moreover, we define the b-ball around $\dot{\theta}$ as $\mathcal{B}(\dot{\theta}, b) = \{\theta : |\theta - \dot{\theta}|_2 < b\}$, where $\dot{\theta}$ is a point within the interior of Θ, and b is a positive constant. Based on these definitions, we impose the following assumptions:

B.1 (Data generating) The sequence $\{\mathbf{r}_t(\mathbf{w}_t, \theta_0)\}_{t\geq 1}$ is a stationary and ergodic MDS with respect to (w.r.t.) $\mathcal{F}_{t-1}$, where $\mathcal{F}_t = \sigma(\{y_i, \mathbf{x}_i\}_{i\leq t})$ is the sigma field generated by $\{y_i, \mathbf{x}_i\}_{i\leq t}$. The dynamic completeness assumptions (3) and (4) hold, such that $\mathrm{E}[\mathbf{r}_t(\mathbf{w}_t, \theta_0)|\mathbf{x}_t] = 0$.

B.2 (i) (Differentiability) There exists a positive constant b_1, such that for all $(\theta, \kappa) \in \Theta \times \Gamma$ that satisfy $|\theta - \theta_0|_2 < b_1$ and $|\kappa - \kappa_0|_2 < b_1$, it holds that each element of $\mathbf{r}_t(\theta)$ is measurable and twice continuously differentiable w.r.t. θ, each element of $\mathbf{D}_t(\mathbf{x}_t(\theta), \theta, \kappa)^{-1}$ is continuously differentiable w.r.t. θ and κ, and each element of

Journal of Time Series Analysis published by John Wiley & Sons Ltd. DOI: 10.1111/jtsa.12770

$\mathbf{B}_0(\theta)$ is continuously differentiable w.r.t. θ. (ii) (Invertibility) $\mathbf{B}_0$ and $\mathbf{C}_0$ are positive definite matrices, and their maximum eigenvalues are bounded.

B.3 (i) (Uniform convergence) For some $b > 0$, $\sup_{\theta \in B(\theta_0, b)} \left(\check{\mathbf{B}}_T(\theta) - \check{\mathbf{B}}_0(\theta) \right) = o_p(1)$; (ii) (Stochastic equicontinuity) $\check{\mathbf{A}}_T(\theta_0) - \mathbf{A}_T(\theta_0) = o_p(1/\sqrt{T})$.

Remark. Assumption B.1 reiterates the fundamental conditions applied to the conditional moment function evaluated at the true θ_0, as previously detailed in Section 2. Assumption B.2 introduces the continuity and differentiability conditions that can be satisfied by many models of conditional means and variances. For example, the analytical forms of $\mathbf{r}_t(\theta)$ and $\mathbf{D}_t(\mathbf{x}_t(\theta), \theta, \kappa)^{-1}$ of the GARCH(1,1) model, as presented in Section 2.2, satisfy Assumption B.2. Assumption B.3(i) is required to ensure that $\check{\mathbf{B}}_T(\theta)$ eventually converges to $\mathbf{B}_0$; this assumption can be verified by similar arguments presented in the proof of lemma B.5 of Richter *et al.* (2023). B.3(ii) is adapted from assumption 5.2 of Newey (1994). It specifies the general stochastic equicontinuity condition that is required for the $\sqrt{T}$-consistency of a two-stage Z-estimator.

Theorem 3. Under Assumptions A.1–A.4 and B.1–B.3,

$$\sqrt{T}(\hat{\theta} - \theta_0) \to_d N(\mathbf{0}, \mathbf{B}_0^{-1}\mathbf{C}_0\mathbf{B}_0^{-1}). \tag{22}$$

Moreover, if $\mathrm{E}(\mathbf{r}_t(\theta_0)\mathbf{r}_t'(\theta_0)|\mathcal{F}_{t-1}) = \mathbf{D}_t(\mathbf{x}_t, \boldsymbol{\theta}_0, \boldsymbol{\kappa}_0)$, it holds that $\sqrt{T}(\hat{\theta} - \theta_0) \to_d N(\mathbf{0}, \mathbf{B}_0^{-1})$.

The proof can be found in Appendix A.2.

3.3. One-step estimation

Here, we show the asymptotic equivalence between the one-step estimator $\overline{\theta}$ and the ideal estimator $\hat{\theta}$. The former is obtained by updating any $\sqrt{T}$-consistent estimator $\check{\theta}$, as presented in (19); the latter is given by (14). To proceed, we impose:

C.1 Both $\hat{\theta}$ and $\check{\theta}$ are $\sqrt{T}$-consistent estimators of θ_0.

Theorem 4. Under Assumptions A.4, B.2, B.3, and C.1, it holds that $\sqrt{T}(\overline{\theta} - \hat{\theta}) = o_p(1)$.

The proof is in Appendix A.3.

4. THE MULTI-VARIATE CASE

We now describe the estimation strategy in the multi-variate case. For a $d \times 1$ vector $\mathbf{y}_t$, the conditional mean and variance functions are specified as

$$\mathrm{E}(\mathbf{y}_t|\mathbf{x}_t) = \mathbf{m}_t(\mathbf{x}_t, \boldsymbol{\theta}_0), \tag{23}$$

$$\mathrm{Var}(\mathbf{y}_t|\mathbf{x}_t) = \boldsymbol{\Sigma}_t(\mathbf{x}_t, \boldsymbol{\theta}_0). \tag{24}$$

The $d \times 1$ vector of error terms is denoted by $\boldsymbol{u}_t(\theta_0) = \mathbf{y}_t - \mathbf{m}_t(\mathbf{x}_t, \theta_0)$. Furthermore, we can suppress the dependence of the functions $\mathbf{m}_t(\mathbf{x}_t, \boldsymbol{\theta}_0)$ and $\boldsymbol{\Sigma}_t(\mathbf{x}_t, \boldsymbol{\theta}_0)$ on $\mathbf{x}_t$, and write the standardized vector as $\mathbf{e}_t \overset{\text{def}}{=} \boldsymbol{\Sigma}_t(\theta_0)^{-1/2}(\mathbf{y}_t - \mathbf{m}_t(\boldsymbol{\theta}_0)) = \boldsymbol{\Sigma}_t(\theta_0)^{-1/2}\boldsymbol{u}_t(\theta_0)$. Note that $\mathbf{e}_t$ is a $d-$dimensional random vector, and θ is a P-dimensional parameter. In the multi-variate case, we look at the moment vector $\mathbf{r}_t(\theta_0)_{(d+d(d+1)/2)\times 1} = (\boldsymbol{u}_t(\theta_0)', \mathrm{Vech}(\boldsymbol{u}_t(\theta_0)\boldsymbol{u}_t'(\theta_0) - \boldsymbol{\Sigma}_t(\theta_0))')'$. It should be noted that Vec is the vectorization of a matrix, and Vech is denoted as a half vectorization. The switching between Vec and Vech is via a duplication operator D_n and an elimination operator L_n, such that for an $n \times n$ symmetric matrix $\mathbf{A}$, we have $D_n\mathrm{Vech}(\mathbf{A}) = \mathrm{Vec}(\mathbf{A})$ and $L_n\mathrm{Vec}(\mathbf{A}) = \mathrm{Vech}(\mathbf{A})$. Now, we have the following conditional

DOI: 10.1111/jtsa.12770 *Journal of Time Series Analysis* published by John Wiley & Sons Ltd.

moment conditions:

$$\mathrm{E}\left(\boldsymbol{u}_t(\theta_0)|\mathbf{x}_t\right) = \mathbf{0},\ \mathrm{Var}\left(\boldsymbol{u}_t(\theta_0)|\mathbf{x}_t\right) = \boldsymbol{\Sigma}_t(\mathbf{x}_t, \boldsymbol{\theta}_0),$$
$$\mathrm{E}\left(\mathbf{e}_t|\mathbf{x}_t\right) = \mathbf{0},\ \mathrm{Vec}\left(\mathrm{Var}(\mathbf{e}_t|\mathbf{x}_t)\right) = \mathrm{E}(\mathbf{e}_t \otimes \mathbf{e}_t|\mathbf{x}_t) = \mathrm{Vec}(\mathbb{I}_d),$$

where $\mathbb{I}_d$ denotes the d-dimensional identity matrix. We define the third and fourth conditional moments of $\mathbf{e}_t$ as $\mathbf{K}_3 = \mathrm{E}(\mathbf{e}_t \otimes \mathbf{e}_t\mathbf{e}_t'|\mathbf{x}_t)$ and $\mathbf{K}_4 = \mathrm{E}(\mathbf{e}_t \otimes \mathbf{e}_t\mathbf{e}_t' \otimes \mathbf{e}_t'|\mathbf{x}_t)$ respectively. These two matrices are the multi-variate versions of (7) and (8). In particular, for an i.i.d. standard normally distributed $\mathbf{e}_t$, we have $\mathbf{K}_3 = \mathbf{0}$ and $\mathbf{K}_4 = \mathbb{I}_{d^2} + \overline{\mathbf{K}}_d + \mathrm{Vec}(\mathbb{I}_d)\mathrm{Vec}(\mathbb{I}_d)'$, where $\overline{\mathbf{K}}_d$ is a commutation matrix defined, for example, in Magnus and Neudecker (1979).

Now, we discuss the estimation strategy, which is in general by applying the ones discussed in Sections 2.1 and 2.2 to the multi-variate case. In the following, we describe the case with fully observed $\mathbf{x}_t$. The case with partially observed $\mathbf{x}_t$ can be obtained by replacing $\mathbf{x}_t$ with $\tilde{\mathbf{x}}_t(\theta)$, as described in Section 2.2. We start by describing how $\mathbf{K}_3$ and $\mathbf{K}_4$ can be similarly estimated from a sample. Let j_1, j_2, j_3, j_4, i_1, and i_2 be integers taking values from 1 to d. Then, we have

$$\check{\mathbf{K}}_3 \text{ is } d^2 \times d \qquad \text{with } [\check{\mathbf{K}}_3]_{(j_1-1)d+j_2,j_3} = \frac{1}{T}\sum_{t=1}^{T} \check{e}_{tj_1}\check{e}_{tj_2}\check{e}_{tj_3},$$

$$\check{\mathbf{K}}_4 \text{ is } d^2 \times d^2 \qquad \text{with } [\check{\mathbf{K}}_4]_{(j_1-1)d+j_2,(j_3-1)d+j_4} = \frac{1}{T}\sum_{t=1}^{T} \check{e}_{tj_1}\check{e}_{tj_2}\check{e}_{tj_3}\check{e}_{tj_4},$$

where $\check{e}_{tj}$ is the jth element of $\check{\mathbf{e}}_t$, and $\check{\mathbf{e}}_t = \boldsymbol{\Sigma}_t(\check{\theta})^{-1/2}(\mathbf{y}_t - \mathbf{m}(\mathbf{x}_t, \check{\theta}))$ for an initial estimator $\check{\theta}$.

For constructing the working variance-covariance matrix, we can preserve certain properties of a Gaussian distributed random vector by imposing that the elements of $\mathbf{e}_t$ are independent of each other and have zero means. It is important to emphasize that this condition, just like conditions (7) and (8) in the univariate case, is not an actual assumption essential to our theoretical framework. Instead, it serves as a rationale for constructing the working variance-covariance matrix for our proposed method. If $\mathbf{e}_t$ is normally distributed, these conditions will naturally hold. Even without the normality of $\mathbf{e}_t$, the working variance-covariance matrix can still capture certain characteristics of $\mathbf{e}_t$ that the GQMLE might overlook, potentially resulting in superior performance compared to the GQMLE. Under this condition, an estimator of $((j_1 - 1)d + j_2) \times j_3$th element of $\mathbf{K}_3$ can be $\frac{1}{T}\sum_{t=1}^{T}\check{e}_{tj}^3$ for $j_1 = j_2 = j_3 = j$, and otherwise 0 (because $\mathrm{E}(e_{j_1}e_{j_2}e_{j_3}) = 0$ if two of the elements in (j_1,j_2,j_3) are unequal.) The $\{(j_1 - 1)d + j_2, (j_3 - 1)d + j_4\}$th element in $\check{K}_4$ is non-zero when $j_1 = j_2 = i_1$ and $j_3 = j_4 = i_2, j_1 = j_4 = i_1$ and $j_2 = j_3 = i_2, j_1 = j_3 = i_1$ and $j_2 = j_4 = i_2$, or $j_1 = j_2 = j_3 = j_4 = j$. We can then estimate the non-zero element by $(\frac{1}{T}\sum_{t=1}^{T}\check{e}_{ti_1}^2)(\frac{1}{T}\sum_{t=1}^{T}\check{e}_{ti_2}^2)$ for the first three cases, and by $\frac{1}{T}\sum_{t=1}^{T}\check{e}_{tj}^4$ for the case of $j_1 = j_2 = j_3 = j_4 = j$.

The matrix $\mathbf{D}_t$ in Section 2 becomes

$$\mathbf{D}_t \overset{\text{def}}{=} \mathbf{D}_t(\mathbf{x}_t, \boldsymbol{\theta}_0) = \begin{pmatrix} \boldsymbol{\Sigma}_t & \boldsymbol{\Sigma}_{12,t} \\ \boldsymbol{\Sigma}_{12,t}' & \boldsymbol{\Sigma}_{22,t} \end{pmatrix}, \tag{25}$$

where $\boldsymbol{\Sigma}_t$ is the abbreviation of $\boldsymbol{\Sigma}_t(\boldsymbol{\theta}_0)$, and

$$\boldsymbol{\Sigma}_{12t}' = L_d\boldsymbol{\Sigma}_t(\theta_0)^{1/2} \otimes \boldsymbol{\Sigma}_t(\theta_0)^{1/2}\mathbf{K}_3\boldsymbol{\Sigma}_t(\theta_0)^{1/2},$$

$$\boldsymbol{\Sigma}_{22t} = L_d(\boldsymbol{\Sigma}_t(\theta_0)^{1/2} \otimes \boldsymbol{\Sigma}_t(\theta_0)^{1/2}\mathbf{K}_4\boldsymbol{\Sigma}_t(\theta_0)^{1/2} \otimes \boldsymbol{\Sigma}_t(\theta_0)^{1/2} - \mathrm{Vec}\ \boldsymbol{\Sigma}_t(\theta_0)\mathrm{Vec}\boldsymbol{\Sigma}_t(\theta_0)')$$
$$L_d\left(\boldsymbol{\Sigma}_t(\theta_0)^{1/2} \otimes \boldsymbol{\Sigma}_t(\theta_0)^{1/2}\mathbf{K}_4\boldsymbol{\Sigma}_t(\theta_0)^{1/2} \otimes \boldsymbol{\Sigma}_t(\theta_0)^{1/2} - \mathrm{Vec}\ \boldsymbol{\Sigma}_t(\theta_0)\mathrm{Vec}\boldsymbol{\Sigma}_t(\theta_0)'\right)'.$$

Journal of Time Series Analysis published by John Wiley & Sons Ltd. DOI: 10.1111/jtsa.12770

The gradient of the moment functions is a $(d + d(d+1)/2) \times P$ matrix:

$$\nabla \mathbf{r}_t(\mathbf{w}_t, \theta_0) = (-\nabla \mathbf{m}_t(\mathbf{x}_t, \theta_0)', (L_d \left[- \left(\boldsymbol{u}_t(\boldsymbol{\theta}_0) \otimes \mathbb{I}_d + \mathbb{I}_d \otimes \boldsymbol{u}_t(\boldsymbol{\theta}_0) \right) \nabla \mathbf{m}_t(\mathbf{x}_t, \theta_0) - \nabla \text{Vec } \boldsymbol{\Sigma}_t(\theta_0) \right])')'.$$

Define $\mathbf{R}_t(\theta) = \text{E}(\nabla \mathbf{r}_t(\mathbf{w}_t, \theta) | \mathcal{F}_{t-1})$, and $\check{\mathbf{R}}_t = \mathbf{R}_t(\check{\theta})$. The optimal instrument matrix is $\check{\mathbf{Z}}_t = \check{\mathbf{D}}_t^{-1} \check{\mathbf{R}}_t$, where $\check{\mathbf{D}}_t$ is a variation of $\mathbf{D}_t$ that is evaluated at $\check{\theta}$, $\check{\mathbf{K}}_3$, and $\check{\mathbf{K}}_4$. Then, we can estimate θ_0 as we previously did for equation (14).

