This article proposes the asymmetric linear double autoregression, which jointly models the conditional mean and conditional heteroscedasticity characterized by asymmetric effects. A sufficient condition is established for the existence of a strictly stationary solution. With a quasi-maximum likelihood estimation (QMLE) procedure introduced, a Bayesian information criterion (BIC) and its modified version are proposed for model selection. To detect asymmetric effects in the volatility, the Wald, Lagrange multiplier and quasi-likelihood ratio test statistics are put forward, and their limiting distributions are established under both null and local alternative hypotheses. Moreover, a mixed portmanteau test is constructed to check the adequacy of the fitted model. All asymptotic properties of inference tools including QMLE, BICs, asymmetric tests and the mixed portmanteau test, are established without any moment condition on the data process, which makes the new model and its inference tools applicable for heavy-tailed data. Simulation studies indicate that the proposed methods perform well in finite samples, and an empirical application to S&P500 index illustrates the usefulness of the new model.

Received 28 April 2021; Accepted 16 August 2021

Keywords: Asymmetry tests; autoregressive time series model; model selection; portmanteau test; quasi-maximum likelihood estimation; stationary solution.

MOS subject classification: 62M10; 62F03; 62F12.

1. INTRODUCTION

Volatility clustering is a major feature of financial time series, and the ability to forecast volatility is of vital importance for the pricing and risk management of financial assets. To capture the time-varying volatility, many conditional heteroscedastic models are proposed, and among them, the autoregressive conditional heteroscedastic (ARCH) and the generalized autoregressive conditional heteroscedastic (GARCH) models (Engle, 1982; Bollerslev, 1986) are very successful specifications. However, empirical facts indicate that the autocorrelation and volatility dynamics usually coexist in time series; see, for example, the daily returns of NASDAQ composite index in Kuester et al. (2006) and the weekly or monthly returns of S&P500 index in Linton and Mammen (2005). As a result, to better capture the volatility dynamics in the presence of data autocorrelations, it is necessary to jointly model the conditional mean and volatility (Li et al., 2002). The autoregressive moving average models with GARCH errors (ARMA-GARCH) and double autoregressive (DAR) models are popular specifications for this purpose.

In financial applications, ARMA-GARCH models are commonly used to fit return series (Francq and Zakoian, 2019). Many researchers have studied the estimation for ARMA-GARCH models; see among others, Francq and Zakoian (2004) and Zhu and Ling (2011). Francq and Zakoian (2004) has shown that, a finite fourth moment on the data process is required to establish the asymptotic normality of the Gaussian quasi-maximum likelihood estimator (QMLE) for the ARMA-GARCH model. This largely narrows down the application of ARMA-GARCH models since the heavy-tailedness is very common for financial data. Meanwhile, the DAR model proposed by

*Correspondence to: Qianqian Zhu, School of Statistics and Management, Shanghai University of Finance and Economics, Shanghai, China. E-mail: zhu.qianqian@mail.shufe.edu.cn

© 2021 John Wiley & Sons Ltd
Ling (2007) has recently attracted growing attention. The DAR model of order \( p \) is defined as

\[
y_t = \sum_{i=1}^{p} \phi_i y_{t-i} + \epsilon_t \sqrt{\omega + \sum_{i=1}^{p} \beta_i^2 y_{t-i}^2}, \tag{1.1}
\]

where \( \omega > 0, \beta_i \geq 0 \) for \( 1 \leq i \leq p \), and \( \{\epsilon_t\} \) are i.i.d. innovations with zero mean and unit variance. In contrast to the ARMA-GARCH model, the Gaussian QMLE of model (1.1) is asymptotically normal provided that \( y_t \) has a fractional moment (Ling, 2007). This important property makes model (1.1) suitable for handling heavy-tailed data in application. In addition, the DAR model with order \( p = 1 \) can still be stationary even if \( |\phi_1| > 1 \) or \( \beta_1 > 1 \), thus it enjoys a larger parameter space than conventional AR and AR-ARCH models.

Many variants of DAR models have been widely proposed and studied, such as the threshold DAR (Li et al., 2016), the mixture DAR (Li et al., 2017), the linear DAR (Zhu et al., 2018) and the augmented DAR (Jiang et al., 2020) models. Specifically, the linear DAR model of order \( p \) has the form of

\[
y_t = \sum_{i=1}^{p} \phi_i y_{t-i} + \epsilon_t \left( \omega + \sum_{i=1}^{p} \beta_i |y_{t-i}| \right), \tag{1.2}
\]

where the innovations \( \{\epsilon_t\} \) and parameters are defined as in model (1.1). Model (1.2) assumes that the conditional standard deviation rather than the conditional variance of \( y_t \) is in a linear structure, which can lead to more robust inference than model (1.1); see Taylor (2008) and Zhu et al. (2018). As shown by Zhu et al. (2018), the linear DAR model has a larger parameter space than conventional AR and AR-ARCH models as for DAR models. Moreover, the asymptotic normality of the Gaussian QMLE can also be established for model (1.2) without any moment restrictions on \( y_t \); see Liu et al. (2020). As a result, the linear DAR model enjoys the important property of DAR models and hence can also be used to fit heavy-tailed data.

It is well known that financial time series are usually characterized by asymmetry (leverage) effects, in the sense that the volatility of financial returns tends to be higher after a decrease than an equal increase. The leverage effect was documented by many authors as a stylized fact of stock returns; see, for example, Black (1976), Rabemananjara and Zakoian (1993) and Francq and Zakoian (2013). To account for the leverage phenomenon, many variants of classical GARCH models are introduced and studied, such as the exponential GARCH (Nelson, 1991), the threshold GARCH (Zakoian, 1994) and the power GARCH (Pan et al., 2008) models. Specifically, Engle and Ng (1993) defined the news impact curve to measure how new information is incorporated into volatility estimates, and based on this curve they provided diagnostic tests to detect asymmetric effects of news on volatility. However, limited literatures investigate the leverage effect in the presence of the conditional mean structure, even rare in the framework of DAR models. To fill this gap, we propose an asymmetric linear DAR model, which can be regarded as a modification of the linear DAR model along the lines of the threshold GARCH model. It is expected that the new model can preserve the advantages of DAR type models in handling with heavy-tailed data and meanwhile be able to capture the asymmetric effect successfully. The main contributions of this article are listed as follows.

First, Section 2 introduces an asymmetric linear DAR model to capture leverage effects, where the coefficients for positive and negative parts of \( y_t \) in the conditional standard deviation can be different. We establish a sufficient condition for the strict stationarity and ergodicity of the new model by showing that the Markov chain \( Y_t = (y_t, \ldots, y_{t-p+1})' \) is \( \nu_p \)-irreducible and satisfies Tweedie’s drift criterion (Tweedie, 1983); see also Zhu et al. (2018). It is shown that the stationary region of the proposed process depends on the moment condition of innovations and the degree of asymmetry. Moreover, the proposed process can still be stationary even if some AR coefficient is greater than one, leading to a large stationary region as for the DAR and linear DAR models.

Second, Section 3 propose the Gaussian QMLE for the new model and establishes its consistency and asymptotic normality. Particularly, the consistency is carefully considered to avoid the identification problem due to the coexistence of the positive and negative parts of \( y_t \), and the asymptotic normality is established without any moment condition on \( y_t \). Hence, the new model can be used to fit heavy-tailed data. Moreover, based on the QMLE,
a Bayesian information criterion (BIC) is proposed for model selection, and a modified BIC is introduced to improve the finite-sample performance. It is shown that both BICs enjoy the selection consistency without any moment condition on the process \( \{y_t\} \), and the modified BIC usually outperforms the unmodified BIC especially for small and moderate samples. As a result, the first two stages of Box–Jenkins’ procedure including model specification and estimation are constructed for the new model in a robust way, which facilitates its application to financial time series.

Third, to detect the asymmetry in volatility, the Wald, Lagrange multiplier (LM) and quasi-likelihood ratio (QLR) tests are constructed in Section 4. Under the null and local alternative hypotheses, the Wald and LM test statistics are shown to have the same limiting distributions, while the QLR test statistic converges to weighted sums of i.i.d. central and non-central Chi-squared random variables respectively. It is noteworthy that, to show the asymptotic distributions of three test statistics under local alternatives, we need to verify the local asymptotic normality (LAN) of the proposed model. However, without the normal assumption on the innovation term, the Le Cam’s third lemma cannot be employed to show the LAN property (van der Vaart, 2000). Alternatively, we show that the new model satisfies the LAN in a direct way; see also Jiang et al. (2020). Moreover, we conduct simulation experiments to compare the local power of all three tests in finite samples, and simulation results further support the theoretical findings.

