Electronic Journal of Statistics
Vol. 18 (2024) 1970–2020
ISSN: 1935-7524
https://doi.org/10.1214/24-EJS2244

Robust inference in AR-G/GARCH models under model uncertainty

Shang-Yuan Shiu

Department of Mathematics, National Central University, Taoyuan City 320317, Taiwan
e-mail: shiu@math.ncu.edu.tw

and

Hsin-Chieh Wong*

Department of Statistics & Fintech and Green Finance Center, National Taipei University, New Taipei City 237303, Taiwan
e-mail: hcwong@gm.ntpu.edu.tw

Abstract: This paper provides a robust test for a function of the autoregressive parameters in AR models driven by G/GARCH noise under model uncertainty in an asymptotic framework. To address this method, we adopt the model average and choose weights based on our Mallows-type methods. We next present a valid confidence interval by dividing the sample into a fixed number of groups to form a normalized estimator which is asymptotically related to the Student's t-distribution. We derive asymptotic results that are not only interesting in their own right, but contribute to the theoretical foundations. These results include limiting distributions of the proposed Mallows-type model averaging and selection estimators. The proposed averaging estimators are stable-family distributions and are yet to be precisely characterized; hence they cannot be implemented by simulation. Through simulation experiments, our method yields outstanding numerical performance, especially for testing the quotient of coefficients in finite-sample tests.

MSC2020 subject classifications: Primary 62M10, 62F03; secondary 62F12.
Keywords and phrases: AR-G/GARCH, model averaging, model selection, heavy tails, tail behavior, stable distribution, Mallows-type criteria.

Received June 2022.

1. Introduction

Since the 1960s, empirical evidence has led many to reject the normal assumption in favor of heavy-tailed alternatives. This suggests that the normal assumption might not be suitable in some real applications; see, e.g., [20, Chapter 6], [6], and [55], among others. Many studies show that the *heavy-tailed noise* assumption better fits empirical data, which suggests that heavy-tailed distributions

*Corresponding author.

are necessary to model certain economic variables and stock price changes. Interested readers may refer to [23], [53], [54], [18], [9], [8], [2], [59], [26], [7], and [76].

To make the idea of heavy-tailed distributions concrete, one considers that noise is generated by the generalized autoregressive conditional heteroskedasticity (GARCH) model, as proposed in the seminal works of [21] and [4]. This model captures heavy-tailed features, and its sample autocorrelation function is not $\sqrt{T}$-consistent, and hence not asymptotically normal, where T denotes the sample size. However, how the GARCH parameters affect the asymptotic behavior of the estimator is far from clear, and only a few references can be found, e.g., in [57], [46], [80], and the references therein. In practice, GARCH-type noise has attracted a growing strand of literature as noise added to the autoregressive moving average (ARMA) model; we term this combined structure the ARMA-GARCH model. For example, [4] uses an $AR(4)$-$GARCH(1,1)$ model to study the US gross national product (GNP) series. [25], [73], [79], and [42] also use GARCH-type noise; these models include the $AR(1)$-$GJR(1,1)$ model, where the GJR model is introduced by [28]; several stock market index models $MA(3)$-$GARCH(1,1)$; and the AR-GARCH model on the return rate of the simulated Dow Jones Industrial Average, to name a few. In the literature, there are basically two types of GARCH models: the polynomial GARCH and the exponential GARCH. These can be unified as the augmented GARCH model proposed by [17], in which the general type of polynomial model is called the general GARCH model proposed by [36]. Along that line, the seminal paper of [1] presents a necessary and sufficient condition for the stationary solution of the augmented GARCH model.

In ARMA models, setting up dependency within these time intervals (or sequences) is one of the main topics in the model selection procedure. As the name suggests, model selection is choosing a model from a set of candidate models. A typical way to select a model appeals to information criteria, for example, the Akaike information criterion (AIC), the Bayesian information criterion (BIC), and the Hannan–Quinn (HQ) principle. In the past few decades, the trend has moved from model selection to model averaging for model uncertainty. Unlike model selection, which picks a single model from all candidate models, in model averaging all available information is incorporated by averaging over all potential models. The main problem in model averaging is hence deciding the weights for all basic models.

Mallows model averaging (MMA), proposed by [30], is a common technique in model averaging if compared to the Bayesian model averaging approach in [41] or [65]. [30] demonstrates via simulations that the MMA estimator outperforms AIC and BIC model selection methods and other averaging methods in the sense of less expected squared error. The approach of [30] is an ordinary least squares (OLS) based model averaging estimator with the weights selected by minimizing a criterion inspired by [51]. [31] also considers the asymptotic properties of a least-squares forecast averaging method based on the MMA criteria for stationary time-series observations. Since then, the MMA methodology has been widely applied to other regression models. For example, [75] investigate nonnested models, [35] study a jackknife averaging approach under heteroskedastic error settings, [11] discuss factor augmented regression models, and [77] provide the MMA criterion for the MIxed DAta Sampling (MIDAS) model. Other work can be found on the time-series noise framework, e.g., [81] and [12].

Another central question is how to form inferences based on model selection or model averaging estimators. One way to address the inference problem is to provide asymptotic distributions of the estimators for all candidate models. For related work, see [50], [64], and [32, 33, 34], among others. These studies focus on a two-model case under the homoskedastic framework. A result of [49] then permits applications to cross-section, panel, and time-series data. However, few attempts have been made at time-series error by using model averaging.

An AR model driven by GARCH-type noise has been considered an empirical setup, and therefore is an important model. In this paper we address the inference problem based on the AR models driven by general GARCH noise (AR-G/GARCH models, for short) under model uncertainty. Based on the ideas inspired by [43] and [62], we propose a robust test for a function of the autoregressive parameters in AR-augmented GARCH models under model uncertainty. Under the asymptotic distributions presented by [46] and [80], [62] verifies that the sufficient conditions of [43] hold based on an AR model driven by general GARCH noise (AR-G/GARCH model, for short). However, [46], [80], and [62] do not consider model uncertainty.

For asymptotic inference without model uncertainty, [46] and [80] provide asymptotic distributions of autoregressive parameters for the OLS estimator. In finite samples under model uncertainty, adding regressors reduces model bias but increases its variance. This implies a trade-off between bias and variance in a finite-sample setting if one desires an approximation. To overcome this obstacle, interested readers may refer to [39] and [13], who discuss the asymptotic distribution in a local asymptotic framework, where the scaling of convergence in the related literature is $T^{1/2}$. The realism of the local asymptotic framework can be found in [40] and [66]. Of course, the scaling differs from ours since we consider heavy-tailed noise. We determine that the tail behavior of noise strongly impacts the scaling of convergence; not surprisingly, such scaling is upon our local asymptotic framework, which exhibits four different levels as suggested in asymptotic results by [80].

Another challenge is that the asymptotic distribution of the OLS estimator for the autoregressive parameters is not asymptotically pivotal. This rules out a traditional method as confidence intervals are constructed by inverting the t-statistic of all candidate models, which ought to lead to a distorted inference. Here we combine these two approaches to solve the issues mentioned above and in the previous paragraph. First, we show that OLS-based averaging estimators with fixed weights are asymptotic stable-family distributions in a local asymptotic framework and then we present a model averaging estimator based on the Mallows model averaging proposed by [30]; see Section 4 for details. Second, in light of [43] and [62], we provide a robust test of the autoregressive parameters by normalization of subsamples obtained by dividing the original sample into a fixed number of groups related to the Student's t-statistic.

We summarize our contributions as follows. Based on the ideas inspired by [43] and [62], we propose a robust test for AR-augmented GARCH models, and perform inference using models with heavy-tailed noise as the presence of conditional heterogeneity under model uncertainty. We also present the limiting distribution of OLS-based estimators for the submodels of an AR-G/GARCH model under the local asymptotic assumption with different scaling and then derive the limiting distribution of the estimator of the averaging model with fixed weights. Finally, we offer a model averaging estimator based on the Mallows model averaging ([30]) which we here term "Mallows-type model averaging."

This paper is organized in the following order: We present a robust test in Section 2. Section 3 contains the theoretical foundations needed to develop our method. The proposed model averaging estimator and asymptotic results are presented in Section 4, and in Section 5, we conduct simulations to demonstrate the feasibility of our method for finite samples. We conclude in Section 6. All technical proofs are deferred to the Appendix.

2. Robust inference

Consider the following $AR(p+q)$ model:

$$y_t = \rho_1 y_{t-1} + \cdots + \rho_p y_{t-p} + \gamma_1 y_{t-p-1} + \cdots + \gamma_q y_{t-p-q} + \epsilon_t,$$

where $\{\epsilon_t\}$ is a sequence of yet unspecified random variables. Let

$$\bar{y}_t = (y_{t-1}, \ldots, y_{t-p}, y_{t-p-1}, \ldots, y_{t-(p+q)})', \tag{1}$$
$$\bar{y}_{t,p} = (y_{t-1}, \ldots, y_{t-p})', \tag{2}$$
$$\bar{y}_{t-p,q} = (y_{t-p-1}, \ldots, y_{t-(p+q)})', \tag{3}$$

which are $(p+q) \times 1$, $p \times 1$, and $q \times 1$ matrices (or column vectors) of regressors, respectively. We use the prime ($'$) to denote the transpose of a matrix or vector and assume that all vectors are column vectors throughout.

We call $\bar{y}_{t,p}$ the core lag period regressors; they are necessarily included. $\bar{y}_{t-p,q}$ are additional lag period regressors that are optional to the model. Many references indicate that several financial market indexes fit the AR-GARCH model well; see, e.g., [42]. In empirical study, statisticians or econometricians test models for stationarity by applying a unit-root test, e.g., an augmented Dickey-Fuller (ADF) test. This ADF formulation allows higher-period autoregressive processes, and the lag period p must be determined when applying the test. When the above unit root hypothesis is rejected, the next question is to estimate parameters to perform statistical inference. For higher periods from $p+1$ to $p+q$, one may consider a more general stationary $AR(r)$ model, where $r > p$. Since submodels cause bias, the determination of the lag period is a long-standing and significant problem. One way to solve this is to take the average of the submodels; details are in Section 2.2.

Denote $\boldsymbol{\vartheta} = (\boldsymbol{\rho}', \boldsymbol{\gamma}')'$ to be the $(p+q) \times 1$ matrix of the regression coefficients where $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_p)'$ and $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_q)'$. Furthermore, we consider that parameter $\mu(\boldsymbol{\vartheta})$ is a smooth real-valued function.[1]

The purpose of this section is to provide a hypothesis test on $\boldsymbol{\vartheta}$, or more generally $\mu(\boldsymbol{\vartheta})$, instead of model fitting,

$$H_0 : \mu(\boldsymbol{\vartheta}) = \mu(\boldsymbol{\vartheta}_0)$$

against $H_1 : \mu(\boldsymbol{\vartheta}) \neq \mu(\boldsymbol{\vartheta}_0)$. Moreover, we will show that this hypothesis test is robust.

The outline of this section is the following: in Section 2.1 we introduce the AR $(p+q)$-$G/GARCH(1,1)$ model and the local-to-zero assumption. Based on the above setup, we consider model averaging in Section 2.2 and use the Mallows-type criterion to determine weights. Finally, we apply [62]'s theorem to verify that our method is robust.

2.1. Model setup

We consider the following AR $(p+q)$-$G/GARCH(1,1)$ model:

$$y_t = \rho_1 y_{t-1} + \cdots + \rho_p y_{t-p} + \gamma_1 y_{t-p-1} + \cdots + \gamma_q y_{t-p-q} + \epsilon_t, \tag{4}$$

$$\epsilon_t = h_t z_t, \text{ and } \Lambda(h_t^2) = g(z_{t-1}) + c(z_{t-1})\Lambda(h_{t-1}^2), \tag{5}$$

where $-\infty < t < \infty$, $\Lambda(x) = x^{\delta/2}$ for $\delta > 0$, $g(\cdot)$ and $c(\cdot)$ are real-valued functions such that $P[\Lambda(h_t^2) > 0] = 1$ and $c(0) < 1$, and $\{z_t\}_{t=\infty}^{\infty}$ is a sequence of independent and identically distributed (*i.i.d.*) symmetric noise. Model (5) is called the general $GARCH(1,1)$ (or $G/GARCH$ (1,1)) process proposed by [36]. Many well-known models are special cases of this model, e.g., the $GARCH(1,1)$ in [4], the absolute value $GARCH(1,1)$ model in [74] and [71], the nonlinear $GARCH(1,1)$ model in [22], the volatility switching $GARCH(1,1)$ model in [24], the threshold $GARCH(1,1)$ model in [78], and the generalized quadratic $ARCH(1,1)$ model in [72].

We could relax assumption $\Lambda(x) = x^{\delta/2}$ by a more general one: the existence of Λ^{-1}. Under this setup, model (5) is called an augmented $GARCH(1,1)$ process, as proposed by [17]. This model includes the exponential GARCH model, see, e.g., the multiplicative $GARCH(1,1)$ model in [27] and the exponential $GARCH(1,1)$ (or $EGARCH(1,1)$) model in [60].

We seek to find conditions sufficient to make the process $\{y_t\}$ stationary. This can be done by the following steps.

We say that $\{y_t\}$ is causal if there exists a sequence $\{\phi_j\}$ with $\sum_{j=0}^{\infty} |\phi_j| < \infty$ such that

$$y_t = \sum_{j=0}^{\infty} \phi_j \epsilon_{t-j}, \; -\infty < t < \infty. \tag{6}$$

[1] For example, $\mu(\boldsymbol{\vartheta})$ is an individual parameter or a ratio of two parameters of regressors. To form a long-run impact, one may consider $\mu(\boldsymbol{\vartheta})$ the summation of all parameters.

This property together with the stationarity of $\{\epsilon_t\}$—which we will show later—imply that $\{y_t\}$ is stationary.

We first tackle the stationarity of $\{\epsilon_t\}$. Thanks to Theorem 2.3 in [1], if

$$E \log^+ |g(z_1)| < \infty, \tag{7}$$

$$E \log^+ |c(z_1)| < \infty, \text{ and } E \log |c(z_1)| < 0, \tag{8}$$

then $\{\epsilon_t\}$ is stationary. Here we define $\log^+(x) = \max\{\log x, 0\}$.

It remains to show that $\{y_t\}$ is causal. Recall that $\boldsymbol{\vartheta} = (\boldsymbol{\rho}', \boldsymbol{\gamma}')'$ is the $(p+q) \times 1$ matrix of the regression coefficients where $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_p)'$ and $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_q)'$. One can show that if

$$\begin{aligned} \vartheta(z) &= 1 - \sum_{j=1}^{p} \rho_j z^j - \sum_{j=1}^{q} \gamma_j z^{p+j} && (9) \\ &= 1 - \boldsymbol{\vartheta}'(z^1, z^2, \ldots, z^{p+q})' \neq 0 \text{ for all } z \in \mathbb{C} \text{ such that } |z| \leq 1, && (10) \end{aligned}$$

then $\{y_t\}$ is causal. The above proof follows by the arguments in the proof of Theorem 3.1.1 on page 85 and Proposition 3.1.1 on page 83 in the book by [5] without assuming the existence of the autocovariance function; nevertheless, the arguments require the condition $\sup_t E\epsilon_t < \infty$. Since z_t and h_t are independent, $E[\epsilon_t] = E[h_t z_t] = Eh_t Ez_t < \infty$ if $Eh_t < \infty$. Condition $\sup_t E\epsilon_t < \infty$ can be fulfilled by using Theorem 2.2 in [1], which states that if $\nu > 0$,

$$E \log |c(z_1)| < 0,$$
$$E|g(z_1)|^\nu < \infty \text{ and } E|c(z_1)|^\nu < 1,$$

then

$$E|\Lambda(h_t^2)|^\nu < \infty.$$

Recall that $\Lambda(x) = x^{\delta/2}$, so if $\delta\nu \geq 1$, then $Eh_t < \infty$. Moreover, under the same assumptions as above, $\{h_t\}$ is stationary by Theorem 2.3 in [1]. All these ensure that

$$\sup_t E\epsilon_t = E\epsilon_1 < \infty.$$

Lemma 2.1 in [80] shows that ϵ_1 is heavy-tailed and still holds for the augmented $GARCH(1,1)$ model, if we carefully check the arguments. For the sake of completeness, we include the lemma below.