The case with partially observed $\mathbf{x}_t$ can be handled by replacing $\mathbf{x}_t$ with $\tilde{\mathbf{x}}_t(\theta)$. The approach aligns with what has been presented in Section 2.2.

5. ESTIMATION PERFORMANCE

5.1. When is the GQMLE asymptotically efficient?

To understand better the circumstances under which we can improve on the GQMLE, we study in this section the factors that impact the efficiency of the GQMLE. In particular, we study the effect of the third and fourth moments on the asymptotic variance–covariance matrix. We follow Bollerslev and Wooldridge (1992) to introduce the setup of the GQMLE. Again, it should be noted that the discussion in this section pertains to the scenario where $\mathbf{x}_t$ is observable. For cases involving partially observed $\mathbf{x}_t$, one can substitute $\mathbf{x}_t$ with $\tilde{\mathbf{x}}_t(\theta)$, as outlined in Section 2.2.

The GQMLE is defined to be

$$\check{\theta} = \text{argmax}_{\theta \in \Theta} \frac{1}{T} \sum_{t=1}^{T} \ell_t(\theta; \mathbf{y}_t, \mathbf{x}_t).$$

The contribution of observation t to the quasi-log-likelihood function is

$$\ell_t(\theta; \mathbf{y}_t, \mathbf{x}_t) = -1/2 \log |\boldsymbol{\Sigma}_t(\mathbf{x}_t, \theta)| - 1/2 (\boldsymbol{u}_t(\mathbf{w}_t, \theta)') \boldsymbol{\Sigma}_t^{-1}(\mathbf{x}_t, \theta) \boldsymbol{u}_t(\mathbf{w}_t, \theta).$$

By abbreviating $\boldsymbol{u}_t(\mathbf{w}_t, \theta)$ as $\boldsymbol{u}_t(\theta)$, we can express the contribution of observation t to the score function as

$$\mathbf{s}_t(\theta) = \nabla_\theta \mathbf{m}_t(\theta)' \boldsymbol{\Sigma}_t(\theta)^{-1} \boldsymbol{u}_t(\theta) + 1/2 \nabla_\theta \boldsymbol{\Sigma}_t(\theta)' (\boldsymbol{\Sigma}_t^{-1}(\theta) \otimes \boldsymbol{\Sigma}_t^{-1}(\theta)) \text{Vec}(\boldsymbol{u}_t(\theta) \boldsymbol{u}_t(\theta)' - \boldsymbol{\Sigma}_t(\theta)),$$

where $\nabla_\theta \boldsymbol{\Sigma}_t(\theta) = \nabla_{\theta'} \text{Vec}(\boldsymbol{\Sigma}_t(\theta))$ is a $d^2 \times P$ matrix, and $\nabla_\theta \mathbf{m}_t(\theta)' = -\nabla_\theta \mathbf{u}_t(\theta)'$ is a $P \times d$ matrix.

To study the negative Hessian matrix evaluated at the θ_0, we write $\nabla_\theta \mathbf{s}_t(\theta_0)$ as

$$\mathbf{I}_t(\theta_0) = \nabla_\theta \mathbf{m}_t(\theta_0)' \boldsymbol{\Sigma}_t(\theta_0)^{-1} \nabla_\theta \boldsymbol{u}_t(\theta_0) + 1/2 \nabla_\theta \boldsymbol{\Sigma}_t(\theta_0)' (\boldsymbol{\Sigma}_t^{-1}(\theta_0) \otimes \boldsymbol{\Sigma}_t^{-1}(\theta_0)) \nabla_\theta \boldsymbol{\Sigma}_t(\theta_0).$$

Let us define $\mathbf{J}_t\left(\theta_0\right) = \text{E}[\mathbf{s}_t(\theta_0) \mathbf{s}_t'(\theta_0) | \mathbf{x}_t]$. Through some rearrangements, we get:

$$\begin{aligned}
\mathbf{J}_t(\theta_0) &= \nabla_\theta \boldsymbol{u}_t(\theta_0)' \boldsymbol{\Sigma}_t^{-1}(\theta_0) \nabla_\theta \boldsymbol{u}_t(\theta_0) \\
&\quad + 1/2 \nabla_\theta \boldsymbol{\Sigma}_t(\theta_0)' \boldsymbol{\Sigma}_t(\theta_0)^{-1/2} \otimes \boldsymbol{\Sigma}_t(\theta_0)^{-1/2} \mathbf{K}_3 \boldsymbol{\Sigma}_t(\theta_0)^{-1/2} \nabla \boldsymbol{u}_t(\theta_0) \\
&\quad + 1/2 \{ \nabla_\theta \boldsymbol{\Sigma}_t(\theta_0)' \boldsymbol{\Sigma}_t(\theta_0)^{-1/2} \otimes \boldsymbol{\Sigma}_t(\theta_0)^{-1/2} \mathbf{K}_3 \boldsymbol{\Sigma}_t(\theta_0)^{-1/2} \nabla \boldsymbol{u}_t(\theta_0) \}' \\
&\quad + 1/4 \nabla_\theta \boldsymbol{\Sigma}_t(\theta_0)' \boldsymbol{\Sigma}_t(\theta_0)^{-1/2} \otimes \boldsymbol{\Sigma}_t(\theta_0)^{-1/2} \mathbf{K}_4 \boldsymbol{\Sigma}_t(\theta_0)^{-1/2} \otimes \boldsymbol{\Sigma}_t(\theta_0)^{-1/2} \nabla_\theta \boldsymbol{\Sigma}_t(\theta_0) \\
&\quad - 1/4 \nabla_\theta \boldsymbol{\Sigma}_t(\theta_0)' \text{Vec}(\boldsymbol{\Sigma}_t(\theta_0)^{-1}) \text{Vec}'(\boldsymbol{\Sigma}_t(\theta_0)^{-1}) \nabla_\theta \boldsymbol{\Sigma}_t(\theta_0) \\
&= \mathbf{J}_{t1} + \mathbf{J}_{t2} + \mathbf{J}_{t2}' + \mathbf{J}_{t3} + \mathbf{J}_{t4}.
\end{aligned}$$

Let 'plim' denote the probability limit. Then, we define $\mathbf{J}_0(\theta) = \text{plim}_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \mathbf{J}_t(\theta)$, $\mathbf{J}_0 = \mathbf{J}_0(\theta_0)$, $\mathbf{I}_0(\theta) = \text{plim}_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \mathbf{I}_t(\theta)$, and $\mathbf{I}_0 = \mathbf{I}_0(\theta_0)$. Under some regularity conditions, the asymptotic variance of the GQMLE

DOI: 10.1111/jtsa.12770 *Journal of Time Series Analysis* published by John Wiley & Sons Ltd.

is $\mathbf{I}_0^{-1}\mathbf{J}_0\mathbf{I}_0^{-1}$ (see, for example, Theorem 6 in Appendix A.4). If the conditional distribution of $\mathbf{e}_t$ is indeed a standardized multi-variate Gaussian, we have $\mathbf{I}_0 = \mathbf{J}_0$, and the asymptotic variance would hit the lower bound $\mathbf{J}_0^{-1}$.

The variance of a GQMLE depends on the analytical form of $\mathbf{J}_t(\theta_0)$ and $\mathbf{I}_t(\theta_0)$. In the one-dimensional case of a GARCH(p, q) model, $\mathbf{m}_t(\theta) = 0$, and thus $\mathbf{J}_{t1} = \mathbf{0}$ and $\mathbf{J}_{t2} = \mathbf{0}$. We can see that $\mathbf{J}_t(\theta_0) = \frac{\kappa_4^0 - 1}{2}\mathbf{I}_t(\theta_0)$. If κ_4^0, the fourth moment of a Gaussian random variable, equals 3, then $\mathbf{J}_t(\theta_0) = \mathbf{I}_t(\theta_0)$, and the asymptotic variance will hit the lower bound. This corresponds to the findings of Francq and Zakoïan (2004). Now we define

$$\mathbf{J}_{20} + \mathbf{J}_{20}' + \mathbf{J}_{30} + \mathbf{J}_{40} = \text{plim}_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}(\mathbf{J}_{t2} + \mathbf{J}_{t2}' + \mathbf{J}_{t3} + \mathbf{J}_{t4}).$$

The following proposition explains how, in the case where the model specified in (23) and (24) is not generated from standard Gaussian-distributed innovations but is estimated using the GQMLE, the deviation from the optimal variance matrix can be attributed specifically to the skewness matrix and the fourth-moment matrix.

Proposition 5. Let us assume that (i) the model satisfies equations (23) and (24), (ii) the model is dynamically complete and has a consistent GQMLE, (iii) the innovation $\mathbf{e}_t$ has a finite fourth moment, and (iv) the GQMLE is asymptotically normal, then the efficiency loss of the GQMLE is $\mathbf{V}_0 \overset{\text{def}}{=} \mathbf{I}_0^{-1}(\mathbf{J}_{20} + \mathbf{J}_{20}' + \mathbf{J}_{30} + \mathbf{J}_{40} - \mathbf{I}_0)\mathbf{I}_0^{-1}$, with $\mathbf{V}_3 \overset{\text{def}}{=} \mathbf{I}_0^{-1}(\mathbf{J}_{20} + \mathbf{J}_{20}')\mathbf{I}_0^{-1}$ being associated with the skewness matrix, and $\mathbf{V}_4 \overset{\text{def}}{=} \mathbf{I}_0^{-1}\mathbf{J}_{30}\mathbf{I}_0^{-1}$ being associated with the fourth-moment matrix.

Based on Proposition 5, we can analyze scenarios regarding the overlap of mean and variance parameters. Let us first look at the case when the mean parameter and the variance parameter do not overlap, i.e., $\theta = (\beta', \gamma')'$, where β is a $P_1 \times 1$ mean parameter, and γ is a $P_2 \times 1$ variance parameter. Then, the item $\nabla_\theta \boldsymbol{u}_t(\theta_0)$ consists of $\nabla_\beta \boldsymbol{u}_t(\theta_0)_{d\times P_1}$ and $\nabla_\gamma \boldsymbol{u}_t(\theta_0)_{d\times P_2}$ $(= \mathbf{0})$, and the item $\nabla_\theta \boldsymbol{\Sigma}_t(\theta_0)$ contains $\nabla_\beta \boldsymbol{\Sigma}_t(\theta_0)(= \mathbf{0})$ and $\nabla_\gamma \boldsymbol{\Sigma}_t(\theta_0)$. Then, we have

$$\begin{aligned}
&\begin{pmatrix} \mathbf{I}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_2 \end{pmatrix} \overset{\text{def}}{=} \mathbf{I}_0 = \text{plim}_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\mathbf{I}_t(\theta_0) \\
&= \begin{pmatrix} \text{plim}_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\nabla_\beta \mathbf{u}_t(\theta_0)'\boldsymbol{\Sigma}_t(\theta_0)^{-1}\nabla_\beta \mathbf{u}_t(\theta_0) & \mathbf{0} \\ \mathbf{0} & \text{plim}_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}1/2\nabla_\gamma \boldsymbol{\Sigma}_t(\theta_0)'\mathbf{O}_{2t}\nabla_\gamma \boldsymbol{\Sigma}_t(\theta_0) \end{pmatrix},
\end{aligned}$$

where $\mathbf{O}_{2t} \overset{\text{def}}{=} \boldsymbol{\Sigma}_t^{-1}(\theta_0) \otimes \boldsymbol{\Sigma}_t^{-1}(\theta_0)$.

To investigate the role of the third and fourth moments of $\mathbf{e}_t$ on the variance of the estimator, we look at $\mathbf{J}_{t2}$ concerning the third moment matrix $\mathbf{K}_3$ and $\mathbf{J}_{t3}$ concerning the fourth moment matrix $\mathbf{K}_4$. Ideally, if the distribution of an innovation is symmetric, we should have $\mathbf{K}_3 = \mathbf{0}$, which is true for a Gaussian vector. Any deviation from $\mathbf{K}_3 = \mathbf{0}$ would contribute to the deviation of $\mathbf{J}_0$ from $\mathbf{I}_0$. Furthermore, we define

$$\text{plim}_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}(\mathbf{J}_{t2} + \mathbf{J}_{t2}') = \begin{pmatrix} \mathbf{0} & \mathbf{D}_{\kappa_3^0} \\ \mathbf{D}_{\kappa_3^0}' & \mathbf{0} \end{pmatrix}, \text{plim}_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\mathbf{J}_{t3} = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{E}_{\kappa_4^0} \end{pmatrix},$$

where $\mathbf{D}_{\kappa_3^0} = \text{plim}_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}1/2\nabla_\gamma \boldsymbol{\Sigma}_t(\theta_0)'\boldsymbol{\Sigma}_t(\theta_0)^{-1/2} \otimes \boldsymbol{\Sigma}_t(\theta_0)^{-1/2}\mathbf{K}_3\boldsymbol{\Sigma}_t(\theta_0)^{-1/2}\nabla_\beta \boldsymbol{u}_t(\theta_0)$ is a $P_1 \times P_2$ matrix, and $\mathbf{E}_{\kappa_4^0} = \text{plim}_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}1/4\nabla_\gamma \boldsymbol{\Sigma}_t(\theta_0)'\boldsymbol{\Sigma}_t(\theta_0)^{-1/2} \otimes \boldsymbol{\Sigma}_t(\theta_0)^{-1/2}\mathbf{K}_4\boldsymbol{\Sigma}_t(\theta_0)^{-1/2} \otimes \boldsymbol{\Sigma}_t(\theta_0)^{-1/2}\nabla_\gamma \boldsymbol{\Sigma}_t(\theta_0)$ is a $P_2 \times P_2$ matrix.

