Hacettepe Journal of Mathematics & Statistics

Hacet. J. Math. Stat.
Volume 51 (1) (2022), 253 – 272
DOI : 10.15672/hujms.802601

Research Article

Distribution of test statistics under parameter uncertainty for time series data: an application to testing skewness, kurtosis and normality

Anil K. Bera[1], Osman Doğan*[1], Süleyman Taşpınar[2]

[1] *Department of Economics, University of Illinois at Urbana-Champaign (UIUC), U.S.A.*
[2] *Department of Economics, Queens College, The City University of New York, U.S.A.*

Abstract

In this paper, we provide a general result under some high level assumptions that shows how to account for the parameter uncertainty problem in test statistics formulated with the quasi maximum likelihood (QML) estimator. We use our general result to develop various test statistics for testing skewness, kurtosis and normality for time series data. We show that the asymptotic distributions of our test statistics coincide with the asymptotic distributions of some tests suggested in the literature. Thus, our general result provides a unified approach for test statistics formulated with the QML estimator for time series data.

Mathematics Subject Classification (2020). 62F03, 62F05, 62F35, 62M10

Keywords. Normality tests, skewness, kurtosis, asymptotic variance, MLE, QMLE, inference, test statistics, parameter uncertainty

1. Introduction

The parameter uncertainty problem arises when the asymptotic distribution of a given test statistic $\sqrt{n}T_n(y^n, \hat{\theta}_n)$, where $\hat{\theta}_n$ is a consistent estimator of the true parameter vector θ_0, does not coincide with the asymptotic distribution of the unfeasible version $\sqrt{n}T_n(y^n, \theta_0)$ [4, 20]. To obtain the asymptotic distribution of $\sqrt{n}T_n(y^n, \hat{\theta}_n)$, where $\hat{\theta}_n$ is the maximum likelihood estimator (MLE) of θ_0, Pierce [26] provides a simple correction method that shows how to adjust the asymptotic distribution of $\sqrt{n}T_n(y^n, \theta_0)$ when the expectation of $T_n(y^n, \theta_0)$ is free of θ_0. His method also provides a condition under which the parameter uncertainty problem is asymptotically irrelevant for inference about $\sqrt{n}T_n(y^n, \theta_0)$, i.e., the asymptotic distribution of $\sqrt{n}T_n(y^n, \hat{\theta}_n)$ coincides with that of $\sqrt{n}T_n(y^n, \theta_0)$. However, the Pierce correction may not hold in the quasi maximum likelihood (QML) setting considered in [35, 37], and therefore can lead to incorrect inference about $\sqrt{n}T_n(y^n, \theta_0)$.

There are alternative methods in the literature to account for the parameter uncertainty problem [4–6, 8, 19, 22, 23, 29, 31–34, 36, 38, 39]. Randles [29] studies the parameter

*Corresponding Author.
Email addresses: abera@illinois.edu (A.K. Bera), odogan@illinois.edu (O. Doğan), staspinar@qc.cuny.edu (S. Taşpınar)
Received: 04.10.2020; Accepted: 12.10.2021

uncertainty problem through the classical delta method and show when the problem is asymptotically irrelevant for the U-statistics. For the popular specification tests such as the Lagrange Multiplier test for nested hypotheses, the versions of Hausman [13]'s specification tests and White [35]'s information matrix test, the parameter uncertainty problem is accounted for by implementing these tests alternatively through ordinary least-squares regressions [18, 22, 23, 30, 31, 34, 36, 38]. As elaborated in [38], it is important to note that the validity of regression based procedures relies on certain auxiliary assumptions holding in addition to the relevant null hypothesis. Moreover, the finite sample properties of regression based procedures can be poor and highly misleading in some cases [11, 25]. The parameter uncertainty problem can also affect out-of-sample inference regarding the moments of functions of out-of-sample forecasts and forecast errors in parametric forecasting models. In these models, the parameter uncertainty problem is asymptotically irrelevant only when the expected value of gradient of moment functions is zero or the limiting ratio of the size of the prediction sample to that of regression sample is zero [19, 32, 33].

In the generalized method of moments (GMM) framework, the moment conditions can be adjusted so that they become robust against the parameter uncertainty problem [4–6]. Bontemps and Meddahi [5] show that the empirical moment functions formulated as linear combinations of Hermite polynomials are robust against the parameter uncertainty problems. Moreover, the Hermite polynomials associated with the distribution of a random variable have zero mean if and only if the random variable has a standard normal distribution [5, 9]. These results suggest that the empirical moments based on the Hermite polynomials can be used in the GMM framework to test the null hypothesis of normality. In particular, the JB test of [14] coincides with the joint test based on the third and fourth Hermite polynomials [5, 15]. More recently, Bontemps [4] suggests a method based on the oblique projection for transforming any moment function into a robust moment function, i.e., a moment function that is robust against the parameter uncertainty problem. The approach in [4] is only valid for the moment functions that satisfy an information matrix-type equality.* Though these approaches provide moment based tests that are simple to implement, it is not clear how to choose the number of moment functions that can lead to an optimal test.

Recently, Bera et al. [3] revisit the Pierce correction method and show how it can be extended to the QML setting considered in [35, 37] under some primitive conditions imposed on the density function and the test statistics. In this paper, we derive their main result under some high-level assumptions. Our general result indicates that the parameter uncertainty problem is asymptotically irrelevant, i.e., both $\sqrt{n}T_n(y^n, \hat{\theta}_n)$ and $\sqrt{n}T_n(y^n, \theta_0)$ have the same asymptotic distribution, when the expectation of gradient of test statistic is zero. We then use our result to develop various test statistics for testing skewness, kurtosis and normality for time-series data. We compare our tests with those suggested in [2], and analytically show that the asymptotic distributions of our tests coincide with the asymptotic distributions of their tests. Thus, our analysis demonstrates that various test statistics designed for testing skewness, kurtosis and normality fall under one category and our general result can be applied to all of them.

The rest of this paper proceeds as follows. In Section 2, we revisit the QML framework considered in [35, 37] and define the QML estimator (QMLE) under some high level assumptions. In this section, following [3], we revisit the Pierce correction method and show how to adjust it in the QML setting for certain type of test statistics. In Section 3, we revisit the data generating process (DGP) considered in [2] and [14], and use our result

*In terms of our notation in Section 2, this information matrix-type equality is stated as $\mathcal{P}_n'(\theta_0, \psi_0) = -\mathcal{D}_n(\theta_0)$. Also note that the analysis in [4] requires that the moment functions identify the parameter vector and satisfy a CLT condition.

to develop various test statistics for testing normality, skewness and kurtosis for time series data. In Section 4, we consider a Monte Carlo study to investigate the finite sample size and power properties of our suggested tests. In Section 5, we conclude with some directions for future studies. We collect some technical results in an appendix.

2. The asymptotic variance formula in the QML setting

In this section, following [3], we state a general result in the QMLE framework for the asymptotic distribution of certain test statistics. We will then use this result to derive our suggested test statistics for skewness, kurtosis and normality in Section 3. For completeness, we first state the assumptions that are required to define the QMLE. The DGP is characterized by the following assumption.

Assumption 2.1. *Let $(\Omega, \mathfrak{F}, P_0)$ be a complete probability space, where $\Omega = \mathbb{R}^{\nu\infty} \equiv \times_{t=1}^{\infty}\mathbb{R}^{\nu}$, $\nu \in \mathbb{N}$, and $\mathfrak{F}$ is the Borel σ-field generated by the finite dimensional cylinder sets of Ω. The observed data are a realization of the stochastic process defined by $Y = \{Y_t : \Omega \to \mathbb{R}^{\nu},\ t = 1, 2, \ldots\}$.*

We use $Y^n = (Y_1^{'}, Y_2^{'}, \ldots, Y_n^{'})^{'}$ to denote a random sample of size n, and $y^n = (y_1^{'}, y_2^{'}, \ldots, y_n^{'})^{'}$ to denote a realization of Y^n. The probability measure P_0^n governing the behavior of Y^n is defined as the restriction of P_0 to the measurable space $(\mathbb{R}^{\nu n}, \mathfrak{B}(\mathbb{R}^{\nu n}))$ by $\mathrm{P}_0^n(B) = \mathrm{P}_0(Y^n \in B)$, where $B \in \mathfrak{B}(\mathbb{R}^{\nu n})$ and $\mathfrak{B}(\mathbb{R}^{\nu n})$ is the Borel σ-field generated by the open sets of $\mathbb{R}^{\nu n} \equiv \times_{t=1}^{n}\mathbb{R}^{\nu}$. P_0^n assumes a Radon-Nikodým density under the following assumption.

Assumption 2.2. *Let μ^n be a σ-finite measure defined on $(\mathbb{R}^{\nu n}, \mathfrak{B}(\mathbb{R}^{\nu n}))$ for $n \in \mathbb{N}$. Then, P_0^n is absolutely continuous with respect to μ^n.*

Under Assumption 2.2, the Radon-Nikodým theorem ensures the existence of a measurable non-negative Radon-Nikodým density $g^n = \mathrm{d}\mathrm{P}_0^n/\mathrm{d}\mu^n$ such that $\mathrm{P}_0^n(B) = \int_B g^n \mathrm{d}\mu^n$ for all $B \in \mathfrak{B}(\mathbb{R}^{\nu n})$. Thus, given μ^n, P_0^n will be known if we know g^n. To this end, we assume an approximation to g^n based on the parametric stochastic specification defined by $\mathcal{S} = \{f_t : \mathbb{R}^{\nu t} \times \Theta \to \mathbb{R}^{+},\ \Theta \subseteq \mathbb{R}^p,\ p \in \mathbb{N},\ t = 1, 2, \ldots\}$, where $f_t(\cdot, \theta)$ is measurable-$\mathfrak{B}(\mathbb{R}^{\nu t})$ for all $\theta \in \Theta$. Here, $\mathcal{S}$ is called a "specification for Y", and is assumed to satisfy the following assumption.

Assumption 2.3. *For each t, the function $f_t : \mathbb{R}^{\nu t} \times \Theta \to \mathbb{R}^{+}$ satisfies the following conditions: (i) $f_t(\cdot, \theta)$ is measurable-$\mathfrak{B}(\mathbb{R}^{\nu t})$ for all $\theta \in \Theta$, where Θ is a compact subset of $\mathbb{R}^p$, and (ii) $f_t(Y^t, \cdot)$ is continuous on Θ a.s. -P_0, i.e., there exists a set $B_t \in \mathfrak{B}(\mathbb{R}^{\nu t})$ such that $f_t(y^t, \cdot)$ is continuous on Θ for all $y^t \in B_t$ and $P_0^t(B_t) = 1$.*

Under Assumption 2.3, $f^n(y^n, \theta) = \prod_{t=1}^{n} f_t(y^t, \theta)$ is called the quasi likelihood function generated by $\mathcal{S}$ and can be viewed as an approximation to $g^n(y^n)$. The divergence or discrepancy of f^n from g^n can be measured by the Kullback-Leibler Information Criterion (KLIC) given by

$$
\begin{aligned}
\mathbb{I}(g^n : f^n; \theta) &= \int_{S^n} \left(\ln \frac{g^n(y^n)}{f^n(y^n, \theta)} \right) g^n(y^n) \mathrm{d}\mu^n(y^n) \\
&= \int_{S^n} \left(\ln g^n(y^n) \right) g^n(y^n) \mathrm{d}\mu^n(y^n) - \int_{S^n} \left(\ln f^n(y^n, \theta) \right) g^n(y^n) \mathrm{d}\mu^n(y^n) \\
&= \mathbb{E}\left(\ln g^n(Y^n) \right) - \mathbb{E}\left(\ln f^n(Y^n, \theta) \right), \qquad (2.1)
\end{aligned}
$$

where $S^n = \{y^n : g^n(y^n) > 0\}$. The result in Equation (2.1) indicates that the KLIC minimizer θ is the value that maximizes $\mathbb{E}\left(\ln f^n(Y^n, \theta)\right)$. Thus, we can defined the QMLE $\hat{\theta}_n$

as the parameter vector that maximizes the estimated version of $\mathbb{E}\left(\ln f^n(Y^n,\theta)\right)$, namely

$$\hat{\theta}_n = \text{argmax}_{\theta\in\Theta} \ln f^n(Y^n,\theta) = \text{argmax}_{\theta\in\Theta} \sum_{t=1}^{n} \ln f_t(Y^t,\theta). \tag{2.2}$$

Assumptions 2.1 and 2.2 ensure that $\hat{\theta}_n$ exists almost surely. To establish the large sample properties, namely, the consistency and asymptotic normality of $\hat{\theta}_n$, we adopt the following assumptions.