Lemma 1. *Let Λ be a monotone increasing function. Assume the following:*

1. *z_1 has a density with respect to the Lebesgue measure on $\mathbb{R}$ that is bounded away from zero and infinity on compact sets.*
2. *$E[\log(c(z_1))] < 0$.*
3. *There exists a $k_0 > 0$ such that*

$$E[(c(z_1))^{k_0}] \geq 1,$$

$$E[(c(z_1))^{k_0} \log^+(c(z_1))] < \infty,$$

and

$$E[(g(z_1) + |z_1|)^{k_0}] < \infty,$$

where $\log^+(x) = \max\{0, \log(x)\}$.

Then there exists a unique $\alpha \in (0, k_0]$ *such that* $E[(c(z_1))^\alpha] = 1$ *and*

$$P[|\epsilon_1| > x] \sim c_0 E|z_1|^\alpha \left[\Lambda(x^2)\right]^{-\alpha}$$

for large x*, where*

$$c_0 = \frac{E\left[\left(g(z_1) + c(z_t)\Lambda(h_1^2)\right)^\alpha - \left(c(z_t)\Lambda(h_1^2)\right)^\alpha\right]}{\alpha E\left[(c(z_{t-1}))^\alpha \log^+(c(z_t))\right]}.$$

Note that even if we start with weak noise such as Gaussian innovations $\{z_t\}$, the process $\{\epsilon_t\}$ is heavy-tailed through the GARCH machinery, see [3].

We summarize the above assumptions, which can also be found in [46] and [80], below.

Assumption 1. *1.* z_1 *has a density with respect to the Lebesgue measure on* $\mathbb{R}$ *that is bounded away from zero and infinity on compact sets.*
2. $E[\log(c(z_1))] < 0$.
3. There exists a $k_0 > 0$ *such that*

$$E[(c(z_1))^{k_0}] \geq 1,$$
$$E[(c(z_1))^{k_0} \log^+(c(z_1))] < \infty,$$

and

$$E[(g(z_1) + |z_1|^\delta)^{k_0}] < \infty,$$

where $\log^+(x) = \max\{0, \log(x)\}$.
4. $\vartheta(\mathbf{z}) = 1 - \boldsymbol{\vartheta}'(z^1, z^2, \ldots, z^{p+q})' \neq 0$, $|z| \leq 1$.

Remark 1. *The first condition in Assumption 1 is a regular condition for the density function of* z_1 *and is a technical condition for proving Lemma 1.*

The second condition in Assumption 1 is a necessary and sufficient condition for which there exists a stationary solution of h_t^2*, by [60]. Therefore,* $G/GARCH(1,1)$ *has a strictly stationary solution. For example, if we consider for* $GARCH(1,1)$ *that*

$$h_t^2 := \omega + \alpha\epsilon_{t-1}^2 + \beta h_{t-1}^2 = \omega + (\alpha z_t^2 + \beta)h_{t-1}^2,$$

then the second condition in Assumption 1 can be rewritten as

$$E[\log(\alpha z_t^2 + \beta)] < 0.$$

The third condition in Assumption 1 implies that h_t is not a constant and hence excludes that $\{h_t\}$ are i.i.d. *Again if we consider the GARCH$(1,1)$ as an example, then $c(x)$ and $g(x)$ in Assumption 1 are $\alpha x^2+\beta$ and ω respectively.*

Assuming that the first three conditions in Assumption 1 hold, Lemma 1 can be restated as follows if we set $\lambda=\alpha\delta$ and $\Lambda(s)=x^\delta$: there exists a unique $\lambda\in(0,\delta k_0]$ such that $E[(c(z_t))^{\lambda/\delta}]=1$ and

$$P[|\epsilon_1|>x]\sim c_0^{(\lambda)}E[|z_1|^\lambda]x^{-\lambda},$$

where

$$c_0^{(\lambda)}=\frac{E\left[\left(g(z_1)+c(z_t)h_1^\delta\right)^{\lambda/\delta}-\left(c(z_t)h_1^\delta\right)^{\lambda/\delta}\right]}{\lambda E\left[(c(z_t))^{\lambda/\delta}\log^+(c(z_t))\right]}. \tag{11}$$

Recall that the fourth condition in Assumption 1 is used to ensure that y_t in (4) is stationary. Moreover, $\Sigma\equiv E[\bar{y}_t\bar{y}_t']$ exists and is positive definite when $\lambda>2$.

Under some assumptions similar to Assumption 1, all finite-dimensional vectors $(y_t,\dots,y_{t+k})$ have regularly varying tails defined in [68] with a tail index which is the same as the tail index λ in the G/GARCH process; see [46] and [80].

Apart from Assumption 1, we assume that the auxiliary regressor coefficients $\boldsymbol{\gamma}$ satisfy the local-to-zero framework. More precisely,

$$\boldsymbol{\gamma}=\boldsymbol{\delta}/f(T)$$

where $\boldsymbol{\delta}$ is a constant, $f(T)$ is a non-decreasing function in T, and function $f(T)$ depends on λ which is a tail index in Remark 1. The local-to-zero framework for auxiliary regressor coefficients is given in Assumption 2.

2.2. Model averaging estimator

Let

$$\mathbf{Y}=(y_1,\dots,y_T)',\quad \bar{\mathbf{Y}}_{1,p}=(\bar{y}_{1,p},\dots,\bar{y}_{T,p})',\quad \bar{\mathbf{Y}}_{1-p,q}=(\bar{y}_{1-p,q},\dots,\bar{y}_{T-p,q})'$$

and

$$\boldsymbol{\epsilon}=(\epsilon_1,\dots,\epsilon_T)'.$$

In matrix notation,

$$\mathbf{Y}=\bar{\mathbf{Y}}_{1,p}\boldsymbol{\rho}+\bar{\mathbf{Y}}_{1-p,q}\boldsymbol{\gamma}+\boldsymbol{\epsilon}=\mathbf{X}\boldsymbol{\vartheta}+\boldsymbol{\epsilon},$$

where $\mathbf{X}=(\bar{\mathbf{Y}}_{1,p},\bar{\mathbf{Y}}_{1-p,q})$ is a $T\times(p+q)$ matrix.

To study the model averaging method, we consider a set of M submodels and label a submodel by m, where $1\le m\le M$. Let $\boldsymbol{\Pi}_m$ be the $q_m\times q$ selection matrix that is used to select the partial auxiliary lag period regressors. Each

m-th submodel includes all core lag period regressors $\bar{\mathbf{Y}}_{1,p}$ and a subset of auxiliary lag period regressors $\bar{\mathbf{Y}}_{1-p,q}\mathbf{\Pi}'_m$ so that the m-th submodel has $p+q_m$ regressors. For example, we consider a sequence of nested models, that is, a sequence of submodels which have increasing auxiliary lag period regressors from no auxiliary lag period regressors to full auxiliary lag period regressors; in this case, $M = q + 1$. If we consider all submodels (all subsets of additional regressors), then $M = 2^q$.

We define $\mathbf{I}_k$ to be the $k \times k$ identity matrix; $\mathbf{0}$ stands for the zero matrix, in which all entries are zeros. Let

$$\mathbf{S}_0 = \begin{pmatrix} \mathbf{0}_{p\times q} \\ \mathbf{I}_q \end{pmatrix} \quad \text{and} \quad \mathbf{S}_m = \begin{pmatrix} \mathbf{I}_p & \mathbf{0}_{p\times q_m} \\ \mathbf{0}_{q\times p} & \mathbf{\Pi}'_m \end{pmatrix}$$

be selection matrices of dimension $(p+q)\times q$ and $(p+q)\times(p+q_m)$, respectively.

The OLS estimator of $\boldsymbol{\vartheta}$ for the full model—i.e., all auxiliary regressors are included in the model—is given by

$$\hat{\boldsymbol{\vartheta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y},$$

and the estimator of $\boldsymbol{\vartheta}_m = \mathbf{S}'_m\boldsymbol{\vartheta} = (\boldsymbol{\rho}', \boldsymbol{\gamma}'\mathbf{\Pi}'_m)' = (\boldsymbol{\rho}', \boldsymbol{\gamma}'_m)'$ for submodel m is

$$\hat{\boldsymbol{\vartheta}}_m = (\mathbf{X}'_m\mathbf{X}_m)^{-1}\mathbf{X}'_m\mathbf{Y},$$

where $\mathbf{X}_m = (\bar{\mathbf{Y}}_{1,p}, \bar{\mathbf{Y}}_{1-p,q}\mathbf{\Pi}'_m)(= \mathbf{X}\mathbf{S}_m)$ is a $T \times (p+q_m)$ matrix. If $\mathbf{\Pi}_m = \mathbf{I}_q$ ($\mathbf{S}_m = \mathbf{I}_{p+q}$), then we have $\hat{\boldsymbol{\vartheta}}_m = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \hat{\boldsymbol{\vartheta}}$, which is the OLS estimator for the full model. If $\mathbf{\Pi}_m = \mathbf{0}$, then we have $\hat{\boldsymbol{\vartheta}}_m = \left(\bar{\mathbf{Y}}'_{1,p}\bar{\mathbf{Y}}_{1,p}\right)^{-1}\bar{\mathbf{Y}}'_{1,p}\mathbf{Y}$, which is the OLS estimator for the narrow model; i.e., the model without auxiliary regressor coefficients.

We here define the averaging estimator of the parameters $\mu(\boldsymbol{\vartheta})$ with fixed weight. Let $\mathbf{w} = (w_1, \ldots, w_M)'$ be a weight vector with $w_m \geq 0$ and $\sum_{m=1}^M w_m = 1$. In other words, a weight vector lies in the unit simplex in $\mathbb{R}^M$. We collect those weights as

$$\mathcal{H}_T = \left\{ \mathbf{w} \in [0,1]^M : \sum_{m=1}^M w_m = 1 \right\}.$$

The averaging estimator of $\mu(\boldsymbol{\vartheta})$ is defined by

$$\hat{\mu}(\mathbf{w}) = \sum_{m=1}^M w_m \mu(\hat{\boldsymbol{\vartheta}}_m).$$

Through this definition, we observe that model selection is included in model averaging. More specifically, let $\mathbf{w}_m^0$ be an $M \times 1$ vector, in which the m-th element is one and the rest are zeros, and each $\mathbf{w}_m^0$ is in $\mathcal{H}_T$ for $m = 1, 2, \ldots, M$.

To determine the weights, we use Mallow's type criterion. However, we leave a detailed discussion on Mallow's type criterion to Section 4 and proceed directly to present our methodology.

2.3. Methodology

Consider the AR-G/GARCH model defined in (4)–(5). Recall that we are interested in testing the null hypothesis

$$H_0 : \mu(\boldsymbol{\vartheta}) = \mu(\boldsymbol{\vartheta}_0)$$

against $H_1 : \mu(\boldsymbol{\vartheta}) \neq \mu(\boldsymbol{\vartheta}_0)$, where $\boldsymbol{\vartheta}_0 = (\boldsymbol{\rho}_0', \boldsymbol{\gamma}_0')'$ with $\boldsymbol{\rho}_0 = (\rho_{1,0}, \ldots, \rho_{p,0})'$ and $\boldsymbol{\gamma}_0 = (\gamma_{1,0}, \ldots, \gamma_{q,0})'$. First, we divide the time series of size T into a fixed number $Q \geq 2$ of equal-sized subsequences. More precisely, for $i = 1, \ldots, Q$,

$$\begin{aligned} \mathbf{Y}^{(i)} &= (y_{1+(i-1)\lfloor T/Q \rfloor}, \ldots, y_{i\lfloor T/Q \rfloor})', \\ \bar{\mathbf{Y}}^{(i)}_{1,p} &= (\bar{y}_{1+(i-1)\lfloor T/Q \rfloor,p}, \ldots, \bar{y}_{i\lfloor T/Q \rfloor,p})', \\ \bar{\mathbf{Y}}^{(i)}_{1-p,q} &= (\bar{y}_{1+(i-1)\lfloor T/Q \rfloor-p,q}, \ldots, \bar{y}_{i\lfloor T/Q \rfloor-p,q})', \end{aligned}$$

where $\lfloor x \rfloor$ denotes the integer part of $x \in \mathbb{R}$. For each subsequence, we give an OLS estimator for $\mu(\boldsymbol{\vartheta})$ and apply the method by [62]. Define

$$Z^{(i)} := \hat{\boldsymbol{\vartheta}}^{(i)} - \boldsymbol{\vartheta}_0,$$

where

$$\hat{\boldsymbol{\vartheta}}^{(i)} = \left[(\mathbf{X}^{(i)})'\mathbf{X}^{(i)}\right]^{-1}(\mathbf{X}^{(i)})'\mathbf{Y}^{(i)} \quad \text{and} \quad \mathbf{X}^{(i)} = (\bar{\mathbf{Y}}^{(i)}_{1,p}, \bar{\mathbf{Y}}^{(i)}_{1-p,q}).$$

Hence, the corresponding t-statistic based on Q observation is given by

$$\tau_{H_0}(full) := \sqrt{Q}\frac{\bar{Z}}{s}, \tag{12}$$

where $\bar{Z} = \sum_{i=1}^{Q} Z^{(i)}/Q$ and $s^2 = \sum_{i=1}^{Q}(Z^{(i)} - \bar{Z})^2/(Q-1)$. Let T_{Q-1} be a random variable with a Student's t-distribution with $Q-1$ degrees of freedom; $\mathrm{cv}_Q(\alpha)$ denotes the two-sided quantile, namely, $P[|T_{Q-1}| > \mathrm{cv}_Q(\alpha)] = \alpha$. The robustness of our method is motivated by the following theorem in [62] which is based on t-statistic (12).

Proposition 1 (Theorem 3.4 of [62])**.** *Suppose that random variable T_{Q-1} has a Student's t-distribution with $Q-1$ degrees of freedom and that $\mathrm{cv}_Q(\alpha)$ satisfies $P[|T_{Q-1}| > \mathrm{cv}_Q(\alpha)] = \alpha$. If $\alpha \leq 0.05$, then*

$$\limsup_{T\to\infty} P[|\tau_{H_0}(full)| > \mathrm{cv}_Q(\alpha)] \leq \alpha$$

for $\lambda \geq 2$.