Then, the impacts arising from the skewness matrix and the fourth-moment matrix are expressed as follows:

$$\mathbf{V}_3 = \begin{pmatrix} \mathbf{0} & \mathbf{I}_1^{-1}\mathbf{D}_{\kappa_3^0}\mathbf{I}_2^{-1} \\ \mathbf{I}_1^{-1}\mathbf{D}_{\kappa_3^0}'\mathbf{I}_2^{-1} & \mathbf{0} \end{pmatrix}, \qquad \mathbf{V}_4 = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_2^{-1}\mathbf{E}_{\kappa_4^0}\mathbf{I}_2^{-1} \end{pmatrix}.$$

Journal of Time Series Analysis published by John Wiley & Sons Ltd. DOI: 10.1111/jtsa.12770

In sum, we can see that if the parameters do not overlap, the third moment does not affect the asymptotic variance of the estimated mean parameter $\hat{\beta}$ and variance parameter $\hat{\gamma}$, but plays a role in their covariance. The fourth moment plays a role only in the variance of the estimated variance parameter $\hat{\gamma}$. However, if they overlap, both the third and fourth moments of errors play a role in the variance–covariance matrix of the GQMLE.

Again, we take the one-dimensional case as an example: If the mean and variance parameters overlap, both κ_3^0 and κ_4^0 generally play a role; if the mean and variance parameters vary separately, κ_3^0 does not affect the variance of the GQMLE. Particularly, in the case where no conditional mean is presented, such as in a GARCH(p, q) model, we have $\nabla_\theta \mathbf{m}_t(\theta) = -\nabla_\theta \boldsymbol{u}_t(\theta) = \mathbf{0}$. Thus, it holds that $\mathbf{J}_t(\theta_0) = \mathbf{J}_{t3} + \mathbf{J}_{t4}$, and the parameter β does not exist anymore. In this case, $\theta = \gamma$, $\mathbf{I}_0 = \mathbf{I}_2$, and $\text{plim}_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathbf{J}_{t3} = \mathbf{E}_{\kappa_4^0}$. Consequently, only the fourth moment plays a role in the asymptotic variance of $\hat{\gamma}$.

5.2. Comparison with semi-parametric methods of dynamic models

Here, we compare the proposed WOPIV method with the semi-parametric method of dynamic models of Anatolyev (2003, 2007), Hafner and Rombouts (2007) and Di and Gangopadhyay (2011). See Section 6.3 for the comparison of Monte Carlo simulation results.

5.2.1. Comparison with Anatolyev (2003) and Anatolyev (2007)

These two papers utilize the full information of higher-order moments to specify the OPIV. In the case of the model described by (1) and (2), the correct specifications of the third and fourth moments are needed, whereas our working instrument matrix only requires the knowledge of the first two moments. Consequently, our estimator could be less efficient than the one proposed by Anatolyev (2003) if the third and fourth moments are known and deviate from the specifications in (9) and (10). Nonetheless, the WOPIV does not depend on the correct specification of the higher-order moments of the innovation terms.

5.2.2. Comparison with Hafner and Rombouts (2007) and Di and Gangopadhyay (2011)

These two methods are similar in that they both estimate the innovation term's density function non-parametrically, subsequently plugging the estimates into a second-stage likelihood function and solving for the plug-in MLE. In the second stage, both methods need to deal with the plug-in bias arising from the first-stage non-parametric estimation. While Hafner and Rombouts (2007) add correction terms to the semiparametric likelihood score function (Proposition 3 of Hafner and Rombouts (2007), on p. 259), Di and Gangopadhyay (2011) directly subtract an estimated bias term from the semiparametric MLE (Theorem 3.3 of Di and Gangopadhyay (2011), on p. 262). As both methods employ non-parametric techniques to estimate the unknown density function of the innovation, they are, to a certain extent, immune to model misspecification. If their first-stage non-parametric estimators of the innovation's density function, along with the corresponding estimated correction terms, fulfill certain regularity conditions, their methods can achieve the semi-parametric efficiency bound and cannot be beaten by other regular estimators. In comparison, we employ a working variance-covariance matrix. Deriving this matrix is computationally lighter since we estimate only the skewness and kurtosis of the unknown innovation, rather than its density function or the bias correction term. Although our variance–covariance matrix may be misspecified, as discussed in the article, any potential misspecification should not impact the consistency. The efficiency improvement over the GQMLE hinges on how closely the working variance–covariance matrix approximates the true one. Simulation results presented in Section 6 demonstrate that our proposed WOPIV performs effectively. Furthermore, as mentioned in the introduction, both semi-parametric methods impose an i.i.d. restriction on standardized innovations. In contrast, our proposed method only requires the innovations to be MDS.

In addition, the performance of the semi-parametric estimators proposed by Hafner and Rombouts (2007) and Di and Gangopadhyay (2011) depends on proper choices of tuning parameters, while the WOPIV is tuning-parameter-free. In Section 6.3, we present a Monte Carlo comparison that both WOPIV and the semiparametric method proposed in Hafner and Rombouts (2007) (HR hereafter) can outperform the GQMLE. In an ideal

setting, HR can achieve the best performance among the three methods. However, this is contingent on proper choices of tuning parameters. In comparison, our WOPIV method is tuning-parameter-free and easy to implement. In simulations, it outperforms the GQMLE by a considerable margin if the underlying innovation is skew normally distributed.

6. SIMULATIONS

To assess the performance of our method, we run four Monte Carlo simulation experiments, comparing the finite sample performances of the WOPIV, the GQMLE, and the semi-parametric method proposed in HR. In the following Sections 6.1 and 6.2, we show that, for both univariate models and multi-variate models, our method significantly outperforms the GQMLE when the normality assumption is violated. When the innovation terms η_t are normally distributed, the two methods perform similarly. Our experiments encompass sample sizes ranging from 200 to 2000, with 1000 Monte Carlo replications. We employ the function *rsnorm* from the R package *fGarch* to generate series of the skew normal distribution. The resulting series has a skewness $\kappa_3^0 \approx 0.78$ and a kurtosis $\kappa_4^0 \approx 3.49$ (the arguments of the function *rsnorm* in the package fGarch are chosen as mean = 0, SD = 1, and $xi = 2$), compared with the standard normal distribution where $\kappa_3^0 = 0$ and $\kappa_4^0 = 3$.

6.1. Univariate models

We first show the Monte Carlo results for the univariate GARCH model.

6.1.1. *Data generating process*

Case 1: GARCH(1,1). As discussed in Example 1, we have

$$\begin{aligned} \varepsilon_t &= \sigma_t \eta_t, \\ \sigma_t^2 &= \omega_0 + \alpha_0 \varepsilon_{t-1}^2 + \beta_0 \sigma_{t-1}^2, \end{aligned} \tag{26}$$

where the noise term $\eta_t \overset{\text{i.i.d.}}{\sim} (0,1)$. In the notations of the general model specified in (1) and (2), we have $y_t = \varepsilon_t$ and $\mathbf{x}_t = (\varepsilon_{t-1}, \sigma_{t-1})'$. The parameter of interest is $\theta_0 = (\omega_0, \alpha_0, \beta_0)'$. Based on our discussion in Section 2.2, the estimating equation (18) is replicated here:

$$-\frac{1}{\tilde{c}_\kappa}\sum_{t=1}^{T}\begin{pmatrix} \frac{\varepsilon_t^2-(\omega+\alpha\varepsilon_{t-1}^2+\beta\tilde{\sigma}_{t-1}^2(\theta))}{\tilde{\sigma}_t^4(\check{\theta})} - \tilde{\kappa}_3\tilde{\sigma}_t^{-3}(\check{\theta})\varepsilon_t \\ \frac{\varepsilon_t^2-(\omega+\alpha\varepsilon_{t-1}^2+\beta\tilde{\sigma}_{t-1}^2(\theta))}{\tilde{\sigma}_t^4(\check{\theta})}\varepsilon_{t-1}^2 - \tilde{\kappa}_3\tilde{\sigma}_t^{-3}(\check{\theta})\varepsilon_{t-1}^2\varepsilon_t \\ \frac{\varepsilon_t^2-(\omega+\alpha\varepsilon_{t-1}^2+\beta\tilde{\sigma}_{t-1}^2(\theta))}{\tilde{\sigma}_t^4(\check{\theta})}\tilde{\sigma}_{t-1}^2(\check{\theta}) - \tilde{\kappa}_3\tilde{\sigma}_t^{-3}(\check{\theta})\tilde{\sigma}_{t-1}^2(\check{\theta})\varepsilon_t \end{pmatrix} = \mathbf{0}. \tag{27}$$

In comparison, the GQMLE score function of the GARCH (1,1) is (see, e.g., p. 143 of Francq and Zakoian (2010))

$$\sum_{t=1}^{T}\begin{pmatrix} \frac{\varepsilon_t^2-(\omega+\alpha\varepsilon_{t-1}^2+\beta\tilde{\sigma}_{t-1}^2(\theta))}{\tilde{\sigma}_t^4(\theta)} \\ \frac{\varepsilon_t^2-(\omega+\alpha\varepsilon_{t-1}^2+\beta\tilde{\sigma}_{t-1}^2(\theta))}{\tilde{\sigma}_t^4(\theta)}\varepsilon_{t-1}^2 \\ \frac{\varepsilon_t^2-(\omega+\alpha\varepsilon_{t-1}^2+\beta\tilde{\sigma}_{t-1}^2(\theta))}{\tilde{\sigma}_t^4(\theta)}\tilde{\sigma}_{t-1}^2(\theta) \end{pmatrix} = \mathbf{0}. \tag{28}$$

Journal of Time Series Analysis published by John Wiley & Sons Ltd. DOI: 10.1111/jtsa.12770

Table I. Simulation results of GARCH(1,1), with $\omega_0 = 0.1$, $\alpha_0 = 0.3$, and $\beta_0 = 0.6$

T	Methods	$\hat{\mu}_\omega$	$\hat{\mu}_\alpha$	$\hat{\mu}_\beta$	MSE_ω	MSE_α	MSE_β
η_t is standard normal							
500	GQMLE	0.1125	0.2956	0.5860	1.5010	3.0996	6.5017
	WOPIV	0.1118	0.2935	0.5862	1.4907	3.0848	6.5737
1000	GQMLE	0.1059	0.2973	0.5939	1.0612	3.1335	5.5281
	WOPIV	0.1056	0.2965	0.5938	1.0682	3.1256	5.5803
2000	GQMLE	0.1030	0.2989	0.5962	0.8951	2.9028	5.0283
	WOPIV	0.1029	0.2985	0.5962	0.8927	2.8907	5.0296
η_t is skew normal							
500	GQMLE	0.1101	0.2921	0.5917	1.3161	3.8766	6.6624
	WOPIV	0.1066	0.2920	0.5947	**0.9969**	**3.3153**	**5.5099**
1000	GQMLE	0.1072	0.2988	0.5900	1.2716	3.9706	6.8525
	WOPIV	0.1042	0.2965	0.5954	**0.9202**	**2.9511**	**4.9964**
2000	GQMLE	0.1049	0.3019	0.5920	1.1439	4.0527	6.5567
	WOPIV	0.1034	0.3007	0.5943	**0.8121**	**3.0785**	**4.8870**

6.1.2. Simulation results

Table I presents the Monte Carlo averages and mean square errors (MSE), multiplied by T, for each of the 1000 estimates in Case 1. The true values are set to be $(\omega_0, \alpha_0, \beta_0) = (0.1, 0.3, 0.6)$. Throughout this section, we use $\hat{\mu}_\theta$ and MSE_θ to denote the averages and MSEs of the Monte Carlo estimates for the respective parameters in each case.

In Case 1, to understand the different performances of WOPIV and GQMLE, we can compare the score functions in (27) and (28). Note that the term $-\frac{1}{\tilde{c}_\kappa}$ does not influence the solution of (27) and can be ignored. The score function of the WOPIV (27) contains an additional element in each row. These elements are products of the estimated skewness κ_3^0 and some zero-mean terms. If the original distribution is symmetric, such as the normal distribution, we have $\kappa_3^0 = 0$. In this situation, the WOPIV should be approximately equivalent to the GQMLE. If the skewness of η_t is non-zero, the WOPIV should capture this information and beat the GQMLE in terms of efficiency. Table I supports this conjecture.

By comparing the Monte Carlo averages to the true parameters, we observe that both GQMLE and WOPIV generally appear to be asymptotically unbiased. The upper panel shows the simulation results where η_t follows the standard normal distribution, which implies $\kappa_3^0 = 0$. In this case, the estimation performances of the GQMLE and our method become very similar. The lower panel shows the case where η_t follows the skew normal distribution and $\kappa_3^0 \approx 0.78$. With a non-zero κ_3^0, the GQMLE is no longer efficient, and we see that our method has achieved improved performance in terms of MSE for every sample size and every parameter (highlighted in bold). For example, in the sample size $n = 2000$, the Monte Carlo MSE of ω has dropped from 1.14 to 0.81, and the Monte Carlo MSE of α has dropped from 4.05 to 3.08.

6.2. Multi-variate models

6.2.1. Data generating process

Here, we present that our methods can perform well in multi-variate models. For simplicity, we consider two-dimensional cases. Let $\eta_t = (\eta_{1,t}, \eta_{2,t})'$ be an i.i.d. two-dimensional random vector with a variance-covariance matrix $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$; for $i = 1, 2$, we have $\kappa_3^0 = \text{E}(\eta_{i,t}^3)$, and $\kappa_4^0 = \text{E}(\eta_{i,t}^4)$.