Assumption 2.4. *(i) $\mathbb{E}\left(\ln f_t(Y^t,\theta)\right)$ exists and is finite for all t. (ii) $\mathbb{E}\left(\ln f_t(Y^t,\theta)\right)$ is continuous on Θ for all t. (iii) The sequence $\{\ln f_t(Y^t,\theta)\}$ obeys the strong uniform law of large numbers (ULLN).*

Assumption 2.5. *(i) The sequence $\{\mathbb{E}\left(n^{-1}\ln f^n(Y^n,\theta)\right)\}$ is $O(1)$ uniformly on Θ. (ii) $\{\mathbb{E}\left(n^{-1}\ln f^n(Y^n,\theta)\right)\}$ has the identifiably unique maximizers $\theta^\star \equiv \{\theta_n^\star\}$. (iii) $\theta^\star \equiv \{\theta_n^\star\}$ lie in the interior of Θ uniformly in n.*

Assumption 2.6. *(i) $f_t(Y^t,\theta)$ is continuously differentiable of order 2 on Θ a.s.-P_0 for all t, i.e., there exists a set $F_t \in \mathfrak{B}(\mathbb{R}^{\nu t})$ such that $f_t(y^t,\cdot)$ is continuously differentiable of order 2 on Θ for all $y^t \in F_t$ and $P_0^t(F_t) = 1$ for each t. (ii) $\mathbb{E}\left(n^{-1}\nabla \ln f^n(Y^n,\theta)\right) < \infty$ for all n and $\theta \in \Theta$, where ∇ is the gradient operator with respect to θ.*

Assumption 2.7. *(i) $\mathbb{E}\left(n^{-1}\nabla^2 \ln f^n(Y^n,\theta)\right) < \infty$ for all n and $\theta \in \Theta$. (ii) $\mathbb{E}\left(n^{-1}\nabla^2 \ln f^n(Y^n,\cdot)\right)$ is continuous on Θ uniformly in n. (iii) The sequence $\{\nabla^2 \ln f_t(Y^t,\theta)\}$ obeys the strong ULLN.*

Assumption 2.8. *The sequence $\{n^{-1/2}\nabla \ln f_t(Y^t,\theta^\star)\}$ obeys the central limit theorem (CLT) with the covariance matrix $\left\{\mathcal{B}_n(\theta^\star) \equiv \text{Var}\left(n^{-1/2}\sum_{t=1}^n \nabla \ln f_t(Y^t,\theta^\star)\right)\right\}$, where $\{\mathcal{B}_n(\theta^\star)\}$ is $O(1)$ and positive definite uniformly in n.*

Assumption 2.9. *$\{\mathcal{A}_n(\theta^\star) \equiv -\mathbb{E}\left(n^{-1}\nabla^2 \ln f^n(Y^n,\theta^\star)\right)\}$ is $O(1)$ and positive definite uniformly in n.*

The strong consistency result, namely, $\hat{\theta}_n - \theta^\star \to 0$ a.s.-P_0, follows from Assumptions 2.1, 2.3, 2.4 and 2.5 (ii). It follows from Assumption 2.8 that $\mathcal{B}_n^{-1/2}(\theta^\star) n^{-1/2}\sum_{t=1}^n \nabla \ln f_t(Y^t,\theta^\star) \overset{A}{\sim} N[0,I_p]$, where $\overset{A}{\sim}$ denotes the asymptotic distribution and I_p is the $p\times p$ identity matrix. Then, the asymptotic normality property of QMLE follows from Assumptions 2.1, 2.3 and 2.4-2.9.* If there exists θ_0 in Θ such that $f^n(y^n,\theta_0) = g^n(y^n)$ for all $y^n \in \mathbb{R}^{\nu n}$, then the parametric stochastic specification $\mathcal{S}$ is said to be correct in its entirety for Y on Θ with respect to μ^n [37]. When $\mathcal{S}$ is correct in its entirety, $\hat{\theta}_n$ defined in Equation (2.2) is called the MLE, and the information matrix equality $\mathcal{A}_n(\theta_0) = \mathcal{B}_n(\theta_0)$ holds.

Next, we describe the test statistic considered in this paper. We assume that the test statistic has the following form

$$T_n(y^n, \hat{\theta}_n) = \frac{1}{n}\sum_{t=1}^{n} \psi_t(y^t, \hat{\theta}_n), \tag{2.3}$$

where the vector-valued test indicator function $\psi_t : \mathbb{R}^{\nu t} \times \Theta \to \mathbb{R}^q$ satisfies the conditions in the following assumptions.

Assumption 2.10. *(i) $\psi_t(\cdot,\theta)$ is measurable-$\mathfrak{B}(\mathbb{R}^{\nu t})$ for all t and $\theta \in \Theta$, where Θ is a compact subset of $\mathbb{R}^p$. (ii) $\psi_t(Y^t,\cdot)$ is continuous on Θ a.s.-P_0, i.e., there exists a set $A_t \in \mathfrak{B}(\mathbb{R}^{\nu t})$ such that $\psi_t(y^t,\cdot)$ is continuous on Θ for all $y^t \in A_t$ and $P_0^t(A_t) = 1$. (iii) $\mathbb{E}\left(\psi_t(Y^t,\theta^\star)\right) = \psi^\star$ is independent of $\theta^\star$ for all t, where $\psi^\star \in \mathbb{R}^q$.*

*The literature provides various primitive conditions imposed on $\{f_t\}$ and $\{Y_t\}$ for ensuring these theoretical properties. For a summary of these results, the reader is referred to [7,27,28,37].

Assumption 2.11. *The function $\psi_t(Y^t,\theta)$ is continuously differentiable on Θ a.s.-P_0 for all t, i.e., there exists a set $K_t \in \mathfrak{B}(\mathbb{R}^{\nu t})$ such that $\psi_t(y^t,\cdot)$ is continuously differentiable on Θ for all $y^t \in K_t$ and $P_0^t(K_t)=1$ for each t.*

Assumption 2.12. *(i) $\mathbb{E}\left(\psi_t(Y^t,\theta)\right)$ exists and is finite for all t. (ii) $\mathbb{E}\left(\psi_t(Y^t,\theta)\right)$ is continuous on Θ for all t. (iii) The sequence $\{\psi_t(Y^t,\theta)\}$ obeys the strong ULLN.*

Assumption 2.13. *(i) $\mathbb{E}\left(\nabla\psi_t(Y^t,\theta)\right)<\infty$ for all t and $\theta\in\Theta$. (ii) $\mathbb{E}\left(\nabla\psi_t(Y^t,\cdot)\right)$ is continuous on Θ uniformly in n. (iii) The sequence $\{\nabla\psi_t(Y^t,\theta)\}$ obeys the strong ULLN.*

Assumption 2.14. *The sequence $\{n^{-1/2}\nabla\psi_t(Y^t,\theta^\star)-\psi^\star\}$ obeys the central limit theorem (CLT) with the covariance matrix $\left\{\mathcal{C}_n(\theta^\star,\psi^\star)\equiv \operatorname{Var}\left(n^{-1/2}\sum_{t=1}^n(\psi_t(Y^t,\theta^\star)-\psi^\star)\right)\right\}$, where $\{\mathcal{C}_n(\theta^\star,\psi^\star)\}$ is $O(1)$ and positive definite uniformly in n.*

Assumptions 2.10-2.14, except 2.10(iii), are counterparts to those assumed for $\{f_t(Y^t,\theta)\}$ and ensure the asymptotic normality for the test statistic. Under these assumptions, our test indicator function can be augmented with the score functions to form a vector of estimating equations. Thus, we can determine the asymptotic distribution of our test statistic as a by-product of the likelihood estimation. Pierce [26] suggests Assumption 2.10(iii) to reach a simple variance formula for the test statistic in the ML setting. Our ensuing analysis will show that this assumption is not enough to obtain the Pierce formula in the QML setting, because the information matrix equality does not hold in the QML setting. When $\mathcal{S}$ is correct in its entirety, we express Assumption 2.10(iii) as $\mathbb{E}\left(\psi_t(Y^t,\theta_0)\right)=\psi_0$, where $\psi_0\in\mathbb{R}^q$. Assumption 2.12 and Lemma A.1 in Appendix A ensure that $\hat\psi_n-\psi^\star\to 0$ a.s.-P_0, where $\hat\psi_n=T_n(Y^n,\hat\theta_n)$.

Let

$$\mathcal{P}_n(\theta^\star,\psi^\star)=\mathbb{E}\left(\left(\frac{1}{\sqrt{n}}\sum_{t=1}^n\frac{\partial\ln f_t(Y^t,\theta^\star)}{\partial\theta}\right)\times\left(\frac{1}{\sqrt{n}}\sum_{t=1}^n\left(\psi_t(Y^t,\theta^\star)-\psi^\star\right)\right)'\right)$$

and

$$\mathcal{D}_n(\theta^\star)=\mathbb{E}\left(n^{-1}\sum_{t=1}^n\frac{\partial\psi_t(Y^t,\theta^\star)}{\partial\theta'}\right).$$

In the following proposition, we provide a general result on the joint asymptotic distribution of $\sqrt{n}(T_n(Y^n,\hat\theta_n)-\psi^\star)$ and $\sqrt{n}(\hat\theta_n-\theta^\star)$ in the QML setting.

Proposition 2.15. *Under Assumptions 2.1, 2.3 and 2.4-2.14, the asymptotic joint distribution of $\sqrt{n}(T_n(Y^n,\hat\theta_n)-\psi^\star)$ and $\sqrt{n}(\hat\theta_n-\theta^\star)$ is given by*

$$\begin{pmatrix}\sqrt{n}(\hat\theta_n-\theta^\star)\\ \sqrt{n}(T_n(Y^n,\hat\theta_n)-\psi^\star)\end{pmatrix}\overset{A}{\sim}N\left[0,\begin{pmatrix}\mathcal{A}_n^{-1}(\theta^\star)\mathcal{B}_n(\theta^\star)\mathcal{A}_n^{-1}(\theta^\star) & \mathcal{V}_n'(\theta^\star,\psi^\star)\\ \mathcal{V}_n(\theta^\star,\psi^\star) & \mathcal{S}_n(\theta^\star,\psi^\star)\end{pmatrix}\right], \tag{2.4}$$

where

$$\mathcal{V}_n(\theta^\star,\psi^\star)=\mathcal{D}_n(\theta^\star)\mathcal{A}_n^{-1}(\theta^\star)\mathcal{B}_n(\theta^\star)\mathcal{A}_n^{-1}(\theta^\star)+\mathcal{P}_n'(\theta^\star,\psi^\star)\mathcal{A}_n^{-1}(\theta^\star), \tag{2.5}$$

$$\begin{aligned}\mathcal{S}_n(\theta^\star,\psi^\star)=\;&\mathcal{C}_n(\theta^\star,\psi^\star)+\mathcal{D}_n(\theta^\star)\mathcal{A}_n^{-1}(\theta^\star)\mathcal{B}(\theta^\star)\mathcal{A}_n^{-1}(\theta^\star)\mathcal{D}_n'(\theta^\star)+\mathcal{P}_n'(\theta^\star,\psi^\star)\mathcal{A}_n^{-1}(\theta^\star)\mathcal{D}_n'(\theta^\star)\\&+\mathcal{D}_n(\theta^\star)\mathcal{A}_n^{-1}(\theta^\star)\mathcal{P}_n(\theta^\star,\psi^\star).\end{aligned} \tag{2.6}$$

Proof. See Appendix A. □

Proposition 2.15 extends [3] to our setting, and thus provides a generalization of the asymptotic variance formula suggested by [26] to the QML setting. When $\mathcal{S}$ is correct in its entirety, our asymptotic variance formula in Equation(2.6) reduces to the Pierce formula under Assumption 2.10(iii). To see this, consider $\frac{\partial \mathbb{E}(T(Y^n,\theta))}{\partial \theta'}\big|_{\theta_0} = 0$, which can be expressed as

$$\begin{aligned}\frac{\partial \mathbb{E}\left(T(Y^n,\theta)\right)}{\partial \theta'}\bigg|_{\theta_0} &= n^{-1}\sum_{t=1}^{n}\frac{\partial \mathbb{E}(\psi_t(Y^t,\theta))}{\partial \theta'}\bigg|_{\theta_0}\\ &= \int n^{-1}\sum_{t=1}^{n}\frac{\partial \psi_t(y^t,\theta)}{\partial \theta'}\bigg|_{\theta_0} \times f^n(y^n,\theta_0)\mathrm{d}\mu^n(y^n)\\ &+ \int\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\psi_t(y^t,\theta)\right)\left(\frac{1}{\sqrt{n}}\frac{\partial \ln f^n(y^n,\theta)}{\partial \theta}\bigg|_{\theta_0}\right)' f^n(y^n,\theta_0)\mathrm{d}\mu^n(y^n) = 0.\end{aligned} \tag{2.7}$$

Since $\mathbb{E}\left(\frac{\partial \log f^n(Y^n,\theta)}{\partial \theta}\big|_{\theta_0}\right) = 0$ holds under Assumption 2.5, Equation (2.7) can be expressed as

$$\begin{aligned}&\int n^{-1}\sum_{t=1}^{n}\frac{\partial \psi_t(y_n^t,\theta)}{\partial \theta'}\bigg|_{\theta_0} \times f^n(y^n,\theta_0)\mathrm{d}\mu^n(y^n)\\ &+ \int\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\psi_t(y^t,\theta)-\psi_0)\right)\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial \ln f_t(y^t,\theta)}{\partial \theta}\bigg|_{\theta_0}\right)' f^n(y^n,\theta_0)\mathrm{d}\mu^n(y^n) = 0.\end{aligned} \tag{2.8}$$

The result in Equation (2.8) gives the following information matrix type equality [21, p.217, Equation (14)]

$$\mathcal{P}_n'(\theta_0,\psi_0) = -\mathcal{D}_n(\theta_0). \tag{2.9}$$

Also, if $\mathcal{S}$ is correct in its entirety, then it follows from the information matrix equality that

$$\begin{aligned}\mathcal{S}_n(\theta_0,\psi_0) = \mathcal{C}_n(\theta_0,\psi_0) &+ \mathcal{D}_n(\theta_0)\mathcal{A}_n^{-1}(\theta_0)\mathcal{D}_n'(\theta_0) + \mathcal{P}_n'(\theta_0,\psi_0)\mathcal{A}_n^{-1}(\theta_0)\mathcal{D}_n'(\theta_0)\\ &+ \mathcal{D}_n(\theta_0)\mathcal{A}_n^{-1}(\theta_0)\mathcal{P}_n(\theta_0,\psi_0).\end{aligned} \tag{2.10}$$

Then, using Equations (2.9) in (2.10), we obtain the simple asymptotic variance formula suggested by [26] in the ML setting:

$$\mathcal{S}_n(\theta_0,\psi_0) = \mathcal{C}_n(\theta_0,\psi_0) - \mathcal{D}_n(\theta_0)\mathcal{A}_n^{-1}(\theta_0)\mathcal{D}_n'(\theta_0). \tag{2.11}$$

Remark 2.16. In the QML setting, $\mathcal{S}_n(\theta^\star,\psi^\star)$ in Equation (2.6) indicates that the parameter uncertainty problem is asymptotically irrelevant for inference about $\sqrt{n}T_n(y^n,\theta^\star)$ when $\mathcal{D}_n(\theta^\star) = 0$ holds. Similarly, when $\mathcal{D}_n(\theta_0) = 0$ holds in Equation (2.11), the asymptotic covariance of both $\sqrt{n}T_n(y^n,\hat{\theta}_n)$ and $\sqrt{n}T_n(y^n,\theta_0)$ is given by $\mathcal{C}_n(\theta_0,\psi_0)$ in the ML setting.