Finally, we investigate diagnostic checking for fitted models using a mixed portmanteau test in Section 5. The portmanteau test for pure mean models (Ljung and Box, 1978) is constructed using the sample autocorrelation functions (ACFs) of residuals, while that for volatility models (Li and Li, 2008) employs the ACFs of squared or absolute residuals. To detect disparity of the fitted model in both the conditional mean and conditional variance, a mixed portmanteau test using the ACFs of residuals and squared residuals is introduced by Wong and Ling (2005) for location-scale models, and a mixed portmanteau test via the ACFs of residuals and absolute residuals is proposed by Zhu (2013) for ARMA-GARCH models. As validated by simulation findings of Li and Li (2005), the portmanteau test based on absolute residuals is more powerful than that based on squared residuals under heavy-tailed situations. As a result, this article proposes a mixed portmanteau test using the ACFs of residuals and absolute residuals to detect inadequacy in the conditional mean and volatility of the fitted new model. The joint limiting distribution for ACFs of residuals and absolute residuals is established without any moment condition on the process \( \{y_t\} \). Simulation results validate that the mixed test can detect inadequacy of fitted models due to either the conditional mean or the volatility. Therefore, as the last stage of Box–Jenkins’ procedure, the diagnostic checking tool is successfully constructed in a robust way as well.

In addition, Section 6 conducts simulation studies to evaluate the finite-sample performance of all inference tools for the proposed model. Section 7 illustrates the usefulness of the new model by analyzing the S&P500 index and demonstrates the forecasting superiority over its counterparts especially for the skewed and heavy-tailed data. The conclusion and discussion appear in Section 8. All technical details are relegated to the Supporting information. Throughout the article, \( \mathbb{N} \) denotes the integer, \( \rightarrow_p \) and \( \rightarrow_d \) denote the convergences in probability and in distribution, respectively, and \( o_p(1) \) denotes a sequence of random variables converging to zero in probability.

### 2. ASYMMETRIC LINEAR DOUBLE AUTOREGRESSION

Consider the asymmetric linear double autoregressive (DAR) model of order \( p \),

\[
y_t = \sum_{i=1}^{p} \alpha_i y_{t-i} + \eta_t \left( \omega + \sum_{i=1}^{p} \left( \beta^+_i y^+_{t-i} - \beta^-_i y^-_{t-i} \right) \right), \tag{2.1}
\]

where \( \omega > 0, \beta^+_i, \beta^-_i \geq 0 \) for \( 1 \leq i \leq p \), \( y^+_t = \max \{0, y_t\} \) and \( y^-_t = \min \{0, y_t\} \) are positive and negative parts of \( \{y_t\} \), respectively, and \( \{\eta_t\} \) is a sequence of i.i.d. random variables with mean zero and variance one. The asymmetric linear DAR model in (2.1) is an extension of the linear DAR model (Zhu et al., 2018) along the lines of the threshold GARCH model (Zakoian, 1994). Although the linear DAR model can be extended to allow for...
asymmetries in both the conditional mean and conditional heteroscedasticity, this article focuses on model (2.1) to take account for the asymmetry in volatilities. The real example of stock index returns in Section 7 provides evidence for this motivation.

For general distributions of \( \eta_t \), it is difficult to derive a necessary and sufficient condition for the strict stationarity due to the nonlinearity of model (2.1); see also Li et al. (2016) and Zhu et al. (2018). Alternatively, a sufficient condition is provided below.

**Assumption 1.** The density function of \( \eta_t \) is continuous and positive everywhere on \( \mathbb{R} \), and \( E(|\eta_t|^\kappa) < \infty \) for some \( \kappa > 0 \).

**Theorem 1.** Under Assumption 1, if either of the following conditions holds:

(i) for \( 0 < \kappa \leq 1 \), \( \sum_{i=1}^p \max \{E(|\alpha_i - \beta_i \eta_t|^\kappa), E(|\alpha_i + \beta_i \eta_t|^\kappa)\} < 1 \);

(ii) for \( \kappa \in \{2, 3, 4, \ldots\} \), \( E \left[ \left( \sum_{i=1}^p \max \{|\alpha_i + \beta_i \eta_t|, |\alpha_i - \beta_i \eta_t|\} \right)^\kappa \right] < 1 \);

then there exists a strictly stationary solution \( \{y_t\} \) to model (2.1), and this solution is unique and geometrically ergodic with \( E(|y_t|^\kappa) < \infty \).

The stationarity region in Theorem 1 depends on the distribution of \( \eta_t \) and implies a moment condition on \( y_t \). In addition, when \( \eta_t \) has a symmetric distribution and the asymmetric linear DAR model reduces to a linear DAR model, that is \( \beta_i^- = \beta_i^+ \), then it simplifies to \( \sum_{i=1}^p E(|\alpha_i + \beta_i \eta_t|^\kappa) < 1 \) for \( 0 < \kappa \leq 1 \), and to \( E \left[ \sum_{i=1}^p (|\alpha_i + \beta_i \eta_t|)^\kappa \right] < 1 \) for \( \kappa \in \{2, 3, 4, \ldots\} \). Since the stationarity region of model (2.1) is at least three-dimensional, for illustration, we provide the stationarity regions of model (2.1) of order one, and consider \( \beta_i^+ = d \beta_i^- \) with the constant \( d \) being different positive values. Figure 1(a) indicates that model (2.1) of order one can be stationary if \( |\alpha_i| \geq 1 \), hence model (2.1) preserves a large parameter space as DAR and linear DAR.

![Figure 1](https://wileyonlinelibrary.com/journal/jtsa)
models. As shown in Figure 1(b), a larger value of $\kappa$ in Theorem 1 leads to a higher moment of $y_t$, and hence results in a narrower stationarity region. Moreover, Figure 1(c) shows that the stationarity region gets smaller as the asymmetry in volatilities becomes greater.

**Remark 1.** To derive the sufficient condition for the geometric ergodicity and existence of $k$th moment of $\{y_t\}$, we can alternatively use the piggyback method in Cline and Pu (2004) and obtain a more sharp condition than that in Theorem 1. However, this method requires an extra moment condition $\sup_x (1 + |x|) f(x) < \infty$ on the density of $\eta_t$. Moreover, the resulting sufficient condition is more complicated to verify than that in Theorem 1. As a result, the sufficient condition in Theorem 1 is suggested for simplicity.

**Remark 2.** The order of model (2.1) can be different for the conditional mean and volatility, that is, we can consider the asymmetric linear DAR model of order $(p_1, p_2)$ as follows:

$$y_t = \sum_{i=1}^{p_1} a_i y_{t-i} + \eta_t \left( \omega + \sum_{i=1}^{p_1} \left( \beta_{i+} y_{t-i} + \beta_{i-} y_{t-i}^+ \right) \right),$$

where $p_1$ and $p_2$ are positive integers. For the strict stationarity of this general model setting, Theorem 1 still holds by letting $p = \max\{p_1, p_2\}$ with $a_i = 0$ for $i > p_1$ and $\beta_{i+} = \beta_{i-} = 0$ for $i > p_2$. However, if $p_1 > p_2$, the conditional scale structure $\omega + \sum_{i=1}^{p_2} \left( \beta_{i+} y_{t-i}^+ - \beta_{i-} y_{t-i}^+ \right)$ cannot be used to reduce the moment condition on $y_t$ for showing the asymptotic normality of the quasi-maximum likelihood estimator in Section 3. To establish the asymptotic normality without any moment condition on $y_t$, other estimation methods such as the self-weighted approach (Ling, 2005) should be considered. We leave this extension for future research, and consider the same order $p = p_1 = p_2$ in this article.

3. MODEL ESTIMATION

3.1. Quasi-maximum Likelihood Estimation

Let $\theta = (\alpha', \beta')'$ be the parameter vector of model (2.1), where $\alpha = (a_1, a_2, \ldots, a_p)'$, $\beta = (\omega, \beta_1^+, \beta_2^+, \ldots, \beta_p^+, \beta_1^-, \beta_2^-, \ldots, \beta_p^-)'$. Denote the true parameter vector by $\theta_0 = (\alpha_0', \beta_0')'$ and the parameter space by $\Theta$, where $\Theta$ is a compact subset of $\mathbb{R}^p \times \mathbb{R}_{+}^{2p+1}$ with $\mathbb{R}_{+} = (0, \infty)$.

Let $Y_t = (y_1, \ldots, y_{n+p})'$ and $X_t = (1, Y_t', -Y_t')'$, where $Y_{t+} = (y_{t+1}, \ldots, y_{t+p})'$ and $Y_{t-} = (y_{t-1}, \ldots, y_{t-p-1})'$. The conditional log-likelihood function (ignoring a constant) can be written as

$$L_n(\theta) = \sum_{t=p+1}^{n} \ell_t(\theta) \quad \text{and} \quad \ell_t(\theta) = -\ln \left( \beta' X_{t-1} \right) - \frac{(y_t - \alpha' Y_{t-1})^2}{2 \left( \beta' X_{t-1} \right)^2}. \quad (3.1)$$

Then the quasi-maximum likelihood estimator (QMLE) of $\theta_0$ can be defined as

$$\hat{\theta}_n = \arg\max_{\theta \in \Theta} L_n(\theta). \quad (3.2)$$

**Assumption 2.** $\{y_t : t \in \mathcal{N}\}$ is strictly stationary and ergodic with $E(|y_t|^\kappa) < \infty$ for some $\kappa > 0$.