Now we are ready for the robust test on model averaging. To introduce the model averaging approach, we move from the full model approach introduced above to the model averaging approach under model uncertainty. We first introduce some notation. Denote that

$$\hat{\boldsymbol{\vartheta}}^{(i)}_m = \left[(\mathbf{X}^{(i)}_m)'\mathbf{X}^{(i)}_m\right]^{-1}(\mathbf{X}^{(i)}_m)'\mathbf{Y}^{(i)}$$

with $\mathbf{X}_m^{(i)} = (\bar{\mathbf{Y}}_{1,p}^{(i)}, \bar{\mathbf{Y}}_{1-p,q}^{(i)}\mathbf{\Pi}'_m)$, and that

$$Z_m^{(i)} := \mu(\hat{\boldsymbol{\vartheta}}_m^{(i)}) - \mu(\boldsymbol{\vartheta}_0)$$

and

$$Z^{(i)}(\mathbf{w}) := \sum_{m=1}^{M} w_m Z_m^{(i)}$$

for $i = 1, \ldots, Q$. The t-statistic based on Q observations is given by

$$\tau_{H_0}(m) := \sqrt{Q}\frac{\bar{Z}_m}{s_m} \quad \text{and} \quad \tau_{H_0}(\mathbf{w}) := \sqrt{Q}\frac{\bar{Z}(\mathbf{w})}{s(\mathbf{w})},$$

where $\bar{Z}_m = \sum_{i=1}^{Q} Z_m^{(i)}/Q$, $\bar{Z}(\mathbf{w}) = \sum_{i=1}^{Q} Z^{(i)}(\mathbf{w})/Q$, $s_m^2 = \sum_{i=1}^{Q}(Z_m^{(i)} - \bar{Z}_m)^2/(Q-1)$, and $s^2(\mathbf{w}) = \sum_{i=1}^{Q}(Z^{(i)}(\mathbf{w}) - \bar{Z}(\mathbf{w}))^2/(Q-1)$. According to Proposition 1, we expect a similar result; our method highly relies on this result. More specifically, if $\alpha \leq 0.05$, then

$$\limsup_{T\to\infty} P[|\tau_{H_0}(\mathbf{w})| > \mathrm{cv}_Q(\alpha)] \leq \alpha \tag{13}$$

for $\lambda \geq 2$. In addition, by simulation experiments, result (13) works well for different model weights of interest.

3. Asymptotic framework

In this section, we introduce the local-to-zero assumption on auxiliary regressor coefficients that is a part of our model setup. We also present results on asymptotic distributions of the model averaging estimator under $G/GARCH(1,1)$ noise.

We state the asymptotic results for the OLS estimators of parameters $\boldsymbol{\vartheta}$ for the full model, and then provide the limiting distribution of the OLS estimators for submodels.

Proposition 2 (Theorem 2.1 of [80])**.** *Suppose that Assumption 1 holds. Let λ be given as in Remark 1. Then, as $T \to \infty$ we have*

(a) *when $\lambda \in (0, 2)$,*

$$\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta} \xrightarrow{d} \left(\Sigma^{(\lambda/2)}\right)^{-1} S_1(\lambda/2),$$

where $S_1(\lambda/2)$ is a $\lambda/2$ stable random vector on $\mathbb{R}^{p+q}$ and $\Sigma^{(\lambda/2)}$ is an $(p+q)\times(p+q)$ matrix whose elements are composed of $\lambda/2$ stable variables;

(b) *when $\lambda = 2$,*

$$\log T(\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}) \xrightarrow{d} \left(\Sigma^{(1)}\right)^{-1} S_1(1),$$

where $\Sigma^{(1)}$ is an $(p+q)\times(p+q)$ matrix with the (i, j)-th element

$$\sum_{l=0}^{\infty} \varphi_l \varphi_{l+|i-j|},$$

where φ_l is defined in (6).

(c) *when* $\lambda \in (2, 4)$,

$$T^{1-2/\lambda}(\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}) \xrightarrow{d} \Sigma^{-1} S_1(\lambda/2),$$

where $\Sigma = E[\bar{y}_t \bar{y}_t']$.

(d) *when* $\lambda = 4$,

$$\sqrt{T/\log T}(\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}) \xrightarrow{d} \mathbf{N}(\mathbf{0}, \Sigma^{-1}\Omega\Sigma^{-1}),$$

where $\Omega = \left(c_0^{(4)} E[z_1^2]\right)(a_{ij})_{(p+q)\times(p+q)}$ *is positively defined with* $a_{ij} = \lim_{N\to\infty} E[u_{t,i,N} u_{t,j,N}]$ *and*

$$u_{t,i,N} = \sum_{l=i}^{N} \varphi_{l-i} z_{t-l} \prod_{i=1}^{l} (c(z_{t-i}; \boldsymbol{\phi}_2))^{1/\delta} \prod_{k=l+1}^{N} (c(z_{t-k}; \boldsymbol{\phi}_2))^{2/\delta}.$$

Our method requires that the tail probability be approximately $x^{-\lambda}$ for some λ, since this index λ is relevant to the scaling of our asymptotic theory in Section 3 which matters to our robust test. It is possible to extend our asymptotic theories from G/GARCH to augmented GARCH, but we as yet do not know the proper scaling for our asymptotic theory if we consider the augmented GARCH model.

Remark 2. *We consider a special case in which*

$$h_t^2(\boldsymbol{\theta}) = \omega + \alpha\epsilon_t^2 + \beta h_{t-1}^2(\boldsymbol{\theta}),$$

where $\omega > 0$*,* $\alpha > 0$*, and* $\beta > 0$*. Case* (c) *in Proposition 2 was obtained from [46].*

Remark 3. *We summarize recent asymptotic results of estimators when the noise* ϵ_t *is* i.i.d. *with respect to three moment assumptions:*

(i) $E[\epsilon_1^2] < \infty$*;*
(ii) $E[\epsilon_1^2] = \infty$ *with a constraint condition on the tail probability;*[2] *for more detail see [76]; and*
(iii) $E[\epsilon_1^2] = \infty$ *with tail index* $\lambda \in (0, 2)$*.*

(i) Under $E[\epsilon_1^2] < \infty$*, the least-squares estimator, the least absolute deviation estimator, and the* M*-estimator are all* $\sqrt{T}$*-consistent and asymptotically normal. (ii) [76] shows that the least-squares estimator is* $L(T)\sqrt{T}$*-consistent and asymptotically normal, where* $L(T)$ *is a slowly varying function. They also provide a robust test for predictability in the predictive regression model with heavy-tailed noise, in which the predictive variable is persistent and its noise is highly correlated with returns. (iii) In the last case, the least-squares estimator, the least absolute deviation estimator, and the* M*-estimator are all* $L(T)T^{1/\lambda}$*-consistent and converge to a stable law; see [15], [56], and [14]. For noise that is not* i.i.d.*, few references address asymptotic behavior, e.g., [57], [46], and [80].*

[2] Two examples are the Pareto distribution and the t-distribution with 2 degrees of freedom.

Due to model uncertainty, one does not know which lag period regressors should be included in the real model. In our submodels, all core lag period regressors must be included. The auxiliary lag period regressors are partially, or even fully, included. We observe that the scaling error of the submodel—but not that of the full model—yields bias. If parameter γ does not depend on T, the asymptotic bias tends to infinity for $\lambda \geq 2$. Therefore, γ ought to converge to zero when T goes to infinity. Due to this technical issue, we adopt the local-to-zero asymptotic framework, which is stated as follows:

Assumption 2. *Suppose that for $\lambda \in (0, 4]$, $\boldsymbol{\gamma}$ is chosen as follows:*

(a) *When $\lambda \in (0, 2)$, $\boldsymbol{\gamma} = \boldsymbol{\gamma}^{(\lambda)} = \delta^{(\lambda)}$ is an unknown constant vector which is independent of sample size T.*
(b) *When $\lambda = 2$, $\boldsymbol{\gamma} = \boldsymbol{\gamma}^{(2)} = \boldsymbol{\delta}^{(2)}/\log T$, where $\boldsymbol{\delta}^{(2)}$ is an unknown constant vector.*
(c) *When $\lambda \in (2, 4)$, $\boldsymbol{\gamma} = \boldsymbol{\gamma}^{(\lambda)} = \boldsymbol{\delta}^{(\lambda)}/T^{1-2/\lambda}$, where $\boldsymbol{\delta}^{(\lambda)}$ is an unknown constant vector.*
(d) *When $\lambda = 4$, $\boldsymbol{\gamma} = \boldsymbol{\gamma}^{(4)} = \boldsymbol{\delta}^{(4)}/\sqrt{T/\log T}$, where $\boldsymbol{\delta}^{(4)}$ is an unknown constant vector.*

This local-to-zero asymptotic framework ([39]) is to ensure that the asymptotic mean squared error of the averaging estimator remains finite. Along this line, many studies analyze the asymptotic and finite sample properties of the model selection and averaging estimator, see for example the papers by [47], [64], [19], [34], and [49], among others. The order of T in Assumption 2 comes from Proposition 2, the reason for which is given by first considering the case of $\lambda \in (2, 4)$, since the other cases merely mirror the following arguments. The OLS estimator for submodel m can be decomposed as

$$\begin{aligned}\hat{\boldsymbol{\vartheta}}_m &= (\mathbf{X}_m'\mathbf{X}_m)^{-1}\mathbf{X}_m'\mathbf{Y} \\ &= (\mathbf{X}_m'\mathbf{X}_m)^{-1}\mathbf{X}_m'[\mathbf{X}\boldsymbol{\vartheta} + \boldsymbol{\epsilon}] \\ &= \boldsymbol{\vartheta}_m + (\mathbf{X}_m'\mathbf{X}_m)^{-1}\mathbf{X}_m'\bar{\mathbf{Y}}_{1-p,q}(\mathbf{I}_q - \boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m)\boldsymbol{\gamma} + (\mathbf{X}_m'\mathbf{X}_m)^{-1}\mathbf{X}_m'\boldsymbol{\epsilon}. \quad (14)\end{aligned}$$

One can consider the second term in (14) as a bias where $(\mathbf{I}_q - \boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m)$ is a selection matrix that keeps the non-selected auxiliary regressors in the m-th model; the estimator variance is from the third term in (14). The following explains why Assumption 2 appears. If $\boldsymbol{\gamma}^{(\lambda)}$ converges to 0 slower than $T^{1-2/\lambda}$, the asymptotic bias goes to infinity, except for the full model, which suggests that we should use the full model. If $\boldsymbol{\gamma}^{(\lambda)}$ converges to 0 faster than $T^{2/\lambda-1}$, the asymptotic bias goes to zero, which shows that all submodels are acceptable. Since we prefer low variance, the narrow model wins.

$O\left(T^{2/\lambda-1}\right)$ is the order in our local-to-zero framework because we know that the estimator variances[3] are of the order $O(T^{-1})$, and the order of the

[3]By Remark 1 of [46], it holds that as $T \to \infty$,

$$\left(\frac{1}{T}\mathbf{X}'\mathbf{X}\right) \overset{a.s.}{\to} \Sigma,$$

where $\Sigma = E[\bar{y}_t\bar{y}_t']$.

asymptotic distribution[4] is $O\left(T^{-2/\lambda}\right)$.

Now we are ready to present the scaling limit of the difference between $\hat{\boldsymbol{\vartheta}}_m$ and $\boldsymbol{\vartheta}_m$ for submodel m in the following form:

$$n(T)(\hat{\boldsymbol{\vartheta}}_m - \boldsymbol{\vartheta}_m) \xrightarrow{d} \text{deterministic column vector} + \text{matrix} \times \text{stable distribution},$$

where $n(\cdot)$ is a non-decreasing function on $\{1, 2, 3, \ldots\}$. The asymptotic distribution of the difference between the estimator and the true parameter is stated as follows. The purpose of this is because of inference.

Lemma 2. *Suppose that Assumptions 1 and 2 hold. Then, as $T \to \infty$, we have*

(a) *when $\lambda \in (0, 2)$,*

$$\hat{\boldsymbol{\vartheta}}_m - \boldsymbol{\vartheta}_m \xrightarrow{d} \mathbf{A}_m^{(\lambda/2)} \boldsymbol{\gamma}^{(\lambda)} + \mathbf{B}_m^{(\lambda/2)} S_1(\lambda/2),$$

where $\mathbf{A}_m^{(\lambda/2)} = (\Sigma_m^{(\lambda/2)})^{-1} \mathbf{S}'_m \Sigma^{(\lambda/2)} \mathbf{S}_0 (\mathbf{I}_q - \mathbf{\Pi}'_m \mathbf{\Pi}_m)$, $\mathbf{B}_m^{(\lambda/2)} = (\Sigma_m^{(\lambda/2)})^{-1} \mathbf{S}'_m$, *and* $\Sigma_m^{(\lambda/2)} = \mathbf{S}'_m \Sigma^{(\lambda/2)} \mathbf{S}_m$.

(b) *when $\lambda = 2$,*

$$(\log T)(\hat{\boldsymbol{\vartheta}}_m - \boldsymbol{\vartheta}_m) \xrightarrow{d} \mathbf{A}_m^{(1)} \boldsymbol{\delta}^{(2)} + \mathbf{B}_m^{(1)} S_1(1),$$

where $\mathbf{A}_m^{(1)} = (\Sigma_m^{(1)})^{-1} \mathbf{S}'_m \Sigma^{(1)} \mathbf{S}_0 (\mathbf{I}_q - \mathbf{\Pi}'_m \mathbf{\Pi}_m)$, $\mathbf{B}_m^{(1)} = (\Sigma_m^{(1)})^{-1} \mathbf{S}'_m$, *and* $\Sigma_m^{(1)} = \mathbf{S}'_m \Sigma^{(1)} \mathbf{S}_m$.

(c) *when $\lambda \in (2, 4)$,*

$$T^{1-2/\lambda}(\hat{\boldsymbol{\vartheta}}_m - \boldsymbol{\vartheta}_m) \xrightarrow{d} \mathbf{A}_m \boldsymbol{\delta}^{(\lambda)} + \mathbf{B}_m S_1(\lambda/2),$$

where $\mathbf{A}_m = (\Sigma_m)^{-1} \mathbf{S}'_m \Sigma \mathbf{S}_0 (\mathbf{I}_q - \mathbf{\Pi}'_m \mathbf{\Pi}_m)$, $\mathbf{B}_m = (\Sigma_m)^{-1} \mathbf{S}'_m$, *and* $\Sigma_m = \mathbf{S}'_m \Sigma \mathbf{S}_m$.

(d) *when $\lambda = 4$,*

$$\sqrt{T/\log T}(\hat{\boldsymbol{\vartheta}}_m - \boldsymbol{\vartheta}_m) \xrightarrow{d} \mathbf{N}(\mathbf{A}_m \boldsymbol{\delta}^{(4)}, (\Sigma_m)^{-1} \Omega_m (\Sigma_m)^{-1}) = \mathbf{A}_m \boldsymbol{\delta}^{(4)} + \mathbf{B}_m \mathbf{N}(\mathbf{0}, \boldsymbol{\Omega}),$$

where $\Omega_m = \mathbf{S}'_m \Omega \mathbf{S}_m$.

In Lemma 2, $\mathbf{A}_m \boldsymbol{\delta}^{(\lambda)}$ represents the asymptotic bias of the submodel estimators. If the parameters of the auxiliary regressors are all zeros ($\boldsymbol{\gamma} = \mathbf{0}$) or the auxiliary regressors are uncorrelated (Σ is a diagonal matrix), the asymptotic bias of the submodels is zero. Those two components contribute the asymptotic bias. Note that $\mathbf{A}_m^{(\lambda/2)}$ and $\mathbf{B}_m^{(\lambda/2)}$ are not continuous at $\lambda = 2$; that is, as $\lambda \uparrow 2$,

$$\mathbf{A}_m^{(\lambda/2)} \neq \mathbf{A}_m^{(1)} + o(1) \quad \text{and} \quad \mathbf{B}_m^{(\lambda/2)} \neq \mathbf{B}_m^{(1)} + o(1).$$

[4]By Theorem 2 of [46] (also Proposition 3.3 in [16]), as $T \to \infty$,

$$T^{-2/\lambda} \mathbf{X}' \boldsymbol{\epsilon} \xrightarrow{d} S_1(\lambda/2).$$

We now consider that the parameter $\mu(\boldsymbol{\vartheta})$ is a smooth real-valued function. We are more interested in estimating $\mu(\boldsymbol{\vartheta})$ than in model fitting, which is more commonly studied in traditional model selection or model averaging approaches. Let

$$\mathbf{D}_{\boldsymbol{\vartheta}} = \left(\mathbf{D}'_{\boldsymbol{\rho}}, \mathbf{D}'_{\boldsymbol{\gamma}}\right)', \quad \mathbf{D}_{\boldsymbol{\vartheta}_m} = \left(\mathbf{D}'_{\boldsymbol{\rho}}, \mathbf{D}'_{\boldsymbol{\gamma}_m}\right)',$$
$$\mathbf{D}'_{\boldsymbol{\rho}} = \frac{\partial \mu}{\partial \boldsymbol{\rho}}(\boldsymbol{\rho}, \boldsymbol{\gamma}_m^{(\lambda)}, \mathbf{0}), \quad \mathbf{D}'_{\boldsymbol{\gamma}} = \frac{\partial \mu}{\partial \boldsymbol{\gamma}}(\boldsymbol{\rho}, \boldsymbol{\gamma}_m^{(\lambda)}, \mathbf{0}), \text{ and } \mathbf{D}'_{\boldsymbol{\gamma}_m} = \frac{\partial \mu}{\partial \boldsymbol{\gamma}_m^c}(\boldsymbol{\rho}, \boldsymbol{\gamma}_m^{(\lambda)}, \mathbf{0}). \tag{15}$$

Assume that the partial derivatives are continuous in the neighborhood of the null point. The following lemma gives more general results than Lemma 2 and is proved by the delta method.