Remark 2.17. To estimate the elements of the covariance matrix in Proposition 2.15, we need consistent estimators of $\mathcal{A}_n(\theta^\star)$, $\mathcal{B}_n(\theta^\star)$, $\mathcal{C}_n(\theta^\star,\psi^\star)$, $\mathcal{D}_n(\theta^\star)$ and $\mathcal{P}_n(\theta^\star,\psi^\star)$. We can use the plug-in method for $\mathcal{A}_n(\theta^\star)$ and $\mathcal{D}_n(\theta^\star)$, and a kernel type estimator [1, 10, 24] for $\mathcal{B}_n(\theta^\star)$, $\mathcal{C}_n(\theta^\star,\psi^\star)$ and $\mathcal{P}_n(\theta^\star,\psi^\star)$.

3. Testing skewness, kurtosis and normality

In this section, we show how our result can be used to determine the asymptotic distribution of the omnibus test for normality, the skewness test statistic in the presence of excess kurtosis, and the kurtosis test statistic in the presence of asymmetry. Following [2], we consider the following DGP

$$y_t = \mu_0 + \varepsilon_t, \quad t = 1,\ldots,n, \tag{3.1}$$

where μ_0 is the unknown mean of y_t and ε_t is an ergodic strong stationary process with mean zero and variance σ_0^2. Let $\theta_0 = (\mu_0, \sigma_0^2)'$ be the true parameter vector and $\theta = (\mu, \sigma^2)'$ be an arbitrary value in the parameter space. The misspecified model assumes that ε_t's are i.i.d normal random variables with mean zero and variance σ_0^2.* Then, the quasi log-likelihood function of an observation can be expressed as

$$\ln f(y_t, \theta) = -\frac{1}{2}\ln 2\pi - \frac{1}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\varepsilon_t^2(\theta) \tag{3.2}$$

where $\varepsilon_t(\theta) = y_t - \mu$. The first and the second order conditions are

$$\frac{\partial \ln f(y_t,\theta)}{\partial\mu} = \frac{1}{\sigma^2}\varepsilon_t(\theta), \quad \frac{\partial \ln f(y_t,\theta)}{\partial\sigma^2} = -\frac{1}{2\sigma^2} + \frac{1}{2\sigma^4}\varepsilon_t^2(\theta), \quad \frac{\partial^2 \ln f(y_t,\theta)}{\partial\mu^2} = -\frac{1}{\sigma^2},$$
$$\frac{\partial^2 \ln f(y_t,\theta)}{\partial\mu\partial\sigma^2} = -\frac{1}{\sigma^4}\varepsilon_t(\theta), \quad \frac{\partial^2 \ln f(y_t,\theta)}{\partial\sigma^2\partial\sigma^2} = \frac{1}{2\sigma^4} - \frac{1}{\sigma^6}\varepsilon_t^2(\theta). \tag{3.3}$$

The QMLE $\hat{\theta}_n$ is defined by $\hat{\theta}_n = \operatorname{argmax}_{\theta\in\Theta}\sum_{t=1}^n \ln f(y_t,\theta)$. The first order conditions yield $\hat{\mu}_n = n^{-1}\sum_{t=1}^n y_t$ and $\hat{\sigma}_n^2 = n^{-1}\sum_{t=1}^n \hat{\varepsilon}_t^2$, where $\hat{\varepsilon}_t = y_t - \hat{\mu}_n$. Using the second order conditions, it follows that

$$\mathcal{A}_n(\theta_0) = \begin{pmatrix} 1/\sigma_0^2 & 0 \\ 0 & 1/2\sigma_0^4 \end{pmatrix}. \tag{3.4}$$

On the other hand, due to the presence of serial correlation, $\mathcal{B}_n(\theta_0)$ takes the following form

$$\mathcal{B}_n(\theta_0) = \mathbb{E}\left(g_1(y_t,\theta_0)g_1'(y_t,\theta_0)\right) + \textstyle\sum_{s=1}^\infty \left(\mathbb{E}\left(g_1(y_t,\theta_0)g_1'(y_{t-s},\theta_0)\right) + \mathbb{E}\left(g_1(y_{t-s},\theta_0)g_1'(y_t,\theta_0)\right)\right), \tag{3.5}$$

where $g_1(y_t,\theta_0) = \left(\varepsilon_t/\sigma_0^2,\ \varepsilon_t^2/2\sigma_0^4 - 1/2\sigma_0^2\right)'$. This long-run covariance matrix can be estimated by the kernel type estimators [1,10,24]. Depending on the specification adopted for the test statistic, our subsequent analysis will also require the long-run covariances $\mathcal{C}_n(\theta_0)$ and $\mathcal{P}_n(\theta_0)$. Define $g_2(y_t,\theta_0) = \varepsilon_t^3/\sigma_0^3 - \mu_3\sigma_0^{-3}$, $g_3(y_t,\theta_0) = (\varepsilon_t^{r_1} - \mu_{r_1},\ \varepsilon_t^{r_2} - \mu_{r_2})'$, $g_4(y_t,\theta_0) = \varepsilon_t^4/\sigma_0^4 - \mu_4/\sigma_0^4$, $g_5(y_t,\theta_0) = (\varepsilon_t^3,\ \varepsilon_t^4 - 3\sigma_0^4)'$, where $\mu_r = \mathbb{E}(y_t-\mu_0)^r$, and r_1, r_2 are two positive odd numbers. Consider the following long-run covariance matrix

$$\mathcal{H}_n(\theta_0) = \mathbb{E}\left(h(y_t,\theta_0)h'(y_t,\theta_0)\right) + \textstyle\sum_{s=1}^\infty \left(\mathbb{E}\left(h(y_t,\theta_0)h'(y_{t-s},\theta_0)\right) + \mathbb{E}\left(h(y_{t-s},\theta_0)h'(y_t,\theta_0)\right)\right), \tag{3.6}$$

where $h(y_t,\theta_0) = \left(g_1'(y_t,\theta_0),\ g_2(y_t,\theta_0),\ g_3'(y_t,\theta_0),\ g_4(y_t,\theta_0),\ g_5'(y_t,\theta_0)\right)'$. Consider the partition of $\mathcal{H}_n(\theta_0)$ into sub-matrices $\mathcal{H}_{ij,n}(\theta_0)$, for $i,j = 1,2,\ldots 5$, corresponding to the long-run covariance between $g_i(y_t,\theta_0)$ and $g_j(y_t,\theta_0)$. We will use these sub-matrices to derive expressions for $\mathcal{C}_n(\theta_0)$ and $\mathcal{P}_n(\theta_0)$ in the subsequent sections.

Remark 3.1. Note that when disturbance terms are independent, $\mathcal{B}_n(\theta_0)$ in Equation (3.5) simplifies to

$$\mathcal{B}_n(\theta_0) = \begin{pmatrix} \frac{1}{\sigma_0^2} & \frac{\mu_3}{2\sigma_0^6} \\ \frac{\mu_3}{2\sigma_0^6} & \frac{\mu_4-\sigma_0^4}{4\sigma_0^8} \end{pmatrix}. \tag{3.7}$$

Then, using Equations (B.2) and (3.7), we obtain

$$\mathcal{A}_n^{-1}(\theta_0)\mathcal{B}_n(\theta_0)\mathcal{A}_n^{-1}(\theta_0) = \begin{pmatrix} \sigma_0^2 & \mu_3 \\ \mu_3 & \mu_4 - \sigma_0^4 \end{pmatrix}. \tag{3.8}$$

*For notational simplicity, we denote the parameter vector with $\theta_0 = (\mu_0, \sigma_0^2)'$ even the model is misspecified.

3.1. Testing skewness

To test for skewness, we consider the following test statistic

$$T_{3,n}(y^n,\,\hat{\theta}_n) = \frac{1}{n}\sum_{t=1}^{n}\psi(y_t,\hat{\theta}_n), \quad \psi(y_t,\hat{\theta}_n) = \hat{\varepsilon}_t^3/\hat{\sigma}_n^3 - \mu_3\sigma_0^{-3}. \tag{3.9}$$

Simple calculations show that

$$\mathcal{D}_{3,n}(\theta_0) = \mathbb{E}\left(n^{-1}\sum_{t=1}^{n}\frac{\partial\psi(y_t,\theta)}{\partial\theta'}|_{\theta_0}\right) = \begin{pmatrix}-\frac{3}{\sigma_0} & -\frac{3\mu_3}{2\sigma_0^5}\end{pmatrix}. \tag{3.10}$$

Then, Proposition 2.15 yields the following corollary.

Corollary 3.2. *The asymptotic distribution of* $\sqrt{n}T_{3,n}(y^n,\,\hat{\theta}_n)$ *is*

$$\sqrt{n}T_{3,n}(y^n,\,\hat{\theta}_n) \overset{A}{\sim} N\left[0,\mathcal{S}_{3,n}(\theta_0)\right], \tag{3.11}$$

where

$$\begin{aligned}\mathcal{S}_{3,n}(\theta_0) &= \mathcal{H}_{22,n}(\theta_0) + \mathcal{D}_{3,n}(\theta_0)\mathcal{A}_n^{-1}(\theta_0)\mathcal{H}_{11,n}(\theta_0)\mathcal{A}_n^{-1}(\theta_0)\mathcal{D}_{3,n}^{'}(\theta_0)\\ &\quad + \mathcal{H}_{12,n}^{'}(\theta_0)\mathcal{A}_n^{-1}(\theta_0)\mathcal{D}_{3,n}^{'}(\theta_0) + \mathcal{D}_{3,n}(\theta_0)\mathcal{A}_n^{-1}(\theta_0)\mathcal{H}_{12,n}(\theta_0).\end{aligned}$$

Proof. See Appendix B. □

In Corollary 3.2, note that $\mathcal{H}_{11,n}(\theta_0) = \mathcal{B}_n(\theta_0)$ and the explicit expressions for $\mathcal{H}_{22,n}(\theta_0)$ and $\mathcal{H}_{12,n}(\theta_0)$ are given by

$$\begin{aligned}\mathcal{H}_{22,n}(\theta_0) &= \mathbb{E}\left(g_2^2(y_t,\theta_0)\right) + 2\sum_{s=1}^{\infty}\mathbb{E}\left(g_2(y_t,\theta_0)g_2(y_{t-s},\theta_0)\right), \qquad (3.12)\\ \mathcal{H}_{12,n}(\theta_0) &= \mathbb{E}\left(g_1(y_t,\theta_0)g_2(y_t,\theta_0)\right)\\ &\quad + \sum_{s=1}^{\infty}\left(\mathbb{E}\left(g_1(y_t,\theta_0)g_2(y_{t-s},\theta_0)\right) + \mathbb{E}\left(g_1(y_{t-s},\theta_0)g_2(y_t,\theta_0)\right)\right),\end{aligned}$$

where $g_2(y_t,\theta_0) = \psi(y_t,\theta_0) = \varepsilon_t^3/\sigma_0^3 - \mu_3\sigma_0^{-3}$.