**Assumption 3.** The density function of $\eta_t$ is continuous and positive everywhere on $\mathbb{R}$.

**Assumption 4.** The parameter space $\Theta$ is compact with $\underline{\omega} \leq \omega \leq \bar{\omega}$, $\underline{\beta} \leq \beta_{i-}, \beta_{i+} \leq \bar{\beta}$ for $i = 1, \ldots, p$, where $\underline{\omega}, \bar{\omega}, \underline{\beta}, \bar{\beta}$ are some positive constants. The true parameter vector $\theta_0$ is an interior point in $\Theta$. 

*J. Time Ser. Anal.* 43: 371–388 (2022) © 2021 John Wiley & Sons Ltd wileyonlinelibrary.com/journal/jtsa DOI: 10.1111/jtsa.12618
For the strict stationarity of \( \{y_t\} \) in Assumption 2, a sufficient condition is given in Theorem 1. Assumption 3 is imposed for identifying the unique maximizer of \( E[\ell_1(\theta)] \) at \( \theta_0 \); see also Francq and Zakoian (2012). Assumption 4 is required to ensure the log-likelihood function, score function and information matrix to be bounded without any moment restrictions on \( y_t \); see also Ling (2007). As a result, the model based on the QMLE can be applied to heavy-tailed data.

Let \( \kappa_1 = E(\eta_t^4) \) and \( \kappa_2 = E(\eta_t^6) - 1 \). Define the \((3p + 1) \times (3p + 1)\) matrices

\[
\Omega = E \left[ \frac{\partial \ell_i(\theta_0)}{\partial \theta} \frac{\partial \ell_i(\theta_0)}{\partial \theta'} \right] = E \left[ \begin{bmatrix} Y_{t-1}'Y_{t-1}' & \kappa_1 Y_{t-1}'X_{t-1}' & \kappa_2 X_{t-1}'X_{t-1}' \\ \kappa_1 X_{t-1}'Y_{t-1}' & \kappa_2 X_{t-1}'X_{t-1}' & \kappa_3 Y_{t-1}'X_{t-1}' \\ \kappa_2 X_{t-1}'X_{t-1}' & \kappa_3 Y_{t-1}'X_{t-1}' & \kappa_4 \end{bmatrix} \right],
\]

and

\[
\Sigma = -E \left[ \frac{\partial^2 \ell_i(\theta_0)}{\partial \theta \partial \theta'} \right] = \text{diag} \left\{ E \left[ \begin{bmatrix} Y_{t-1}'Y_{t-1}' & \kappa_1 Y_{t-1}'X_{t-1}' \\ \kappa_1 X_{t-1}'Y_{t-1}' & \kappa_2 X_{t-1}'X_{t-1}' \\ \kappa_2 X_{t-1}'X_{t-1}' & \kappa_3 Y_{t-1}'X_{t-1}' \end{bmatrix} \right], E \left[ X_{t-1}'X_{t-1}' \right] \right\}.
\]

**Theorem 2.** Suppose that Assumptions 2–4 hold. Then,

(i) \( \hat{\theta}_n \to_p \theta_0 \) as \( n \to \infty \);

(ii) furthermore, if \( E(\eta_t^4) < \infty \) and the matrix \( D = \begin{bmatrix} 1 & \kappa_1 \\ \kappa_1 & \kappa_2 \end{bmatrix} \) is positive definite, then \( \sqrt{n}(\hat{\theta}_n - \theta_0) \to \mathcal{L} N(0, \Sigma) \) as \( n \to \infty \), where \( \Sigma = \Sigma^{-1} \Omega \Sigma^{-1} \).

Note that the positive definiteness of \( D \) is satisfied for continuous \( \eta_t \) with \( E(\eta_t^4) < \infty \); see Jiang et al. (2020). If \( \eta_t \) is normal, then \( \kappa_1 = 0, \kappa_2 = 2 \) and \( \Omega = \Sigma \), thus the QMLE reduces to the MLE and its asymptotic normality in Theorem 2 can be simplified to \( \sqrt{n}(\hat{\theta}_n - \theta_0) \to \mathcal{L} N(0, \Sigma^{-1}) \) as \( n \to \infty \). To calculate the asymptotic covariance of \( \hat{\theta}_n \), we use sample averages to replace matrices \( \Omega \) and \( \Sigma \), and the QMLE \( \hat{\theta}_n \) to replace \( \theta_0 \).

### 3.2. Model Selection

This subsection considers the selection of order \( p \) for model (2.1) in practice. We first introduce the Bayesian information criterion (BIC) below to select the order \( p \),

\[
\text{BIC}_1(p) = -2L_n(\hat{\theta}_n^p) + (3p + 1) \ln(n - p),
\]

where \( \hat{\theta}_n^p \) is the QMLE when the order is set to \( p \), and \( L_n(\hat{\theta}_n^p) \) is the log-likelihood evaluated at \( \hat{\theta}_n^p \). However, model (2.1) is fitted by the Gaussian QMLE, hence the model misspecification should be considered in deriving the asymptotic expansion of the Bayesian principle, which leads to the modified BIC below

\[
\text{BIC}_2(p) = -2L_n(\hat{\theta}_n^p) + (3p + 1) \ln\left( \frac{n - p}{2\pi} \right) + \ln(\det(\hat{\Sigma}^p)),
\]

where \( \hat{\Sigma}^p \) is a consistent estimator of \( \Sigma \) defined as in (3.3) at order \( p \), and \( \det(\hat{\Sigma}^p) \) is its determinant. In practice, \( \hat{\Sigma}^p \) can be calculated with \( \theta_0 \) replaced by \( \hat{\theta}_n^p \) and the expectation approximated by the sample average. The modified BIC in (3.5) is adaptive to the generalized BIC proposed by Lv and Liu (2014), where the additional term of their generalized BIC is introduced to account for model misspecifications. For more details of the derivation of the above BICs, please refer to the Supporting information.
Let $\hat{p}_n = \arg\min_{1 \leq p \leq p_{\max}} \text{BIC}_i(p)$ for $i = 1$ and $2$, where $p_{\max}$ is a predetermined positive integer. Note that, when the sample size $n$ is sufficiently large, the additional terms $-(3p + 1) \ln(2\pi)$ and $\ln(\det(\hat{\Sigma}))$ can be ignored as they are $O(1)$. As a result, $\text{BIC}_1(p)$ and $\text{BIC}_2(p)$ are asymptotically equivalent in order selection. Simulation results in Section 6 indicate that the modified BIC in (3.5) performs better than the original BIC in (3.4) for moderate and small samples, although the two BICs have very similar performance for large samples. Hence, for moderate and small samples, we suggest to use (3.5). The following theorem verifies their selection consistency.

**Theorem 3.** Under the assumptions of Theorem 2, if $p_{\max} \geq p_0$, then as $n \to \infty$,

$$P(\hat{p}_n = p_0) \to 1 \quad \text{and} \quad P(\hat{p}_{2n} = p_0) \to 1,$$

where $p_0$ is the true order, and $p_{\max}$ is a predetermined positive integer.

4. TESTING FOR ASYMMETRY

This section studies the Wald, LM and quasi-likelihood ratio (QLR) tests to detect the asymmetry (leverage) effect of news on volatilities.

4.1. Asymmetry Tests

For model (2.1), the asymmetry testing is of the form

$$H_0 : \beta_{\theta+} = \beta_{\theta-}, \text{for all } i \quad \text{against} \quad H_1 : \beta_{\theta+} \neq \beta_{\theta-} \text{ for some } i,$$

where $i \in \{1, \ldots, p\}$. Let $R = (0_{m \times (p+1)}, I_p, -I_p)$ be the $p \times (3p + 1)$ matrix, where $0_{max}$ is the $m \times n$ zero matrix and $I_p$ is the $p \times p$ identity matrix. Then the null hypothesis can be represented as $H_0 : R\theta_0 = 0_p$, where $\theta_0$ is the true parameter vector and $0_p$ is a $p$-dimensional zero vector. Hence, the Wald, LM and QLR test statistics are defined as

$$W_n = n\hat{\theta}_n^\prime R(\hat{R}^\prime \hat{R})^{-1} \hat{R}^\prime \hat{\theta}_n,$$

$$L_n = \frac{1}{n} \left[ \frac{\partial L_n(\hat{\theta}_n)}{\partial \theta} \right] \Sigma^{-1} \hat{R}^\prime \hat{R}^{-1} \hat{R} \Sigma^{-1} \frac{\partial L_n(\hat{\theta}_n)}{\partial \theta},$$

$$Q_n = -2 \left[ L_n(\hat{\theta}_n) - L_n(0_p) \right],$$

respectively, where $\hat{\theta}_n$ is the restricted QMLE under $H_0$, while $\hat{\theta}_n$ is the unrestricted QMLE, $\hat{\Sigma}$ is the sample estimate of $\Sigma$ with $\theta_0$ estimated by $\hat{\theta}_n$, and the expectation replaced by sample average, while $\hat{\Sigma}$ (or $\hat{\Sigma}$) is the sample estimate of $\Sigma$ (or $\Sigma$) with $\theta_0$ estimated by $\hat{\theta}_n$ and the expectation replaced by sample average.