In Section 4, we have defined

$$\mathbf{D}_t \equiv \text{Var}[\mathbf{r}_t(\mathbf{w}_t, \theta_0)|\mathbf{x}_t] = \begin{pmatrix} \boldsymbol{\Sigma}_t & \boldsymbol{\Sigma}_{12,t} \\ \boldsymbol{\Sigma}'_{12,t} & \boldsymbol{\Sigma}_{22,t} \end{pmatrix},$$

where $\mathbf{\Sigma}_t$ is the abbreviation of $\mathbf{\Sigma}_t(\mathbf{x}_t, \theta_0)$ defined in Section 4, $\mathbf{\Sigma}'_{12t} = L_2\mathbf{\Sigma}_t^{1/2} \otimes \mathbf{\Sigma}_t^{1/2}\mathbf{K}_3\mathbf{\Sigma}_t^{1/2}$, and $\mathbf{\Sigma}_{22t} = L_2(\mathbf{\Sigma}_t^{1/2} \otimes \mathbf{\Sigma}_t^{1/2}\mathbf{K}_4\mathbf{\Sigma}_t^{1/2} \otimes \mathbf{\Sigma}_t^{1/2} - \text{Vec}\ \mathbf{\Sigma}_t\text{Vec}\mathbf{\Sigma}'_t)\, L_2\left(\mathbf{\Sigma}_t^{1/2} \otimes \mathbf{\Sigma}_t^{1/2}\mathbf{K}_4\mathbf{\Sigma}_t^{1/2} \otimes \mathbf{\Sigma}_t^{1/2} - \text{Vec}\ \mathbf{\Sigma}_t\text{Vec}\mathbf{\Sigma}'_t\right)'$. For the i.i.d. two-dimensional innovation η_t, we have

$$L_2 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1/2 & 1/2 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

$$\mathbf{K}_3 = E(\eta_t \otimes \eta_t\eta_t') = \begin{pmatrix} \kappa_3^0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & \kappa_3^0 \end{pmatrix},\ \mathbf{K}_4 = E(\eta_t \otimes \eta_t\eta_t' \otimes \eta_t') = \begin{pmatrix} \kappa_4^0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & \kappa_4^0 \end{pmatrix}.$$

6.2.2. Case 2: CCC-GARCH

Next, let $\varepsilon_t = (\varepsilon_{1,t}, \varepsilon_{2,t})'$ be a two-dimensional vector, and we have the Constant Conditional Correlations model:

$$\begin{cases} \varepsilon_t = \mathbf{\Sigma}_t^{1/2}\eta_t \\ \mathbf{\Sigma}_t = \mathbf{\Lambda}_t\mathbf{\Gamma}_t\mathbf{\Lambda}_t \\ \sigma_{1,t}^2 = \omega_1 + \alpha_1\varepsilon_{1,t-1}^2 + \beta_1\sigma_{1,t-1}^2 \\ \sigma_{2,t}^2 = \omega_2 + \alpha_2\varepsilon_{2,t-1}^2 + \beta_2\sigma_{2,t-1}^2, \end{cases}$$

where $\mathbf{\Lambda}_t = \begin{pmatrix} \sigma_{1,t} & 0 \\ 0 & \sigma_{2,t} \end{pmatrix}$ and $\mathbf{\Gamma}_t = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$. In the notations of the general model specified in (23) and (24), we have $\mathbf{y_t} = \varepsilon_t$ and $\mathbf{x}_t = (\varepsilon_{1,t-1}, \sigma_{1,t-1}, \varepsilon_{2,t-1}, \sigma_{2,t-1})'$. We reparameterize $\rho = \sin\delta$ (and $\delta = \arcsin\rho$) to ensure that $\mathbf{\Gamma}_t$ is a well-defined correlation matrix.

Then we have

$$\mathbf{r}_t(\mathbf{w}_t, \theta_0) = \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \varepsilon_{1,t}^2 - (\omega_1 + \alpha_1\varepsilon_{1,t-1}^2 + \beta_1\sigma_{1,t-1}^2) \\ \varepsilon_{1,t}\varepsilon_{2,t} - \sin\delta\sigma_{1,t}\sigma_{2,t} \\ \varepsilon_{2,t}^2 - (\omega_2 + \alpha_2\varepsilon_{2,t-1}^2 + \beta_2\sigma_{2,t-1}^2) \end{pmatrix},$$

where the parameter of interest are $\theta_0 = (\omega_1, \alpha_1, \beta_1, \omega_2, \alpha_2, \beta_2, \delta)'$, and

$$\mathbf{R}_t = -\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & \varepsilon_{1,t-1}^2 & \sigma_{1,t-1}^2 & 0 & 0 & 0 & 0 \\ \frac{\sin\delta\sigma_{2,t}}{2\sigma_{1,t}} & \frac{\sin\delta\sigma_{2,t}}{2\sigma_{1,t}}\varepsilon_{1,t-1}^2 & \frac{\sin\delta\sigma_{2,t}}{2\sigma_{1,t}}\sigma_{1,t-1}^2 & \frac{\sin\delta\sigma_{1,t}}{2\sigma_{2,t}} & \frac{\sin\delta\sigma_{1,t}}{2\sigma_{2,t}}\varepsilon_{2,t-1}^2 & \frac{\sin\delta\sigma_{1,t}}{2\sigma_{2,t}}\sigma_{2,t-1}^2 & \sigma_{1,t}\sigma_{2,t}\cos\delta \\ 0 & 0 & 0 & 1 & \varepsilon_{2,t-1}^2 & \sigma_{2,t-1}^2 & 0 \end{pmatrix}.$$

With $\mathbf{D}_t$ defined in (25) and following the procedure described in Section 2.2, we can derive the sample moment condition $\sum_{t=1}^{T}\tilde{\mathbf{Z}}_t'\mathbf{r}_t(\tilde{\mathbf{w}}_t(\theta), \theta) = \mathbf{0}$.

DOI: 10.1111/jtsa.12770

6.2.3. Case 3: BEKK-GARCH

Let us consider the BEKK-GARCH model

$$\begin{cases} \varepsilon_t = \mathbf{\Sigma}_t^{1/2}\eta_t \\ \mathbf{\Sigma}_t = \mathbf{C} + \mathbf{A}\varepsilon_{t-1}\varepsilon_{t-1}'\mathbf{A}' + \mathbf{B}\mathbf{\Sigma}_{t-1}\mathbf{B}'. \end{cases}$$

In the Monte Carlo simulation, we set

$$\mathbf{C} = \begin{pmatrix} c_{11} & c_{12} \\ c_{12} & c_{22} \end{pmatrix}, \mathbf{A} = \begin{pmatrix} a_{11} & 0 \\ 0 & a_{22} \end{pmatrix}, \text{and } \mathbf{B} = \mathbf{0}.$$

For this model, we have

$$\mathbf{r}_t(\mathbf{w}_t, \theta_0) = \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \varepsilon_{1,t}^2 - (c_{11} + a_{11}^2\varepsilon_{1,t-1}^2) \\ \varepsilon_{1,t}\varepsilon_{2,t} - (c_{12} + \varepsilon_{1,t-1}\varepsilon_{2,t-1}a_{11}a_{22}) \\ \varepsilon_{2,t}^2 - (c_{22} + a_{22}^2\varepsilon_{2,t-1}^2) \end{pmatrix},$$

where the parameters of interest are $\theta_0 = (c_{11}, c_{12}, c_{22}, a_{11}, a_{22})'$. In the notations of the general model specified in (23) and (24), we have $\mathbf{y_t} = \varepsilon_t$ and $\mathbf{x}_t = \varepsilon_{t-1}$, and

$$\mathbf{R}_t = -\begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 2a_{11}\varepsilon_{1,t-1}^2 & 0 \\ 0 & 1 & 0 & a_{22}\varepsilon_{1,t-1}\varepsilon_{2,t-1} & a_{11}\varepsilon_{1,t-1}\varepsilon_{2,t-1} \\ 0 & 0 & 1 & 0 & 2a_{22}\varepsilon_{2,t-1}^2 \end{pmatrix}.$$

With $\mathbf{D}_t$ defined in (25) and following the procedure described in Section 2.1, we can derive the sample moment condition $\sum_{t=1}^{T}\check{\mathbf{Z}}_t'\mathbf{r}_t(\mathbf{w}_t, \theta) = \mathbf{0}$.

6.2.4. Simulation results

Here, we show the results of the simulation study. Tables II and III present the Monte Carlo averages and MSEs (multiplied by T) from 1000 estimates for Case 2 and Case 3 respectively. The true values are $(\omega_1, \alpha_1, \beta_1, \omega_2, \alpha_2, \beta_2, \rho) = (0.4, 0.8, 0.15, 0.2, 0.7, 0.2, 0.7)$ for the CCC-GARCH(1,1) model, and $(c_{11}, c_{12}, c_{22}, a_{11}, a_{22}) = (0.8, 0.5, 0.7, 0.6, 0.5)$ for the BEKK-GARCH model.

In Case 2, both GQMLE and WOPIV appear to be consistent and asymptotically unbiased. We note that in both the upper panel and lower panel of Table II, the Monte Carlo means of estimates, $\hat{\mu}$'s, converge to the true values as sample sizes increase, and the magnitudes of the biases are generally minimal. Therefore, we can focus on the comparison of variances. As demonstrated in Section 4, our proposed method for multi-variate cases involves estimating both $\mathbf{K}_3$ and $\mathbf{K}_4$. The upper panel of Table II shows the simulation results with $\eta_t \overset{\text{i.i.d.}}{\sim} \text{N}(\mathbf{0}, \mathbb{I}_2)$, which implies

$$\mathbf{K}_3 = \underset{(4\times 2)}{\mathbf{0}} \text{ and } \mathbf{K}_4 = \begin{pmatrix} 3 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 3 \end{pmatrix}.$$

Table II. Simulation results of CCC-GARCH (1,1), with $(\omega_1, \alpha_1, \beta_1, \omega_2, \alpha_2, \beta_2, \rho) = (0.4, 0.8, 0.15, 0.2, 0.7, 0.2, 0.7)$

Mean								
T	Methods	$\hat{\mu}_{\omega_1}$	$\hat{\mu}_{\alpha_1}$	$\hat{\mu}_{\beta_1}$	$\hat{\mu}_{\omega_2}$	$\hat{\mu}_{\alpha_2}$	$\hat{\mu}_{\beta_2}$	$\hat{\mu}_{\rho}$
500	GQMLE	0.4063	0.7962	0.1444	0.2034	0.6953	0.1951	0.6998
	WOPIV	0.4045	0.7920	0.1445	0.2025	0.6895	0.1957	0.6997
1000	GQMLE	0.4042	0.7959	0.1490	0.2022	0.7002	0.1974	0.7006
	WOPIV	0.4023	0.7924	0.1497	0.2014	0.6971	0.1979	0.7004
2000	GQMLE	0.4022	0.7995	0.1489	0.2014	0.6989	0.1992	0.7011
	WOPIV	0.4012	0.7972	0.1491	0.2011	0.6975	0.1994	0.7011
Variance								
T	Methods	$\hat{\sigma}^2_{\omega_1}$	$\hat{\sigma}^2_{\alpha_1}$	$\hat{\sigma}^2_{\beta_1}$	$\hat{\sigma}^2_{\omega_2}$	$\hat{\sigma}^2_{\alpha_2}$	$\hat{\sigma}^2_{\beta_2}$	$\hat{\sigma}^2_{\rho}$
500	GQMLE	1.7891	4.3450	1.0231	0.5621	3.8848	1.4554	0.2770
	WOPIV	1.8507	4.7392	1.0811	0.6008	4.4032	1.6261	0.2837
1000	GQMLE	1.8539	4.3591	1.0416	0.5766	3.6734	1.3152	0.2620
	WOPIV	1.9925	4.7354	1.2147	0.5882	3.9364	1.4830	0.2726
2000	GQMLE	1.8187	4.1762	0.9328	0.5284	3.5566	1.2467	0.2553
	WOPIV	1.9441	4.8180	1.0558	0.5683	3.8627	1.3947	0.2675
η_t is skew normal								
Mean								
T	Methods	$\hat{\mu}_{\omega_1}$	$\hat{\mu}_{\alpha_1}$	$\hat{\mu}_{\beta_1}$	$\hat{\mu}_{\omega_2}$	$\hat{\mu}_{\alpha_2}$	$\hat{\mu}_{\beta_2}$	$\hat{\mu}_{\rho}$
500	GQMLE	0.4086	0.7995	0.1473	0.2022	0.6945	0.1992	0.7011
	WOPIV	0.4058	0.7994	0.1473	0.2009	0.6961	0.1993	0.7012
1000	GQMLE	0.4068	0.8010	0.1464	0.2038	0.6979	0.1954	0.7002
	WOPIV	0.4052	0.8025	0.1454	0.2032	0.6998	0.1945	0.7002
2000	GQMLE	0.4042	0.7970	0.1488	0.2022	0.7007	0.1976	0.7009
	WOPIV	0.4030	0.7971	0.1490	0.2019	0.7012	0.1973	0.7008
Variance								
T	Methods	$\hat{\sigma}^2_{\omega_1}$	$\hat{\sigma}^2_{\alpha_1}$	$\hat{\sigma}^2_{\beta_1}$	$\hat{\sigma}^2_{\omega_2}$	$\hat{\sigma}^2_{\alpha_2}$	$\hat{\sigma}^2_{\beta_2}$	$\hat{\sigma}^2_{\rho}$
500	GQMLE	2.5501	5.4769	1.4274	0.6263	3.9147	1.7264	0.2603
	WOPIV	**2.0396**	**4.5859**	1.**2229**	**0.5260**	**3.6312**	**1.5843**	0.2632
1000	GQMLE	2.3926	5.4903	1.2257	0.6864	4.4586	1.7915	0.2454
	WOPIV	**1.8615**	**4.8035**	**1.0580**	**0.5784**	**3.9711**	**1.5975**	0.2478
2000	GQMLE	2.3407	4.7825	1.1856	0.6335	4.4659	1.5817	0.2344
	WOPIV	**1.9055**	**4.3984**	**1.0963**	**0.5701**	**3.9433**	**1.4392**	0.2409

Across all sample sizes, the Monte Carlo variances of WOPIV and GQMLE are very close, indicating that the two methods exhibit similar performances when the innovations follow a standard normal distribution.

The lower panel of Table II shows the case where η_t follows the skew normal distribution with

$$\mathbf{K}_3 = \begin{pmatrix} \kappa_3^0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & \kappa_3^0 \end{pmatrix} \text{ and } \mathbf{K}_4 = \begin{pmatrix} \kappa_4^0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & \kappa_4^0 \end{pmatrix},$$

where $\kappa_3^0 \approx 0.78$ and $\kappa_4^0 \approx 3.49$. As predicted, the GQMLE is no longer efficient due to the model misspecification. We see that our method has achieved smaller variances in every sample size and almost every coefficient, as highlighted in bold.