Since the odd moments of a symmetric distribution are zero, a test based on the several odd moments can have more power. Following [2], we consider an alternative test statistic based on two odd moments. This test statistic takes the following form*

$$T_{35,n}(y^n,\,\hat{\theta}_n) = \frac{1}{n}\sum_{t=1}^{n}\psi(y_t,\hat{\theta}_n), \quad \psi(y_t,\hat{\theta}_n) = \begin{pmatrix}\hat{\varepsilon}_t^{r_1} - \mu_{r_1}\\ \hat{\varepsilon}_t^{r_2} - \mu_{r_2}\end{pmatrix}. \tag{3.13}$$

Using Proposition 2.15, we can determine the asymptotic covariance of $\sqrt{n}T_{35,n}(y^n,\hat{\theta}_n)$ as

$$\begin{aligned}\mathcal{S}_{35,n}(\theta_0) &= \mathcal{H}_{33,n}(\theta_0) + \mathcal{D}_{35,n}(\theta_0)\mathcal{A}_n^{-1}(\theta_0)\mathcal{H}_{11,n}(\theta_0)\mathcal{A}_n^{-1}(\theta_0)\mathcal{D}_{35,n}^{'}(\theta_0)\\ &\quad + \mathcal{H}_{13,n}^{'}(\theta_0)\mathcal{A}_n^{-1}(\theta_0)\mathcal{D}_{35,n}^{'}(\theta_0) + \mathcal{D}_{35,n}(\theta_0)\mathcal{A}_n^{-1}(\theta_0)\mathcal{H}_{13,n}(\theta_0), \qquad (3.14)\end{aligned}$$

where

$$\mathcal{D}_{35,n}(\theta_0) = \mathbb{E}\left(n^{-1}\sum_{t=1}^{n}\frac{\partial\psi(y_t,\theta)}{\partial\theta'}|_{\theta_0}\right) = \begin{pmatrix}-r_1\mu_{r_1-1} & 0\\ -r_2\mu_{r_2-1} & 0\end{pmatrix}. \tag{3.15}$$

Let $\mathcal{S}_{35,n}(\hat{\theta}_n)$ be a consistent estimator of $\mathcal{S}_{35,n}(\theta_0)$. Then, the following result follows from Proposition 2.15.

*Note that we can form a joint test of several odd moment conditions in a similar fashion. For example, a joint test based on three odd moment conditions can be based on $\psi(y_t,\hat{\theta}_n) = (\hat{\varepsilon}_t^{r_1} - \mu_{r_1}, \hat{\varepsilon}_t^{r_2} - \mu_{r_2}, \hat{\varepsilon}_t^{r_3} - \mu_{r_3})^{'}$, where r_1, r_2 and r_3 are three positive odd numbers.

Corollary 3.3. *Under the null hypothesis of no skewness, it follows that*

$$n\,T'_{35,n}(y^n,\hat{\theta}_n)\left(\mathcal{S}_{35,n}(\hat{\theta}_n)\right)^{-1}T_{35,n}(y^n,\hat{\theta}_n)\overset{A}{\sim}\chi^2_2. \tag{3.16}$$

Proof. See Appendix B. □

In Corollaries 3.2 and 3.3, we account for the serial dependence in data through the log-run covariances $\mathcal{H}_{11,n}(\theta_0)$, $\mathcal{H}_{22,n}(\theta_0)$, $\mathcal{H}_{12,n}(\theta_0)$, $\mathcal{H}_{33,n}(\theta_0)$ and $\mathcal{H}_{13,n}(\theta_0)$. To account for the serial dependence, Bai and Ng [2] use three dimensional long-run covariances for $\sqrt{n}T_{3,n}(y^n,\,\hat{\theta}_n)$ and $\sqrt{n}T_{35,n}(y^n,\,\hat{\theta}_n)$. Let $Z_{3,t}=(\varepsilon^3_t-\mu_3,\,\varepsilon_t,\,\varepsilon^2_t-\sigma^2_0)'$ and $\alpha_3=\left(1,\,-3\sigma^2_0,\,-\frac{3\sigma_0}{2}\frac{\mu_3}{\sigma^3_0}\right)$. Under the assumption that $\frac{1}{\sqrt{n}}\sum_{t=1}^{n}Z_{3,t}\overset{A}{\sim}N[0,\Gamma_{3,n}]$, where $\Gamma_{3,n}$ is the long-run covariance matrix of $\frac{1}{\sqrt{n}}\sum_{t=1}^{n}Z_{3,t}$, Bai and Ng [2, Theorem 1] show that $\sqrt{n}T_{3,n}(y^n,\,\hat{\theta}_n)\overset{A}{\sim}N[0,\,\alpha_3\Gamma_{3,n}\alpha'_3/\sigma^6_0]$. To determine the asymptotic distribution of $T_{35,n}(y^n,\,\hat{\theta}_n)$, let

$$\alpha_{35}=\begin{pmatrix}1 & 0 & -r_1\mu_{r_1-1}\\ 0 & 1 & -r_2\mu_{r_2-1}\end{pmatrix},\quad\text{and}\quad Z_{35,t}=\begin{pmatrix}\varepsilon^{r_1}-\mu_{r_1}\\ \varepsilon^{r_2}-\mu_{r_2}\\ \varepsilon_t\end{pmatrix}. \tag{3.17}$$

Then, Bai and Ng [2] show that $\sqrt{n}T_{35,n}(y^n,\,\hat{\theta}_n)\overset{A}{\sim}N[0,\,\alpha_{35}\Gamma_{35,n}\alpha'_{35}]$, where $\Gamma_{35,n}$ is the long-run covariance matrix of $\frac{1}{\sqrt{n}}\sum_{t=1}^{n}Z_{35,t}$. In the following corollary, we show that our results are the same as with those derived in [2].

Corollary 3.4. *It follows that* $\alpha_3\Gamma_{3,n}\alpha'_3/\sigma^6_0=\mathcal{S}_{3,n}(\theta_0)$ *and* $\alpha_{35}\Gamma_{35,n}\alpha'_{35}=\mathcal{S}_{35,n}(\theta_0)$.

Proof. See Appendix B. □

Remark 3.5. There are two reasons why our results coincide with those in [2]. First, the QMLE $\hat{\theta}_n$ in the context of Equation (3.1) coincides with the sample mean and variance. Secondly, in both approaches, the asymptotic distributions of test statistics are based on the mean value expansions. To illustrate the second point, we consider the skewness statistic. The mean value expansions of $\hat{\mu}_3=\frac{1}{n}\sum_{t=1}^{n}\hat{\varepsilon}^3_t$ and $\sqrt{n}\,(\hat{\sigma}^2_n)^{3/2}$ around μ_0 and σ^2_0, respectively, give the following results:

$$\sqrt{n}\hat{\mu}_3=\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(y_t-\hat{\mu})^3=\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\varepsilon^3_t-3\sigma^2_0\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\varepsilon_t+o_{P_0}(1), \tag{3.18}$$

$$\sqrt{n}\left(\hat{\sigma}^2_n\right)^{3/2}=\sqrt{n}\left(\sigma^2_0\right)^{3/2}+\frac{3}{2}\left(\sigma^2_0\right)^{1/2}\sqrt{n}\left(\hat{\sigma}^2_n-\sigma^2_0\right)+o_{P_0}(1). \tag{3.19}$$

Note that $T_{3,n}(y^n,\,\hat{\theta}_n)$ can be expressed as

$$T_{3,n}(y^n,\,\hat{\theta}_n)=\frac{\hat{\mu}_3}{\hat{\sigma}^3_n}-\frac{\mu_3}{\sigma^3_0}=\frac{\hat{\mu}_3-\mu_3}{\hat{\sigma}^3_n}-\frac{\mu_3}{\sigma^3_0}\frac{\hat{\sigma}^3_n-\sigma^3_0}{\hat{\sigma}^3_n}. \tag{3.20}$$

Then, using Equations (3.18) and (3.19) in (3.20), we obtain

$$T_{3,n}(y^n,\,\hat{\theta}_n)=\frac{\alpha_3}{\hat{\sigma}^3_n}\frac{1}{n}\sum_{t=1}^{n}Z_{3,t}+o_{P_0}(1). \tag{3.21}$$

Then, the result $\sqrt{n}T_{3,n}(y^n,\,\hat{\theta}_n)\overset{A}{\sim}N[0,\,\alpha_3\Gamma_{3,n}\alpha'_3/\sigma^6_0]$ stated in [2] directly follows from Equation (3.21). Similarly, the proof of Proposition 2.15 in Appendix A shows that our general variance formula is also based on the mean value expansions of the test statistic and the score functions.

3.2. Testing kurtosis

The kurtosis statistic is defined by

$$T_{4,n}(y^n,\, \hat{\theta}_n) = \frac{1}{n}\sum_{i=1}^{n}\psi(y_t, \hat{\theta}), \quad \text{where} \quad \psi(y_t, \hat{\theta}) = \hat{\varepsilon}_t^4/\hat{\sigma}_n^4 - \mu_4/\sigma_0^4, \tag{3.22}$$

The following corollary gives the asymptotic distribution of $\sqrt{n}T_{4,n}(y^n,\, \hat{\theta}_n)$.

Corollary 3.6. *The asymptotic distribution of* $\sqrt{n}T_{4,n}(y^n,\, \hat{\theta}_n)$ *is*

$$\sqrt{n}T_{4,n}(y^n,\, \hat{\theta}_n) \overset{A}{\sim} N\left[0, \mathcal{S}_{4,n}(\theta_0)\right], \tag{3.23}$$

where

$$\begin{aligned}\mathcal{S}_{4,n}(\theta_0) &= \mathcal{H}_{44,n}(\theta_0) + \mathcal{D}_{4,n}(\theta_0)\mathcal{A}_n^{-1}(\theta_0)\mathcal{H}_{11,n}(\theta_0)\mathcal{A}_n^{-1}(\theta_0)\mathcal{D}_{4,n}^{'}(\theta_0)\\ &\quad + \mathcal{H}_{14,n}^{'}(\theta_0)\mathcal{A}_n^{-1}(\theta_0)\mathcal{D}_{4,n}^{'}(\theta_0) + \mathcal{D}_{4,n}(\theta_0)\mathcal{A}_n^{-1}(\theta_0)\mathcal{H}_{14,n}(\theta_0)\end{aligned}$$

and

$$\mathcal{D}_{4,n}(\theta_0) = \mathbb{E}\left(n^{-1}\sum_{t=1}^{n}\frac{\partial\psi(y_t,\theta)}{\partial\theta'}|_{\theta_0}\right) = \left(-\frac{4\mu_3}{\sigma_0^4}, -\frac{2\mu_4}{\sigma_0^6}\right).$$

Proof. See Appendix B. □

Alternatively, Bai and Ng [2] show that $\sqrt{n}T_{4,n}(y^n,\, \hat{\theta}_n) \overset{A}{\sim} N\left[0, \alpha_4\Gamma_{4,n}\alpha_4^{'}/\sigma_0^8\right]$, where $\alpha_4 = (1,\, -4\mu_3,\, -2\mu_4/\sigma_0^2)$ and $\Gamma_{4,n}$ is the long-run covariance matrix of $\frac{1}{\sqrt{n}}\sum_{t=1}^{n} Z_{4,t}$ with $Z_{4,t} = \left(\varepsilon_t^4 - \mu_4,\, \varepsilon_t,\, \varepsilon_t^2 - \sigma_0^2\right)^{'}$. The following corollary shows that our result coincides with their result.

Corollary 3.7. *It follows that* $\alpha_4\Gamma_{4,n}\alpha_4^{'}/\sigma_0^8 = \mathcal{S}_{4,n}(\theta_0)$.

Proof. See Appendix B. □

3.3. Testing normality

To test the null hypothesis of normality, we use both skewness and kurtosis statistics to formulate the following test statistic:

$$T_{34,n}(y^n, \hat{\theta}_n) = \frac{1}{n}\sum_{t=1}^{n}\psi(y_t, \hat{\theta}_n), \quad \text{where} \quad \psi(y_t, \hat{\theta}_n) = \begin{pmatrix}\hat{\varepsilon}_t^3/\hat{\sigma}_n^3\\ \hat{\varepsilon}_t^4/\hat{\sigma}_n^4 - 3\end{pmatrix}. \tag{3.24}$$

Under the null hypothesis of normality, $T_{3,n}(y^n,\, \hat{\theta}_n)$ and $T_{4,n}(y^n,\, \hat{\theta}_n)$ are asymptotically independent even for time series data [2, 16, 17]. Thus, under the null hypothesis of normality, a generalization of JB test of Jarque and Bera [14] to dependent data is

$$\begin{aligned} &n\, T_{34,n}^{'}(y^n, \hat{\theta}_n)\left(\operatorname{Var}\left(\sqrt{n}T_{34,n}(y^n, \hat{\theta}_n)\right)\right)^{-1} T_{34,n}(y^n, \hat{\theta}_n) \\ &\quad = n\, T_{3,n}^2(y^n, \hat{\theta}_n)/\operatorname{Var}\left(\sqrt{n}T_{3,n}(y^n, \hat{\theta}_n)\right) + n\, T_{4,n}^2(y^n, \hat{\theta}_n)/\operatorname{Var}\left(\sqrt{n}T_{4,n}(y^n, \hat{\theta}_n)\right).\end{aligned} \tag{3.25}$$

Following Bai and Ng [2], we also consider an alternative test statistic based on the third and fourth central moments. This test statistic is defined as

$$T_{34,n}^{\mu}(y^n, \hat{\theta}_n) = \frac{1}{n}\sum_{t=1}^{n}\psi(y_t, \hat{\theta}_n), \quad \text{where} \quad \psi(y_t, \hat{\theta}_n) = \begin{pmatrix}\hat{\varepsilon}_t^3\\ \hat{\varepsilon}_t^4 - 3\hat{\sigma}_n^4\end{pmatrix}. \tag{3.26}$$