Let $\chi^2_v$ be the Chi-squared distribution with $v$ degrees of freedom. Define the $p \times p$ matrix $\Psi = \Delta^{-1/2} R\hat{\Sigma} \Delta^{-1/2}$ with $\Delta = R\Sigma^{-1} R^\prime$. For $j = 1, \ldots, p$, let $e_j$ be the eigenvalues of $\Psi$, and $x_j$ be the i.i.d. random variables following the $\chi^2_1$ distribution. The following theorem gives the limiting distributions of three test statistics under $H_0$.

**Theorem 4.** Suppose the assumptions of Theorem 2 hold. Then, under $H_0$, as $n \to \infty$,

(i) $W_n \rightarrow \chi^2_p$;  (ii) $L_n \rightarrow \chi^2_p$;  (iii) $Q_n \rightarrow \chi^2_p$.

where $Q = \sum_{j=1}^p e_j x_j$.

Theorem 4 shows that the limiting null distribution of $Q_n$ is not the usual $\chi^2_p$ distribution but a distribution of the weighted sum of i.i.d. $\chi^2_1$ random variables. In fact, $W_n$, $LM_n$ and $Q_n$ can be rewritten into forms of $z_i^2 \Delta \tau_i + o_p(1)$ for
Suppose that Assumptions 3–5 hold and Theorem 5.

Remark 3 (Calculation of p-values for the QLR test). First, calculate $\mu_Q = c_1$, $\sigma_Q = \sqrt{2c_1}$ and $l = c_k^2/c_1^2$, where $c_k = \sum_{j=1}^k c_j^2$ for $k = 1, 2, 3$. Then, the p-value of the QLR test is approximated by $P(X^2 > (Q_n - \mu_Q)\sqrt{2l/\sigma_Q + l})$, where $Q_n$ is the observed value of the QLR test statistic.

4.2. Power Analysis

We next discuss the efficiency of the proposed asymmetry tests through Pitman analysis. Note that $\theta_0 = (\alpha_0, \omega_0, \beta_{00}^-, \beta_{00}+)’$ under $H_0$. Denote $h = (h_a’, h_p’, h_+’, h_-’)’ \in \mathbb{R}^{3p+1}$, where $h_a = (h_1, \ldots, h_p, 1)’, h_+ = (h_1, \ldots, h_p)’, h_- = (h_1, \ldots, h_p)’$ and $h_+ \neq h_-$. Let $\theta_0 = \gamma_0 + h/\sqrt{n}$ such that $\theta_n \in \Theta$ for sufficiently large $n$. Consider the local alternatives, that is, for each $n$, the observed time series $\{y_{t+1,n}, \ldots, y_{n,n}\}$ are generated by

$$H_{ln} : y_{n,n} = \left(\alpha_0 + \frac{h_a}{\sqrt{n}}\right) Y_{t-1,n} + \eta_n \left(\beta_0 + \frac{h_0}{\sqrt{n}}\right)^’ X_{t-1,n},$$

where the subscript $n$ is used to emphasize the dependence of $y_{t,n}$ on $n$, $\eta_n$ is defined as in the model (2.1), $Y_{t,n} = (y_{t,n}, \ldots, y_{t-p+1,n})’, X_{t,n} = (1, Y_{t,n}’, -Y_{t,n}’)’$ with $Y_{t+1,n} = (y_{t+1,n}, \ldots, y_{t-p+1,n})’$ and $Y_{t,n} = (y_{t,n}, \ldots, y_{t-p+1,n})’$. $\{y_{t,n}\}$ satisfies the condition below.

Assumption 5. There exists a positive integer $n_0$ such that for $n \geq n_0$, $\{y_{t,n} : t \in \mathcal{N}\}$ is strictly stationary and geometrically ergodic with $E(|y_{t,n}|^r) < \infty$ for some $k > 0$.

Based on $\{y_{1,n}, \ldots, y_{n,n}\}$, the QMLE under $H_{ln}$ can be defined as

$$\hat{\theta}_{n,h} = \arg\max_{\theta \in \Theta} \sum_{t=p+1}^n \ell_{t,n}(\theta),$$

where

$$\ell_{t,n}(\theta) = -\ln \left(\beta’ X_{t-1,n}\right) - \frac{(y_{t,n} - \alpha’ Y_{t-1,n})^2}{2 \left(\beta’ X_{t-1,n}\right)^2}.$$

Denote $P_{n,h}$ as the law of $y_{t,n}$. The asymptotic distribution of $\hat{\theta}_{n,h}$ under sequences of local alternatives is given below.

Theorem 5. Suppose that Assumptions 3–5 hold and $E(\eta_n^4) < \infty$, then, under $P_{n,h}$, $\sqrt{n}(\hat{\theta}_{n,h} - \theta_0) \to \mathcal{L} \mathcal{N}(h, \Xi)$ as $n \to \infty$, where $\Xi$ is defined as in Theorem 2.

DOI: 10.1111/jtsa.12618
Theorem 5 verifies that model (4.3) is locally asymptotically normal (van der Vaart, 2000) at \( \theta_0 \). If \( \eta \) follows a normal distribution, we can show Theorem 5 by Le Cam’s third lemma. However, when \( \eta \) is not normal, the sequences \( \mathbb{P}_{n,0} \) and \( \mathbb{P}_{n,h} \) are not mutually contiguous; see Example 6.5 in van der Vaart (2000) for more discussions. Therefore, we show Theorem 5 in a direct way; see also Jiang et al. (2020).

Denote \( D = \langle R\Xi R' \rangle_1 / \Delta_1 \langle R\Xi R' \rangle_1 \), and define \( \Gamma \) as an orthogonal matrix such that \( \Gamma D \Gamma^* = \text{diag} \{ e_1^*, \ldots, e_p^* \} \), where \( e_j^* \)'s are eigenvalues of \( D \). Denote \( v = \Gamma \langle R\Xi R' \rangle^{-1//2} R \h_1 \in \mathbb{R}^p \), and let \( v_j \) be its jth component for \( j = 1, \ldots, p \). Let \( \chi_n^2(c) \) be the non-central Chi-squared distribution with degrees of freedom \( v \) and non-centrality parameter \( c \), and \( \chi_n^2(c) \) be its rth quantile.

**Theorem 6.** Suppose the assumptions of Theorem 5 hold and \( \theta_0 = 0 \). Then, under \( \mathbb{P}_{n,h} \), as \( n \to \infty \),

\[
(i) \ W_n \to \chi^2_p(\delta); \quad (ii) \ LM_n \to \chi^2_p(\delta); \quad (iii) \ Q_n \to \sum_{j=1}^p e^*_j x_{j,j}^2 ;
\]

where \( \delta = h' R \langle R\Xi R' \rangle^{-1} R \h_1 \), and \( x_{j,j} \)'s are independent random variables following the \( \chi^2(v_j) \) distribution for \( j = 1, \ldots, p \).

Theorem 6 obtains the asymptotic distributions of three test statistics under the local alternatives, which shows that the Wald and LM tests have the same local asymptotic powers. Under \( \mathbb{P}_{n,h} \), \( W_n \), \( LM_n \) and \( Q_n \) have the forms of \( z_{1,h} A_1 \h_1 + \alpha_p(1) \) for \( i = 1, 2, 3 \) respectively as \( n \to \infty \), where \( z_{1,h} = \eta_{3,h} = \sqrt{n} \bar{R} \hat{\theta}_{n,h}, \bar{z}_{2,h} = n^{-1//2} \bar{R} \Sigma^{-1} \partial L_{\hat{\theta}}(\hat{\theta}_{n,h}) / \partial \theta \), \( A_1 = A_2 = \langle R\Xi R' \rangle^{-1} \) and \( A_3 = \langle R\Sigma^{-1} R \rangle^{-1} \) are defined as previous. It can be verified that, under \( H_{1_n} \), we have \( z_{1,h} \to \epsilon \text{N}(R \h_1, R \langle R\Xi R' \rangle) \) for \( i = 1, 2, 3 \) as \( n \to \infty \). Therefore, \( W_n \) and \( LM_n \) have the same limiting non-central Chi-squared distribution under \( H_{1_n} \), whereas \( Q_n \) has a non-standard limiting distribution as \( \Xi \neq \Sigma \) for non-normal \( \eta \). If \( \eta \) is normally distributed, then \( \sum_{j=1}^p e^*_j x_{j,j}^2 \) reduces to a \( \chi^2(\delta) \) distribution, and the QLR test is as efficient as the Wald and LM tests. Moreover, note that \( P(\chi^2(\delta) \geq \chi^2_{1,1-r}(\delta)) = P(e^*_1 \chi^2(v_1) \geq e^*_1 \chi^2_{1,1-r}(v_1)) \) holds for \( p = 1 \), then it follows that the proposed three tests are equivalent in local asymptotic power when \( p = 1 \). For general cases of \( \eta \) with \( p > 1 \), it is difficult to compare the local asymptotic power of the QLR test with the other two tests. Alternatively, the simulation study in Section 6.3 compares the local power of all three tests in finite samples, and it is found that three tests perform very similarly when the sample size is as large as 2000.