Journal of Time Series Analysis published by John Wiley & Sons Ltd. DOI: 10.1111/jtsa.12770

Table III. Simulation results of BEKK-GARCH, with $c_{11} = 0.8, c_{12} = 0.5, c_{22} = 0.7, a_{11} = 0.6$, and $a_{22} = 0.5$

T	Methods	$\hat{\mu}_{c_{11}}$	$\hat{\mu}_{c_{12}}$	$\hat{\mu}_{c_{22}}$	$\hat{\mu}_{a_{11}}$	$\hat{\mu}_{a_{22}}$	MSE_{c11}	MSE_{c12}	MSE_{c22}	MSE_{a11}	MSE_{a22}
η_t is standard normal											
500	GQMLE	0.8018	0.5029	0.7023	0.5966	0.4928	2.3033	1.1891	1.3228	1.3367	1.3493
	WOPIV	0.7973	0.4999	0.6980	0.5944	0.4909	2.3343	1.1811	1.3259	1.3410	1.3698
1000	GQMLE	0.7987	0.4991	0.6987	0.5985	0.4971	2.3419	1.2341	1.4147	1.2844	1.3237
	WOPIV	0.7962	0.4977	0.6967	0.5974	0.4962	2.3638	1.2487	1.4470	1.2915	1.3235
2000	GQMLE	0.7996	0.5002	0.7001	0.5990	0.4991	2.2034	1.2320	1.4746	1.2845	1.2059
	WOPIV	0.7985	0.4995	0.6990	0.5986	0.4987	2.2051	1.2259	1.4719	1.2869	1.2081
η_t is skew normal											
500	GQMLE	0.8020	0.5032	0.7016	0.5936	0.4930	2.5672	1.2991	1.6204	1.5426	1.4141
	WOPIV	0.7977	0.5008	0.6995	0.5930	0.4924	**2.0405**	**1.1562**	**1.3030**	**1.3291**	**1.2210**
1000	GQMLE	0.7981	0.4988	0.6975	0.5975	0.4996	2.5126	1.3372	1.8252	1.5768	1.4566
	WOPIV	0.7964	0.4976	0.6960	0.5978	0.4988	**2.1277**	**1.2239**	**1.5045**	**1.3145**	**1.2225**
2000	GQMLE	0.7997	0.5003	0.6999	0.5973	0.4970	2.7286	1.3321	1.8529	1.6249	1.5519
	WOPIV	0.7999	0.5001	0.6997	0.5968	0.4971	**2.2430**	**1.2109**	**1.5833**	**1.3074**	**1.2955**

In Case 3 for the BEKK-GARCH model, we see the same pattern: our methods and the GQMLE have similar performances if the underline η_t is correctly specified as $N(\mathbf{0}, \mathbb{I}_2)$. If η_t is drawn from a skew normal distribution, our method outperforms GQMLE in every sample size and every parameter.

To summarize, in both univariate models and multi-variate models, the simulation results show that our method has good performance. While the GQMLE and our methods perform similarly in the case of normally distributed η_t, our method outperforms the GQMLE in the case of skew normally distributed η_t.

6.3. Comparison with a semi-parametric estimator

Here, we compare our proposed WOPIV, the GQMLE, and the method proposed by Hafner and Rombouts (2007) (HR). The data generating process is similar to Case 1: For each sample size $n = 200, 500$ and 1000, we generate 1000 Monte Carlo samples from a GARCH(1,1) model with $(\omega_0, \alpha_0, \beta_0) = (0.1, 0.3, 0.6)$. For the HR, we try two different bandwidths: $r \cdot n^{-2/5}$ and $r \cdot n^{-1/5}$, where r is the range of the normalized residuals. See section 3.2 of Hafner and Rombouts (2007) (pp. 261–263) for more details on the implementation of their method (Table IV). If η_t is standard normally distributed, the performance of the three methods is similar. The GQMLE has a slight edge over the other two, which is reasonable given that it is efficient in this case. If η_t is skew normally distributed, both WOPIV and HR outperform the GQMLE across all sample sizes and parameters. For the HR, the choice of bandwidth plays an important role: When selecting a bandwidth of $h = r \cdot n^{-2/5}$, the HR demonstrates the best performance, with the WOPIV exhibiting larger MSEs than the HR but still significantly outperforming the GQMLE. When opting for a bandwidth of $h = r \cdot n^{-1/5}$, the WOPIV emerges as the superior method. In this case, the HR has larger MSEs than the WOPIV and only marginally surpasses the GQMLE. In sum, the HR can achieve the best performance when the bandwidth is appropriately chosen. Meanwhile, the WOPIV offers a tuning-parameter-free, easy-to-implement method that can outperform the GQMLE when the innovation is not Gaussian.

7. APPLICATION

Here, we illustrate the application of our methodology by analyzing the stock price series of Amazon (AMZN). We have gathered daily observations from January 2013 to December 2017, resulting in a total of 1500 data points. The data source is Yahoo Finance, https://finance.yahoo.com. Figure 1 displays the original series, illustrating a

DOI: 10.1111/jtsa.12770 *Journal of Time Series Analysis* published by John Wiley & Sons Ltd.

Table IV. Comparison between GQMLE, WOPIV and HR for GARCH(1,1) with $\omega_0 = 0.1$, $\alpha_0 = 0.3$ and $\beta_0 = 0.6$

T	Methods	h	$\hat{\mu}_\omega$	$\hat{\mu}_\alpha$	$\hat{\mu}_\beta$	MSE_ω	MSE_α	MSE_β
η_t is standard normal								
200	GQMLE	—	0.1417	0.2832	0.5540	3.5848	3.0821	9.7503
	WOPIV	—	0.1410	0.2793	0.5514	3.6418	3.1293	10.6301
	HR	$O_P(n^{-2/5})$	0.1400	0.2817	0.5523	3.3366	3.2169	10.3751
	HR	$O_P(n^{-1/5})$	0.1401	0.2803	0.5529	3.3672	3.1070	10.2860
500	GQMLE	—	0.1125	0.2956	0.5860	1.5010	3.0996	6.5017
	WOPIV	—	0.1118	0.2935	0.5862	1.4907	3.0848	6.5737
	HR	$O_P(n^{-2/5})$	0.1121	0.2941	0.5863	1.5222	3.2294	6.7409
	HR	$O_P(n^{-1/5})$	0.1120	0.2942	0.5861	1.4875	3.0699	6.5164
1000	GQMLE	—	0.1059	0.2973	0.5939	1.0612	3.1335	5.5281
	WOPIV	—	0.1056	0.2965	0.5938	1.0682	3.1256	5.5803
	HR	$O_P(n^{-2/5})$	0.1058	0.2969	0.5936	1.1150	3.3126	5.8068
	HR	$O_P(n^{-1/5})$	0.1057	0.2968	0.5938	1.0564	3.1294	5.5321
η_t is skew normal								
200	GQMLE	—	0.1508	0.2895	0.5334	3.9382	3.8102	11.2614
	WOPIV	—	0.1444	0.2927	0.5376	3.7112	3.1911	10.2598
	HR	$O_P(n^{-2/5})$	0.1439	0.2930	0.5387	3.5037	2.8889	9.6033
	HR	$O_P(n^{-1/5})$	0.1508	0.2895	0.5322	3.9975	3.8124	11.3887
500	GQMLE	—	0.1101	0.2921	0.5917	1.3161	3.8766	6.6624
	WOPIV	—	0.1066	0.2920	0.5947	0.9969	3.3153	5.5099
	HR	$O_P(n^{-2/5})$	0.1075	0.2947	0.5916	0.8342	2.4666	4.4345
	HR	$O_P(n^{-1/5})$	0.1098	0.2917	0.5913	1.2667	3.7642	6.6063
1000	GQMLE	—	0.1072	0.2988	0.5900	1.2716	3.9706	6.8525
	WOPIV	—	0.1042	0.2965	0.5954	0.9202	2.9511	4.9964
	HR	$O_P(n^{-2/5})$	0.1037	0.2968	0.5960	0.6682	2.2406	3.7268
	HR	$O_P(n^{-1/5})$	0.1072	0.2990	0.5900	1.2451	3.9116	6.7115

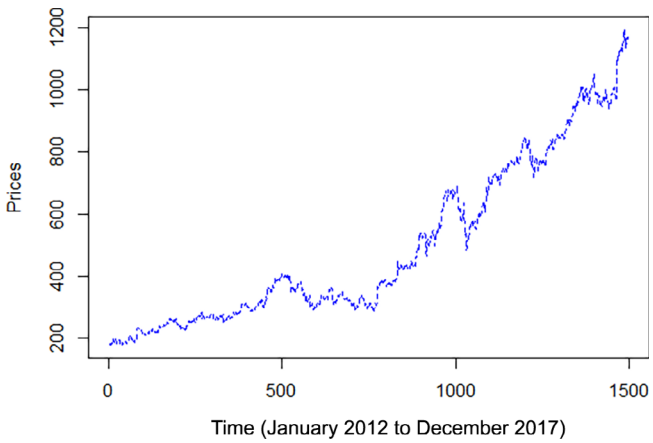

Figure 1. Series of the stock prices of AMZN

DOI: 10.1111/jtsa.12770

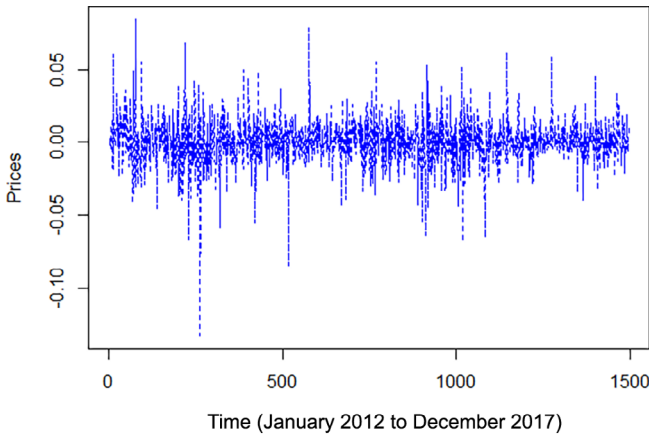

Figure 2. First difference in log prices of AMZN

Table V. Summary statistics of the AR(3) residual of the differenced logarithmic series of stock price

Series	Mean	SD	Min.	Max	Standard skewness	Standard kurtosis
Original	495.4334	266.8574	175.93	1195.83	0.7854	2.3485
Residual	0.0000	0.0157	−0.1330	0.0841	−0.3704	9.3648

Table VI. Estimation results of GARCH(1,1) for GQMLE and WOPIV

Methods	$\hat{\omega}$	$\hat{\alpha}$	$\hat{\beta}$
QMLE	1.0814	0.2519	0.3534
	(0.0850)	(0.0142)	(0.0342)
WOPIV	1.0161	0.2657	0.3706
	(0.0791)	(0.0143)	(0.0325)

clear indication of the presence of trends and potential autocorrelation. To remove these trends and autocorrelation, we work with the AR(3) residual of the logarithmic first-difference series of these prices. Figure 2 presents these series (re-scaled by 100), and Table V provides a summary of their descriptive statistics.

Compared to the original price series, the time series after pre-processing no longer exhibits non-stationarity. However, the sample skewness and kurtosis of both series suggest clear deviations from the normal distribution. For both QMLE and our proposed method, we fit the GARCH (1,1) model. The parameters of interest are $\theta = (\omega_0, \alpha_0, \beta_0)$, with the model specified as

$$\mathbf{r}_t(\mathbf{w}_t, \theta_0) = \begin{pmatrix} \varepsilon_t \\ \varepsilon_t^2 - (\omega_0 + \alpha_0 \varepsilon_{t-1}^2 + \beta_0 \sigma_{t-1}^2) \end{pmatrix}.$$

Table VI summarizes the estimation results for both the QMLE and the WOPIV. The estimates of $(\omega_0, \alpha_0, \beta_0)$ are reported, along with their SDs in the bracket. We observe that the estimates from the QMLE and the WOPIV are close to each other. All the coefficients are highly significant, and the standard deviation of our proposed method is generally smaller than those of the QMLE. Overall, the outcomes from both simulation and application support our estimation strategy.

ACKNOWLEDGMENT

Weining Wang's research is partially supported by the ESRC (Grant Reference: ES/T01573X/1). Open Access funding enabled and organized by Projekt DEAL.

DATA AVAILABILITY STATEMENT

The data supporting the findings of this article are openly available on Yahoo Finance at https://finance.yahoo.com. Here, we illustrate the use of our methods by modeling the processes of the stock price series of Amazon (AMZN). The observations are dated from January 2013 to December 2017. We have 1500 observations. The data is downloaded from Yahoo Finance: https://finance.yahoo.com.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

REFERENCES

Anatolyev S. 2003. The form of the optimal nonlinear instrument for multiperiod conditional moment restrictions. *Econometric Theory* **19**(4):602–609.
Anatolyev S. 2007. Optimal instruments in time series: a survey. *Journal of Economic Surveys* **21**(1):143–173.
Billingsley P. 1961. The Lindeberg-Levy theorem for martingales. *Proceedings of the American Mathematical Society* **12**(5):788–792.
Bollerslev T. 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* **31**(3):307–327.
Bollerslev T. 1987. A conditionally heteroskedastic time series model for speculative prices and rates of return. *The Review of Economics and Statistics* **69**(3):542–547.
Bollerslev T, Wooldridge JM. 1992. Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric Reviews* **11**(2):143–172.
Chandra SA, Taniguchi M. 2001. Estimating functions for nonlinear time series models. *Annals of the Institute of Statistical Mathematics* **53**(1):125–141.
Di J, Gangopadhyay A. 2011. On the efficiency of a semi-parametric GARCH model. *The Econometrics Journal* **14**(2):257–277.
Drost FC, Klaassen CA, Werker BJ. 1997. Adaptive estimation in time-series models. *The Annals of Statistics* **25**(2):786–817.
Durbin J. 1960. The fitting of time-series models. *Revue de l'Institut International de Statistique* **28**(3):233–244.
Engle RF, Gonzalez-Rivera G. 1991. Semiparametric arch models. *Journal of Business & Economic Statistics* **9**(4):345–359.
Engle RF, Lilien DM, Robins RP. 1987. Estimating time varying risk premia in the term structure: the arch-m model. *Econometrica: Journal of the Econometric Society* **55**(2):391–407.
Fan J, Qi L, Xiu D. 2014. Quasi-maximum likelihood estimation of GARCH models with heavy-tailed likelihoods. *Journal of Business & Economic Statistics* **32**(2):178–191.
Francq C, Zakoïan J-M. 2004. Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli* **10**(4):605–637.
Francq C, Zakoian J-M. 2010. *GARCH Models: Structure, Statistical Inference and Financial Applications*. Chichester: John Wiley & Sons.
Godambe V. 1985. The foundations of finite sample estimation in stochastic processes. *Biometrika* **72**(2):419–428.
Godambe VP, Heyde CC. 2010. Quasi-likelihood and optimal estimation. In *Selected works of C.C. Heyde*, New York: Springer; 386–399.
Hafner CM, Rombouts JV. 2007. Semiparametric multivariate volatility models. *Econometric Theory* **23**(2):251–280.
Harvey CR, Siddique A. 1999. Autoregressive conditional skewness. *Journal of Financial and Quantitative Analysis* **34**(4):465–487.
Heyde CC. 1997. *Quasi-Likelihood and Its Application: A General Approach to Optimal Parameter Estimation*. New York: Springer.
Im KS, Schmidt P. 2008. More efficient estimation under non-normality when higher moments do not depend on the regressors, using residual augmented least squares. *Journal of Econometrics* **144**(1):219–233.
Jacod J, Sørensen M. 2018. A review of asymptotic theory of estimating functions. *Statistical Inference for Stochastic Processes* **21**(2):415–434.
Komunjer I, Vuong Q. 2010. Semiparametric efficiency bound in time-series models for conditional quantiles. *Econometric Theory* **26**(2):383–405.