Let $\mathcal{S}_{34,n}^{\mu}(\theta_0)$ be the asymptotic covariance of $\sqrt{n}T_{34,n}^{\mu}(y^n, \hat{\theta}_n)$. Then, under the null hypothesis of normality, Proposition 2.15 gives

$$\begin{aligned}\mathcal{S}_{34,n}^{\mu}(\theta_0) &= \mathcal{H}_{55,n}(\theta_0) + \mathcal{D}_{34,n}^{\mu}(\theta_0)\mathcal{A}_n^{-1}(\theta_0)\mathcal{H}_{11,n}(\theta_0)\mathcal{A}_n^{-1}(\theta_0)\mathcal{D}_{34,n}^{\mu'}(\theta_0)\\ &\quad + \mathcal{H}_{15,n}^{'}(\theta_0)\mathcal{A}_n^{-1}(\theta_0)\mathcal{D}_{34,n}^{\mu'}(\theta_0) + \mathcal{D}_{34,n}^{\mu}(\theta_0)\mathcal{A}_n^{-1}(\theta_0)\mathcal{H}_{15,n}(\theta_0),\end{aligned} \tag{3.27}$$

where

$$\mathcal{D}^{\mu}_{34,n}(\theta_0) = \mathbb{E}\left(n^{-1}\sum_{t=1}^{n}\frac{\partial\psi(y_t,\theta)}{\partial\theta'}|_{\theta_0}\right) = \begin{pmatrix} -3\sigma_0^2 & 0 \\ 0 & -6\sigma_0^2 \end{pmatrix}. \tag{3.28}$$

The following corollary gives the asymptotic distributions of $\sqrt{n}T_{34,n}(y^n,\hat{\theta}_n)$ and $\sqrt{n}T^{\mu}_{34,n}(y^n,\hat{\theta}_n)$.

Corollary 3.8. (1) *Under the null hypothesis of normality, we have*

$$\begin{aligned} &n\,T'_{34,n}(y^n,\hat{\theta}_n)\left(\text{Var}\left(\sqrt{n}T_{34,n}(y^n,\hat{\theta}_n)\right)\right)^{-1}T_{34,n}(y^n,\hat{\theta}_n) \\ &= n\,T^2_{3,n}(y^n,\hat{\theta}_n)/\mathcal{S}_{3,n}(\hat{\theta}_n) + n\,T^2_{4,n}(y^n,\hat{\theta}_n)/\mathcal{S}_{4,n}(\hat{\theta}_n) \overset{A}{\sim} \chi^2_2, \end{aligned} \tag{3.29}$$

where $\mathcal{S}_{3,n}(\hat{\theta}_n)$ and $\mathcal{S}_{4,n}(\hat{\theta}_n)$ are consistent estimators of $\mathcal{S}_{3,n}(\theta_0)$ and $\mathcal{S}_{4,n}(\theta_0)$, respectively.

(2) *Under the null hypothesis of normality, it follows that*

$$n\,T^{\mu'}_{34,n}(y^n,\hat{\theta}_n)\left(\mathcal{S}^{\mu}_{34,n}(\hat{\theta}_n)\right)^{-1}T^{\mu}_{34,n}(y^n,\hat{\theta}_n) \overset{A}{\sim} \chi^2_2, \tag{3.30}$$

where $\mathcal{S}^{\mu}_{34,n}(\hat{\theta}_n)$ is a consistent estimator of $\mathcal{S}^{\mu}_{34,n}(\theta_0)$.

Proof. See Appendix B. □

In the case of $T^{\mu}_{34,n}(y^n,\hat{\theta}_n)$, Bai and Ng [2] alternatively show that

$$n\,T^{\mu'}_{34,n}(y^n,\hat{\theta}_n)\left(\alpha_{34}\Gamma_{34,n}\alpha'_{34}\right)^{-1}T^{\mu}_{34,n}(y^n,\hat{\theta}_n) \overset{A}{\sim} \chi^2_2, \tag{3.31}$$

where $\Gamma_{34,n}$ is the long-run covariance matrix of $\frac{1}{\sqrt{n}}\sum_{t=1}^{n} Z_{34,t}$ with $Z_{34,t} = (\varepsilon_t,\, \varepsilon_t^2 - \sigma_0^2, \varepsilon_t^3,\, \varepsilon_t^4 - 3\sigma_0^4,\,)'$ and

$$\alpha_{34} = \begin{pmatrix} -3\sigma_0^2 & 0 & 1 & 0 \\ 0 & -6\sigma_0^2 & 0 & 1 \end{pmatrix}. \tag{3.32}$$

Then, the following corollary shows that our results coincide with the results of [2].

Corollary 3.9. *It follows that $\alpha_{34}\Gamma_{34,n}\alpha'_{34} = \mathcal{S}^{\mu}_{34,n}(\theta_0)$.*

Proof. See Appendix B. □

Remark 3.10. When the disturbance terms are i.i.d, the asymptotic variance of the unfeasible version $\sqrt{n}T_{34,n}(y^n,\theta_0)$ under the null hypothesis of normality can be derived as

$$\begin{aligned} \mathcal{C}_{34,n}(\theta_0) &= \mathbb{E}\left(\psi(y_t,\theta_0)\psi'(y_t,\theta_0)\right) = \mathbb{E}\begin{pmatrix} \varepsilon_t^6/\sigma_0^6 & \varepsilon_t^7/\sigma_0^7 - 3\,\varepsilon_t^3/\sigma_0^3 \\ \varepsilon_t^7/\sigma_0^7 - 3\,\varepsilon_t^3/\sigma_0^3 & \varepsilon_t^8/\sigma_0^8 - 6\,\varepsilon_t^4/\sigma_0^4 + 9 \end{pmatrix} \\ &= \begin{pmatrix} 15 & 0 \\ 0 & 96 \end{pmatrix}. \end{aligned} \tag{3.33}$$

Then, using Equation (2.11), it can be shown that $\sqrt{n}T_{34,n}(y^n,\hat{\theta}_n)$ has the following asymptotic covariance under the null hypothesis of normality

$$\text{Var}\left(\sqrt{n}T_{34,n}(y^n,\hat{\theta}_n)\right) = \begin{pmatrix} 6 & 0 \\ 0 & 24 \end{pmatrix}. \tag{3.34}$$

Then, the omnibus test statistic derived in Jarque and Bera [14] can alternatively be derived as

$$\begin{aligned} \text{JB} &= n\,T'_{34,n}(y^n,\hat{\theta}_n)\left(\text{Var}\left(\sqrt{n}T_{34,n}(y^n,\hat{\theta}_n)\right)\right)^{-1}T_{34,n}(y^n,\hat{\theta}_n) \\ &= \frac{n}{6}\left(\frac{1}{n}\sum_{i=1}^{n}\hat{\varepsilon}_i^3/\hat{\sigma}_n^3\right)^2 + \frac{n}{24}\left(\frac{1}{n}\sum_{i=1}^{n}\hat{\varepsilon}_i^4/\hat{\sigma}_n^4 - 3\right)^2 \overset{A}{\sim} \chi^2_2. \end{aligned} \tag{3.35}$$

It is important to note that the JB is derived under under the null hypothesis of normality, and its components cannot be used separately to test symmetry or excess kurtosis.

4. A Monte Carlo simulation

In this section, we design a Monte Carlo simulation to investigate the finite sample properties of test statistics derived in Section 3. Following [2], we generate y_t according to $y_t = \rho_0 y_{t-1} + u_t$, where $\rho_0 \in \{0, 0.5, 0.8\}$, and u_t's are i.i.d. random variables generated from the symmetric and asymmetric distributions. We consider 3 symmetric and 3 asymmetric distributions listed in Table 1. We consider $n \in \{100, 500, 1000\}$, and set the nominal size to 0.05 and the number of re-sampling to 2000 in all cases.*

Table 1. Distributions.

	Symmetric Distributions
1	Standard normal distribution: $N(0,1)$
2	Student's t distribution: t_5
3	$e_1\mathbf{1}(z \le 0.5) + e_2\mathbf{1}(z > 0.5)$, where $z \sim U(0,1)$, $e_1 \sim N(-1,1)$ and $e_2 \sim N(1,1)$.
	Asymmetric Distributions
4	Lognormal: $\exp(e)$, $e \sim N(0,1)$
5	Chi-squared distribution: χ^2_2
6	Exponential: Exp(1)

We use $\hat{\mu}_n = n^{-1}\sum_{t=1}^n y_t$ and $\hat{\sigma}_n^2 = n^{-1}\sum_{t=1}^n \hat{\varepsilon}_t^2$, where $\hat{\varepsilon}_t = y_t - \hat{\mu}_n$ to compute our test statistics given in Section 3. Also, we need to formulate consistent estimators of $\mathcal{A}_n(\theta_0)$, $\mathcal{D}_{3,n}(\theta_0)$, $\mathcal{D}_{4,n}(\theta_0)$, $\mathcal{D}_{35,n}(\theta_0)$, $\mathcal{D}^{\mu}_{34,n}(\theta_0)$ and $\mathcal{H}_n(\theta_0)$. We use the plug-in estimators for $\mathcal{A}_n(\theta_0)$, $\mathcal{D}_{3,n}(\theta_0)$, $\mathcal{D}_{4,n}(\theta_0)$, $\mathcal{D}_{35,n}(\theta_0)$, and $\mathcal{D}^{\mu}_{34,n}(\theta_0)$, and the following estimator for $\mathcal{H}_n(\theta_0)$,

$$\mathcal{H}_n(\hat{\theta}) = \hat{\Gamma}_0 + \sum_{i=1}^{P}\omega_i\left(\hat{\Gamma}_i + \hat{\Gamma}_i^{'}\right), \tag{4.1}$$

where $\hat{\Gamma}_i = \frac{1}{n}\sum_{t=i+1}^n h(y_t, \hat{\theta})h^{'}(y_{t-i}, \hat{\theta})$ for $i = 0, 1, 2, \ldots, P$, P is the maximum lag length and ω_i's are weights chosen to ensure that $\mathcal{H}_n(\hat{\theta})$ is positive definite and consistent. We use three methods to choose ω_i and P: (i) the Bartlett method, (ii) the Parzen method and (iii) the quadratic spectral method.† Table 2 shows how P and ω_i can be determined according to each method. Using Table 2, we use a three-step approach to compute $\mathcal{H}_n(\hat{\theta})$. For a given method, in the first step, we choose the initial P value given in the first column of Table 2. In the second step, we compute $\hat{H}_0 = \hat{\Gamma}_0 + \sum_{i=1}^P\left(\hat{\Gamma}_i + \hat{\Gamma}_i^{'}\right)$, $\hat{H}_1 = 2\sum_{i=1}^P i\hat{\Gamma}_i$, $\hat{H}_2 = 2\sum_{i=1}^P i^2\hat{\Gamma}_i$ and $\nu_i = l^{'}\hat{H}_i l$ for $i = 0, 1, 2$, where l is a matching column vector of ones. We use these quantities to update P according to the second column of Table 2. In the third step, we use the updated P value to compute the weights given in the third column of Table 2. We then use these quantities to compute $\mathcal{H}_n(\hat{\theta})$ in Equation (4.1).

The simulation results are presented in Tables 3-5. The results for testing the null hypothesis of symmetry are given in Table 3. The results in the first three rows of each panel show the empirical size properties, while those in the remaining rows show the empirical power properties. In the case of $T_{35,n}$, we set $r_1 = 3$ and $r_2 = 5$. The results in Table 3 show that $T_{35,n}$ is severely under-sized, and generally has low power in all cases.

*The simulation results in this section can be replicated by using the code available at https://sites.google.com/view/osmandogan/software.

†The simulation results presented in [2] are based on the Bartlett method.

On the other hand, $T_{3,n}$ performs better than $T_{35,n}$ in terms of both empirical size and power properties. Although the size of $T_{3,n}$ is robust to the degree of serial correlation in data, there are few cases where this test is over-sized (e.g., under the case where we have the quadratic method, $\rho_0 = 0.8$ and $n = 100$). In general, the power decreases when ρ_0 increases to 0.8, especially when $n = 100$. Overall, our results in Table 3 are consistent with those presented in [2] (e.g., see $\hat{\pi}_3^*$ and $\hat{\mu}_{35}$ in their Table 1).

Table 2. Choosing lag length and weights.