5. MODEL CHECKING

To check adequacy of the fitted asymmetric linear DAR model, we construct a mixed portmanteau test (Wong and Ling, 2005) to detect misspecifications in the conditional mean and standard deviation. In the literature, diagnostic checking the conditional mean and standard deviation, can be conducted by checking the significant of sample autocorrelation functions (ACFs) of residuals and absolute residuals, respectively; see Zhu (2013). To check adequacy of the fitted asymmetric linear DAR model, we construct a mixed portmanteau test (Wong and Ling, 2005) to detect misspecifications in the conditional mean and standard deviation. In the literature, diagnostic checking the conditional mean and standard deviation, can be conducted by checking the significant of sample autocorrelation functions (ACFs) of residuals and absolute residuals, respectively; see Zhu (2013).

The ACFs of \( \{ \eta_t \} \) and \( \{ |\eta_t| \} \) at lag \( k \) can be defined by \( \rho_k = \text{cov}(\eta_t, \eta_{t-k}) / \text{var}(\eta_t) \) and \( \gamma_k = \text{cov}(|\eta_t|, |\eta_{t-k}|) / \text{var}(|\eta_t|) \) respectively. If the data generating process is correctly specified by model (2.1), then \( \{ \eta_t \} \) and \( \{ |\eta_t| \} \) are i.i.d. such that \( \rho_k = 0 \) and \( \gamma_k = 0 \) hold for any \( k \geq 1 \). For model (2.1) fitted by the QMLE, the corresponding residuals can be defined as \( \hat{\eta}_t = (\hat{\gamma}_t - \hat{\bar{\gamma}}_t \hat{Y}_{t-1}) / (\hat{\tilde{p}} \hat{X}_{t-1}) \), and then the residual ACF and absolute residual ACF at lag \( k \) can be calculated as

\[
\hat{\rho}_k = \frac{\sum_{t=p+k+1}^n (\hat{\eta}_t - \hat{\eta}_t)(\hat{\eta}_{t-k} - \hat{\eta}_k)}{\sum_{t=p+1}^n (\hat{\eta}_t - \hat{\eta}_t)^2} \quad \text{and} \quad \hat{\gamma}_k = \frac{\sum_{t=p+k+1}^n |\hat{\eta}_t| - \hat{\eta}_t)(|\hat{\eta}_{t-k}| - |\hat{\eta}_k|)}{\sum_{t=p+1}^n (|\hat{\eta}_t| - \hat{\eta}_t)^2}.
\]

respectively, where \( \hat{\eta}_t = (n-p)^{-1} \sum_{t=p+1}^n \hat{\eta}_t \) and \( \hat{\gamma}_t = (n-p)^{-1} \sum_{t=p+1}^n |\hat{\eta}_t| \). Clearly, \( \hat{\rho}_k \) (or \( \hat{\gamma}_k \)) is the sample version of \( \rho_k \) (or \( \gamma_k \)). Accordingly, if the value of \( \hat{\rho}_k \) (or \( \hat{\gamma}_k \)) deviates from zero significantly, it indicates that the conditional mean (or standard deviation) structure in model (2.1) is misspecified.
Let $\hat{\rho} = (\hat{\rho}_1, \ldots, \hat{\rho}_M)'$ and $\hat{\gamma} = (\hat{\gamma}_1, \ldots, \hat{\gamma}_M)'$, where $M$ is a predetermined positive integer. Denote $\tau_1 = E[\text{sgn}(\eta_1)]$ and $\tau_2 = E(|\eta_1|).$ Let $\xi_t = |\eta_t| - \tau_2$, then $E(\xi_t) = 0$ and $\sigma_\xi^2 = \text{var}(\xi_t) = 1 - \tau_2^2$. Define the $M \times (3p + 1)$ matrices $U_\rho = (U_{\rho 1}', \ldots, U_{\rho M}')'$ and $U_\gamma = (U_{\gamma 1}', \ldots, U_{\gamma M}')'$, where

$$U_{\rho k} = -\left( E\left( \frac{\eta_{t-k} Y_{t-1}}{\beta_0 X_{t-1}} \right), 0_{1 \times (2p+1)} \right)$$

and

$$U_{\gamma k} = -\left( \tau_1 E\left( \frac{\xi_{t-k} Y_{t-1}'}{\beta_0' X_{t-1}} \right), \tau_2 E\left( \frac{\xi_{t-k} X_{t-1}'}{\beta_0' X_{t-1}} \right) \right).$$

Denote the $2M \times (2M + 3p + 1)$ matrix below

$$V = \begin{pmatrix} I_M & 0_{M \times M} & U_{\rho} \\ 0_{M \times M} & I_M & U_{\gamma} \sigma_\xi^2 \end{pmatrix}.$$

Let $v_t = (\eta_{t-1}, \ldots, \eta_{t-M}, \xi_{t-1}' / \sigma_\xi, \ldots, \xi_{t-M-1}' / \sigma_\xi, \partial \gamma / \partial \theta, \partial \gamma / \partial \theta')'$, and $G = E(v, v')$.

**Theorem 7.** Suppose the assumptions of Theorem 2 hold. If model (2.1) is correctly specified, then $\sqrt{n}(\hat{\rho}', \hat{\gamma}') \rightarrow_d N(0, VGV')$ as $n \rightarrow \infty$.

Theorem 7 can be used to check the significance of $\hat{\rho}_k$ or $\hat{\gamma}_k$ individually. We can construct consistent estimators of $V$ and $G$ using sample averages and plug-in method, which are denoted by $\hat{G}$ and $\hat{V}$ respectively. Then we can approximate the asymptotic distribution in Theorem 7, and obtain confidence intervals for $\rho_k$ and $\gamma_k$.

To check the first $M$ lags jointly, we construct a portmanteau test statistic below

$$Q(M) = n \left( \frac{\hat{\rho}}{\hat{\gamma}} \right)' \left( \hat{V} \hat{G} \hat{V}' \right)^{-1} \left( \frac{\hat{\rho}}{\hat{\gamma}} \right). \quad (5.1)$$

Theorem 7 and the continuous mapping theorem imply that, $Q(M) \rightarrow_d \chi^2_{2M}$ as $n \rightarrow \infty$. Therefore, we reject the null hypothesis that $\rho_k$ and $\gamma_k$ $(1 \leq k \leq M)$ are jointly insignificant at level $\alpha$, if $Q(M)$ exceeds the $(1 - \alpha)$th quantile of $\chi^2_{2M}$ distribution. To fully detect short-term or long-term dependence among residuals and absolute residuals due to misspecification, it is suggested to consider multiple choices of $M$ such as 6, 12, 18 and so on for portmanteau tests; see detailed simulation findings in Section 6.4.

6. SIMULATION EXPERIMENTS

This section presents four simulation experiments to evaluate the finite-sample performance of the proposed QMLE, model selection method, three asymmetry tests and the mixed portmanteau test.

6.1. Model Estimation

The first experiment aims to examine the finite-sample performance of the QMLE $\hat{\theta}_a$, for which the data generating process is

$$\text{DGP1} : \gamma_t = 0.5y_{t-1} + \eta_t (0.4 + 0.4y_{t-1} - 0.6y_{t-1}^+),$$

where $\{\eta_t\}$ are standard normal, or follow standardized Student $t_d$ distribution with unit variance, or standardized skewed $t$ distribution, denoted by $st_{5-1.2}$, with unit variance and skew parameter $-1.2$ (Jiang et al., 2020).
sample size is set to \( n = 500, 1000 \) or 2000, with 1000 replications for each sample size. The projected Newton method (Bertsekas, 1982) is employed for solving the optimization (3.2) in each replication. Table I lists the biases, empirical standard deviations (ESDs) and asymptotic standard deviations (ASDs) of \( \hat{\theta}_n \) for different innovation distributions and sample sizes. As the sample size increases, most of the biases, ESDs and ASDs become smaller, and the ESDs get closer to the corresponding ASDs. Moreover, when the distribution of \( \eta_t \) gets more heavy-tailed or skewed, all ESDs and ASDs increase. This is as expected since either heavier tails or severer skewness of \( \{ \eta_t \} \) will lead to lower efficiency of the QMLE. In addition, we also consider other parameter settings for the data generating process, and the simulation findings are similar as previous.

6.2. Model Selection

In the second experiment, we evaluate the performance of the proposed model selection method in Section 3.2, and compare the BIC\(_1\) and its modified version BIC\(_2\) in finite samples. The data generating process is

\[
DGP2 : y_t = 0.3y_{t-1} - 0.2y_{t-2} + \eta_t (0.4 + 0.2y_{t-1}^+ + 0.2y_{t-2}^+ - 0.2y_{t-1}^- - 0.1y_{t-2}^-),
\]

where the innovations \( \{ \eta_t \} \) are defined as in the previous experiment. Three sample sizes, \( n = 200, 500 \) and 1000, are considered, and 1000 replications are generated for each sample size. The BIC\(_1\) in (3.4) and BIC\(_2\) in (3.5) are employed to select the order \( p \) with \( p_{\max} = 5 \). For \( i = 1 \) or 2, the cases of underfitting, correct selection and overfitting by BIC\(_i\) correspond to \( \hat{p}_i \) being 1, 2 and greater than 2 respectively.