Lemma 3. *Under the same assumptions as in Lemma 2,*

(a) *when* $\lambda \in (0, 2)$*, there exists a function* $h : \mathbb{R}^{p+q} \mapsto \mathbb{R}$ *with*

$$\lim_{\mathbf{x} \to (\boldsymbol{\rho}', \boldsymbol{\gamma}'_m, \mathbf{0}')'} h(\mathbf{x}) = 0$$

such that as $T \to \infty$*,*

$$\mu(\hat{\boldsymbol{\vartheta}}_m) - \mu(\boldsymbol{\vartheta}) \xrightarrow{d} \mathbf{D}'_{\boldsymbol{\vartheta}} \mathbf{C}_m^{(\lambda/2)} \boldsymbol{\gamma}^{(\lambda)} - h(\boldsymbol{\vartheta}) |\boldsymbol{\Pi}_{m^c} \boldsymbol{\gamma}^{(\lambda)}| + \mathbf{D}'_{\boldsymbol{\vartheta}} \mathbf{V}_m^{(\lambda/2)} S_1(\lambda/2),$$

where $\mathbf{C}_m^{(\lambda/2)} = (\mathbf{V}_m \Sigma^{(\lambda/2)} - \mathbf{I}_{p+q}) \mathbf{S}_0 (\mathbf{I}_q - \boldsymbol{\Pi}'_m \boldsymbol{\Pi}_m)$ *and* $\mathbf{V}_m^{(\lambda/2)} = \mathbf{S}_m (\mathbf{S}'_m \Sigma^{(\lambda/2)} \mathbf{S}_m)^{-1} \mathbf{S}'_m$.

(b) *when* $\lambda = 2$*, as* $T \to \infty$*,*

$$\log T \left(\mu(\hat{\boldsymbol{\vartheta}}_m) - \mu(\boldsymbol{\vartheta})\right) \xrightarrow{d} \mathbf{D}'_{\boldsymbol{\vartheta}} \mathbf{C}_m^{(1)} \boldsymbol{\delta}^{(2)} + \mathbf{D}'_{\boldsymbol{\vartheta}} \mathbf{V}_m^{(1)} S_1(1),$$

where $\mathbf{C}_m^{(1)} = (\mathbf{V}_m \Sigma^{(1)} - \mathbf{I}_{p+q}) \mathbf{S}_0 (\mathbf{I}_q - \boldsymbol{\Pi}'_m \boldsymbol{\Pi}_m)$ *and* $\mathbf{V}_m^{(1)} = \mathbf{S}_m (\mathbf{S}'_m \Sigma^{(1)} \mathbf{S}_m)^{-1} \mathbf{S}'_m$.

(c) *when* $\lambda \in (2, 4)$*, as* $T \to \infty$*,*

$$T^{1-2/\lambda} \left(\mu(\hat{\boldsymbol{\vartheta}}_m) - \mu(\boldsymbol{\vartheta})\right) \xrightarrow{d} \mathbf{D}'_{\boldsymbol{\vartheta}} \mathbf{C}_m \boldsymbol{\delta}^{(\lambda)} + \mathbf{D}'_{\boldsymbol{\vartheta}} \mathbf{V}_m S_1(\lambda/2),$$

where $\mathbf{C}_m = (\mathbf{V}_m \Sigma - \mathbf{I}_{p+q}) \mathbf{S}_0 (\mathbf{I}_q - \boldsymbol{\Pi}'_m \boldsymbol{\Pi}_m)$ *and* $\mathbf{V}_m = \mathbf{S}_m (\mathbf{S}'_m \Sigma \mathbf{S}_m)^{-1} \mathbf{S}'_m$.

(d) *when* $\lambda = 4$*, as* $T \to \infty$*,*

$$\sqrt{T / \log T} \left(\mu(\hat{\boldsymbol{\vartheta}}_m) - \mu(\boldsymbol{\vartheta})\right) \xrightarrow{d} N(\mathbf{D}'_{\boldsymbol{\vartheta}} \mathbf{C}_m \boldsymbol{\delta}^{(4)}, \mathbf{D}'_{\boldsymbol{\vartheta}} \mathbf{V}_m \Omega \mathbf{V}_m \mathbf{D}_{\boldsymbol{\vartheta}}).$$

4. Mallows-type criteria

Model uncertainty is pervasive in empirical economic/finance applications. One way to address this is to use model selection, where we consider a group of candidate models and pick the best among them according to a criterion. A more general way to handle model uncertainty is called model averaging. Instead of selecting one candidate model as in model selection, model averaging incorporates all available information by taking the average over all basic models. Many studies show that model averaging is more robust than model selection. The fundamental problem in model averaging is selecting the weights for all basic models. In what follows, under G/GARCH noise, we provide a model averaging estimator based on the Mallows averaging model proposed by [30]. We also construct a model selection estimator by reformulating traditional Mallows model selection.

First, we consider an averaging model for which the weights are fixed and define the averaging estimator of parameters $\mu(\boldsymbol{\vartheta})$. Recall that $\mathbf{w} = (w_1, \ldots, w_M)'$ is a weight vector with $w_m \geq 0$ and $\sum_{m=1}^{M} w_m = 1$,

$$\mathcal{H}_T = \left\{ \mathbf{w} \in [0,1]^M : \sum_{m=1}^{M} w_m = 1 \right\},$$

and the averaging estimator of $\mu(\boldsymbol{\vartheta})$ is defined by

$$\hat{\mu}(\mathbf{w}) = \sum_{m=1}^{M} w_m \mu(\hat{\boldsymbol{\vartheta}}_m).$$

The following lemma provides the asymptotic distribution of the averaging estimator with fixed weights; this result covers the case of model selection.

Lemma 4. *Under the assumptions of Lemma 2,*

(a) *when* $\lambda \in (0, 2)$*, as* $T \to \infty$*,*

$$\hat{\mu}(\mathbf{w}) - \mu(\boldsymbol{\vartheta}) \xrightarrow{d} \mathbf{D}_{\boldsymbol{\vartheta}}' \mathbf{C}^{(\lambda/2)}(\mathbf{w}) \boldsymbol{\gamma}^{(\lambda)} - h(\boldsymbol{\vartheta}) \sum_{m=1}^{M} w_m |\boldsymbol{\Pi}_{m^c} \boldsymbol{\gamma}^{(\lambda)}| + \mathbf{D}_{\boldsymbol{\vartheta}}' \mathbf{V}^{(\lambda/2)}(\mathbf{w}) S_1(\lambda/2),$$

where $\mathbf{C}^{(\lambda/2)}(\mathbf{w}) = \sum_{m=1}^{M} w_m \mathbf{C}_m^{(\lambda/2)}$*,* $\mathbf{V}^{(\lambda/2)}(\mathbf{w}) = \sum_{m=1}^{M} w_m \mathbf{V}_m^{(\lambda/2)}$*, and* h *is defined in Lemma 3.*

(b) *when* $\lambda = 2$*, as* $T \to \infty$*,*

$$\log T \left(\hat{\mu}(\mathbf{w}) - \mu(\boldsymbol{\vartheta}) \right) \xrightarrow{d} \mathbf{D}_{\boldsymbol{\vartheta}}' \mathbf{C}^{(1)}(\mathbf{w}) \boldsymbol{\delta}^{(2)} + \mathbf{D}_{\boldsymbol{\vartheta}}' \mathbf{V}^{(1)}(\mathbf{w}) S_1(1),$$

where $\mathbf{C}^{(1)}(\mathbf{w}) = \sum_{m=1}^{M} w_m \mathbf{C}_m^{(1)}$ *and* $\mathbf{V}^{(1)}(\mathbf{w}) = \sum_{m=1}^{M} w_m \mathbf{V}_m^{(1)}$*.*

(c) *when* $\lambda \in (2, 4)$*, as* $T \to \infty$*,*

$$T^{1-2/\lambda} \left(\hat{\mu}(\mathbf{w}) - \mu(\boldsymbol{\vartheta}) \right) \xrightarrow{d} \mathbf{D}_{\boldsymbol{\vartheta}}' \mathbf{C}(\mathbf{w}) \boldsymbol{\delta}^{(\lambda)} + \mathbf{D}_{\boldsymbol{\vartheta}}' \mathbf{V}(\mathbf{w}) S_1(\lambda/2),$$

where $\mathbf{C}(\mathbf{w}) = \sum_{m=1}^{M} w_m \mathbf{C}_m$ *and* $\mathbf{V}(\mathbf{w}) = \sum_{m=1}^{M} w_m \mathbf{V}_m$*.*

(d) *when* $\lambda = 4$*, as* $T \to \infty$*,*

$$\sqrt{T/\log T} \left(\hat{\mu}(\mathbf{w}) - \mu(\boldsymbol{\vartheta}) \right) \xrightarrow{d} N(\mathbf{D}_{\boldsymbol{\vartheta}}' \mathbf{C}(\mathbf{w}) \boldsymbol{\delta}^{(4)}, V(\mathbf{w}))$$

where $V(\mathbf{w}) = \sum_{m=1}^{M} w_m^2 \mathbf{D}_{\boldsymbol{\vartheta}}' \mathbf{V}_m \Omega \mathbf{V}_m' \mathbf{D}_{\boldsymbol{\vartheta}} + 2 \sum\sum_{m \neq k} w_m w_k \mathbf{D}_{\boldsymbol{\vartheta}}' \mathbf{V}_m \Omega \mathbf{V}_k' \mathbf{D}_{\boldsymbol{\vartheta}}$*.*

Remark 4. *Recall that the limiting distributions in cases* (b), (c), *and* (d) *are stable distributions since a linear transformation of a stable distribution is a stable distribution if the stability parameter is greater than or equal to 1. This is a direct result of Theorems 2.1.2 and 2.1.5* (c) *in [70].*

Remark 5. *When $\lambda = 4$, it is clear that $V(\mathbf{w})$ can be rewritten as*

$$\mathbf{D}'_{\boldsymbol{\vartheta}}\mathbf{V}(\mathbf{w})\Omega\mathbf{V}(\mathbf{w})\mathbf{D}_{\boldsymbol{\vartheta}},$$

where $\mathbf{V}(\mathbf{w})$ is a symmetric matrix since $\mathbf{V}_m$ is a symmetric matrix for $m = 1, \ldots, M$.

Remark 6. *In this paper, we propose a robust test based on our Lemma 4 and Theorem 1, and the idea of [62] for AR-G/GARCH models under model uncertainty. Note that [49, Theorem 6] corrects bias from the submodels and estimates weights to construct a valid confidence interval. Although the limiting distributions of errors in [49] are all Gaussian, the limiting distributions of errors under G/GARCH noise have stable laws and are yet to be completely characterized. To account for this and obtain a valid confidence interval, we use the idea of [62] to construct a valid confidence interval and in so doing make new progress in handling model uncertainty. Our method can also be used for model selection.*

Thanks to Lemma 4, $a^{(\lambda)}_{\lfloor T/Q \rfloor} Z^{(i)}(\mathbf{w})$ is asymptotically stable and hence asymptotically mixed Gaussian ([70, Proposition 1.3.1]). Thanks to the argument in Lemma 3.2 of [62], we know that $a^{(\lambda)}_{\lfloor T/Q \rfloor} Z^{(i)}(\mathbf{w})$ and $a^{(\lambda)}_{\lfloor T/Q \rfloor} Z^{(j)}(\mathbf{w})$ are asymptotically independent for $i \neq j$. Therefore, in the proposed method of Section 2.3, a valid $100(1-\alpha)\%$ confidence interval for $\mu(\boldsymbol{\vartheta})$ is $[-\mathrm{cv}_Q(\alpha), \mathrm{cv}_Q(\alpha)]$. Through our simulation experiments, we note that the result (13) likely still holds even if we withdraw the zero median condition.

4.1. Mallows-type model averaging estimator

The celebrated MMA approach of [30] is an OLS-based model averaging estimator whose weights are selected by minimizing a criterion in the spirit of Mallows' C_p ([51]), where the MMA estimator is asymptotically optimal in the sense of achieving the lowest squared error with a penalty over all weights in the unit simplex $\mathbb{R}^M$. MMA has been widely applied to other regression models. Note that [30] extends this asymptotic optimality from model selection in [48] to model averaging, and it has been established that the average squared error of the MMA estimator is asymptotically equivalent to the lowest expected squared error. To consider the asymptotic optimality for time-series noise, readers are referred to [12]. In this paper, we do not study asymptotic optimality because[5] (i) the lag period $p + q$ is finite instead of infinite and (ii) high-order moments of noise do not exist in our setting.

[5] The reader is referred to [30], [75], and [77], among others.

Let us introduce the classical MMA criterion: define

$$\hat{\boldsymbol{\epsilon}}(\mathbf{w}) = \mathbf{Y} - \mathbf{X}\bar{\boldsymbol{\vartheta}}(\mathbf{w}) = \mathbf{Y} - \sum_{m=1}^{M} (w_m \mathbf{X}_m \hat{\boldsymbol{\vartheta}}_m) = \sum_{m=1}^{M} w_m \hat{\boldsymbol{\epsilon}}_m$$

to be the residual vector, where the averaging estimator is

$$\bar{\boldsymbol{\vartheta}}(\mathbf{w}) = \sum_{m=1}^{M} w_m \mathbf{S}_m \hat{\boldsymbol{\vartheta}}_m.$$

Under the homoskedastic linear regression model,[6] [30] suggests selecting the weights for all basic models by minimizing the following Mallows criterion:

$$C(\mathbf{w}) = \hat{\boldsymbol{\epsilon}}(\mathbf{w})'\hat{\boldsymbol{\epsilon}}(\mathbf{w}) + 2\sigma^2 \mathbf{k}'\mathbf{w}, \tag{16}$$

where $\mathbf{k} = (k_1, \ldots, k_M)'$ with $k_m = p + q_m$ for $1 \leq m \leq M$. Here σ^2 in (16) is the variance of the noise.

In this section, we propose the Mallows-type model averaging (MTMA) criterion in AR-G/GARCH models. One aim of this paper is to study asymptotic inference under an assumption of heavy-tailed noise. Therefore, the first step is to rescale the criteria to cause the asymptotic distribution to exist. The second step is to modify the penalty term.

The first term on the right-hand side in (16) must be scaled such that the sample size goes to infinity. To obtain an asymptotic distribution, we multiply the scaling by a constant to the first term in (16) to yield

(a) when $\lambda \in (0, 2)$,

$$\left(a_T^{(\lambda)}\right)^{-2} \hat{\boldsymbol{\epsilon}}(\mathbf{w})'\hat{\boldsymbol{\epsilon}}(\mathbf{w}),$$

where $a_T^{(\lambda)} = \left(c_0^{(\lambda)} E|z_1|^\lambda T\right)^{1/\lambda}$ and $c_0^{(\lambda)}$ is defined in (11).