Li DX, Turtle HJ. 2000. Semiparametric arch models: an estimating function approach. *Journal of Business & Economic Statistics* **18**(2):174–186.
Liang K-Y, Zeger SL. 1986. Longitudinal data analysis using generalized linear models. *Biometrika* **73**(1):13–22.
Magnus JR, Neudecker H. 1979. The commutation matrix: some properties and applications. *The Annals of Statistics* **7**(2):381–394.
Meddahi N, Renault É. 1998. Quadratic M-estimators for ARCH-type processes.
Newey WK. 1994. The asymptotic variance of semiparametric estimators. *Econometrica: Journal of the Econometric Society* **62**(6):1349–1382.
Newey WK, Steigerwald DG. 1997. Asymptotic bias for quasi-maximum-likelihood estimators in conditional heteroskedasticity models. *Econometrica: Journal of the Econometric Society* **65**(3):587–599.
Patton AJ. 2006. Modelling asymmetric exchange rate dependence. *International Economic Review* **47**(2):527–556.
Prono T. 2010. Simple GMM estimation of the semi-strong GARCH(1,1) model. Technical Report, University Library of Munich, Germany.
Richter S, Wang W, Wu WB. 2023. Testing for parameter change epochs in GARCH time series. *The Econometrics Journal* **26**(3):utad006.
Van der Vaart AW. 2000. *Asymptotic Statistics*, Vol. 3. Cambridge: Cambridge university press.
Wooldridge JM. 1994. Estimation and inference for dependent processes. *Handbook of Econometrics* **4**:2639–2738.
Zeger SL, Liang K-Y. 1986. Longitudinal data analysis for discrete and continuous outcomes. *Biometrics* **42**(1):121–130.

APPENDIX A: PROOFS AND OTHER RESULTS

Here, we present the proofs of the main theorems. In addition, the asymptotic results of the GQMLE are also included.

A.1. Consistency

A.1.1. Proof of Theorem 1

Proof. By A.4, we have $Q_T\left(\check{\theta}, \check{\kappa}, \hat{\theta}\right) = 0$. Thus,

$$\begin{aligned} |Q_T(\check{\theta}, \check{\kappa}, \hat{\theta})|_2 \leq & |Q_T(\check{\theta}, \check{\kappa}, \theta_0)|_2 \\ = & |Q_T(\theta_0, \kappa_0, \theta_0)|_2 + o_p(1) \\ = & |Q_\infty(\theta_0, \kappa_0, \theta_0)|_2 + o_p(1), \end{aligned} \tag{A1}$$

where the first inequality follows from A.4, the first equality follows from A.1, and the second equality follows from A.2. Subtracting $|Q_\infty(\theta_0, \kappa_0, \hat{\theta})|_2$ from both sides of (A1) yields

$$\begin{aligned} |Q_T(\check{\theta}, \check{\kappa}, \hat{\theta})|_2 - |Q_\infty(\theta_0, \kappa_0, \hat{\theta})|_2 &\leq |Q_\infty(\theta_0, \kappa_0, \theta_0)|_2 - |Q_\infty(\theta_0, \kappa_0, \hat{\theta})|_2 + o_p(1), \\ |Q_T(\theta_0, \kappa_0, \hat{\theta})|_2 + o_p(1) - |Q_\infty(\theta_0, \kappa_0, \hat{\theta})|_2 &\leq 0 - |Q_\infty(\theta_0, \kappa_0, \hat{\theta})|_2 + o_p(1), \\ o_p(1) + o_p(1) &\leq -|Q_\infty(\theta_0, \kappa_0, \hat{\theta})|_2 + o_p(1), \\ |Q_\infty(\theta_0, \kappa_0, \hat{\theta})|_2 &\leq o_p(1), \end{aligned} \tag{A2}$$

where the second inequality follows from A.1 and A.3, and the third inequality follows from A.2. By A.3, for any ε, there exists η such that for any $\theta \in \Theta$, it holds $|\theta - \theta_0|_2 \geq \varepsilon \Rightarrow |Q_\infty(\theta_0, \kappa_0, \theta)|_2 > \eta$. By taking $\theta = \hat{\theta}$ in the preceding expression, we obtain

$$\mathrm{P}\left(|\hat{\theta} - \theta_0|_2 \geq \varepsilon\right) \leq \mathrm{P}\left(|Q_\infty(\theta_0, \kappa_0, \hat{\theta})|_2 > \eta\right) \to 0,$$

where the $\to$ follows from (A2). Thus, $|\hat{\theta} - \theta_0|_2 = o_p(1)$. ∎

DOI: 10.1111/jtsa.12770 *Journal of Time Series Analysis* published by John Wiley & Sons Ltd.

A.1.2. Proof of Theorem 2

Proof. Recall that the b-ball around $\dot{\theta}$ is defined as $\mathcal{B}(\dot{\theta}, b) = \{\theta : |\theta - \dot{\theta}|_2 < b\}$, where $\dot{\theta}$ is a point within the interior of Θ, and b is a positive constant. For $k = 1, 2, \ldots$, let $\mathcal{B}(\dot{\theta}, 1/k)$ be a sequence of shrinking open balls centered around $\dot{\theta}$. For any $\theta_1 \in \Theta$ that satisfies $\theta_1 \neq \theta_0$, we have

$$\lim_{k\to\infty} \inf_{\theta\in\mathcal{B}(\theta_1,1/k)} \lim_{T\to\infty} |\mathrm{E}Q_T(\theta_0, \kappa_0, \theta)|_2 \uparrow |Q_\infty(\theta_0, \kappa_0, \theta_1)|_2 > 0 = |Q_\infty(\theta_0, \kappa_0, \theta_0)|_2, \tag{A3}$$

where the $\uparrow$ follows by the monotone convergence theorem, and the inequality follows by A.3.

Thus, for this θ_1, we can find a large enough positive integer $k(\theta_1)$ such that

$$\inf_{\theta\in\mathcal{B}(\theta_1,1/k(\theta_1))} \lim_{T\to\infty} |\mathrm{E}Q_T(\theta_0, \kappa_0, \theta)|_2 > 0 = |Q_\infty(\theta_0, \kappa_0, \theta_0)|_2. \tag{A4}$$

Considering $\Theta_\varepsilon \overset{\text{def}}{=} \{\theta : |\theta - \theta_0|_2 \geq \varepsilon\}$ for a positive constant ε, by A.2'(ii), we can find a finite set of l points, such that $\{\theta_i\}_{i=1}^l \in \Theta_\varepsilon$ and $\Theta_\varepsilon \subset \cup_{i=1}^l \mathcal{B}(\theta_i, 1/k(\theta_i))$. Furthermore, let us define

$$\delta = \min[\inf_{i\in 1,\cdots,l} \lim \inf_{\theta\in\mathcal{B}(\theta_i,1/k(\theta_i))} \lim_{T\to\infty} |\mathrm{E}Q_T(\theta_0, \kappa_0, \theta)|_2, 1]. \tag{A5}$$

By (A4) and $l < \infty$, it holds that $\delta > 0$. By A.4 and the definition of $\{\theta_i\}_{i=1}^l$, we have $\mathrm{P}(|\hat{\theta} - \theta_0| \geq \varepsilon) \leq \mathrm{P}(\inf_{i=1,\cdots,l} \inf_{\theta\in\mathcal{B}(\theta_i,1/k(\theta_i))} |Q_T(\check{\theta}, \check{\kappa}, \theta)|_2 - |Q_T(\check{\theta}, \check{\kappa}, \theta_0)|_2 < 0)$. Next, we show that the right-hand side of this equation is $o(1)$.

By A.1, A.2'(i), (A3), and (A5), it holds that $\mathrm{P}[\inf_{i=1,\cdots,l} \inf_{\theta\in\mathcal{B}(\theta_i,1/k(\theta_i))} |Q_T(\check{\theta}, \check{\kappa}, \theta)|_2 \leq \delta/2] = o(1)$. By A.1 and A.2'(iii), it holds that $\mathrm{P}(|Q_T(\check{\theta}, \check{\kappa}, \theta_0) - Q_\infty(\theta_0, \kappa_0, \theta_0)|_2 \geq \delta/2) = \mathrm{P}(|Q_T(\check{\theta}, \check{\kappa}, \theta_0)|_2 \geq \delta/2) = o(1)$. Consequently, $\mathrm{P}(|\hat{\theta} - \theta_0| \geq \varepsilon) \leq \mathrm{P}(\inf_{i=1,\cdots,l} \inf_{\theta\in\mathcal{B}(\theta_i,1/k(\theta_i))} |Q_T(\check{\theta}, \check{\kappa}, \theta)|_2 - |Q_T(\check{\theta}, \check{\kappa}, \theta_0)|_2 < 0) = o(1)$. ∎

A.2. Proof of Theorem 3

Before we begin the proof, let us first review the notations defined in Section 3.2. For any $\theta, \tilde{\theta} \in \Theta$ and initial estimators $(\check{\theta}, \check{\kappa})$, we have

$$\begin{aligned}
&\mathbf{r}_t(\theta) = \mathbf{r}_t(\mathbf{w}_t(\theta), \theta),\\
&\mathbf{R}_t(\tilde{\theta}) = \nabla \mathbf{r}_t(\tilde{\theta}) = \partial \mathbf{r}_t(\theta)/\partial\theta'|_{\theta=\tilde{\theta}},\\
&\check{\mathbf{R}}_t = \mathbf{R}_t(\check{\theta}), \mathbf{R}_t = \mathbf{R}_t(\theta_0), \mathbf{D}_t = \mathbf{D}_t(\mathbf{x}_t, \boldsymbol{\theta}_0, \boldsymbol{\kappa}_0).
\end{aligned}$$

Further,

$$\begin{aligned}
\check{\mathbf{A}}_T(\theta) &= \frac{1}{T}\sum_{t=1}^T \check{\mathbf{R}}_t' \mathbf{D}_t(\mathbf{x}_t(\check{\theta}), \check{\boldsymbol{\theta}}, \check{\boldsymbol{\kappa}})^{-1} \mathbf{r}_t(\theta),\\
\mathbf{A}_T(\theta) &= \frac{1}{T}\sum_{t=1}^T \mathbf{R}_t(\theta_0)' \mathbf{D}_t(\mathbf{x}_t, \boldsymbol{\theta}_0, \boldsymbol{\kappa}_0)^{-1} \mathbf{r}_t(\theta),
\end{aligned}$$

and

$$\check{\mathbf{B}}_T(\theta) = \frac{1}{T}\sum_{t=1}^T \check{\mathbf{R}}_t' \mathbf{D}_t(\mathbf{x}_t(\check{\theta}), \check{\boldsymbol{\theta}}, \check{\boldsymbol{\kappa}})^{-1} \mathbf{R}_t(\theta),$$

Journal of Time Series Analysis published by John Wiley & Sons Ltd. DOI: 10.1111/jtsa.12770

$$\begin{aligned}
\check{\mathbf{B}}_0(\theta) &= \mathrm{E}(\check{\mathbf{R}}_t'\mathbf{D}_t(\mathbf{x}_t(\check{\theta}), \check{\boldsymbol{\theta}}, \check{\boldsymbol{\kappa}})^{-1}\mathbf{R}_t(\theta)),\\
\mathbf{B}_0(\theta) &= \mathrm{E}(\mathbf{R}_t(\theta_0)'\mathbf{D}_t(\mathbf{x}_t, \boldsymbol{\theta}_0, \boldsymbol{\kappa}_0)^{-1}\mathbf{R}_t(\theta)),\\
\mathbf{B}_0 &= \mathbf{B}_0(\theta_0),\\
\mathbf{C}_0 &= \mathrm{E}(\mathbf{R}_t(\theta_0)'\mathbf{D}_t(\mathbf{x}_t, \boldsymbol{\theta}_0, \boldsymbol{\kappa}_0)^{-1}\mathbf{r}_t(\theta_0)\mathbf{r}_t(\theta_0)'\mathbf{D}_t(\mathbf{x}_t, \boldsymbol{\theta}_0, \boldsymbol{\kappa}_0)^{-1}\mathbf{R}_t(\theta_0)).
\end{aligned} \tag{A6}$$

Proof. We start with equation (21),

$$\frac{1}{T}\sum_{t=1}^{T}\check{\mathbf{R}}_t'\mathbf{D}_t(\mathbf{x}_t(\check{\theta}), \check{\boldsymbol{\theta}}, \check{\boldsymbol{\kappa}})^{-1}\mathbf{r}_t(\hat{\theta}) = 0.$$

We are interested in $\hat{\theta}$, the root of this score function. A Taylor expansion leads to

$$\begin{aligned}
0 &= \frac{1}{T}\sum_{t=1}^{T}\check{\mathbf{R}}_t'\mathbf{D}_t\left(\mathbf{x}_t\left(\check{\theta}\right), \check{\theta}, \check{\kappa}\right)^{-1}\mathbf{r}_t(\hat{\theta})\\
&= \frac{1}{T}\sum_{t=1}^{T}\check{\mathbf{R}}_t'\mathbf{D}_t\left(\mathbf{x}_t\left(\check{\theta}\right), \check{\theta}, \check{\kappa}\right)^{-1}\mathbf{r}_t\left(\theta_0\right) + \frac{1}{T}\sum_{t=1}^{T}\check{\mathbf{R}}_t'\mathbf{D}_t\left(\mathbf{x}_t\left(\check{\theta}\right), \check{\theta}, \check{\kappa}\right)^{-1}\nabla\mathbf{r}_t(\tilde{\theta})\left(\hat{\theta} - \theta_0\right),
\end{aligned}$$

where $\tilde{\theta}$ is a vector that is between $\hat{\theta}$ and θ_0 (in an element-wise sense). By B.2(ii) and some rearrangements, we have

$$\begin{aligned}
\sqrt{T}\left(\hat{\theta} - \theta_0\right) =& -\mathbf{B}_0^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{R}_t'\mathbf{D}_t\left(\mathbf{x}_t, \theta_0, \kappa_0\right)^{-1}\mathbf{r}_t\left(\theta_0\right)\\
&- \mathbf{B}_0^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\check{\mathbf{R}}_t'\mathbf{D}_t\left(\mathbf{x}_t\left(\check{\theta}\right), \check{\theta}, \check{\kappa}\right)^{-1} - \mathbf{R}_t'\mathbf{D}_t\left(\mathbf{x}_t, \theta_0, \kappa_0\right)^{-1}\right]\mathbf{r}_t\left(\theta_0\right)\\
&- \mathbf{B}_0^{-1}\left(\check{\mathbf{B}}_T(\tilde{\theta}) - \mathbf{B}_0\right)\sqrt{T}\left(\hat{\theta} - \theta_0\right)\\
=& \; I + II + III.
\end{aligned}$$

By B.2(i) and B.3(ii), it holds that $II = o_p(1)$. By A.1–A.4, we have $\hat{\theta} \to_p \theta_0$, which implies $\tilde{\theta} \to_p \theta_0$. As a result, $\check{\mathbf{B}}_T(\tilde{\theta}) - \mathbf{B}_0 = o_p(1)$ holds by B.2(i) and B.3(i), leading to $III = o_p\left(\sqrt{T}(\hat{\theta} - \theta_0)\right)$. Thus,

$$\sqrt{T}\left(\hat{\theta} - \theta_0\right) = -\frac{1}{\sqrt{T}}\mathbf{B}_0^{-1}\sum_{t=1}^{T}\mathbf{R}_t'\mathbf{D}_t\left(\mathbf{x}_t, \theta_0, \kappa_0\right)^{-1}\mathbf{r}_t\left(\theta_0\right) + o_p\left(\sqrt{T}(\hat{\theta} - \theta_0)\right). \tag{A7}$$

Consequently, the first term of the right-hand side of (A7) is the leading term.