Table II reports the percentages of underfitted, correctly selected and overfitted models by the two information criteria. The performance of both information criteria gets better when the sample size increases, while that

---

Table I. Biases (\( \times 10 \)), ESDs (\( \times 10 \)) and ASDs (\( \times 10 \)) of the QMLE \( \hat{\theta}_n \) when the innovations follow the standard normal, standardized Student \( t_5 \) or standardized skewed Student \( st_{5,-1.2} \) distribution

| \( n \) | N(0,1) Bias | ESD | ASD | \( t_5 \) Bias | ESD | ASD | \( st_{5,-1.2} \) Bias | ESD | ASD |
|---|---|---|---|---|---|---|---|---|---|
| \( \alpha_1 \) | 500 | -0.024 | 0.515 | 0.520 | -0.022 | 0.570 | 0.548 | -0.024 | 0.568 | 0.554 |
| | 1000 | 0.022 | 0.371 | 0.396 | -0.019 | 0.378 | 0.392 | -0.001 | 0.401 | 0.395 |
| | 2000 | 0.040 | 0.271 | 0.264 | 0.017 | 0.426 | 0.386 | 0.009 | 0.427 | 0.403 |
| \( \omega \) | 500 | 0.016 | 0.193 | 0.187 | 0.016 | 0.324 | 0.290 | 0.008 | 0.333 | 0.311 |
| | 1000 | 0.010 | 0.129 | 0.132 | -0.001 | 0.214 | 0.213 | -0.003 | 0.231 | 0.223 |
| | 2000 | 0.014 | 0.188 | 0.195 | -0.005 | 0.274 | 0.278 | -0.009 | 0.291 | 0.280 |
| \( \beta_{1+} \) | 500 | -0.088 | 0.640 | 0.587 | -0.013 | 1.049 | 0.957 | -0.039 | 1.191 | 1.031 |
| | 1000 | -0.051 | 0.418 | 0.416 | -0.006 | 0.819 | 0.724 | -0.022 | 0.860 | 0.709 |
| | 2000 | -0.020 | 0.295 | 0.295 | 0.001 | 0.565 | 0.534 | -0.008 | 0.619 | 0.570 |
| \( \beta_{1-} \) | 500 | -0.147 | 0.736 | 0.699 | -0.109 | 1.262 | 1.145 | -0.156 | 1.274 | 1.182 |
| | 1000 | -0.045 | 0.509 | 0.498 | -0.059 | 0.896 | 0.860 | -0.075 | 0.962 | 0.914 |
| | 2000 | -0.037 | 0.352 | 0.352 | -0.037 | 0.687 | 0.636 | -0.049 | 0.725 | 0.660 |

---

Table II. Percentages of underfitted, correctly selected and overfitted models by BIC\(_1\) and BIC\(_2\) when the innovations follow the standard normal, standardized Student \( t_5 \) or standardized skewed Student \( st_{5,-1.2} \) distribution

| \( n \) | BIC\(_1\) Under | Exact | Over | BIC\(_1\) Under | Exact | Over | BIC\(_2\) Under | Exact | Over |
|---|---|---|---|---|---|---|---|---|---|
| 200 | 0 | 49.9 | 50.1 | 0.3 | 55.2 | 44.5 | 0.6 | 53.6 | 45.8 |
| 300 | 0.1 | 94.7 | 5.2 | 2.1 | 91.4 | 6.5 | 2.2 | 91.1 | 6.7 |
| 1000 | 0 | 100 | 0 | 1.8 | 98.1 | 0.1 | 2.0 | 98.0 | 0 |
| BIC\(_2\) | 200 | 1.1 | 89.8 | 9.1 | 2.7 | 88.3 | 9.0 | 3.8 | 86.6 | 9.6 |
| 300 | 0.4 | 99.5 | 0.1 | 5.1 | 94.6 | 0.3 | 5.4 | 93.6 | 0.1 |
| 1000 | 0 | 100 | 0 | 4.4 | 95.6 | 0 | 6.4 | 93.6 | 0 |
becomes slightly worse as the distribution of \( \eta_t \) gets more heavy-tailed or more skewed. For the comparison between \( \text{BIC}_1 \) and \( \text{BIC}_2 \), it can be seen that the modified BIC (\( \text{BIC}_2 \)) selects the correct model in most of the replications when the sample size is as small as \( n = 200 \), while \( \text{BIC}_1 \) has comparable performance when the sample size is as large as \( n = 500 \). Overall, \( \text{BIC}_2 \) has better performance in model selection than that of \( \text{BIC}_1 \), especially for small and moderate samples. This indicates the necessity of the modified BIC in finite samples.

### 6.3. Asymmetry Tests

The third experiment examines the empirical size and power of the proposed asymmetry test statistics \( W_n, L_n \) and \( Q_n \). The data are generated from

\[
\text{DGP3} : y_t = 0.4 y_{t-1} + \eta_t [0.4 + 0.5 y_{t-1}^+ -(0.5+k)y_{t-1}^-],
\]

where \( k = h/\sqrt{n} \) with \( h \in \{-10, \ldots, -1, 0, 1, \ldots, 10\} \) and \( n \) being the sample size, and the innovations \( \{\eta_t\} \) are defined as in the first experiment. The null hypothesis of the asymmetry test is \( H_0 : k = 0 \), so that the case of \( k = 0 \) corresponds to the size of the tests, the cases of \( k \neq 0 \) correspond to the local power. Table III reports the empirical sizes of three tests at the significance level 5% with \( n = 500, 1000 \) and 2000. From this table, we can see that, all tests have accurate sizes when the sample size is large. In addition, \( W_n \) and \( Q_n \) are slightly oversized, especially when the sample size is small.

We next compare the local power of all three tests in finite samples at 5% significance level. Figure 2 shows the empirical power of three tests for \( n = 500 \) and 2000. We have the following findings. First, the local powers of \( W_n \) and \( Q_n \) are very similar, and they are slightly higher than that of \( L_n \) when the sample size is small (i.e. \( n = 500 \)), especially when \( \eta_t \) is not normal or \( |h| \) is not large. Second, the local power of three tests is close to each other when the sample size is large (i.e. \( n = 2000 \)), which is consistent to the theoretical comparison in Theorem 6 for \( p = 1 \). Finally, the local power for all three tests gets smaller as the innovations become more heavy-tailed or more skewed. In addition, we also conduct simulation studies for the data generating process with \( p > 1 \), the general findings are unchanged for the empirical size and power.

### 6.4. Portmanteau Test

In the fourth experiment, we study the proposed mixed portmanteau test \( Q(M) \). The data are generated from

\[
\text{DGP4} : y_t = 0.3 y_{t-1} + c_1 y_{t-2} + \eta_t (0.4 + 0.3 y_{t-1}^+ + c_2 y_{t-2}^+ - 0.4 y_{t-1}^- - c_2 y_{t-2}^-),
\]

where the innovations \( \{\eta_t\} \) are defined as in the first experiment. We fit an asymmetric linear DAR model with \( p = 1 \) using the same method as in Section 3.1, so that the case of \( c_1 = c_2 = 0 \) corresponds to the size of the test, the case of \( c_1 \neq 0 \) corresponds to misspecifications in the conditional mean, and the case of \( c_2 > 0 \) corresponds to misspecifications in the conditional standard deviation. Two departure levels, 0.1 and 0.3, are considered for all \( c_1 \) and \( c_2 \).

---

Table III. Empirical sizes of three tests \( W_n, L_n \) and \( Q_n \) at 5% significance level, where the innovations follow the standard normal, standardized Student \( t \) or standardized skewed Student \( \text{st}_{5,0.2} \) distribution.

| \( n \)  | \( W_n \) | \( L_n \) | \( Q_n \) | \( W_n \) | \( L_n \) | \( Q_n \) | \( W_n \) | \( L_n \) | \( Q_n \) |
|-------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| 500   | 0.062   | 0.054   | 0.061   | 0.065   | 0.038   | 0.063   | 0.061   | 0.043   | 0.060   |
| 1000  | 0.061   | 0.058   | 0.059   | 0.056   | 0.041   | 0.058   | 0.060   | 0.040   | 0.059   |
| 2000  | 0.048   | 0.047   | 0.047   | 0.052   | 0.047   | 0.053   | 0.053   | 0.048   | 0.054   |
Figure 2. Local power comparison at the 5% significance level. Upper: \( n = 500 \); bottom: \( n = 2000 \). Circle (\( \circ \)): \( W_n \); triangle (\( \triangle \)): \( L_n \); cross (\( + \)): \( Q_n \).

Table IV reports the rejection rates of \( Q(6) \) and \( Q(12) \) at 5\% significance level based on 1000 replications, for sample size \( n = 500, 1000 \) or 2000. We have the following findings. First, all sizes are close to the nominal level as the sample size \( n \) increases, and most powers improve as \( n \) or the departure level increases. Second, \( Q(6) \) (or \( Q(12) \)) is more powerful in detecting the misspecification in the conditional mean \( (c_1 \neq 0, c_2 = 0) \) than that in the conditional standard deviation \( (c_1 = 0, c_2 > 0) \). Finally, the performance of \( Q(6) \) (or \( Q(12) \)) gets worse as the innovation distribution becomes more heavy-tailed or more skewed. This finding seems to be consistent with the result in the first experiment that, as the innovation distribution becomes more heavy-tailed or skewed, the estimation performance for all parameters tends to worsen. We also consider other choices of \( M \) and another data generating process for the mixed portmanteau test \( Q(M) \). To save space, we relegate their results to the Supporting information. Note that the aforementioned findings still hold for these simulation settings; see the Supporting information for details.