(b) when $\lambda = 2$,

$$\left(\frac{\log T}{c_0^{(2)} T}\right) \hat{\boldsymbol{\epsilon}}(\mathbf{w})'\hat{\boldsymbol{\epsilon}}(\mathbf{w}).$$

(c) when $\lambda \in (2, 4)$,

$$\left(T^{1-4/\lambda}\right) \hat{\boldsymbol{\epsilon}}(\mathbf{w})'\hat{\boldsymbol{\epsilon}}(\mathbf{w}).$$

[6]The homoskedastic linear regression is expressed as

$$y_i = \sum_{j=1}^{\infty} \theta_j x_{ji} + e_i, \quad E[e_i|x_i] = 0, \quad E[e_i^2|x_i] = \sigma^2,$$

where y_i is real-valued whereas $x_i = (x_{1i}, x_{2i}, \ldots)$ is countably infinite.

(d) when $\lambda = 4$,

$$\left(\frac{1}{\log T}\right)\hat{\boldsymbol{\epsilon}}(\mathbf{w})'\hat{\boldsymbol{\epsilon}}(\mathbf{w}).$$

It is known (cf. [80]) that $\sigma^2 := E[\epsilon_1^2]$ exists if $\lambda \in (2,4]$. Thus, the penalty term for $\lambda \in (2,4]$ is still set to $2\sigma^2$, as in the classical version. However, the second moment does not exist for $\lambda \in (0,2]$. The tail behavior shown by [80, Lemma 2.1] infers that the level of the penalty for $\lambda \in (0,2]$ is

$$\begin{cases} 2\left(c_0^{(\lambda)} E[|z_1|^\lambda]\right)^{2/\lambda} & \text{if } \lambda \in (0,2); \\ 2c_0^{(2)} & \text{if } \lambda = 2. \end{cases}$$

Here the criterion varies when λ changes. To indicate this dependency, we use MTMA-λ to denote the MTMA criterion with parameter λ.

We summarize the MTMA-λ criterion for different ranges of λ below:

(a) when $\lambda \in (0,2)$,

$$C^{(\lambda)}(\mathbf{w}) = \left(a_T^{(\lambda)}\right)^{-2}\hat{\boldsymbol{\epsilon}}(\mathbf{w})'\hat{\boldsymbol{\epsilon}}(\mathbf{w}) + 2\left(c_0^{(\lambda)} E[|z_1|^\lambda]\right)^{2/\lambda}\mathbf{k}'\mathbf{w}, \tag{17}$$

where $a_T^{(\lambda)} = \left(c_0^{(\lambda)} E|z_1|^\lambda T\right)^{1/\lambda}$ and $c_0^{(\lambda)}$ is defined in (11).

(b) when $\lambda = 2$,

$$C^{(2)}(\mathbf{w}) = \left(\frac{\log T}{c_0^{(2)} T}\right)\hat{\boldsymbol{\epsilon}}(\mathbf{w})'\hat{\boldsymbol{\epsilon}}(\mathbf{w}) + 2c_0^{(2)}\mathbf{k}'\mathbf{w}. \tag{18}$$

(c) when $\lambda \in (2,4)$,

$$C^{(\lambda)}(\mathbf{w}) = \left(T^{1-4/\lambda}\right)\hat{\boldsymbol{\epsilon}}(\mathbf{w})'\hat{\boldsymbol{\epsilon}}(\mathbf{w}) + 2\sigma^2\mathbf{k}'\mathbf{w}, \tag{19}$$

where $\sigma^2 = E[\epsilon_1^2]$.

(d) when $\lambda = 4$,

$$C^{(4)}(\mathbf{w}) = \left(\frac{1}{\log T}\right)\hat{\boldsymbol{\epsilon}}(\mathbf{w})'\hat{\boldsymbol{\epsilon}}(\mathbf{w}) + 2\sigma^2\mathbf{k}'\mathbf{w}. \tag{20}$$

Remark 7. *Consider an example in a simple setting: for $\lambda \in (2,4)$, consider the GARCH$(1,1)$ model with*

$$h_t^2 = \omega + \alpha\epsilon_{t-1}^2 + \beta h_{t-1}^2,$$

where $\omega > 0$, $\alpha > 0$, and $\beta > 0$. It is well-known that under some regular conditions, we have $E[\epsilon_t^2] = \frac{\omega}{1-(\alpha+\beta)}$ and $E[\epsilon_s\epsilon_r] = 0$ for $s \neq r$. So, the criterion is

$$C_T^{(\lambda)}(\mathbf{w}) = \left(T^{1-4/\lambda}\right)\hat{\boldsymbol{\epsilon}}(\mathbf{w})'\hat{\boldsymbol{\epsilon}}(\mathbf{w}) + 2\frac{\omega}{1-(\alpha+\beta)}\mathbf{k}'\mathbf{w}.$$

We use the case of $\lambda \in (2,4)$ to introduce the idea of MTMA, since other cases are basically the same. To derive the asymptotic distribution of the MTMA-λ estimator, we rewrite the criterion (see Remark 8) as

$$C^{(\lambda)}(\mathbf{w}) = \mathbf{w}'\boldsymbol{\zeta}^{(\lambda)}\mathbf{w} + 2\sigma^2\mathbf{k}'\mathbf{w} + \left(T^{1-4/\lambda}\right)\hat{\boldsymbol{\epsilon}}'\hat{\boldsymbol{\epsilon}}, \tag{21}$$

where $\boldsymbol{\zeta}^{(\lambda)} = \boldsymbol{\zeta}^{(\lambda)}(T)$ is an $M \times M$ matrix whose (m,l)-th element is $\zeta^{(\lambda)}_{m,l} = \zeta^{(\lambda)}_{m,l}(T) = \left(T^{1-4/\lambda}\right)(\hat{\boldsymbol{\epsilon}}_m - \hat{\boldsymbol{\epsilon}})'(\hat{\boldsymbol{\epsilon}}_l - \hat{\boldsymbol{\epsilon}})$ and $\hat{\boldsymbol{\epsilon}}$ and $\hat{\boldsymbol{\epsilon}}_m$ are the residual vectors in the full model and the m-th submodel, respectively.

Remark 8. *The MTMA-λ criterion (19) can be rewritten as the form in (21) because of $\hat{\boldsymbol{\epsilon}}_m'\hat{\boldsymbol{\epsilon}} = \hat{\boldsymbol{\epsilon}}'\hat{\boldsymbol{\epsilon}}$; hence $\hat{\boldsymbol{\epsilon}}_m'\hat{\boldsymbol{\epsilon}}_l - \hat{\boldsymbol{\epsilon}}'\hat{\boldsymbol{\epsilon}} = (\hat{\boldsymbol{\epsilon}}_m - \hat{\boldsymbol{\epsilon}})'(\hat{\boldsymbol{\epsilon}}_l - \hat{\boldsymbol{\epsilon}})$. The arguments will be given in the Appendix.*

Note that $\left(T^{1-4/\lambda}\right)$ is not related to the weight vector $\mathbf{w}$. Thus, minimizing $C_T^{(\lambda)}(\mathbf{w})$ over $\mathbf{w}$ is equivalent to minimizing

$$\tilde{C}^{(\lambda)}(\mathbf{w}) := \mathbf{w}'\boldsymbol{\zeta}^{(\lambda)}\mathbf{w} + 2\sigma^2\mathbf{k}'\mathbf{w}.$$

The following theorem gives the asymptotic distribution of the MTMA estimators.

Theorem 1. *Let $\hat{\mathbf{w}}^{(\lambda)} = \operatorname{argmin}_{\mathbf{w}\in\mathcal{H}_T}\tilde{C}^{(\lambda)}(\mathbf{w})$ be the MTMA-λ weights. Suppose that Assumptions 1 and 2 hold. As $T \to \infty$, we have*

(a) *when $\lambda \in (0,2)$,*

$$\begin{aligned}\tilde{C}^{(\lambda)}(\mathbf{w}) &= \mathbf{w}'\boldsymbol{\zeta}^{(\lambda)}\mathbf{w} + 2\left(c_0^{(\lambda)}E[|z_1|^\lambda]\right)^{2/\lambda}\mathbf{k}'\mathbf{w}\\ &\xrightarrow{d} \mathbf{w}'\boldsymbol{\zeta}^{*(\lambda)}\mathbf{w} + 2\left(c_0^{(\lambda)}E[|z_1|^\lambda]\right)^{2/\lambda}\mathbf{k}'\mathbf{w},\end{aligned}$$

where $\boldsymbol{\zeta}^{*(\lambda)}$ is an $M \times M$ matrix whose (m,l)-th element is

$$\zeta^{*(\lambda)}_{m,l} = \left(\Gamma_m^{(\lambda)}\right)'\Sigma^{(\lambda/2)}\Gamma_l^{(\lambda)},$$

where

$$\Gamma_m^{(\lambda)} = \mathbf{C}_m^{(\lambda/2)}\boldsymbol{\gamma}^{(\lambda)} + \left(\mathbf{V}_m^{(\lambda/2)} - \left(\Sigma^{(\lambda/2)}\right)^{-1}\right)S_1(\lambda/2)$$

and $\mathbf{C}_m^{(\lambda/2)}$ and $\mathbf{V}_m^{(\lambda/2)}$ are defined in Lemma 3.
Moreover, as $T \to \infty$, we have

$$\hat{\mathbf{w}}^{(\lambda)} \xrightarrow{d} \mathbf{w}^{*(\lambda)} = \operatorname{argmin}_{\mathbf{w}\in\mathcal{H}_T}\left(\mathbf{w}'\boldsymbol{\zeta}^{*(\lambda)}\mathbf{w} + 2\left(c_0^{(\lambda)}E[|z_1|^\lambda]\right)^{2/\lambda}\mathbf{k}'\mathbf{w}\right)$$

and

$$\begin{aligned}\hat{\mu}(\hat{\mathbf{w}}^{(\lambda)}) - \mu(\boldsymbol{\vartheta}) \xrightarrow{d} &\ \mathbf{D}_{\boldsymbol{\vartheta}}'\mathbf{C}^{(\lambda/2)}(\mathbf{w}^{*(\lambda)})\boldsymbol{\gamma}^{(\lambda)} - h(\boldsymbol{\vartheta})\sum_{m=1}^{M} w^{*(\lambda)}_m|\boldsymbol{\Pi}_{m^c}\boldsymbol{\gamma}^{(\lambda)}|\\ &+ \mathbf{D}_{\boldsymbol{\vartheta}}'\mathbf{V}^{(\lambda/2)}(\mathbf{w}^{*(\lambda)})S_1(\lambda/2),\end{aligned}$$

where $w^{*(\lambda)}_m$ is the m-th element of $\mathbf{w}^{*(\lambda)}$.

(b) *when* $\lambda = 2$,

$$\tilde{C}^{(2)}(\mathbf{w}) = \mathbf{w}'\boldsymbol{\zeta}^{(2)}\mathbf{w} + 2c_0^{(2)}\mathbf{k}'\mathbf{w} \xrightarrow{d} \mathbf{w}'\boldsymbol{\zeta}^{*(2)}\mathbf{w} + 2c_0^{(2)}\mathbf{k}'\mathbf{w},$$

where $\boldsymbol{\zeta}^{*(2)}$ *is an* $M \times M$ *matrix whose* (m,l)*-th element is*

$$\zeta^{*(2)}_{m,l} = \left(\Gamma_m^{(2)}\right)' \Sigma^{(1)}\Gamma_l^{(2)}$$

where

$$\Gamma_m^{(2)} = \mathbf{C}_m^{(1)}\boldsymbol{\delta}^{(2)} + \left(\mathbf{V}_m^{(1)} - \left(\Sigma^{(1)}\right)^{-1}\right) S_1(1)$$

and $\mathbf{C}_m^{(1)}$ *and* $\mathbf{V}_m^{(1)}$ *are defined in Lemma 3.*
Moreover, as $T \to \infty$, *we have*

$$\hat{\mathbf{w}}^{(2)} \xrightarrow{d} \mathbf{w}^{*(2)} = \text{argmin}_{\mathbf{w}\in\mathcal{H}_T}\left(\mathbf{w}'\boldsymbol{\zeta}_*^{(2)}\mathbf{w} + 2c_0^{(2)}\mathbf{k}'\mathbf{w}\right)$$

and

$$\log T\left(\hat{\mu}(\hat{\mathbf{w}}^{(2)}) - \mu(\boldsymbol{\vartheta})\right) \xrightarrow{d} \mathbf{D}'_{\boldsymbol{\vartheta}}\mathbf{C}^{(1)}(\mathbf{w}^{*(2)})\boldsymbol{\delta}^{(2)} + \mathbf{D}'_{\boldsymbol{\vartheta}}\mathbf{V}^{(1)}(\mathbf{w}^{*(2)})S_1(1).$$

(c) *when* $\lambda \in (2, 4)$,

$$\tilde{C}^{(\lambda)}(\mathbf{w}) = \mathbf{w}'\boldsymbol{\zeta}^{(\lambda)}\mathbf{w} + 2\sigma^2\mathbf{k}'\mathbf{w} \xrightarrow{d} \mathbf{w}'\boldsymbol{\zeta}^{*(\lambda)}\mathbf{w} + 2\sigma^2\mathbf{k}'\mathbf{w},$$

where $\boldsymbol{\zeta}^{*(\lambda)}$ *is an* $M \times M$ *matrix whose* (m,l)*-th element is*

$$\zeta^{*(\lambda)}_{m,l} = \left(\Gamma_m^{(\lambda)}\right)' \Sigma\Gamma_l^{(\lambda)}$$

where

$$\Gamma_m^{(\lambda)} = \mathbf{C}_m\boldsymbol{\delta}^{(\lambda)} + \left(\mathbf{V}_m - \Sigma^{-1}\right) S_1(\lambda/2)$$

and $\mathbf{C}_m$ *and* $\mathbf{V}_m$ *are defined in Lemma 3.*
Moreover, as $T \to \infty$, *we have*

$$\hat{\mathbf{w}}^{(\lambda)} \xrightarrow{d} \mathbf{w}^{*(\lambda)} = \text{argmin}_{\mathbf{w}\in\mathcal{H}_T}\left(\mathbf{w}'\boldsymbol{\zeta}^{*(\lambda)}\mathbf{w} + 2\sigma^2\mathbf{k}'\mathbf{w}\right)$$

and

$$T^{1-2/\lambda}\left(\hat{\mu}(\hat{\mathbf{w}}^{(\lambda)}) - \mu(\boldsymbol{\vartheta})\right) \xrightarrow{d} \mathbf{D}'_{\boldsymbol{\vartheta}}\mathbf{C}(\mathbf{w}^{*(\lambda)})\boldsymbol{\delta}^{(\lambda)} + \mathbf{D}'_{\boldsymbol{\vartheta}}\mathbf{V}(\mathbf{w}^{*(\lambda)})S_1(\lambda/2).$$

(d) *when* $\lambda = 4$,

$$\tilde{C}^{(4)}(\mathbf{w}) = \mathbf{w}'\boldsymbol{\zeta}^{(4)}\mathbf{w} + 2\sigma^2\mathbf{k}'\mathbf{w} \xrightarrow{d} \mathbf{w}'\boldsymbol{\zeta}^{*(4)}\mathbf{w} + 2\sigma^2\mathbf{k}'\mathbf{w},$$

where $\boldsymbol{\zeta}^{*(4)}$ *is an* $M \times M$ *matrix whose* (m,l)*-th element is*

$$\zeta^{*(4)}_{m,l} = \left(\Gamma_m^{(4)}\right)' \Sigma\Gamma_l^{(4)}$$

where

$$\Gamma_m^{(4)} = \mathbf{C}_m \boldsymbol{\delta}^{(4)} + \left(\mathbf{V}_m - \Sigma^{-1}\right) \mathbf{N}(\mathbf{0}, \Omega).$$

Moreover, as $T \to \infty$, we have

$$\hat{\mathbf{w}}^{(4)} \xrightarrow{d} \mathbf{w}^{*(4)} = \text{argmin}_{\mathbf{w} \in \mathcal{H}_T} \left(\mathbf{w}' \boldsymbol{\zeta}^{*(4)} \mathbf{w} + 2\sigma^2 \mathbf{k}' \mathbf{w} \right)$$

and

$$\sqrt{T/\log T} \left(\hat{\mu}(\hat{\mathbf{w}}^{(4)}) - \mu(\boldsymbol{\vartheta}) \right) \xrightarrow{d} N(\mathbf{D}_{\boldsymbol{\vartheta}}' \mathbf{C}(\mathbf{w}^{*(4)}) \boldsymbol{\delta}^{(4)}, V(\mathbf{w}^{*(4)})).$$

4.2. Mallows-type model selection estimator

Here we introduce the classical Mallows model selection (MMS) criterion. Recall that

$$\hat{\boldsymbol{\epsilon}}_m = \mathbf{Y} - \mathbf{X}_m \hat{\boldsymbol{\vartheta}}_m$$

is a residual in the m-th submodel. In traditional model selection, [51] suggests that the criterion for selecting a submodel among all candidate models is

$$\min_{m \in H_T} \left(\hat{\boldsymbol{\epsilon}}_m' \hat{\boldsymbol{\epsilon}}_m + 2\sigma^2 k_m \right) =: \min_{m \in H_T} C(m), \tag{22}$$

where H_T is an index set. We mention two remarkable results in the literature of model selection. [48] demonstrates that a model selected upon Mallows's criterion $C(m)$ is asymptotically optimal in homoskedastic linear regression. Later, [44] show that Akaike's and Mallows's criteria can be used for model selection and result in asymptotic optimality for the out-of-sample forecast. As mentioned earlier, the topic of asymptotic optimality will not be discussed here.