Initial P	Updated P	Weight
	Bartlett	
$\text{int}\left[4\left(\frac{n}{100}\right)^{2/9}\right]$	$\text{int}\left[1.1447\left(\frac{\nu_1^2 n}{\nu_0^2}\right)^{1/3}\right]$	$\omega_i = 1 - \frac{i}{P+1}$
	Parzen	
$\text{int}\left[4\left(\frac{n}{100}\right)^{4/25}\right]$	$\text{int}\left[2.6614\left(\frac{\nu_2^2 n}{\nu_0^2}\right)^{1/5}\right]$	$\omega_i = \begin{cases} 1 - 6(\frac{i}{P+1})^2 - 6(\frac{i}{P+1})^3, \ 0 \le i \le \frac{P+1}{2} \\ 2(1 - \frac{i}{P+1})^3, \text{ otherwise} \end{cases}$
	Quadratic Spectral	
$\text{int}\left[4\left(\frac{n}{100}\right)^{2/25}\right]$	$\text{int}\left[1.3221\left(\frac{\nu_2^2 n}{\nu_0^2}\right)^{1/5}\right]$	$\omega_i = \frac{25}{12\pi^2(\frac{i}{P})^2}\left(\frac{5P}{6\pi i}\sin(\frac{6\pi i}{5P}) - \cos(\frac{6\pi i}{5P})\right)$

Table 3. Testing symmetry.

		n=100						n=500					
		Bartlett		Parzen		Quadratic		Bartlett		Parzen		Quadratic	
ρ_0	dist	$T_{3,n}$	$T_{35,n}$	$T_{3,n}$	$T_{35,n}$	$T_{3,n}$	$T_{35,n}$	$T_{3,n}$	$T_{35,n}$	$T_{3,n}$	$T_{35,n}$	$T_{3,n}$	$T_{35,n}$
0	1	0.034	0.008	0.043	0.018	0.052	0.011	0.043	0.015	0.048	0.015	0.046	0.017
0	2	0.026	0.018	0.040	0.029	0.035	0.025	0.028	0.099	0.030	0.088	0.034	0.106
0	3	0.034	0.006	0.057	0.015	0.056	0.008	0.048	0.015	0.044	0.013	0.051	0.019
0	4	0.385	0.462	0.405	0.455	0.405	0.464	0.678	0.932	0.684	0.925	0.676	0.928
0	5	0.756	0.045	0.800	0.079	0.769	0.064	0.991	0.547	0.989	0.428	0.991	0.474
0	6	0.715	0.173	0.741	0.179	0.727	0.170	0.987	0.900	0.987	0.750	0.983	0.790
0.5	1	0.037	0.004	0.040	0.010	0.066	0.014	0.043	0.009	0.052	0.013	0.189	0.051
0.5	2	0.028	0.007	0.045	0.018	0.059	0.015	0.038	0.009	0.039	0.025	0.113	0.029
0.5	3	0.036	0.003	0.041	0.007	0.068	0.016	0.050	0.010	0.042	0.009	0.205	0.076
0.5	4	0.410	0.175	0.432	0.170	0.436	0.206	0.683	0.732	0.687	0.767	0.711	0.783
0.5	5	0.659	0.036	0.673	0.064	0.699	0.051	0.988	0.080	0.986	0.099	0.991	0.168
0.5	6	0.678	0.026	0.692	0.075	0.698	0.059	0.981	0.117	0.978	0.115	0.985	0.239
0.8	1	0.040	0.006	0.043	0.011	0.315	0.085	0.042	0.004	0.039	0.006	0.006	0.022
0.8	2	0.034	0.015	0.036	0.013	0.346	0.068	0.046	0.007	0.040	0.004	0.032	0.063
0.8	3	0.033	0.006	0.040	0.007	0.311	0.091	0.037	0.009	0.052	0.009	0.003	0.018
0.8	4	0.312	0.030	0.306	0.062	0.664	0.083	0.673	0.117	0.684	0.211	0.755	0.294
0.8	5	0.183	0.005	0.195	0.018	0.606	0.096	0.860	0.259	0.857	0.263	0.631	0.142
0.8	6	0.234	0.009	0.260	0.022	0.670	0.095	0.892	0.246	0.907	0.227	0.726	0.129

The simulation results for testing the null hypothesis of no excess kurtosis are presented in Table 4. In this case, the results in the first row of each panel show the empirical size properties, while those in the remaining rows show the empirical power properties. The results show that $T_{4,n}$ is generally over-sized, especially when we have the quadratic method and $\rho_0 = 0.8$. When $n = 100$, the power is very low in all cases. Moreover, the presence of serial correlation, i.e., when $\rho_0 = 0.5$ or $\rho_0 = 0.8$, further reduces the power. The power generally increases when the sample size increases to 1000 in all cases.

Table 4. Testing excess kurtosis.

		n=100			n=500			n=1000		
		Bartlett	Parzen	Quadratic	Bartlett	Parzen	Quadratic	Bartlett	Parzen	Quadratic
ρ_0	dist	$T_{4,n}$	$T_{4,n}$	$T_{4,n}$	$T_{4,n}$	$T_{4,n}$	$T_{4,n}$	$T_{4,n}$	$T_{4,n}$	$T_{4,n}$
0	1	0.058	0.081	0.080	0.070	0.076	0.079	0.066	0.075	0.061
0	2	0.070	0.099	0.091	0.559	0.562	0.589	0.731	0.719	0.713
0	3	0.342	0.384	0.371	0.899	0.911	0.899	0.990	0.989	0.991
0	4	0.209	0.211	0.205	0.372	0.372	0.376	0.473	0.484	0.451
0	5	0.126	0.143	0.148	0.745	0.729	0.749	0.911	0.910	0.919
0	6	0.207	0.232	0.219	0.759	0.766	0.775	0.899	0.897	0.908
0.5	1	0.041	0.060	0.078	0.086	0.075	0.102	0.077	0.074	0.099
0.5	2	0.045	0.062	0.063	0.401	0.372	0.402	0.654	0.647	0.683
0.5	3	0.129	0.169	0.193	0.532	0.522	0.545	0.726	0.724	0.763
0.5	4	0.166	0.180	0.190	0.378	0.402	0.384	0.472	0.489	0.501
0.5	5	0.061	0.086	0.069	0.552	0.579	0.558	0.853	0.846	0.859
0.5	6	0.120	0.128	0.102	0.669	0.664	0.677	0.862	0.856	0.875
0.8	1	0.039	0.044	0.131	0.065	0.075	0.244	0.080	0.088	0.101
0.8	2	0.029	0.043	0.115	0.064	0.058	0.327	0.158	0.163	0.266
0.8	3	0.040	0.058	0.162	0.117	0.139	0.291	0.156	0.191	0.131
0.8	4	0.045	0.051	0.114	0.307	0.281	0.757	0.425	0.441	0.688
0.8	5	0.020	0.033	0.116	0.073	0.068	0.334	0.197	0.174	0.300
0.8	6	0.029	0.030	0.102	0.130	0.099	0.427	0.334	0.298	0.442

Finally, we evaluate the simulation results presented in Table 5 for testing the null hypothesis of normal distribution. In this table, the results in the first row of each panel show the empirical size properties, while those in the remaining rows indicate the empirical power properties. In Table 5, JB denotes our suggested test stated in Equation (3.25). The simulation results show that $T^{\mu}_{34,n}$ is severely under-sized. In terms of size properties, the JB test generally performs better than $T^{\mu}_{34,n}$ in all cases. However, there are some cases where the JB test reports large size distortions, especially when we have serial correlation in data and the method is quadratic. Both tests have low power when the sample size is small and the distributions are symmetric. When the sample size increases to 1000, both tests report good power (though there are some irregular cases when $\rho_0 = 0.8$).

Table 5. Testing normality.

		n=100						n=500						n=1000					
		Bartlett		Parzen		Quadratic		Bartlett		Parzen		Quadratic		Bartlett		Parzen		Quadratic	
ρ_0	dist	$T^{\mu}_{34,n}$	JB	$T^{\mu}_{34,n}$	JB	$T^{\mu}_{34,n}$	JB	$T^{\mu}_{34,n}$	JB	$T^{\mu}_{34,n}$	JB	$T^{\mu}_{34,n}$	JB	$T^{\mu}_{34,n}$	JB	$T^{\mu}_{34,n}$	JB	$T^{\mu}_{34,n}$	JB
0	1	0.006	0.028	0.022	0.065	0.021	0.063	0.024	0.063	0.038	0.075	0.032	0.071	0.030	0.059	0.050	0.072	0.029	0.059
0	2	0.009	0.048	0.021	0.080	0.013	0.067	0.236	0.396	0.239	0.395	0.239	0.408	0.507	0.599	0.514	0.603	0.525	0.597
0	3	0.047	0.219	0.075	0.274	0.048	0.235	0.496	0.855	0.496	0.851	0.507	0.863	0.812	0.977	0.790	0.977	0.796	0.976
0	4	0.001	0.462	0.011	0.472	0.003	0.472	0.291	0.679	0.310	0.675	0.326	0.686	0.534	0.776	0.527	0.743	0.545	0.774
0	5	0.001	0.723	0.011	0.737	0.001	0.729	0.818	0.988	0.812	0.989	0.813	0.987	0.981	0.999	0.978	0.997	0.980	0.994
0	6	0.104	0.718	0.162	0.732	0.093	0.717	0.992	0.980	0.987	0.975	0.989	0.980	1.000	0.996	1.000	0.998	0.998	0.996
0.5	1	0.003	0.024	0.007	0.043	0.017	0.068	0.009	0.057	0.013	0.070	0.046	0.150	0.013	0.070	0.012	0.073	0.055	0.175
0.5	2	0.007	0.036	0.013	0.068	0.013	0.060	0.079	0.255	0.085	0.247	0.111	0.313	0.266	0.483	0.272	0.496	0.326	0.546
0.5	3	0.006	0.054	0.017	0.091	0.030	0.129	0.073	0.420	0.088	0.442	0.177	0.511	0.173	0.628	0.206	0.634	0.346	0.667
0.5	4	0.000	0.493	0.004	0.468	0.004	0.492	0.219	0.707	0.262	0.685	0.293	0.717	0.424	0.786	0.470	0.784	0.467	0.761
0.5	5	0.002	0.545	0.004	0.598	0.009	0.586	0.251	0.981	0.283	0.977	0.422	0.919	0.885	0.999	0.883	0.996	0.875	0.823
0.5	6	0.002	0.610	0.015	0.642	0.021	0.655	0.747	0.971	0.804	0.977	0.809	0.925	0.986	0.996	0.986	0.997	0.884	0.852
0.8	1	0.002	0.009	0.006	0.032	0.018	0.228	0.005	0.051	0.004	0.064	0.045	0.064	0.006	0.066	0.007	0.070	0.009	0.011
0.8	2	0.003	0.019	0.007	0.029	0.018	0.243	0.012	0.058	0.016	0.060	0.072	0.107	0.031	0.115	0.032	0.111	0.043	0.056
0.8	3	0.004	0.011	0.007	0.029	0.022	0.229	0.009	0.080	0.013	0.095	0.051	0.070	0.011	0.126	0.021	0.145	0.011	0.011
0.8	4	0.000	0.228	0.004	0.230	0.011	0.544	0.002	0.696	0.009	0.684	0.388	0.141	0.132	0.785	0.188	0.789	0.103	0.040
0.8	5	0.002	0.085	0.005	0.107	0.015	0.426	0.000	0.758	0.001	0.773	0.103	0.065	0.000	0.959	0.003	0.967	0.062	0.041
0.8	6	0.001	0.120	0.003	0.134	0.014	0.484	0.000	0.845	0.002	0.863	0.202	0.088	0.007	0.972	0.025	0.978	0.074	0.049

5. Conclusion

In this paper, we showed how the parameter uncertainty problem affects the asymptotic distribution of test statistics formulated with the QMLE. We provided a general result for the distribution of test statistics formulated with the QMLE under some high-level assumptions. We then showed how to use this general result to develop test statistics for testing normality, skewness and kurtosis under parameter uncertainty for time series data. Our results on the asymptotic distribution of skewness coefficient are valid in the presence of excess kurtosis, and vice-versa. We showed that the asymptotic distributions of our tests coincide with those proposed in Bai and Ng [2]. Therefore, our analysis provides a unified approach that can be used to determine the asymptotic distributions of various test statistics designed for testing normality, skewness and kurtosis for time series data. In a Monte Carlo study, we investigated the finite sample size and power properties of our suggested test statistics. Our results show that $T_{3,n}$, $T_{4,n}$ and the JB test ($T_{34,n}$) can be useful for testing the null hypothesis of no skewness, no excess kurtosis and normality, respectively.

Our results suggest several directions for future studies. First, though our Proposition 2.15 indicates that we can determine the asymptotic distribution of any test statistic as long as the QMLE has the standard asymptotic properties, we consider only the test statistics for testing skewness, excess kurtosis and normality. Our result can be used to determine the asymptotic distributions of some other type of test statistics including some well-know test statistics such as the Cox test, the White information matrix test and the Durbin h test in the context of time series data. Second, we developed our suggested tests in Section 3 in the context of a simple scale-location model. It will be interesting to consider some specific models such as ARMA and GARCH type models to develop our suggested tests and study their finite sample properties. Third, our test statistics can have size distortions when the sample size is small since we suggested the critical values based on the corresponding asymptotic distributions. In future studies, the block bootstrap versions of our tests can be considered. In this respect, the unified approach suggested in [12] can be used to develop a block bootstrap version of our Proposition 2.15. Such a result will provide the first-order asymptotic validity of the block bootstrap versions of our suggested tests. Finally, another direction for future studies is to compare the finite sample size and power properties of our suggested test with their block bootstrap versions through simulation studies. We leave all these extensions for future studies.

Acknowledgment. We are grateful to the Editor and two anonymous referees for their constructive comments and suggestions on the earlier versions of this paper.

References

[1] D.W.K. Andrews, *Heteroskedasticity and autocorrelation consistent covariance matrix estimation*, Econometrica **59** (3), 817-858, 1991.