7. AN EMPIRICAL EXAMPLE

We illustrate the proposed inference tools using the weekly closing prices of S&P500, denoted as \( p_t \), span from January 1998 to December 2020, with 1200 observations in total. The data is downloaded from the website of Yahoo Finance (https://hk.finance.yahoo.com). Let \( r_t = 100 \left( \ln p_t - \ln p_{t-1} \right) \) be the log returns in percentage, and denote \( y_t = r_t - n^{-1} \sum_{i=1}^{n} r_i \) as the centered log returns in percentage. The time plot of \( \{y_t\} \) in Figure 3 suggests evident volatility clustering. Table V lists summary statistics of \( \{y_t\} \), where the sample skewness \(-0.90\) indicates possible asymmetries in the volatility, and the sample kurtosis 7.17 implies heavy-tailedness of \( \{y_t\} \). Moreover, the ACFs and partial ACFs of \( \{y_t\} \) and \( \{|y_t|\} \) are significant at the first few lags, which suggests that the autocorrelation
Table IV. Rejection rates of the tests $Q(6)$ and $Q(12)$ at the 5% significance level, where the innovations follow the standard normal, standardized $t_5$ or standardized $st_5, -1.2$ distribution

| $c_1$ | $c_2$ | $N(0,1)$ | $t_5$ | $st_{5,-1.2}$ |
|------|------|------|------|--------------|
|      |      | 500  | 1000 | 2000         |
|      |      |      |      |              |
| 0.0  | 0.0  | 0.059| 0.057| 0.050        |
| 0.1  | 0.0  | 0.049| 0.050| 0.057        |
| 0.3  | 0.0  | 0.097| 1.000| 1.000        |
| 0.0  | 0.1  | 0.096| 0.196| 0.455        |
| 0.0  | 0.3  | 0.683| 0.989| 1.000        |
|      |      | 500  | 1000 | 2000         |
|      |      |      |      |              |
| 0.0  | 0.0  | 0.059| 0.057| 0.050        |
| 0.1  | 0.0  | 0.228| 0.469| 0.855        |
| 0.3  | 0.0  | 0.997| 1.000| 1.000        |
| 0.0  | 0.1  | 0.096| 0.196| 0.455        |
| 0.0  | 0.3  | 0.683| 0.989| 1.000        |

Figure 3. Time plot for centered weekly log returns in percentage (black line) of S&P500 index from January 1998 to December 2021, with 1-week negative VaR forecasts at the level of 5% (red line) from January 2008 to December 2020

Table V. Summary statistics for $\{y_t\}$

|        | Mean | Median | Std.Dev. | Skewness | Kurtosis | Min     | Max     |
|--------|------|--------|----------|----------|----------|---------|---------|
| $y_t$  | 0.00 | 0.12   | 2.53     | -0.90    | 7.17     | -20.20  | 11.30   |

coexists with the conditional heteroscedasticity in $\{y_t\}$. The above findings motivate us to investigate $\{y_t\}$ using our proposed model and inference tools.

Based on $p_{\text{max}} = 20$, the proposed BIC$_1$ and BIC$_2$ both select $p = 4$ with BIC$_1(4) = 3055.329$ and BIC$_2(4) = 3025.046$. By the quasi-maximum likelihood estimation method in Section 3.1, the fitted model is

$$y_t = -0.0800_{0.033}y_{t-1} + 0.0340_{0.030}y_{t-2} + 0.0030_{0.032}y_{t-3} - 0.0140_{0.031}y_{t-4} + 0.9880_{0.096} + 0.0440_{0.050}y_{t-1}^+ - 0.4150_{0.063}y_{t-1}^- + 0.0010_{0.047}y_{t-2}^+ - 0.2480_{0.056}y_{t-2}^- + 0.1600_{0.050}y_{t-3}^+ - 0.2860_{0.057}y_{t-3}^- + 0.1510_{0.047}y_{t-4}^+ - 0.1890_{0.054}y_{t-4}^-,$$ (7.1)
where the subscripts are the standard errors of the estimated coefficients. It can be seen that the coefficients of $\gamma_{t-i}$ and $\gamma_{t-i}'$ are clearly different for $i = 1, 2, 3$, which suggests that there may be asymmetric effects in the conditional volatility of $y_t$. The Wald, LM and QLR tests in Section 4.1 are conducted for model (7.1) and all their $p$-values are less than 0.001, which corroborates the asymmetric effects in the volatility of $y_t$. To check the adequacy of the fitted model (7.1), we perform the mixed portmanteau test $Q(M)$ in Section 5 for $M = 6, 12$ and 18. The $p$-values of portmanteau tests are 0.41, 0.18 and 0.27, respectively, which suggests that the fitted model is adequate. In addition, as shown in Figure 4, most of the residual ACFs $\hat{\rho}_k$ and absolute residual ACFs $\hat{\gamma}_k$ fall within their corresponding 95% confidence bounds at the first 18 lags.

Since Value-at-Risk (VaR) is an important risk measure for financial assets, we use the fitted model to forecast the conditional quantile of $y_t$, that is, the negative VaR. To examine the forecasting performance, we conduct one-step-ahead predictions using a rolling forecasting procedure with a fixed moving window covering 10 years’ data points of size 522. Specifically, we fit an asymmetric linear DAR model of order four (ALDAR) for each moving window, and compute the forecast of the $\tau$th conditional quantile of $y_{t+1}$, given by $Q(y_{t+1} \mid \mathcal{F}_t) = \hat{\mu}_{t+1} + \hat{\sigma}_{t+1} \hat{\beta}_\tau$, where $\hat{\mu}_{t+1}$ and $\hat{\sigma}_{t+1}$ are the predicted conditional mean and standard deviation, respectively, and $\hat{\beta}_\tau$ is the $\tau$th sample quantile of residuals $\{\hat{\eta}_1, \ldots, \hat{\eta}_t\}$. Then we move the window forward by one and repeat the above procedure until all data are used. Finally, we obtain 677 one-week-ahead negative VaRs for each $\tau$. For illustration, the rolling forecasts at $\tau = 5\%$ are displayed in Figure 3, which indicates that the negative VaRs change accordingly to the volatility of the data.

To compare the forecasting performance of the proposed model with other counterparts, we also perform the rolling forecasting procedure using a linear DAR(4) model (LDAR) and an AR(4) model with the threshold GARCH(1, 1) errors (AR-TGARCH). Note that the AR-TGARCH model can depict the asymmetric effect in volatilities, while the LDAR model ignores the asymmetric effect. For comparison, all these models are fitted by the QMLE, and their VaR forecasts are computed in the same way as for the ALDAR model. To evaluate the forecasting performance of each model, we calculate the empirical coverage rate (ECR), and perform VaR backtests for the VaR forecasts at $\tau = 1\%, 5\%, 95\%$ and 99%. Specifically, ECR is calculated as the proportion of observations that fall below the corresponding conditional quantile forecast for the last 677 data points. Two VaR backtests, that is, the likelihood ratio test for correct conditional coverage (CC) in Christoffersen (1998) and the dynamic quantile (DQ) test in Engle and Manganelli (2004) are employed. Denote the hit by $H_t = I(y_t < Q_{\gamma}(\tau \mid \mathcal{F}_{t-1}))$. The null hypothesis of CC test is that, conditional on $\mathcal{F}_{t-1}$, $\{H_t\}$ are i.i.d. Bernoulli random variables with success probability being $\tau$. For the DQ test, following Engle and Manganelli (2004), we regress $H_t$ on regressors including a constant, four lagged hits $H_{t-i}, i = 1, 2, 3, 4$, and the contemporaneous VaR forecast. The null hypothesis of DQ
Table VI. Empirical coverage rate (%) and p-values of two VaR backtests of three models at the 1%, 5%, 95% and 99% conditional quantiles. M1, M2 and M3 represent the ALDAR, LDAR and AR-TGARCH models respectively.

|       | $\tau = 1\%$ |       | $\tau = 5\%$ |       | $\tau = 95\%$ |       | $\tau = 99\%$ |
|-------|---------------|-------|---------------|-------|---------------|-------|---------------|
|       | ECR | CC | DQ | ECR | CC | DQ | ECR | CC | DQ | ECR | CC | DQ |
| M1    | 1.48 | 0.16 | 0.10 | 5.76 | 0.58 | 0.89 | 94.68 | 0.93 | 0.98 | 98.82 | 0.82 | 0.96 |
| M2    | 3.54 | 0.00 | 0.00 | 10.04 | 0.00 | 0.00 | 88.18 | 0.00 | 0.00 | 93.80 | 0.00 | 0.00 |
| M3    | 1.33 | 0.19 | 0.06 | 5.02 | 0.26 | 0.33 | 94.68 | 0.35 | 0.06 | 98.22 | 0.15 | 0.02 |

In addition, based on Theorem 2 we can employ the large-sample z test to check the significance of each parameter in the conditional mean of model (7.1). Denote $s.e. (\hat{\alpha}_i)$ as the standard error of $\hat{\alpha}_i$ and $\Phi(\cdot)$ as the distribution function of a standard normal random variable. If the test statistic $z_i = \frac{\hat{\alpha}_i}{s.e. (\hat{\alpha}_i)}$ satisfies $|z_i| \geq \Phi^{-1}(1 - \tau/2)$ or the p-value $2(1 - \Phi(|z_i|)) \leq \tau$ for a certain $i \in \{1, \ldots, p\}$, then the corresponding parameter $\alpha_i$ is significant (i.e. $\alpha_i \neq 0$) at the significance level of $\tau$. As a result, at 5% significance level $\alpha_{i0}$ is significant while $\alpha_{i1}$'s are insignificant for $i = 2, 3$ and 4. This motivates us to consider a reduced asymmetric linear DAR model of order (1, 4) denoted by ALDAR(1, 4) as in Remark 2. We then refit the data using the reduced ALDAR(1, 4) model, and compare its estimation and forecasting performance with the fitted ALDAR(4, 4) model at (7.1). We find that the coefficient estimates and corresponding standard errors of their common parameters only change slightly, and their forecasting performance is very close to each other; see the Supporting information for details.