We let

$$\mathcal{H}_T^0 = \left\{ \mathbf{w}_1^0, \mathbf{w}_2^0, \ldots, \mathbf{w}_M^0 \right\},$$

where we recall that $\mathbf{w}_m^0$ is an $M \times 1$ vector for $m = 1, 2, \ldots, M$, whose m-th element is one and the others are zeros. We thus have $\mathcal{H}_T^0 \subseteq \mathcal{H}_T$ and (22) can be written as

$$\min_{\mathbf{w} \in \mathcal{H}_T^0} C(\mathbf{w}),$$

where $C(\mathbf{w})$ is defined in (16). Similarly, the proposed Mallows-type model selection (MTMS) criterion is given by

(a) when $\lambda \in (0, 2)$,

$$C^{(\lambda)}(\mathbf{w}) = \left(a_T^{(\lambda)} \right)^{-2} \hat{\boldsymbol{\epsilon}}(\mathbf{w})' \hat{\boldsymbol{\epsilon}}(\mathbf{w}) + 2 \left(c_0^{(\lambda)} E[|z_1|^\lambda] \right)^{2/\lambda} \mathbf{k}' \mathbf{w} \tag{23}$$

for $\mathbf{w} \in \mathcal{H}_T^0$, where $a_T^{(\lambda)} = \left(c_0^{(\lambda)} E|z_1|^\lambda T \right)^{1/\lambda}$ and $c_0^{(\lambda)}$ is defined in (11).

(b) when $\lambda = 2$,

$$C^{(2)}(\mathbf{w}) = \left(\frac{\log T}{c_0^{(2)} T}\right) \hat{\boldsymbol{\epsilon}}(\mathbf{w})'\hat{\boldsymbol{\epsilon}}(\mathbf{w}) + 2c_0^{(2)}\mathbf{k}'\mathbf{w} \tag{24}$$

for $\mathbf{w} \in \mathcal{H}_T^0$.

(c) when $\lambda \in (2, 4)$,

$$C^{(\lambda)}(\mathbf{w}) = \left(T^{1-4/\lambda}\right) \hat{\boldsymbol{\epsilon}}(\mathbf{w})'\hat{\boldsymbol{\epsilon}}(\mathbf{w}) + 2\sigma^2\mathbf{k}'\mathbf{w} \tag{25}$$

for $\mathbf{w} \in \mathcal{H}_T^0$, where $\sigma^2 = E[\epsilon_1^2]$.

(d) when $\lambda = 4$,

$$C^{(4)}(\mathbf{w}) = \left(\frac{1}{\log T}\right) \hat{\boldsymbol{\epsilon}}(\mathbf{w})'\hat{\boldsymbol{\epsilon}}(\mathbf{w}) + 2\sigma^2\mathbf{k}'\mathbf{w} \tag{26}$$

for $\mathbf{w} \in \mathcal{H}_T^0$.

Remark 9. *The advantage of our method is that we have a robust test for $\mu(\boldsymbol{\vartheta})$ in AR-G/GARCH models under model uncertainty according to the MTMA or MTMS criterion. The proposed MTMA approach can be applied to a large class of AR-G/GARCH models, and captures correlation, heterogeneity, and heavy tails (caused by G/GARCH noise). Aside from the above, the main reason that we use the term* robust *is because the value of δ and the sharpness of λ do not affect the performance of the test.*

5. Simulation study

To verify the applicability of the proposed MTMA and MTMS methods, we conduct Monte Carlo experiments through the following five data-generating processes (DGPs):

1. $AR(1+2)$-$GARCH(1,1)$

$$\begin{aligned} &y_t = \rho_1 y_{t-1} + \gamma_1 y_{t-2} + \gamma_2 y_{t-3} + \epsilon_t, \\ &\text{where } \ \epsilon_t = h_t z_t \ \text{ and } \ h_t^2 = \omega + \alpha\epsilon_{t-1}^2 + \beta h_{t-1}^2. \end{aligned} \tag{27}$$

2. $AR(2+2)$-$GARCH(1,1)$

$$\begin{aligned} &y_t = \rho_1 y_{t-1} + \rho_2 y_{t-2} + \gamma_1 y_{t-3} + \gamma_2 y_{t-4} + \epsilon_t, \\ &\text{where } \ \epsilon_t = h_t z_t \ \text{ and } \ h_t^2 = \omega + \alpha\epsilon_{t-1}^2 + \beta h_{t-1}^2. \end{aligned} \tag{28}$$

3. $AR(2+3)$-$GARCH(1,1)$

$$\begin{aligned} &y_t = \rho_1 y_{t-1} + \rho_2 y_{t-2} + \gamma_1 y_{t-3} + \gamma_2 y_{t-4} + \gamma_3 y_{t-5} + \epsilon_t, \\ &\text{where } \ \epsilon_t = h_t z_t \ \text{ and } \ h_t^2 = \omega + \alpha\epsilon_{t-1}^2 + \beta h_{t-1}^2. \end{aligned} \tag{29}$$

4. $AR(2+5)$-$GARCH(1,1)$

$$y_t = \rho_1 y_{t-1} + \rho_2 y_{t-2} + \gamma_1 y_{t-3} + \gamma_2 y_{t-4} + \gamma_3 y_{t-5} + \gamma_4 y_{t-6} + \gamma_5 y_{t-7} + \epsilon_t,$$
$$\text{where} \quad \epsilon_t = h_t z_t \quad \text{and} \quad h_t^2 = \omega + \alpha\epsilon_{t-1}^2 + \beta h_{t-1}^2. \tag{30}$$

5. $AR(2+2)$-$EGARCH(1,1)$

$$\begin{aligned} y_t &= \rho_1 y_{t-1} + \rho_2 y_{t-2} + \gamma_1 y_{t-3} + \gamma_2 y_{t-4} + \epsilon_t(\sigma), \\ \text{where} \quad \epsilon_t &= h_t z_t \quad \text{and} \quad \log h_t^2 = \omega + \alpha\epsilon_{t-1}^2 + \beta \log h_{t-1}^2. \end{aligned} \tag{31}$$

We set $(\omega, \alpha, \beta) = (0.00000858, 0.072, 0.925)$ for the first four DGPs: (27), (28), (29), and (30) as in [58], whose paper estimates the tail index λ to be approximately 3.2. We likewise adopt $\lambda = 3.2$ in our finite sample study without further explanation or discussion concerning the estimation of λ as this is beyond the scope of our paper. For the last DGP (31), we set $(\omega, \alpha, \beta) = (-0.00127, 0.11605, 0.95)$, following [52]. Besides the above parameters, we have to determine Q in (12) due to our method. For the sake of convenience for simulation, here we set $Q = 3$.

In the experiments of null hypothesis test

$$H_0 : \mu(\boldsymbol{\vartheta}) = \mu(\boldsymbol{\vartheta}_0)$$

against alternative $H_1 : \mu(\boldsymbol{\vartheta}) \neq \mu(\boldsymbol{\vartheta}_0)$, all tests are evaluated at the 5% significance level. We consider the null hypotheses to be $\mu(\boldsymbol{\vartheta}_0) = \rho_{1,0}$, $\mu(\boldsymbol{\vartheta}_0) = \rho_{2,0}$, and $\mu(\boldsymbol{\vartheta}_0) = \rho_{1,0} + \rho_{2,0}$ from Tables 2 to 9. Besides the above hypotheses, we test the ratio of coefficients $\mu(\boldsymbol{\vartheta}_0) = \rho_{1,0}/\rho_{2,0}$ in Tables 4, 5, 6, and 7. The reasons for testing those null hypotheses are as follows: Testing an individual coefficient is for understanding a relationship between y_t and $\{y_{t-k} : k \in \mathbb{N}\}$, see e.g., [69]. The sum of coefficients, in macroeconomics, is frequently used for long-run effects from lag periods on the fiscal policy, international aid, or foreign investment, we refer interested reader to, e.g., [67] and [63]. The quotient of coefficients is a type of long-run propensity in economics, see [67], [37], and [38]. The testing results for DGPs (27), (28), (29), (30), and (31) are in Tables 2, 3, 4, 6, and 8, respectively. In Table 5, the true model is $AR(2+3)$-$GARCH(1,1)$ and all testing base on the models $AR(2+q)$-$GARCH(1,1)$, for $0 \leq q \leq 5$. The purpose for the Table 7 is in order to compare the results of models with and without auxiliary regressors by using DGP (28). Table 9 presents the outcomes by adopting DGP (28) when we use the wrong tail index.

5.1. Finite sample study

We now investigate the robustness of the proposed model averaging (or selection) estimator and compare the performance of the MTMA method to other OLS-based estimators through DGPs (27) to (30). Those OLS-based estimators are built in a single-nested-model framework. More specifically, for DGPs (27) and

(28), we consider three OLS-based single-nested models which include all core regressors but have partial (this could be none or full) auxiliary lag period regressors: (i) the narrow model contains no auxiliary lag period regressors; (ii) a middle model includes only one auxiliary lag period regressor; (iii) the full model has all auxiliary lag period regressors. In DGPs (29) and (30), we add one auxiliary regressor to the model each time in order of the indices of the auxiliary regressors. Except for the MTMA estimator, we also consider the equal-weighted (EW) model, for which the weights for all submodels are all equal.

As we mentioned earlier, we consider null hypothesis test

$$H_0 : \mu(\boldsymbol{\vartheta}) = \mu(\boldsymbol{\vartheta}_0)$$

against alternative $H_1 : \mu(\boldsymbol{\vartheta}) \neq \mu(\boldsymbol{\vartheta}_0)$, where $\mu(\boldsymbol{\vartheta}_0) = \rho_{1,0}$, $\mu(\boldsymbol{\vartheta}_0) = \rho_{2,0}$, $\mu(\boldsymbol{\vartheta}_0) = \rho_{1,0}+\rho_{2,0}$ for all tables and $\mu(\boldsymbol{\vartheta}_0) = \rho_{1,0}/\rho_{2,0}$ for Tables through 4 to 7. All tests are evaluated at the 5% significance level and the simulation studies are based on 10,000 Monte Carlo draws of the sample path via DGPs (27)–(30) with the same initial $y_0 = 0$. We show the rejection rates by choosing the confidence interval proposed in Section 2.3. Summary statistics of the rejection rate can be found in Table 1.

TABLE 1
Statistics of rejection rate

MTMA		EW
Panel A: model averaging		
$\frac{1}{N}\sum_{i=1}^{N} \mathbf{1}_{\{\tau_{H_0}(\hat{\mathbf{w}})\notin \mathrm{cv}_Q(\alpha)\}}(i)$		$\frac{1}{N}\sum_{i=1}^{N} \mathbf{1}_{\{\tau_{H_0}(\mathbf{w}_{ew})\notin \mathrm{cv}_Q(\alpha)\}}(i)$
$AR(p+q_1)$	$\cdots$	$AR(p+q_M)$
Panel B: single model		
$\frac{1}{N}\sum_{i=1}^{N} \mathbf{1}_{\{\tau_{H_0}(1)\notin \mathrm{cv}_Q(\alpha)\}}(i)$	$\cdots$	$\frac{1}{N}\sum_{i=1}^{N} \mathbf{1}_{\{\tau_{H_0}(M)\notin \mathrm{cv}_Q(\alpha)\}}(i)$

Notes: N is the number of experiments, $\mathbf{1}_{\{\cdot\}}$ is the indicator of event $\{\cdot\}$, $\mathbf{1}_{\{\cdot\}}(i)$ is related to the i-th experiment, $\hat{\mathbf{w}} = \mathrm{argmin}_{\mathbf{w}\in\mathcal{H}_T}\tilde{C}^{(\lambda)}(\mathbf{w})$, and $\mathbf{w}_{ew} = (\frac{1}{M},\ldots,\frac{1}{M})$.

Tables 2–7 report the finite-sample rejection rates of two-sided tests at the 5% significance level with various p and q under the same $GARCH(1,1)$ noise. The results in these tables show that the model averaging methods (MTMA and EW), MTMS method, and full model are relatively robust (near 0.05) if compared to some other single models.

First, for $\rho_{1,0} = 0$ under the model $AR(1+2)$-$GARCH(1,1)$, rejection rates look the same among all methods, see Panel A in Table 2. Table 3 shows that the full and middle model outperforms the narrow model. In Tables 3 and 4, null

Table 2. *Rejection rates under* $AR(1+2)$-$GARCH(1,1)$ *model*

		Model averaging		Single model and model selection				Model averaging		Single model and model selection			
δ	T	MTMA	EW	Narrow	Middle	Full	MTMS	MTMA	EW	Narrow	Middle	Full	MTMS
		Panel A: $\mu(\boldsymbol{\vartheta}_0)=\rho_{1,0}=0$						Panel B: $\mu(\boldsymbol{\vartheta}_0)=\rho_{1,0}=0.3$					
0.3	100	0.0485	0.0497	0.0525	0.0460	0.0489	0.0487	0.0499	0.0505	0.0564	0.0471	0.0498	0.0498
	400	0.0527	0.0525	0.0537	0.0516	0.0526	0.0522	0.0512	0.0538	0.0560	0.0541	0.0515	0.0520
	1000	0.0473	0.0466	0.0472	0.0465	0.0482	0.0477	0.0498	0.0484	0.0527	0.0469	0.0474	0.0495
0.5	100	0.0492	0.0494	0.0520	0.0469	0.0498	0.0490	0.0497	0.0535	0.0653	0.0488	0.0500	0.0501
	400	0.0524	0.0536	0.0542	0.0530	0.0524	0.0531	0.0515	0.0545	0.0620	0.0529	0.0512	0.0519
	1000	0.0480	0.0469	0.0460	0.0472	0.0477	0.0478	0.0497	0.0498	0.0591	0.0477	0.0481	0.0503
0.9	100	0.0494	0.0552	0.0574	0.0539	0.0496	0.0490	0.0523	0.0625	0.1080	0.0504	0.0530	0.0522
	400	0.0524	0.0551	0.0578	0.0547	0.0522	0.0531	0.0516	0.0600	0.0880	0.0540	0.0513	0.0512
	1000	0.0493	0.0476	0.0481	0.0479	0.0487	0.0494	0.0496	0.0546	0.0807	0.0463	0.0484	0.0499

Notes: This table reports the finite-sample rejection rates of two-sided tests at the 5% significance level. The parameters for DGP in (27) are $(\omega,\alpha,\beta)=(0.00000858,0.072,0.925)$ and $\gamma_1=\gamma_2=\frac{\delta}{T^{1-2/3.2}}$. The estimator of the tail index of [58] is 3.2. The simulation studies are based on 10,000 replications.