[2] J. Bai and S. Ng, *Tests for skewness, kurtosis, and normality for time series data*, J. Bus. Econom. Statist. **23** (1), 49-60, 2005.

[3] A.K. Bera, O. Doğan and S. Taşpınar, *Asymptotic variance of test statistics in the ML and QML frameworks*, J. Stat. Theory Pract. **15** (2), 1-26, 2021.

[4] C. Bontemps, *Moment-based tests under parameter uncertainty*, Rev. Econ. Stat. **101** (1), 146-159, 2019.

[5] C. Bontemps and N. Meddahi, *Testing normality: a GMM approach*, J. Econometrics **124** (1), 149-186, 2005.

[6] C. Bontemps and N. Meddahi, *Testing distributional assumptions: A GMM approach*, J. Appl. Econometrics **27** (6), 978-1012, 2012.

[7] J. Davidson, *Stochastic Limit Theory: An Introduction for Econometricians*, Oxford University Press, 1994.
[8] R. Davidson and J.G. MacKinnon, *Model specification tests based on artificial linear regressions*, Internat. Econom. Rev. **25** (2), 485-502, 1984.
[9] A.R. Gallant, *Explicit estimators of parametric functions in nonlinear regression*, J. Amer. Statist. Assoc. **75** (369), 182-193, 1980.
[10] A.R. Gallant and H. White, *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*, B. Blackwell, 1988.
[11] P. Glewwe, *A test of the normality assumption in ordered probit model*, Econometric Rev. **16** (1), 1-19, 1997.
[12] S. Goncalves and H. White, *Maximum likelihood and the bootstrap for nonlinear dynamic models*, J. Econometrics **119** (1), 199-219, 2004.
[13] J.A. Hausman, *Specification tests in econometrics*, Econometrica **46** (6), 1251-1271, 1978.
[14] C.M. Jarque and A.K. Bera, *A test for normality of observations and regression residuals*, Int. Stat. Rev. **55** (2), 163-172, 1987.
[15] N.M. Kiefer and M. Salmon, *Testing normality in econometric models*, Econom. Lett. **11** (1), 123-127, 1983.
[16] I.N. Lobato and C. Velasco, *A simple test of normality for time series*, Econometric Theory **20** (4), 671-689, 2004.
[17] Z.A. Lomnicki, *Tests for departure from normality in the case of linear stochastic processes*, Metrika **4** (1), 37-62, 1961.
[18] J.G. MacKinnon, *Model specification tests and artificial regressions*, J. Econ. Lit. **30** (1), 102-146, 1992.
[19] M.W. McCracken, *Robust out-of-sample inference*, J. Econometrics **99** (2), 2000.
[20] M.W. McCracken and K.S. West, *Inference about predictive ability*, A companion to economic forecasting, **14**, 299-321, Blackwell Publishing, 2002.
[21] J. Neyman, *Optimal Asymptotic Tests of Composite Statistical Hypotheses*, Probability and Statistics, the Harald Cramer Volume, Wiley, 1959.
[22] W.K. Newey, *Maximum likelihood specification testing and conditional moment tests*, Econometrica **53** (5), 1047-1070, 1985.
[23] W.K. Newey, *Generalized method of moments specification testing*, J. Econometrics **29** (3), 229-256, 1985.
[24] W.K. Newey and K.D. West, *A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix*, Econometrica **55** (3), 703-708, 1987.
[25] A. Pagan and F. Vella, *Diagnostic tests for models based on individual data: a survey*, J. Appl. Econometrics **4** (1), 29-59, 1989.
[26] D.A. Pierce, *The asymptotic effect of substituting estimators for parameters in certain types of statistics*, Ann. Statist. **10** (2), 475-478, 1982.
[27] B.M. Pötscher and I.R. Prucha, *Basic structure of the asymptotic theory in dynamic nonlinear econometric models, part I: consistency and approximation concepts*, Econometric Rev. **10** (2), 125-216, 1991.
[28] B.M. Pötscher and I.R. Prucha, *Basic structure of the asymptotic theory in dynamic nonlinear econometric models, part II: consistency and approximation concepts*, Econometric Rev. **10** (3), 253-325, 1991.
[29] R.H. Randles, *On the asymptotic normality of statistics with estimated parameters*, Ann. Statist. **10** (2), 462-474, 1982.
[30] P.A. Ruud, *Tests of specification in econometrics*, Econometric Rev. **3** (2), 211-242, 1984.
[31] G. Tauchen, *Diagnostic testing and evaluation of maximum likelihood models*, J. Econometrics **30** (1), 415-443, 1985.

[32] K.D. West, *Asymptotic inference about predictive ability*, Econometrica **64** (5), 1067-1084, 1996.
[33] K.D. West, *Forecast Evaluation*, Handbook of Economic Forecasting, North-Holland, 2006.
[34] K.D. West and M.W. McCracken, *Regression-based tests of predictive ability*, Internat. Econom. Rev. **39** (4), 817-840, 1998.
[35] H. White, *Maximum likelihood estimation of misspecified models*, Econometrica **50** (1), 1-25, 1982.
[36] H. White, *Specification Testing in Dynamic Models*, Advances in Econometrics - Fifth World Congress, Cambridge University Press, 1987.
[37] H. White, *Estimation, Inference and Specification Analysis*, Cambridge University Press, 1994.
[38] J.M. Wooldridge, *A unified approach to robust, regression-based specification tests*, Econometric Theory **6** (1), 17-43, 1990.
[39] Z. Yang, Y.K. Tse and Z. Bai, *Statistics with estimated parameters*, Statist. Sinica **17** (2), 817-837, 2007.

Appendix

Appendix A. Proof of proposition 2.15

This proposition can be proved by following the argument given in [3]. To that end, we require the following lemma, which shows how a consistent estimator of a term that obeys the strong ULLN can be formulated.

Lemma A.1. *Let $\{Q_n : \Omega \times \Theta \to \mathbb{R}\}$ be a sequence of continuous function on Θ a.s.-P_0, and let $\{\hat{\theta}_n : \Omega \to \Theta\}$ be a sequence satisfying $\hat{\theta}_n - \theta^* \to 0$ a.s.-P_0. Suppose that $\sup_{\theta\in\Theta} \left|Q_n(\cdot,\theta) - \bar{Q}_n(\theta)\right| \to 0$ a.s.-P_0, where $\{\bar{Q}_n : \Theta \to \mathbb{R}\}$ is continuous on Θ uniformly in n. Then,*

$$Q_n(\cdot,\hat{\theta}_n) - \bar{Q}_n(\theta^*) \to 0 \qquad a.s.\text{-}P_0.$$

Proof. See [37, Corollary 3.8]. □

Define the following vector

$$\varphi_t(Y^t,\theta,\psi) = \begin{pmatrix} \frac{\partial \ln f_t(y^t,\theta)}{\partial\theta} \\ \psi_t(y^t,\theta) - \psi \end{pmatrix}. \tag{A.1}$$

By Lemma A.1 and the fact that $\hat{\theta}_n - \theta^\star \to$ a.s. -P$_0$, we have $\frac{1}{n}\sum_{t=1}^n \varphi_t(Y^t,\hat{\theta}_n,\hat{\psi}_n) = 0$ a.s. -P$_0$. Also, under Assumptions 2.8 and 2.14, we have

$$\mathrm{Var}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^n \varphi_t(Y^t,\theta^\star,\psi^\star)\right) = \begin{pmatrix} \mathcal{B}_n(\theta^\star) & \mathcal{P}_n(\theta^\star,\psi^\star) \\ \mathcal{P}_n'(\theta^\star,\psi^\star) & \mathcal{C}_n(\theta^\star,\psi^\star) \end{pmatrix}, \tag{A.2}$$

where $\mathcal{P}_n(\theta^\star,\psi^\star) = \mathbb{E}\left(\left(\frac{1}{\sqrt{n}}\sum_{t=1}^n \frac{\partial \ln f_t(Y^t,\theta^\star)}{\partial\theta}\right) \times \left(\frac{1}{\sqrt{n}}\sum_{t=1}^n \left(\psi_t(Y^t,\theta^\star) - \psi^\star\right)\right)'\right)$. The CLT results in Assumptions 2.8 and 2.14 ensure that

$$\frac{1}{\sqrt{n}}\sum_{t=1}^n \varphi_t(Y^t,\theta^\star,\psi^\star) \overset{A}{\sim} N\left[0, \begin{pmatrix} \mathcal{A}_n(\theta^\star) & \mathcal{P}_n(\theta^\star,\psi^\star) \\ \mathcal{P}_n'(\theta^\star,\psi^\star) & \mathcal{C}_n(\theta^\star,\psi^\star) \end{pmatrix}\right]. \tag{A.3}$$

Let ∇_δ be the gradient with respect to δ. Taking a mean value expansion of $\varphi_t(Y^t, \hat{\theta}_n, \hat{\psi}_n)$ around $\theta^\star$ and $\psi^\star$ gives

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\varphi_t(Y^t,\theta^\star,\psi^\star) = -\frac{1}{n}\sum_{t=1}^{n}\nabla_{\theta\psi}\varphi_t(Y^t,\tilde{\theta}_n,\tilde{\psi}_n)\begin{pmatrix}\sqrt{n}(\hat{\theta}_n-\theta^\star)\\ \sqrt{n}(T_n(Y^n,\,\hat{\theta}_n)-\psi^\star)\end{pmatrix}\quad \text{a.s.-P}_0, \tag{A.4}$$

where $\tilde{\theta}_n$ and $\tilde{\psi}_n$ are the mean values. Our Assumptions 2.7 and 2.13 ensure that

$$\mathbb{E}\left(\tfrac{1}{n}\textstyle\sum_{t=1}^{n}\nabla_{\theta\psi}\varphi_t(Y^t,\theta^\star,\psi^\star)\right) = \mathbb{E}\begin{pmatrix}\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2\ln f_t(Y^t,\theta^\star)}{\partial\theta\partial\theta'} & 0\\ \frac{1}{n}\sum_{t=1}^{n}\frac{\partial\psi_t(Y^t,\theta^\star)}{\partial\theta'} & -I_q\end{pmatrix} = \begin{pmatrix}-\mathcal{A}_n(\theta^\star) & 0\\ \mathcal{D}_n(\theta^\star) & -I_q\end{pmatrix}, \tag{A.5}$$

where I_q is the $q\times q$ identity matrix. Then, under Assumptions 2.7 and 2.13, Lemma A.1 implies that

$$\left(-\frac{1}{n}\sum_{t=1}^{n}\nabla_{\theta\psi}\varphi_t(Y^t,\tilde{\theta}_n,\tilde{\psi}_n)\right) - \begin{pmatrix}-\mathcal{A}_n(\theta^\star) & 0\\ \mathcal{D}_n(\theta^\star) & -I_q\end{pmatrix} \to 0 \quad \text{a.s.-P}_0. \tag{A.6}$$

The result in Equation (A.6) ensures that $\left(-\frac{1}{n}\sum_{t=1}^{n}\nabla_{\theta\psi}\varphi_t(Y^t,\tilde{\theta}_n,\tilde{\psi}_n)\right)$ is non-singular a.s.-P_0. Then, Equation (A.4) can be written as

$$\begin{pmatrix}\sqrt{n}(\hat{\theta}_n-\theta^\star)\\ \sqrt{n}(T_n(Y^n,\,\hat{\theta}_n)-\psi^\star)\end{pmatrix} = \left(-\tfrac{1}{n}\textstyle\sum_{t=1}^{n}\nabla_{\theta\psi}\varphi_t(Y^t,\tilde{\theta}_n,\tilde{\psi}_n)\right)^{-1}\tfrac{1}{\sqrt{n}}\textstyle\sum_{t=1}^{n}\varphi_t(Y^t,\theta^\star,\psi^\star)\ \text{a.s.-P}_0. \tag{A.7}$$

Then, using Equations (A.3) and (A.6) in Equation (A.7), it follows that

$$\begin{pmatrix}\sqrt{n}(\hat{\theta}_n-\theta^\star)\\ \sqrt{n}(T_n(Y^n,\,\hat{\theta}_n)-\psi^\star)\end{pmatrix} \overset{A}{\sim} N\left[0,\begin{pmatrix}\mathcal{A}_n^{-1}(\theta^\star)\mathcal{B}_n(\theta^\star)\mathcal{A}_n^{-1}(\theta^\star) & \mathcal{V}_n'(\theta^\star,\psi^\star)\\ \mathcal{V}_n(\theta^\star,\psi^\star) & \mathcal{S}_n(\theta^\star,\psi^\star)\end{pmatrix}\right], \tag{A.8}$$

where

$$\begin{pmatrix}\mathcal{A}_n^{-1}(\theta^\star)\mathcal{B}_n(\theta^\star)\mathcal{A}_n^{-1}(\theta^\star) & \mathcal{V}_n'(\theta^\star,\psi^\star)\\ \mathcal{V}_n(\theta^\star,\psi^\star) & \mathcal{S}_n(\theta^\star,\psi^\star)\end{pmatrix}\begin{pmatrix}-\mathcal{A}_n(\theta^\star) & 0\\ \mathcal{D}_n(\theta^\star) & -I_q\end{pmatrix}^{-1}\begin{pmatrix}\mathcal{B}_n(\theta^\star) & \mathcal{P}_n(\theta^\star,\psi^\star)\\ \mathcal{P}_n'(\theta^\star,\psi^\star) & \mathcal{C}_n(\theta^\star,\psi^\star)\end{pmatrix}\begin{pmatrix}-\mathcal{A}_n(\theta^\star) & \mathcal{D}_n'(\theta^\star)\\ 0 & -I_q\end{pmatrix}^{-1}. \tag{A.9}$$