Define $\zeta_t = \mathbf{B}_0^{-1}\mathbf{R}_t'\mathbf{D}_t(\mathbf{x}_t, \boldsymbol{\theta}_0, \boldsymbol{\kappa}_0)^{-1}\mathbf{r}_t(\theta_0)$. For an arbitrary P-dimensional vector c, $\xi_t \overset{\text{def}}{=} c'\zeta_t$ is a stationary and ergodic MDS by B.1, and its variance is finite by B.2(ii). As a result, $\lim_{T\to\infty}\mathrm{E}\left[\left(-\frac{1}{\sqrt{T}}\sum_t \xi_t\right)^2\right] = \mathrm{E}(c'\mathbf{B}_0^{-1}\mathbf{C}_0\mathbf{B}_0^{-1}c)$. Then,

$$\sqrt{T}(\hat{\theta} - \theta_0) \to_L N(\mathbf{0}, \mathbf{B}_0^{-1}\mathbf{C}_0\mathbf{B}_0^{-1}) \tag{A8}$$

follows from the CLT for the MDS (Billingsley (1961); see also corollary A.1 of Francq and Zakoian (2010)), the Cramér Wold device, and (A7).

DOI: 10.1111/jtsa.12770 *Journal of Time Series Analysis* published by John Wiley & Sons Ltd.

Finally, if the variance–covariance matrix is correctly specified, namely, $\mathrm{E}(\mathbf{r}_t(\theta_0)\mathbf{r}_t'(\theta_0)|\mathbf{x}_t) = \mathbf{D}_t(\mathbf{x}_t, \boldsymbol{\theta}_0, \boldsymbol{\kappa}_0)$, then,

$$\begin{aligned}\mathbf{C}_0 &= \mathrm{E}(\mathbf{R}_t(\theta_0)'\mathbf{D}_t(\mathbf{x}_t, \boldsymbol{\theta}_0, \boldsymbol{\kappa}_0)^{-1}\mathbf{r}_t(\theta_0)\mathbf{r}_t'(\theta_0)\mathbf{D}_t(\mathbf{x}_t, \boldsymbol{\theta}_0, \boldsymbol{\kappa}_0)^{-1}\mathbf{R}_t(\theta_0)) \\ &= \mathrm{E}(\mathrm{E}(\mathbf{R}_t(\theta_0)'\mathbf{D}_t(\mathbf{x}_t, \boldsymbol{\theta}_0, \boldsymbol{\kappa}_0)^{-1}\mathbf{r}_t(\theta_0)\mathbf{r}_t'(\theta_0)\mathbf{D}_t(\mathbf{x}_t, \boldsymbol{\theta}_0, \boldsymbol{\kappa}_0)^{-1}\mathbf{R}_t(\theta_0)|\mathbf{x}_t)) \\ &= \mathrm{E}(\mathbf{R}_t(\theta_0)'\mathbf{D}_t(\mathbf{x}_t, \boldsymbol{\theta}_0, \boldsymbol{\kappa}_0)^{-1}\mathrm{E}(\mathbf{r}_t(\theta_0)\mathbf{r}_t'(\theta_0)|\mathbf{x}_t)\mathbf{D}_t(\mathbf{x}_t, \boldsymbol{\theta}_0, \boldsymbol{\kappa}_0)^{-1}\mathbf{R}_t(\theta_0)) \\ &= \mathrm{E}(\mathbf{R}_t(\theta_0)'\mathbf{D}_t(\mathbf{x}_t, \boldsymbol{\theta}_0, \boldsymbol{\kappa}_0)^{-1}\mathbf{R}_t(\theta_0)) = \mathbf{B}_0.\end{aligned}$$

Consequently, the asymptotic variance-covariance matrix of (A8) becomes $\mathbf{B}_0^{-1}$. ■

A.3. Proof of Theorem 4

Proof. Define $\check{Q}_T(\theta) = \frac{1}{T}\sum_{t=1}^{T}\check{\mathbf{R}}_t' \mathbf{D}_t(\mathbf{x}_t(\check{\theta}), \check{\theta}, \check{\kappa})^{-1}\mathbf{r}_t(\theta)$. Under A.4 and B.2(i), equations (14), (19), and (A6) yield the following equalities:

$$\begin{aligned}\check{Q}_T(\hat{\theta}) &= 0, \\ \check{\mathbf{B}}_T(\check{\theta}) &= \partial\check{Q}_T(\theta)/\partial\theta'|_{\theta=\check{\theta}}, \\ \overline{\theta} &= \check{\theta} - \check{\mathbf{B}}_T(\check{\theta})^{-1}\check{Q}_T(\check{\theta}).\end{aligned}$$

Rearranging the last equality, we have

$$-\check{Q}_T(\check{\theta}) = \check{\mathbf{B}}_T(\check{\theta})(\overline{\theta} - \check{\theta}). \tag{A9}$$

Extending $\check{Q}_T(\hat{\theta}) = 0$ around $\check{\theta}$ yields

$$\begin{aligned}0 &= \check{Q}_T(\check{\theta}) + \check{\mathbf{B}}_T(\check{\theta})(\hat{\theta} - \check{\theta}) + o_p(\hat{\theta} - \check{\theta}) \\ &= \check{Q}_T(\check{\theta}) + \check{\mathbf{B}}_T(\check{\theta})(\hat{\theta} - \check{\theta}) + o_p(\frac{1}{\sqrt{T}}),\end{aligned} \tag{A10}$$

where the first equality follows from B.2(i), and the second equality follows from C.1. Rearranging (A10) yields

$$-\check{Q}_T(\check{\theta}) = \check{\mathbf{B}}_T(\check{\theta})(\hat{\theta} - \check{\theta}) + o_p(\frac{1}{\sqrt{T}}). \tag{A11}$$

Combining (A9) and (A11) and rescaling them by $\sqrt{T}$, we have

$$\begin{aligned}o_p(1) &= \sqrt{T}\check{\mathbf{B}}_T(\check{\theta})(\overline{\theta} - \hat{\theta}) \\ &= \sqrt{T}(\mathbf{B}_0 + o_p(1))(\overline{\theta} - \hat{\theta}),\end{aligned}$$

where the second equality follows from B.3(i), B.2(i), and C.1. After multiplying both sides by $\mathbf{B}_0^{-1}$ and some rearrangements, we obtain

$$\begin{aligned}\sqrt{T}(\overline{\theta} - \hat{\theta}) &= \mathbf{B}_0^{-1}o_p(1) - \mathbf{B}_0^{-1}o_p\left(\sqrt{T}(\overline{\theta} - \hat{\theta})\right) \\ &= o_p(1),\end{aligned}$$

where the first equality follows from B.2(ii). This shows the first-order asymptotic equivalence between $\hat{\theta}$ and $\overline{\theta}$. ■

Journal of Time Series Analysis published by John Wiley & Sons Ltd. DOI: 10.1111/jtsa.12770

A.4. Asymptotic properties of the GQMLE

The asymptotic normality and consistency of the GQMLE are well established in the literature.

M.1 θ_0 lies in the interior of Θ.
M.2 $\mathbf{I}_0$ is positive definite.
M.3 $\sup_{\theta\in\Theta} |\frac{1}{T}\sum_{t=1}^{T}\mathbf{I}_t(\theta) - \mathbf{I}_0(\theta)|_2 \to_p 0$.
M.4 $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{s}_t(\theta_0) \to_d N(\mathbf{0}, \mathbf{J}_0)$.

Theorem 6. (Wooldridge (1994))**.** Under Assumptions A.4 to A.4, the GQMLE is asymptotically normal, $\sqrt{T}(\check{\theta} - \theta_0) \to_d N(\mathbf{0}, \mathbf{I}_0^{-1}\mathbf{J}_0\mathbf{I}_0^{-1})$.

A proof can be found in Wooldridge (1994).

APPENDIX B: EVALUATING ASSUMPTIONS WITHIN GARCH(1,1)

Here, we examine the validity of Assumptions A.1–A.4, A.2', and A.0 specifically within the context of the GARCH(1,1) model. Throughout this section, we assume that the time series $\{y_t, \mathbf{x}_t\}_t$ defined for the GARCH(1,1) model described in (5) is both covariance stationary and strict stationary, and is also ergodic. This is satisfied if the innovation term η_t is i.i.d. for all t, $\mathrm{E}(\varepsilon_t^2) < \infty$, and the parameters meet the criteria $\alpha_0, \beta_0 > 0$ and $\alpha_0 + \beta_0 < 1$. We denote this specific subset of the parameter space Θ as Θ_s, where 's' represents 'stationary'. It shall be noted that a GARCH process can be strictly stationary without necessarily being covariance stationary. See, e.g., figure 2.8 in Francq and Zakoian (2010) and follow-up discussions for more details. To streamline our explanation, we assume that the innovation η_t is i.i.d. and its $4+\delta$th moment exists, for some $\delta > 0$.

B.1. On Assumptions A.1–A.4

First, we verify Assumption A.2 within Θ_s. Let Θ_π be a finite grid of Θ, partitioned with sufficient granularity. Then, the uniform convergence condition presented in A.2, can be implied by the pointwise convergence on these finite grid points, i.e., $\sup_{\theta\in\Theta_\pi} |Q_T(\theta_0, \kappa_0, \theta) - Q_\infty(\theta_0, \kappa_0, \theta)|_2 \to_p 0$, and a stochastic equicontinuity condition, i.e., for any $\theta_1, \theta_2 \in \Theta_s$,

$$\mathrm{plim}_{\delta\to 0} \sup_{|\theta_1-\theta_2|_2\le\delta} |Q_T(\theta_0, \kappa_0, \theta_1) - Q_\infty(\theta_0, \kappa_0, \theta_1) - Q_T(\theta_0, \kappa_0, \theta_2) + Q_\infty(\theta_0, \kappa_0, \theta_2)|_2 = 0. \tag{B1}$$

Now, we show the pointwise convergence. Recall that $\mathbf{R}_t'\mathbf{D}_t^{-1}$ is the instrument evaluated at θ_0, σ_t^2 is the true conditional variance evaluated at θ_0, and $\varepsilon_t = \sigma_t\eta_t$. For a given θ, a reasoning similar to that in Section 2.2 yields

$$Q_T(\theta_0, \kappa_0, \theta) = \frac{1}{T}\sum_{t=1}^{T}\mathbf{R}_t'\mathbf{D}_t^{-1}\mathbf{r}_t(\theta) = \frac{1}{T}\sum_{t=1}^{T}\frac{1}{c_\kappa^0}\begin{pmatrix} \frac{\varepsilon_t^2-\sigma_t^2(\theta)}{\sigma_t^4} - \kappa_3^0\sigma_t^{-3}\varepsilon_t \\ \frac{\varepsilon_t^2-\sigma_t^2(\theta)}{\sigma_t^4}\varepsilon_{t-1}^2 - \kappa_3^0\sigma_t^{-3}\varepsilon_{t-1}^2\varepsilon_t \\ \frac{\varepsilon_t^2-\sigma_t^2(\theta)}{\sigma_t^4}\sigma_{t-1}^2 - \kappa_3^0\sigma_t^{-3}\sigma_{t-1}^2\varepsilon_t \end{pmatrix}. \tag{B2}$$

Note that $\frac{1}{T}\sum_{t=1}^{T}[\mathbf{R}_t'\mathbf{D}_t^{-1}\mathbf{r}_t(\theta) - \mathrm{E}(\mathbf{R}_t'\mathbf{D}_t^{-1}\mathbf{r}_t(\theta)|\mathcal{F}_{t-1})]$ is an averaged sum of MDS. It will be an $O_p(\frac{1}{\sqrt{T}})$ term if the innovation η_t is i.i.d. and its second moment exists. Thus, we can focus on the term $\frac{1}{T}\sum_{t=1}^{T}\mathrm{E}(\mathbf{R}_t'\mathbf{D}_t^{-1}\ \mathbf{r}_t(\theta)|\mathcal{F}_{t-1})$.