Table 3. *Rejection rates under $AR(2+2)$-$GARCH(1,1)$ model*

δ	T	Model averaging MTMA	Model averaging EW	Single model and model selection Narrow	Single model and model selection Middle	Single model and model selection Full	Single model and model selection MTMS	Model averaging MTMA	Model averaging EW	Single model and model selection Narrow	Single model and model selection Middle	Single model and model selection Full	Single model and model selection MTMS
Panel A: $\mu(\boldsymbol{\vartheta}_0)=\rho_{1,0}=0.3$ with $\boldsymbol{\vartheta}_0=(\rho_{1,0},\rho_{2,0})'=(0.3,0.3)'$								Panel D: $\mu(\boldsymbol{\vartheta}_0)=\rho_{1,0}=0.5$ with $\boldsymbol{\vartheta}_0=(\rho_{1,0},\rho_{2,0})'=(0.5,0.3)'$					
0.3	100	0.0499	0.0493	0.0482	0.0464	0.0492	0.0494	0.0495	0.0487	0.0486	0.0471	0.0500	0.0498
	400	0.0459	0.0459	0.0473	0.0484	0.0458	0.0463	0.0457	0.0451	0.0465	0.0460	0.0450	0.0467
	1000	0.0512	0.0518	0.0499	0.0507	0.0514	0.0490	0.0488	0.0499	0.0487	0.0521	0.0510	0.0477
0.5	100	0.0480	0.0488	0.0524	0.0492	0.0486	0.0491	0.0489	0.0460	0.0657	0.0466	0.0479	0.0493
	400	0.0457	0.0473	0.0534	0.0468	0.0443	0.0449	0.0438	0.0454	0.0563	0.0450	0.0440	0.0438
	1000	0.0511	0.0502	0.0562	0.0509	0.0514	0.0506	0.0481	0.0504	0.0586	0.0510	0.0512	0.0473
0.9	100	0.0471	0.0560	0.1574	0.0518	0.0464	0.0452	0.0491	0.0556	0.1475	0.0530	0.0499	0.0506
	400	0.0456	0.0518	0.0927	0.0473	0.0463	0.0457	0.0427	0.0532	0.1439	0.0451	0.0440	0.0441
	1000	0.0504	0.0529	0.0883	0.0502	0.0511	0.0519	0.0517	0.0569	0.1139	0.0502	0.0519	0.0510
Panel B: $\mu(\boldsymbol{\vartheta}_0)=\rho_{2,0}=0.3$								Panel E: $\mu(\boldsymbol{\vartheta}_0)=\rho_{2,0}=0.3$					
0.3	100	0.0603	0.0570	0.0519	0.0556	0.0611	0.0611	0.0560	0.0533	0.0567	0.0555	0.0566	0.0559
	400	0.0493	0.0498	0.0504	0.0522	0.0511	0.0496	0.0494	0.0481	0.0562	0.0477	0.0508	0.0524
	1000	0.0490	0.0494	0.0515	0.0534	0.0508	0.0485	0.0502	0.0484	0.0629	0.0513	0.0507	0.0494
0.5	100	0.0609	0.0562	0.0551	0.0556	0.0627	0.0616	0.0520	0.0525	0.0955	0.0523	0.0540	0.0523
	400	0.0490	0.0506	0.0605	0.0515	0.0490	0.0487	0.0519	0.0509	0.0844	0.0492	0.0527	0.0538
	1000	0.0497	0.0512	0.0659	0.0530	0.0499	0.0488	0.0502	0.0533	0.0964	0.0517	0.0496	0.0502
0.9	100	0.0558	0.0601	0.1507	0.0554	0.0557	0.0553	0.0482	0.0588	0.1520	0.0574	0.0487	0.0487
	400	0.0522	0.0562	0.1105	0.0559	0.0508	0.0524	0.0511	0.0712	0.2606	0.0550	0.0519	0.0509
	1000	0.0514	0.0607	0.1134	0.0592	0.0510	0.0518	0.0578	0.0717	0.2218	0.0566	0.0503	0.0567
Panel C: $\mu(\boldsymbol{\vartheta}_0)=\rho_{1,0}+\rho_{2,0}=0.6$								Panel F: $\mu(\boldsymbol{\vartheta}_0)=\rho_{1,0}+\rho_{2,0}=0.8$					
0.3	100	0.0510	0.0456	0.0613	0.0513	0.0512	0.0514	0.0615	0.0463	0.1570	0.0632	0.0630	0.0626
	400	0.0489	0.0481	0.0671	0.0509	0.0488	0.0482	0.0497	0.0465	0.1230	0.0509	0.0515	0.0499
	1000	0.0496	0.0494	0.0674	0.0498	0.0492	0.0502	0.0517	0.0510	0.1205	0.0487	0.0466	0.0547
0.5	100	0.0486	0.0445	0.1127	0.0485	0.0490	0.0493	0.0645	0.0560	0.7662	0.0561	0.0677	0.0676
	400	0.0498	0.0500	0.0997	0.0494	0.0506	0.0495	0.0522	0.0553	0.3036	0.0475	0.0513	0.0533
	1000	0.0531	0.0577	0.1026	0.0501	0.0495	0.0536	0.0652	0.0664	0.2717	0.0492	0.0481	0.0688
0.9	100	0.0549	0.1036	0.7117	0.0570	0.0557	0.0560	0.0571	0.1076	0.9963	0.0541	0.0597	0.0584
	400	0.0493	0.0781	0.2552	0.0545	0.0513	0.0495	0.0525	0.1663	0.9231	0.0657	0.0556	0.0528
	1000	0.0570	0.0821	0.2350	0.0586	0.0509	0.0548	0.0674	0.1367	0.7254	0.0633	0.0505	0.0681

Notes: This table reports the finite-sample rejection rates of two-sided tests at the 5% significance level. The parameters for DGP in (28) are $(\omega,\alpha,\beta)=(0.00000858,0.072,0.925)$ and $\gamma_1=\gamma_2=\frac{\delta}{T^{1-2/3.2}}$. The estimator of the tail index of [58] is 3.2. The simulation studies are based on 10,000 replications.

Table 4
Rejection rates under $AR(2+5)$-$GARCH(1,1)$ model

δ	T	Model averaging MTMA	Model averaging EW	Single model and model selection $AR(2)$	$AR(3)$	$AR(4)$	$AR(5)$	$AR(6)$	$AR(7)$	MTMS
Panel A: $\mu(\boldsymbol{\vartheta}_0)=\rho_{1,0}=0.5$ with $\boldsymbol{\vartheta}_0=(\rho_{1,0},\rho_{2,0})'=(0.5,0.3)'$										
0.3	100	0.0494	0.0514	0.0988	0.0538	0.0502	0.0486	0.0473	0.0494	0.0492
	400	0.0509	0.0525	0.0897	0.0502	0.0522	0.0528	0.0518	0.0513	0.0505
	1000	0.0485	0.0450	0.0764	0.0464	0.0466	0.0460	0.0473	0.0467	0.0479
Panel B: $\mu(\boldsymbol{\vartheta}_0)=\rho_{2,0}=0.3$ with $\boldsymbol{\vartheta}_0=(\rho_{1,0},\rho_{2,0})'=(0.5,0.3)'$										
0.3	100	0.0537	0.0485	0.1311	0.0598	0.0474	0.0480	0.0539	0.0543	0.0549
	400	0.0512	0.0514	0.1438	0.0586	0.0471	0.0480	0.0484	0.0506	0.0520
	1000	0.0530	0.0492	0.1355	0.0554	0.0470	0.0488	0.0484	0.0493	0.0544
Panel C: $\mu(\boldsymbol{\vartheta}_0)=\rho_{1,0}+\rho_{2,0}=0.8$ with $\boldsymbol{\vartheta}_0=(\rho_{1,0},\rho_{2,0})'=(0.5,0.3)'$										
0.3	100	0.0619	0.0496	0.9615	0.0582	0.0529	0.0516	0.0529	0.0647	0.0632
	400	0.0508	0.0547	0.5787	0.0633	0.0509	0.0510	0.0510	0.0518	0.0514
	1000	0.0661	0.0545	0.4304	0.0594	0.0494	0.0521	0.0517	0.0527	0.0694
Panel D: $\mu(\boldsymbol{\vartheta}_0)=\rho_{1,0}/\rho_{2,0}=5/3$ with $\boldsymbol{\vartheta}_0=(\rho_{1,0},\rho_{2,0})'=(0.5,0.3)'$										
0.3	100	0.0516	0.0559	0.0867	0.0643	0.0454	0.0450	0.0488	0.0517	0.0513
	400	0.0476	0.0498	0.0868	0.0586	0.0436	0.0471	0.0462	0.0445	0.0465
	1000	0.0523	0.0504	0.0821	0.0560	0.0453	0.0436	0.0445	0.0448	0.0523

Notes: This table reports the finite-sample rejection rates of two-sided tests at the 5% significance level. The parameters for DGP in (30) are $(\omega,\alpha,\beta)=(0.00000858,0.072,0.925)$ and $\gamma_1=\gamma_2=\gamma_3=\gamma_4=\gamma_5=\frac{\delta}{T^{1-2/3.2}}$. The estimator of the tail index of [58] is 3.2. The simulation studies are based on 10,000 replications.

Table 5
Rejection rates under $AR(2+3)$-$GARCH(1,1)$ model

δ	T	Model averaging MTMA	Model averaging EW	Single model and model selection $AR(2)$	$AR(3)$	$AR(4)$	$AR(5)$	$AR(6)$	$AR(7)$	MTMS
Panel A: $\mu(\boldsymbol{\vartheta}_0)=\rho_{1,0}=0.5$ with $\boldsymbol{\vartheta}_0=(\rho_{1,0},\rho_{2,0})'=(0.5,0.3)'$										
0.3	100	0.0533	0.0534	0.0602	0.0519	0.0520	0.0527	0.0549	0.0536	0.0535
	400	0.0539	0.0539	0.0547	0.0544	0.0532	0.0530	0.0537	0.0525	0.0520
	1000	0.0463	0.0460	0.0533	0.0469	0.0495	0.0474	0.0485	0.0461	0.0472
Panel B: $\mu(\boldsymbol{\vartheta}_0)=\rho_{2,0}=0.3$ with $\boldsymbol{\vartheta}_0=(\rho_{1,0},\rho_{2,0})'=(0.5,0.3)'$										
0.3	100	0.0574	0.0491	0.0881	0.0565	0.0528	0.0535	0.0579	0.0573	0.0580
	400	0.0492	0.0483	0.0735	0.0477	0.0495	0.0481	0.0508	0.0498	0.0492
	1000	0.0498	0.0462	0.0779	0.0492	0.0476	0.0469	0.0477	0.0470	0.0503
Panel C: $\mu(\boldsymbol{\vartheta}_0)=\rho_{1,0}+\rho_{2,0}=0.8$ with $\boldsymbol{\vartheta}_0=(\rho_{1,0},\rho_{2,0})'=(0.5,0.3)'$										
0.3	100	0.0662	0.0604	0.5864	0.0608	0.0652	0.0678	0.0653	0.0690	0.0685
	400	0.0527	0.0486	0.2183	0.0514	0.0538	0.0543	0.0540	0.0530	0.0540
	1000	0.0601	0.0505	0.1912	0.0510	0.0541	0.0528	0.0523	0.0525	0.0637
Panel D: $\mu(\boldsymbol{\vartheta}_0)=\rho_{1,0}/\rho_{2,0}=5/3$ with $\boldsymbol{\vartheta}_0=(\rho_{1,0},\rho_{2,0})'=(0.5,0.3)'$										
0.3	100	0.0509	0.0577	0.0803	0.0598	0.0474	0.0455	0.0512	0.0503	0.0497
	400	0.0454	0.0470	0.0658	0.0524	0.0428	0.0447	0.0441	0.0445	0.0463
	1000	0.0497	0.0478	0.0640	0.0504	0.0471	0.0447	0.0439	0.0448	0.0488

Notes: This table reports the finite-sample rejection rates of two-sided tests at the 5% significance level. The parameters for DGP in (29) are $(\omega,\alpha,\beta)=(0.00000858,0.072,0.925)$ and $\gamma_1=\gamma_2=\gamma_3=\frac{\delta}{T^{1-2/3.2}}$. The estimator of the tail index of [58] is 3.2. The simulation studies are based on 10,000 replications.

Table 6
Rejection rates under $AR(2+3)$-$GARCH(1,1)$ model

		Model averaging		Single model and model selection				
δ	T	MTMA	EW	$AR(2)$	$AR(3)$	$AR(4)$	$AR(5)$	MTMS
Panel A: $\mu(\boldsymbol{\vartheta}_0)=\rho_{1,0}=0.5$ with $\boldsymbol{\vartheta}_0=(\rho_{1,0},\rho_{2,0})'=(0.5,0.3)'$								
0.3	100	0.0549	0.0537	0.0594	0.0539	0.0552	0.0554	0.0543
	400	0.0507	0.0499	0.0584	0.0501	0.0516	0.0497	0.0504
	1000	0.0487	0.0478	0.0550	0.0474	0.0489	0.0489	0.0473
Panel B: $\mu(\boldsymbol{\vartheta}_0)=\rho_{2,0}=0.3$ with $\boldsymbol{\vartheta}_0=(\rho_{1,0},\rho_{2,0})'=(0.5,0.3)'$								
0.3	100	0.0554	0.0545	0.0916	0.0567	0.0566	0.0561	0.0555
	400	0.0484	0.0467	0.0711	0.0488	0.0486	0.0503	0.0477
	1000	0.0494	0.0517	0.0796	0.0487	0.0467	0.0476	0.0479
Panel C: $\mu(\boldsymbol{\vartheta}_0)=\rho_{1,0}+\rho_{2,0}=0.8$ with $\boldsymbol{\vartheta}_0=(\rho_{1,0},\rho_{2,0})'=(0.5,0.3)'$								
0.3	100	0.0676	0.0550	0.5897	0.0605	0.0635	0.0692	0.0688
	400	0.0539	0.0523	0.2193	0.0487	0.0545	0.0566	0.0575
	1000	0.0589	0.0521	0.2003	0.0521	0.0495	0.0509	0.0601
Panel D: $\mu(\boldsymbol{\vartheta}_0)=\rho_{1,0}/\rho_{2,0}=5/3$ with $\boldsymbol{\vartheta}_0=(\rho_{1,0},\rho_{2,0})'=(0.5,0.3)'$								
0.3	100	0.0471	0.0586	0.0823	0.0615	0.0465	0.0454	0.0456
	400	0.0450	0.0490	0.0659	0.0509	0.0434	0.0443	0.0445
	1000	0.0515	0.0505	0.0669	0.0514	0.0460	0.0463	0.0505

Notes: This table reports the finite-sample rejection rates of two-sided tests at the 5% significance level. The parameters for DGP in (29) are $(\omega,\alpha,\beta)=(0.00000858,0.072,0.925)$ and $\gamma_1=\gamma_2=\gamma_3=\frac{\delta}{T^{1-2/3.2}}$. The estimator of the tail index of [58] is 3.2. The simulation studies are based on 10,000 replications.

has been rejected more in the narrow model, especially in the case of $\mu(\boldsymbol{\vartheta}_0)=\rho_{1,0}+\rho_{2,0}$. The simulation for the case of $\mu(\boldsymbol{\vartheta}_0)=\rho_{1,0}+\rho_{2,0}$ is an application of Lemma 3 since $\rho_{1,0}+\rho_{2,0}$ can be written as $s(\rho_{1,0},\,\rho_{2,0})$, where $s(x_1,\,x_2)=x_1+x_2$ is a smooth function.