8. CONCLUSION AND DISCUSSION

This article proposes the asymmetric linear double AR model which takes into account asymmetric effects for conditional heteroscedastic time series in the presence of a conditional mean structure. The strict stationarity of the new model is derived, and inference tools, including a Gaussian QMLE for estimation and a mixed portmanteau test for diagnosis, are constructed without any moment condition on the data. Based on the QMLE, a BIC and its modified version are proposed for order selection, and simulation results suggest that the modified BIC performs better in small and moderate samples. The Wald, LM and quasi-likelihood ratio test statistics are constructed to detect asymmetric effects, and it is shown that the Wald and LM tests are asymptotically equivalent in size and power, while the asymptotics of the quasi-likelihood ratio test become non-standard. The usefulness of the new model is confirmed by our empirical evidence, especially when the data are characterized by skewness and heavy-tailedness which are very common features for financial time series.

The study in this article can be extended in several directions. First, our model can be extended to allow for asymmetric effects in both the conditional mean and the standard deviation, then the proposed symmetry tests could...
adapt to detect the asymmetry from the conditional location and scale separately or jointly. Second, since financial time series can be heavy-tailed such that $E(\eta_t^4) = \infty$, it is also of interest to consider more robust estimation methods than Gaussian QMLE, for example, the quasi-maximum exponential likelihood estimation of Zhu and Ling (2011). Third, the joint modeling of conditional mean and volatility in the presence of asymmetric effects for univariate case can be generalized to multi-variate case. As a result, a vector asymmetric LDAR model is a natural extension and the related inference tools are worth to investigate. We leave these extensions for future research.

ACKNOWLEDGEMENTS

We are deeply grateful to the editor, the co-editor, and two anonymous referees for their valuable comments and suggestions, which led to substantial improvement of this article. Zhu’s research was supported by an NSFC grant 12001355, Shanghai Pujiang Program 2019PJ051 and Chenguang Program 19CG44 supported by Shanghai Education Development Foundation and Shanghai Municipal Education Commission.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in (Yahoo Finance) at (https://hk.finance.yahoo.com).

SUPPORTING INFORMATION

The Supporting information includes technical details for Theorems 1–7 and additional results for simulation and empirical analysis. Additional Supporting Information may be found online in the supporting information tab for this article.

REFERENCES

Bertsekas DP. 1982. Projected newton methods for optimization problems with simple constraints. *SIAM Journal on Control and Optimization* **20**: 221–246.

Black F. 1976. Studies of stock market volatility changes. In *1976 Proceedings of the American Statistical Association Business and Economic Statistics Section*.

Bollerslev T. 1986. Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics* **31**: 307–327.

Christoffersen PF. 1998. Evaluating interval forecasts. *International Economic Review* **39**: 841–862.

Cline DB, Pu H. 2004. Stability and the Lyapunov exponent of threshold AR-ARCH models. *The Annals of Applied Probability* **14**: 1920–1949.

Engle RF. 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* **50**: 987–1007.

Engle RF, Manganelli S. 2004. CAViaR: conditional autoregressive value at risk by regression quantiles. *Journal of Business and Economic Statistics* **22**: 367–381.

Engle RF, Ng VK. 1993. Measuring and testing the impact of news on volatility. *The Journal of Finance* **48**: 1749–1778.

Francq C, Zakoian JM. 2004. Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli* **10**: 605–637.

Francq C, Zakoian JM. 2004. QMLE estimation of a class of multivariate asymmetric GARCH models. *Econometric Theory* **28**: 179–206.

Francq C, Zakoian JM. 2013. Inference in nonstationary asymmetric GARCH models. *The Annals of Statistics* **41**: 1970–1998.

Francq C, Zakoian JM. 2019. *GARCH Models: Structure, Statistical Inference and Financial Applications*, 2nd ed. New York: Wiley.

Imhof JP. 1961. Computing the distribution of quadratic forms in normal variables. *Biometrika* **48**: 419–426.

Jiang F, Li D, Zhu K. 2020. Non-standard inference for augmented double autoregressive models with null volatility coefficients. *Journal of Econometrics* **215**: 165–183.

Kuester K, Mittnik S, Paolella M. 2006. Value-at-risk prediction: a comparison of alternative strategies. *Journal of Financial Econometrics* **4**: 53–89.

Li D, Ling S, Zhang R. 2016. On a threshold double autoregressive model. *Journal of Business and Economic Statistics* **34**: 68–80.
Li G, Li WK. 2005. Diagnostic checking for time series models with conditional heteroscedasticity estimated by the least absolute deviation approach. *Biometrika* **92**: 691–701.
Li G, Li WK. 2008. Testing for threshold moving average with conditional heteroscedasticity. *Statistica Sinica* **18**: 647–665.
Li G, Zhu Q, Liu Z, Li WK. 2017. On mixture double autoregressive time series models. *Journal of Business and Economic Statistics* **35**: 306–317.
Li WK, Ling S, McAleer M. 2002. Recent theoretical results for time series models with GARCH errors. *Journal of Economic Surveys* **16**: 245–269.
Ling S. 2005. Self-weighted least absolute deviation estimation for infinite variance autoregressive models. *Journal of the Royal Statistical Society: Series B* **67**: 381–393.
Ling S. 2007. A double AR(p) model: structure and estimation. *Statistica Sinica* **17**: 161–175.
Linton O, Mammen E. 2005. Estimating semiparametric ARCH (∞) models by kernel smoothing methods. *Econometrica* **73**: 771–836.
Liu H, Tan S, Zhu Q. 2020. Quasi-maximum likelihood inference for linear double autoregressive models, arXiv:2010.06103.
Liu H, Tang Y, Zhang HH. 2009. A new Chi-square approximation to the distribution of non-negative definite quadratic forms in non-central normal variables. *Computational Statistics and Data Analysis* **53**: 853–856.
Ljung GM, Box GEP. 1978. On a measure of lack of fit in time series models. *Biometrika* **65**: 297–303.
Lv J, Liu JS. 2014. Model selection principles in misspecified models. *Journal of the Royal Statistical Society: Series B* **76**: 141–167.
MaCurdy TE. 1981. *Asymptotic properties of quasi-maximum likelihood estimators and test statistics*, NBER Technical Working Paper.
Nelson DB. 1991. Conditional heteroskedasticity in asset returns: a new approach. *Econometrica* **59**: 347–370.
Pan J, Wang H, Tong H. 2008. Estimation and tests for power-transformed and threshold GARCH models. *Journal of Econometrics* **142**: 352–378.
Pearson ES. 1959. Note on an approximation to the distribution of non-central $\chi^2$. *Biometrika* **46**: 364.
Rabemananjara R, Zakoian JM. 1993. Threshold ARCH models and asymmetries in volatility. *Journal of Applied Econometrics* **8**: 31–49.
Taylor SJ. 2008. *Modelling Financial Time Series*. New York: World Scientific.
Tweedie RL. 1983. Criteria for rates of convergence of Markov chains, with application to queueing and storage theory. In *Probability Statistics and Analysis*, Kingman JFC., Reuter GEH. (eds.) Cambridge: Cambridge University Press.
van der Vaart AW. 2000. *Asymptotic Statistics*. Cambridge: Cambridge University Press.
Wong H, Ling S. 2005. Mixed portmanteau tests for time-series models. *Journal of Time Series Analysis* **26**: 569–579.
Zakoian JM. 1994. Threshold heteroskedastic models. *Journal of Economic Dynamics and Control* **18**: 931–955.
Zhu K. 2013. A mixed portmanteau test for ARMA-GARCH models by the quasi-maximum exponential likelihood estimation approach. *Journal of Time Series Analysis* **34**: 230–237.
Zhu K, Ling S. 2011. Global self-weighted and local quasi-maximum exponential likelihood estimators for ARMA-GARCH/IGARCH models. *The Annals of Statistics* **39**: 2131–2163.
Zhu Q, Zheng Y, Li G. 2018. Linear double autoregression. *Journal of Econometrics* **207**: 162–174.