Note that by (B2),

$$\frac{1}{T}\sum_{t=1}^{T}\mathrm{E}(\mathbf{R}_t'\mathbf{D}_t^{-1}\mathbf{r}_t(\theta)|\mathcal{F}_{t-1}) = \frac{1}{T}\sum_{t=1}^{T}\frac{1}{c_\kappa^0}\mathrm{E}\begin{pmatrix}\frac{\varepsilon_t^2-\sigma_t^2(\theta)}{\sigma_t^4}-\kappa_3^0\sigma_t^{-3}\varepsilon_t\\ \frac{\varepsilon_t^2-\sigma_t^2(\theta)}{\sigma_t^4}\varepsilon_{t-1}^2-\kappa_3^0\sigma_t^{-3}\varepsilon_{t-1}^2\varepsilon_t\\ \frac{\varepsilon_t^2-\sigma_t^2(\theta)}{\sigma_t^4}\sigma_{t-1}^2-\kappa_3^0\sigma_t^{-3}\sigma_{t-1}^2\varepsilon_t\end{pmatrix}\Bigg|\mathcal{F}_{t-1}$$
$$= \frac{1}{T}\sum_{t=1}^{T}\frac{1}{c_\kappa^0}\mathrm{E}\begin{pmatrix}\frac{\varepsilon_t^2-\sigma_t^2(\theta)}{\sigma_t^4}\\ \frac{\varepsilon_t^2-\sigma_t^2(\theta)}{\sigma_t^4}\varepsilon_{t-1}^2\\ \frac{\varepsilon_t^2-\sigma_t^2(\theta)}{\sigma_t^4}\sigma_{t-1}^2\end{pmatrix}\Bigg|\mathcal{F}_{t-1}, \tag{B3}$$

where the second equality follows from $\mathrm{E}(\varepsilon_t|\mathcal{F}_{t-1})=0$ and $\sigma_t^{-3}=\left(\omega_0+\alpha_0\varepsilon_{t-1}^2+\beta_0\sigma_{t-1}^2\right)^{-3/2}\in\mathcal{F}_{t-1}$. Recall that $\mathbf{x}_t=(\varepsilon_{t-1},\sigma_{t-1})$. We define

$$f(\ldots\mathbf{x}_{t-1},\mathbf{x}_t)\overset{\text{def}}{=}\mathrm{E}\begin{pmatrix}\frac{\varepsilon_t^2-\sigma_t^2(\theta)}{\sigma_t^4}\\ \frac{\varepsilon_t^2-\sigma_t^2(\theta)}{\sigma_t^4}\varepsilon_{t-1}^2\\ \frac{\varepsilon_t^2-\sigma_t^2(\theta)}{\sigma_t^4}\sigma_{t-1}^2\end{pmatrix}\Bigg|\mathcal{F}_{t-1}\Bigg)=\begin{pmatrix}\frac{\sigma_t^2-\sigma_t^2(\theta)}{\sigma_t^4}\\ \frac{\sigma_t^2-\sigma_t^2(\theta)}{\sigma_t^4}\varepsilon_{t-1}^2\\ \frac{\sigma_t^2-\sigma_t^2(\theta)}{\sigma_t^4}\sigma_{t-1}^2\end{pmatrix}, \tag{B4}$$

where the last equality follows from $\sigma_t^2(\theta)=\omega+\alpha\varepsilon_{t-1}^2+\beta\sigma_{t-1}^2(\theta)$ and $\sigma_t^2\overset{\text{def}}{=}\sigma_t^2(\theta_0)$. Note that $\mathbf{x}_t$ is assumed to be stationary and ergodic; f is measurable by (B4); and $\mathrm{E}|f(\ldots\mathbf{x}_{t-1},\mathbf{x}_t)|<\infty$ is implied by $\mathrm{E}(\varepsilon_t^2)<\infty$, the stationarity of ε_t^2, and $\theta\in\Theta_s$. Thus, by the ergodic theorem (e.g., theorem A.2 of Francq and Zakoian (2010)) and the Law of iterated expectation, the last term of (B3) converges to

$$\frac{1}{c_\kappa^0}\mathrm{E}\begin{pmatrix}\frac{\varepsilon_t^2-\sigma_t^2(\theta)}{\sigma_t^4}\\ \frac{\varepsilon_t^2-\sigma_t^2(\theta)}{\sigma_t^4}\varepsilon_{t-1}^2\\ \frac{\varepsilon_t^2-\sigma_t^2(\theta)}{\sigma_t^4}\sigma_{t-1}^2\end{pmatrix}=Q_\infty(\theta_0,\kappa_0,\theta)=\mathrm{E}[\mathbf{R}_t(\theta)'\mathbf{D}_t(\mathbf{x}_t(\theta),\theta,\boldsymbol{\kappa})^{-1}\mathbf{r}_t(\mathbf{w}_t(\theta),\theta)]. \tag{B5}$$

Thus, we have the pointwise convergence of $|Q_T(\theta_0,\kappa_0,\theta)-Q_\infty(\theta_0,\kappa_0,\theta)|_2$ for the GARCH(1,1) model.

Next, we show the stochastic equicontinuity condition (B1). By iterating $\sigma_t^2(\theta)=\omega+\alpha\varepsilon_{t-1}^2+\beta\sigma_{t-1}^2(\theta)$, we obtain $\sigma_t^2(\theta)=\frac{\omega}{1-\beta}+\alpha\sum_{k=0}^{\infty}\beta^k\varepsilon_{t-1-k}^2$. Then,

$$\begin{aligned}Q_T\left(\theta_0,\kappa_0,\theta_1\right)-Q_T\left(\theta_0,\kappa_0,\theta_2\right)&=\frac{1}{T}\sum_{t=1}^{T}\mathbf{R}_t'\mathbf{D}_t^{-1}\left[\mathbf{r}_t\left(\theta_1\right)-\mathbf{r}_t\left(\theta_2\right)\right]\\&=\frac{1}{T}\sum_{t=1}^{T}\mathbf{R}_t'\mathbf{D}_t^{-1}\begin{pmatrix}0\\ \frac{\omega_1}{1-\beta_1}-\frac{\omega_2}{1-\beta_2}+\alpha_1\sum_{k=0}^{\infty}\beta_1^k\varepsilon_{t-1-k}^2-\alpha_2\sum_{k=0}^{\infty}\beta_2^k\varepsilon_{t-1-k}^2\end{pmatrix}\\&=\left(\frac{\omega_1}{1-\beta_1}-\frac{\omega_2}{1-\beta_2}\right)\frac{1}{T}\sum_{t=1}^{T}\mathbf{R}_t'\mathbf{D}_t^{-1}\begin{pmatrix}0\\1\end{pmatrix}\\&+\sum_{k=0}^{\infty}\left(\alpha_1\beta_1^k-\alpha_2\beta_2^k\right)\left[\frac{1}{T}\sum_{t=1}^{T}\mathbf{R}_t'\mathbf{D}_t^{-1}\begin{pmatrix}0\\ \varepsilon_{t-1-k}^2\end{pmatrix}\right].\end{aligned} \tag{B6}$$

Journal of Time Series Analysis published by John Wiley & Sons Ltd. DOI: 10.1111/jtsa.12770

For any $\theta = (\omega, \alpha, \beta)' \in \Theta_s$, it holds that $\alpha, \beta > 0$ and $\alpha + \beta < 1$ by the definition of Θ_s. Thus, the last equality of (B6) indicates that $Q_T\left(\theta_0, \kappa_0, \theta\right)$ is Lipschitz in $\theta \in \Theta_s$. By the same reasoning, $Q_\infty\left(\theta_0, \kappa_0, \theta\right)$ is also Lipschitz in $\theta \in \Theta_s$, as can be seen by substituting '$\frac{1}{T}\sum_{t=1}^{T}$' with 'E' in (B6). Consequently, equation (B1) holds, and we have Assumption A.2 satisfied.

For Assumption A.3, the equality (B5) illustrates that for the GARCH(1,1) model, $Q_\infty(\theta_0, \kappa_0, \theta)$ is equivalent to the mean of the score function of the GQMLE up to a constant factor (see also (27) and (28)). Thus, A.3 is satisfied under the standard identification conditions for the GQMLE; see, e.g., p. 144 of Francq and Zakoian (2010).

For Assumption A.4, we have commented in a remark after equation (14) that the existence of solutions of nonlinear Z-estimation in a finite sample is a non-trivial problem. We refer to Jacod and Sørensen (2018) for further discussions.

Lastly, we verify Assumption A.1. The consistency of the initial estimator, $\breve{\theta} \to_p \theta_0$, can be achieved by using the GQMLE to obtain $\breve{\theta}$. The convergence of $\breve{\kappa} \equiv \left(\breve{\kappa}_3, \breve{\kappa}_4\right) \to_p \kappa_0$ is implied by the consistency of $\breve{\theta}$ and the existence of the third and fourth moments of η_t. The last statement can be similarly validated using the arguments for the verification of Assumption A.2, except that the Lipschitz argument employed in (B6) is applied to $\mathbf{R}_t'(\mathbf{x}, \theta_1)\mathbf{D}_t(\mathbf{x}_t, \theta_1, \kappa_1)^{-1} - \mathbf{R}_t'(\mathbf{x}, \theta_2)\mathbf{D}_t(\mathbf{x}_t, \theta_2, \kappa_2)^{-1}$ instead of $\mathbf{r}_t\left(\theta_1\right) - \mathbf{r}_t\left(\theta_2\right)$ (see equation (16) for the function form of $\mathbf{R}_t'(\mathbf{x}, \theta_0)\mathbf{D}_t(\mathbf{x}_t, \theta_0, \kappa_0)^{-1}$ for the GARCH(1,1) case.).

B.2. On Assumption A.2'

Recall that $Q_T(\theta_0, \kappa_0, \theta) = \frac{1}{T}\sum_{t=1}^{T}\mathbf{R}_t'\mathbf{D}_t^{-1}\mathbf{r}_t(\theta)$. Assumption A.2'(i) is satisfied if we have $Q_T(\theta_0, \kappa_0, \theta)$ converges to $\mathrm{E}Q_T(\theta_0, \kappa_0, \theta)$, almost surely. For the GARCH(1,1) model, we have validated the convergence in probability in our previous discussion of Assumption A.2. The conditions for almost sure convergence, similar to the ones required here, can be found in lemma B.1 of Richter *et al.* (2023).

For A.2'(ii), a compact parameter set, though relaxable under certain circumstances, is standard in the literature. For the GARCH(1,1) model, where $Q_T(\theta_0, \kappa_0, \theta)$ is given by (B2), the condition $\mathrm{E}\sup_{\theta\in\Theta}|Q_T(\theta_0, \kappa_0, \theta)|_2 < \infty$ is guaranteed by the compactness of Θ and the existence of $4 + \delta$th moment of the innovation terms. Lastly, Assumption A.2'(iii) has been addressed in the preceding discussion of Assumption A.2.

B.3. On Assumption A.0

We now show that for our proposed WOPIV approach, the selection of initial values has negligible impact in the context of the GARCH(1,1) model.

By equation (18) in Section 2.2, we have for the GARCH(1,1) model

$$\begin{aligned}
\tilde{Q}_T(\dot{\theta}, \kappa, \theta) &= \frac{1}{T}\sum_{t=1}^{T}\mathbf{R}_t\left(\tilde{\mathbf{x}}_t(\dot{\theta}), \dot{\theta}\right)\mathbf{D}_t^{-1}\left(\tilde{\mathbf{x}}_t(\dot{\theta}), \dot{\theta}, \kappa\right)\mathbf{r}_t(\tilde{\mathbf{w}}_t(\theta), \theta) \\
&= -\frac{1}{c_\kappa}\cdot\left[\frac{1}{T}\sum_{t=1}^{T}\begin{pmatrix} \frac{\varepsilon_t^2 - (\omega + \alpha\varepsilon_{t-1}^2 + \beta\tilde{\sigma}_{t-1}^2(\theta))}{\tilde{\sigma}_t^4(\dot{\theta})} \\ \frac{\varepsilon_t^2 - (\omega + \alpha\varepsilon_{t-1}^2 + \beta\tilde{\sigma}_{t-1}^2(\theta))}{\tilde{\sigma}_t^4(\dot{\theta})}\varepsilon_{t-1}^2 \\ \frac{\varepsilon_t^2 - (\omega + \alpha\varepsilon_{t-1}^2 + \beta\tilde{\sigma}_{t-1}^2(\theta))}{\tilde{\sigma}_t^4(\dot{\theta})}\tilde{\sigma}_{t-1}^2(\dot{\theta}) \end{pmatrix} - \frac{1}{T}\sum_{t=1}^{T}\begin{pmatrix} \kappa_3\tilde{\sigma}_t^{-3}(\dot{\theta})\varepsilon_t \\ \kappa_3\tilde{\sigma}_t^{-3}(\dot{\theta})\varepsilon_{t-1}^2\varepsilon_t \\ \kappa_3\tilde{\sigma}_t^{-3}(\dot{\theta})\tilde{\sigma}_{t-1}^2(\dot{\theta})\varepsilon_t \end{pmatrix}\right] \\
&= -\frac{1}{c_\kappa}\cdot\left[\tilde{S}_1(\dot{\theta}, \kappa, \theta) - \tilde{S}_2(\dot{\theta}, \kappa, \theta)\right], \qquad \text{(B7)}
\end{aligned}$$

where $c_\kappa = \kappa_4 - 1 - (\kappa_3)^2$. Note that $\tilde{S}_1(\dot{\theta}, \kappa, \theta)$ essentially exhibits the form the score function of the GQMLE (see also the discussion before and after (28)), and $\tilde{S}_2(\dot{\theta}, \kappa, \theta)$ can be regarded as an adjustment term introduced by the WOPIV, to enhance the GQMLE.

Similarly,

$$Q_T(\dot{\theta},\kappa,\theta) = -\frac{1}{c_\kappa}\cdot\left[\frac{1}{T}\sum_{t=1}^{T}\begin{pmatrix}\frac{\epsilon_t^2-(\omega+\alpha\epsilon_{t-1}^2+\beta\sigma_{t-1}^2(\theta))}{\sigma_t^4(\dot{\theta})}\\ \frac{\epsilon_t^2-(\omega+\alpha\epsilon_{t-1}^2+\beta\sigma_{t-1}^2(\theta))}{\sigma_t^4(\dot{\theta})}\epsilon_{t-1}^2\\ \frac{\epsilon_t^2-(\omega+\alpha\epsilon_{t-1}^2+\beta\sigma_{t-1}^2(\theta))}{\sigma_t^4(\dot{\theta})}\sigma_{t-1}^2(\dot{\theta})\end{pmatrix} - \frac{1}{T}\sum_{t=1}^{T}\begin{pmatrix}\kappa_3\sigma_t^{-3}(\dot{\theta})\epsilon_t\\ \kappa_3\sigma_t^{-3}(\dot{\theta})\epsilon_{t-1}^2\epsilon_t\\ \kappa_3\sigma_t^{-3}(\dot{\theta})\sigma_{t-1}^2(\dot{\theta})\epsilon_t\end{pmatrix}\right]$$
$$= -\frac{1}{c_\kappa}\cdot\left[S_1(\dot{\theta},\kappa,\theta) - S_2(\dot{\theta},\kappa,\theta)\right]. \tag{B8}$$

Combining (B7) and (B8) yields

$$\tilde{Q}_T(\dot{\theta},\kappa,\theta) - Q_T(\dot{\theta},\kappa,\theta) = \frac{1}{c_\kappa}\left\{\left[S_1(\dot{\theta},\kappa,\theta) - \tilde{S}_1(\dot{\theta},\kappa,\theta)\right] + \left[\tilde{S}_2(\dot{\theta},\kappa,\theta) - S_2(\dot{\theta},\kappa,\theta)\right]\right\}.$$

By lemma B.2 of Richter *et al.* (2023), it holds that $\sup_{\dot{\theta},\kappa,\theta}|\nabla_\theta^l\tilde{S}_1(\dot{\theta},\kappa,\theta) - \nabla_\theta^l S_1(\dot{\theta},\kappa,\theta)|_2 = O_p\left(\frac{1}{T}\right)$ for $l = 0, 1$. Using a similar reasoning, we can derive $\sup_{\dot{\theta},\kappa,\theta}|\nabla_\theta^l\tilde{S}_2(\dot{\theta},\kappa,\theta) - \nabla_\theta^l S_2(\dot{\theta},\kappa,\theta)|_2 = O_p\left(\frac{1}{T}\right)$. Lastly, equations (B7) and (B8) indicate that for the GARCH(1,1), both $\tilde{Q}_T(\dot{\theta},\kappa,\theta)$ and $Q_T(\dot{\theta},\kappa,\theta)$ are differentiable with respect to θ. Thus, the Assumption A.0 is satisfied.

DOI: 10.1111/jtsa.12770