From the inverse partitioned matrix formula, it easily follows that

$$\begin{pmatrix}-\mathcal{A}_n(\theta^\star) & 0\\ \mathcal{D}_n(\theta^\star) & -I_q\end{pmatrix}^{-1} = \begin{pmatrix}-\mathcal{A}_n^{-1}(\theta^\star) & 0\\ -\mathcal{D}_n(\theta^\star)\mathcal{A}_n^{-1}(\theta^\star) & -I_q\end{pmatrix}. \tag{A.10}$$

Then, using Equation(A.10) in Equation (A.9), it can be shown that

$$\mathcal{V}_n(\theta^\star,\psi^\star) = \mathcal{D}_n(\theta^\star)\mathcal{A}_n^{-1}(\theta^\star)\mathcal{B}_n(\theta^\star)\mathcal{A}_n^{-1}(\theta^\star) + \mathcal{P}_n'(\theta^\star,\psi^\star)\mathcal{A}_n^{-1}(\theta^\star), \tag{A.11}$$

$$\begin{aligned}\mathcal{S}_n(\theta^\star,\psi^\star) &= \mathcal{C}_n(\theta^\star,\psi^\star) + \mathcal{D}_n(\theta^\star)\mathcal{A}_n^{-1}(\theta^\star)\mathcal{B}_n(\theta^\star)\mathcal{A}_n^{-1}(\theta^\star)\mathcal{D}_n'(\theta^\star)\\ &\quad + \mathcal{P}_n'(\theta^\star,\psi^\star)\mathcal{A}_n^{-1}(\theta^\star)\mathcal{D}_n'(\theta^\star) + \mathcal{D}_n(\theta^\star)\mathcal{A}_n^{-1}(\theta^\star)\mathcal{P}_n(\theta^\star,\psi^\star).\end{aligned} \tag{A.12}$$

Appendix B. Proofs of corollaries

Corollaries 3.2, 3.3, 3.6 and 3.8 directly follow from Proposition 2.15. Therefore, we only provide proofs for Corollaries 3.4, 3.7 and 3.9. We start by showing that $\alpha_3\Gamma_{3,n}\alpha_3'/\sigma_0^6 = \mathcal{S}_{3,n}(\theta_0)$. Note that we can express $Z_{3,t}$ in terms of $g_1(y_t,\theta_0) = (\varepsilon_t/\sigma_0^2,\ \varepsilon_t^2/2\sigma_0^4 - 1/2\sigma_0^2)'$ and $g_2(y_t,\theta_0) = \varepsilon_t^3/\sigma_0^3 - \mu_3\sigma_0^{-3}$ as

$$Z_{3,t} = \begin{pmatrix}\sigma_0^3 g_2(y_t,\theta_0)\\ \mathcal{A}_n^{-1}(\theta_0)g_1(y_t,\theta_0)\end{pmatrix} = \begin{pmatrix}\sigma_0^3 & 0_{1\times 2}\\ 0_{2\times 1} & \mathcal{A}_n^{-1}(\theta_0)\end{pmatrix}\begin{pmatrix}g_2(y_t,\theta_0)\\ g_1(y_t,\theta_0)\end{pmatrix}, \tag{B.1}$$

where

$$\mathcal{A}_n(\theta_0) = \begin{pmatrix}1/\sigma_0^2 & 0\\ 0 & 1/2\sigma_0^4\end{pmatrix}. \tag{B.2}$$

Thus, in terms of our notation, $\Gamma_{3,n}$ can be derived as

$$\Gamma_{3,n} = \begin{pmatrix} \sigma_0^3 & 0_{1\times 2} \\ 0_{2\times 1} & \mathcal{A}_n^{-1}(\theta_0) \end{pmatrix} \begin{pmatrix} \mathcal{H}_{22,n}(\theta_0) & \mathcal{H}_{12,n}^{'}(\theta_0) \\ \mathcal{H}_{12,n}(\theta_0) & \mathcal{H}_{11,n}(\theta_0) \end{pmatrix} \begin{pmatrix} \sigma_0^3 & 0_{1\times 2} \\ 0_{2\times 1} & \mathcal{A}_n^{-1}(\theta_0) \end{pmatrix}. \tag{B.3}$$

Also simple calculations shows that

$$\alpha_3 \begin{pmatrix} \sigma_0^3 & 0_{1\times 2} \\ 0_{2\times 1} & \mathcal{A}_n^{-1}(\theta_0) \end{pmatrix} = \sigma_0^3 \left(1,\, \mathcal{D}_{3,n}(\theta_0)\mathcal{A}_n^{-1}(\theta_0)\right). \tag{B.4}$$

Then, using Equations (B.3) and (B.4), it follows that

$$\begin{aligned} \alpha_3\Gamma_{3,n}\alpha_3^{'}/\sigma_0^6 &= \left(1,\, \mathcal{D}_{3,n}(\theta_0)\mathcal{A}_n^{-1}(\theta_0)\right) \begin{pmatrix} \mathcal{H}_{22,n}(\theta_0) & \mathcal{H}_{12,n}^{'}(\theta_0) \\ \mathcal{H}_{12,n}(\theta_0) & \mathcal{H}_{11,n}(\theta_0) \end{pmatrix} \left(1,\, \mathcal{D}_{3,n}(\theta_0)\mathcal{A}_n^{-1}(\theta_0)\right)^{'} \\ &= \mathcal{S}_{3,n}(\theta_0). \end{aligned} \tag{B.5}$$

Next, we show that $\alpha_{35}\Gamma_{35,n}\alpha_{35}^{'} = \mathcal{S}_{35,n}(\theta_0)$. In this case, Z_{35} can be expressed in terms of $g_1(y_t,\theta_0)$ and $g_3(y_t,\theta_0)$ in the following way.

$$Z_{35,t} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \sigma_0^2 & 0 \end{pmatrix} \begin{pmatrix} g_3(y_t,\theta_0) \\ g_1(y_t,\theta_0) \end{pmatrix}. \tag{B.6}$$

Thus, $\Gamma_{35,n}$ can alternatively be derived as

$$\Gamma_{35,n} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \sigma_0^2 & 0 \end{pmatrix} \begin{pmatrix} \mathcal{H}_{33,n}(\theta_0) & \mathcal{H}_{13,n}^{'}(\theta_0) \\ \mathcal{H}_{13,n}(\theta_0) & \mathcal{H}_{11,n}(\theta_0) \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \sigma_0^2 & 0 \end{pmatrix}^{'}. \tag{B.7}$$

Also simple calculations shows that

$$\alpha_{35} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \sigma_0^2 & 0 \end{pmatrix} = \left(I_2 \quad \sigma_0^2\mathcal{D}_{35,n}(\theta_0)\right) = \left(I_2 \quad \mathcal{D}_{35,n}(\theta_0)\mathcal{A}_n^{-1}(\theta_0)\right). \tag{B.8}$$

Then, it follows that

$$\begin{aligned} \alpha_{35}\Gamma_{35,n}\alpha_{35}^{'} &= \left(I_2 \quad \mathcal{D}_{35,n}(\theta_0)\mathcal{A}_n^{-1}(\theta_0)\right) \begin{pmatrix} \mathcal{H}_{33,n}(\theta_0) & \mathcal{H}_{13,n}^{'}(\theta_0) \\ \mathcal{H}_{13,n}(\theta_0) & \mathcal{H}_{11,n}(\theta_0) \end{pmatrix} \left(I_2 \quad \mathcal{D}_{35,n}(\theta_0)\mathcal{A}_n^{-1}(\theta_0)\right)^{'} \\ &= \mathcal{S}_{35,n}(\theta_0). \end{aligned} \tag{B.9}$$

To show $\alpha_4\Gamma_{4,n}\alpha_4^{'}/\sigma_0^8 = \mathcal{S}_{4,n}(\theta_0)$, we write $Z_{4,t}$ as

$$Z_{4,t} = \begin{pmatrix} \sigma_0^4 & 0_{1\times 2} \\ 0_{2\times 1} & \mathcal{A}_n^{-1}(\theta_0) \end{pmatrix} \begin{pmatrix} g_4(y_t,\theta_0) \\ g_1(y_t,\theta_0) \end{pmatrix}. \tag{B.10}$$

Then, it can be shown that

$$\Gamma_{4,n} = \begin{pmatrix} \sigma_0^4 & 0_{1\times 2} \\ 0_{2\times 1} & \mathcal{A}_n^{-1}(\theta_0) \end{pmatrix} \begin{pmatrix} \mathcal{H}_{44,n}(\theta_0) & \mathcal{H}_{14,n}^{'}(\theta_0) \\ \mathcal{H}_{14,n}(\theta_0) & \mathcal{H}_{11,n}(\theta_0) \end{pmatrix} \begin{pmatrix} \sigma_0^4 & 0_{1\times 2} \\ 0_{2\times 1} & \mathcal{A}_n^{-1}(\theta_0) \end{pmatrix}, \tag{B.11}$$

$$\alpha_4 \begin{pmatrix} \sigma_0^4 & 0_{1\times 2} \\ 0_{2\times 1} & \mathcal{A}_n^{-1}(\theta_0) \end{pmatrix} = \sigma_0^4 \left(1,\, \mathcal{D}_{4,n}\mathcal{A}_n^{-1}(\theta_0)\right). \tag{B.12}$$

Then, using Equations (B.11) and (B.12), we obtain the desired result as

$$\begin{aligned} \alpha_4\Gamma_{4,n}\alpha_3^{'}/\sigma_0^8 &= \left(1,\, \mathcal{D}_{4,n}\mathcal{A}_n^{-1}(\theta_0)\right) \begin{pmatrix} \mathcal{H}_{44,n}(\theta_0) & \mathcal{H}_{14,n}^{'}(\theta_0) \\ \mathcal{H}_{14,n}(\theta_0) & \mathcal{H}_{11,n}(\theta_0) \end{pmatrix} \left(1,\, \mathcal{D}_{4,n}\mathcal{A}_n^{-1}(\theta_0)\right)^{'} \\ &= \mathcal{S}_{4,n}(\theta_0). \end{aligned} \tag{B.13}$$

Finally, we show that $\alpha_{34}\Gamma_{34,n}\alpha_{34}^{'} = \mathcal{S}^{\mu}_{34,n}(\theta_0)$. We express $Z_{34,t}$ in terms of $g_1(y_t,\theta_0)$ and $g_5(y_t,\theta_0)$ in the following way

$$Z_{34,t} = \begin{pmatrix} \mathcal{A}_n^{-1}(\theta_0) & 0_{2\times 2} \\ 0_{2\times 2} & I_2 \end{pmatrix} \begin{pmatrix} g_1(y_t,\theta_0) \\ g_5(y_t,\theta_0) \end{pmatrix}. \tag{B.14}$$

Then, the long-run covariance matrix $\Gamma_{34,n}$ can alternatively be expressed as

$$\Gamma_{34,n} = \begin{pmatrix} \mathcal{A}_n^{-1}(\theta_0) & 0_{2\times 2} \\ 0_{2\times 2} & I_2 \end{pmatrix} \begin{pmatrix} \mathcal{H}_{11,n}(\theta_0) & \mathcal{H}_{15,n}(\theta_0) \\ \mathcal{H}_{15,n}^{'}(\theta_0) & \mathcal{H}_{55,n}(\theta_0) \end{pmatrix} \begin{pmatrix} \mathcal{A}_n^{-1}(\theta_0) & 0_{2\times 2} \\ 0_{2\times 2} & I_2 \end{pmatrix}^{'}, \tag{B.15}$$

Also, simple calculations gives

$$\alpha_{34} \begin{pmatrix} \mathcal{A}_n^{-1}(\theta_0) & 0_{2\times 2} \\ 0_{2\times 2} & I_2 \end{pmatrix} = \left(\mathcal{D}^{\mu}_{34,n}(\theta_0)\mathcal{A}_n^{-1}(\theta_0),\, I_2\right). \tag{B.16}$$

Thus, using Equations (B.15) and (B.16), we obtain the desired result as

$$\begin{aligned} \alpha_{34}\Gamma_{34,n}\alpha_{34}^{'} &= \left(\mathcal{D}^{\mu}_{34,n}(\theta_0)\mathcal{A}_n^{-1}(\theta_0),\, I_2\right) \begin{pmatrix} \mathcal{H}_{11,n}(\theta_0) & \mathcal{H}_{15,n}(\theta_0) \\ \mathcal{H}_{15,n}^{'}(\theta_0) & \mathcal{H}_{55,n}(\theta_0) \end{pmatrix} \left(\mathcal{D}^{\mu}_{34,n}(\theta_0)\mathcal{A}_n^{-1}(\theta_0),\, I_2\right)^{'} \\ &= \mathcal{S}^{\mu}_{34,n}(\theta_0). \end{aligned} \tag{B.17}$$