Next, we run the tests based on nested models $AR(2+q)$-$GARCH(1,1)$, for $0\le q\le 3$ in Table 6 when the true model is $AR(2+3)$-$GARCH(1,1)$. In many cases, we do not know what the true model is, namely, we are not able to know q. Therefore, we include more models in the pool of candidate models: $AR(2+q)$-$GARCH(1,1)$, where $0\le q\le 5$. The results by using the above pool for true model $AR(2+3)$-$GARCH(1,1)$ are shown in Table 5. Under these set-ups, MTMS, MTMA, EW, and the richest model (i.e., model with greatest q in the pool of models) perform similarly, and model averaging method seems to be a little bit better than the single model for the case of $\rho_{1,0}/\rho_{2,0}$.

Remind that $\delta=0$ means that the true model has no auxiliary regressors. In Table 7, we consider the cases $\delta=0$ versus $\delta=0.6$. When $\delta=0$, the narrow model performs better; when $\delta=0.6$, the rejection rates of the narrow model works poorly. Overall, when δ is large, the rejection rates of the narrow model are liberal, particularly when $\mu(\boldsymbol{\vartheta}_0)=\rho_{1,0}+\rho_{2,0}$.

One might notice that the rejection rates of MTMA are all near 0.05, which shows the advantage of the model averaging approach and in turn strengthens our Theorem 1. Besides, the rejection rates of MTMS are very close to the 5% significance level. This, together with MTMA, can overcome the model uncer-

TABLE 7
Rejection rates under $AR(2+2)$-$GARCH(1,1)$ model

		Model averaging		Single model and model selection			
δ	T	MTMA	EW	Narrow	Middle	Full	MTMS
Panel A: $\mu(\boldsymbol{\vartheta}_0) = \rho_{1,0} = 0.6$ with $\boldsymbol{\vartheta}_0 = (\rho_{1,0}, \rho_{2,0})' = (0.6, 0.3)'$							
0	100	0.0499	0.0534	0.0564	0.0492	0.0493	0.0496
	400	0.0475	0.0494	0.0512	0.0488	0.0462	0.0461
	1000	0.0468	0.0465	0.0470	0.0491	0.0486	0.0477
0.6	100	0.0452	0.0475	0.0812	0.0484	0.0456	0.0459
	400	0.0512	0.0519	0.0817	0.0493	0.0484	0.0502
	1000	0.0483	0.0471	0.0722	0.0454	0.0482	0.0493
Panel B: $\mu(\boldsymbol{\vartheta}_0) = \rho_{2,0} = 0.3$							
0	100	0.0497	0.0544	0.0605	0.0534	0.0492	0.0495
	400	0.0516	0.0554	0.0553	0.0526	0.0491	0.0515
	1000	0.0437	0.0460	0.0502	0.0477	0.0469	0.0421
0.6	100	0.0447	0.0550	0.1065	0.0551	0.0453	0.0454
	400	0.0508	0.0656	0.1614	0.0567	0.0484	0.0496
	1000	0.0526	0.0628	0.1600	0.0511	0.0473	0.0563
Panel C: $\mu(\boldsymbol{\vartheta}_0) = \rho_{1,0} + \rho_{2,0} = 0.9$							
0	100	0.0637	0.0651	0.0457	0.0742	0.0648	0.0636
	400	0.0513	0.0555	0.0488	0.0589	0.0566	0.0515
	1000	0.0373	0.0495	0.0475	0.0495	0.0503	0.0355
0.6	100	0.0649	0.0574	0.9935	0.0517	0.0667	0.0674
	400	0.0602	0.1038	0.9734	0.0534	0.0547	0.0610
	1000	0.0940	0.0898	0.9138	0.0485	0.0495	0.1056
Panel D: $\mu(\boldsymbol{\vartheta}_0) = \rho_{1,0}/\rho_{2,0} = 2$							
0	100	0.0400	0.0407	0.0406	0.0415	0.0394	0.0402
	400	0.0451	0.0428	0.0456	0.0471	0.0460	0.0458
	1000	0.0458	0.0445	0.0468	0.0456	0.0467	0.0460
0.6	100	0.0473	0.0699	0.0968	0.0624	0.0461	0.0462
	400	0.0488	0.0630	0.1108	0.0586	0.0465	0.0486
	1000	0.0553	0.0589	0.1014	0.0546	0.0460	0.0547

Notes: This table reports the finite-sample rejection rates of two-sided tests at the 5% significance level. The parameters for DGP in (28) are $(\omega, \alpha, \beta) = (0.00000858, 0.072, 0.925)$ and $\gamma_1 = \gamma_2 = \frac{\delta}{T^{1-2/3.2}}$. The tail index is 3.2. The simulation studies are based on 10,000 replications.

tainty. Our method can be applied to more general $\mu(\boldsymbol{\vartheta}_0)$ of AR-G/GARCH models under model uncertainty.

Note the method in [62]; hence our method can be applied only to the case of $\rho_{1,0} = 0$; see Section 4 of [62], given the requirement of the zero median condition. The estimator in each submodel causes bias (non-zero median), as does the averaging model. However, our MTMA method proposed in Section 4.1 works well for DGPs. Therefore, it may be that the zero-median condition is unnecessary. We observe that the performance of MTMA is very close to that of EW in most configurations through experiments. In addition, the performance of MTMS is very close to the performance of MTMA and EW in many cases.

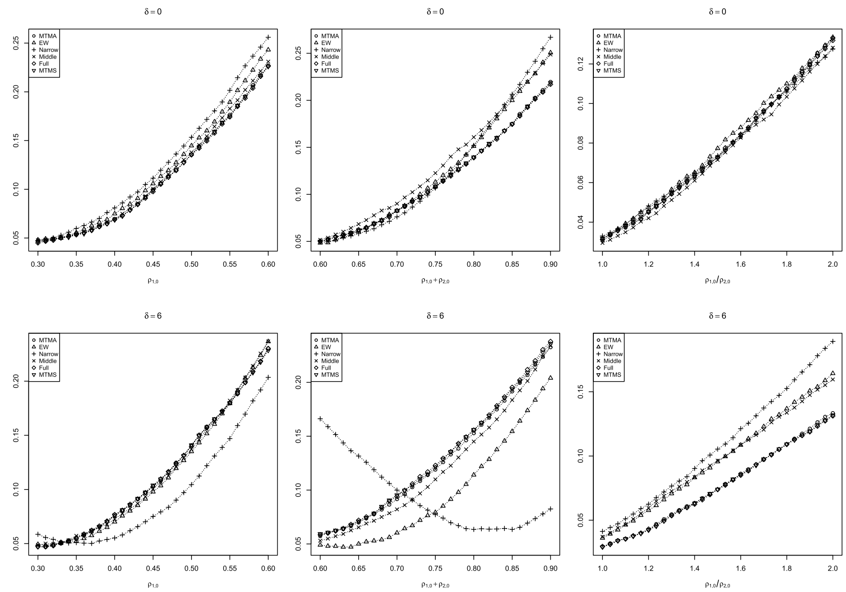

FIG 1. *Power plots under $AR(2+2)$-$GARCH(1,1)$ model when sample size $n = 100$, $\rho_{2,0} = 0.3$, and $\rho_{1,0}$ varies*

Next, we examine the power properties of the statistics. We consider several null $\rho_{1,0}$ taking values from 0.3 to 0.6 with the true coefficients $(\rho_{1,0}, \rho_{2,0}) = (0.3, 0.3)$ of the model $AR(2+2)$-$GARCH(1,1)$. Here we present the results according to sample size $T = 100$, 400 and $\delta = 0$, 0.6. Figures 1 and 2 contain the power curves for MTMA, EW, MTMA, and three signal-model methods. For these results, it seems that the type of the smooth function $\mu(\cdot)$ profoundly impacts the power of test. For example, in the case of $\mu(\boldsymbol{\vartheta}_0) = \rho_{1,0}/\rho_{2,0}$, the narrow model method dominates others, however we have opposite consequences in the case of $\mu(\boldsymbol{\vartheta}_0) = \rho_{1,0}+\rho_{2,0}$. We observe that the power properties are similar under the same hypothesis test except for $\mu(\boldsymbol{\vartheta}_0) = \rho_{1,0}+\rho_{2,0}$ of the narrow model when $\delta = 6$; see Figures 1 and 2. We explain why we have this exception in the following. Remind that the narrow model has no auxiliary regression coefficients, i.e., $\gamma_1 = \gamma_2 = 0$, but the model for DGP $AR(2+2)$-$GARCH(1,1)$ has the local-to-zero auxiliary regression coefficients $\gamma_1 = \gamma_2 = 6/T^{1-2/\lambda}$ where $T = 100$ in Figure 1 and $T = 400$ in Figure 2. As long as we adopt the above DGP to test $\rho_{1,0}+\rho_{2,0}$ by using the narrow model, it probably causes overestimation or underestimation and, in turn, makes power curves so different than other models. Due to the local-to-zero assumption, the effect of the auxiliary regressors decreases if the data size T increases. Therefore, the power curve of the narrow model for $\rho_{1,0}+\rho_{2,0}$ in Figure 1 ($T = 100$) is more discrepant than the power curve of the narrow model for $\rho_{1,0}+\rho_{2,0}$ in Figure 2 ($T = 400$). Moreover, we actually observe that the power curve in the case of $\rho_{1,0}$ is also different from

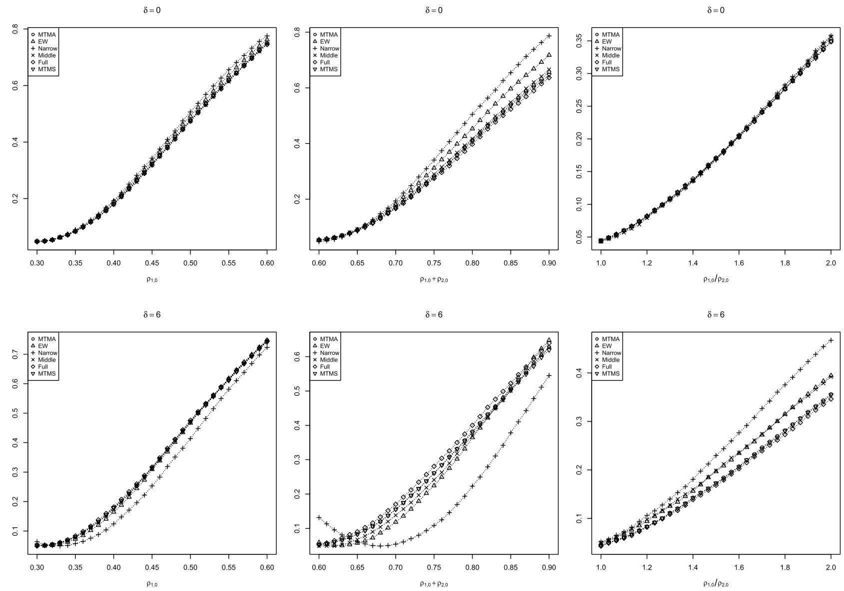

FIG 2. *Power plots under $AR(2+2)$-$GARCH(1,1)$ model when sample size $n = 400$, $\rho_{2,0} = 0.3$, and $\rho_{1,0}$ varies*

other models; the case of $\rho_{2,0}$ should have a similar phenomenon as well. Both together, in synergy, strengthen the effects on the case of $\rho_{1,0} + \rho_{2,0}$. However, we do not see this phenomenon in the case of $\rho_{1,0}/\rho_{2,0}$ probably because of canceling the effects of overestimation or underestimation for $\rho_{1,0}$ and $\rho_{2,0}$.

In summary, the proposed MTMA and MTMS methods are robust in our simulation study under the AR-GARCH model where GARCH noise is known to be heavy-tailed. Moreover, power curves of proposed MTMA and MTMS somehow show the capability and good performance of the test.

5.2. Robustness

Besides the model with polynomial GARCH noise, we also consider exponential GARCH noise. So we employ the DGP (31). The outcomes of simulation by using EGARCH specified in (31) are reported in Table 8. These tables clearly show that the rejection rates using MTMA, EW, and MTMS are very close to 0.05. Through the experiment of DGP (28), we find that MTMA, EW, and MTMS are very close to 0.05 even if we do not know the accurate tail index, see Table 9. Table 8 shows that MTMA and MTMS outperform EW. This observation supports the merit of our proposed criteria in that they will pick a weight informed by the DGP. In addition, in this simulation, the proposed method also accounts for exponential GARCH noise, even if we do not have any tail index information.

Table 8
Rejection rates under $AR(2+2)$-$EGARCH(1,1)$ model

		Model averaging		Single model and model selection			
δ	T	MTMA	EW	Narrow	Middle	Full	MTMS
Panel A: $\mu(\boldsymbol{\vartheta}_0) = \rho_{1,0} = 0.5$ with $\boldsymbol{\vartheta}_0 = (\rho_{1,0}, \rho_{2,0})' = (0.5, 0.3)'$							
0.3	100	0.0492	0.0496	0.0481	0.0480	0.0498	0.0492
	400	0.0428	0.0430	0.0464	0.0420	0.0425	0.0423
	1000	0.0547	0.0537	0.0541	0.0542	0.0548	0.0549
0.5	100	0.0479	0.0489	0.0656	0.0457	0.0475	0.0484
	400	0.0418	0.0428	0.0563	0.0423	0.0421	0.0421
	1000	0.0546	0.0546	0.0650	0.0537	0.0540	0.0538
0.9	100	0.0486	0.0559	0.1433	0.0512	0.0493	0.0494
	400	0.0424	0.0545	0.1490	0.0439	0.0431	0.0428
	1000	0.0539	0.0643	0.1385	0.0546	0.0544	0.0546
Panel B: $\mu(\boldsymbol{\vartheta}_0) = \rho_{2,0} = 0.3$							
0.3	100	0.0566	0.0531	0.0542	0.0534	0.0564	0.0568
	400	0.0505	0.0490	0.0563	0.0472	0.0509	0.0500
	1000	0.0499	0.0495	0.0701	0.0502	0.0500	0.0500
0.5	100	0.0522	0.0521	0.0948	0.0507	0.0523	0.0535
	400	0.0514	0.0548	0.0881	0.0497	0.0525	0.0520
	1000	0.0496	0.0557	0.1123	0.0522	0.0498	0.0497
0.9	100	0.0480	0.0583	0.1524	0.0572	0.0481	0.0480
	400	0.0495	0.0747	0.2632	0.0551	0.0503	0.0506
	1000	0.0494	0.0805	0.2720	0.0582	0.0490	0.0493
Panel C: $\mu(\boldsymbol{\vartheta}_0) = \rho_{1,0} + \rho_{2,0} = 0.8$							
0.3	100	0.0587	0.0439	0.1402	0.0619	0.0599	0.0602
	400	0.0517	0.0498	0.1258	0.0528	0.0516	0.0512
	1000	0.0541	0.0564	0.1471	0.0529	0.0541	0.0548
0.5	100	0.0656	0.0538	0.7430	0.0545	0.0675	0.0680
	400	0.0477	0.0606	0.3012	0.0519	0.0501	0.0494
	1000	0.0543	0.0739	0.3230	0.0529	0.0543	0.0541
0.9	100	0.0593	0.1057	0.9950	0.0511	0.0609	0.0599
	400	0.0498	0.1646	0.8971	0.0653	0.0505	0.0508
	1000	0.0532	0.1638	0.7893	0.0665	0.0536	0.0539

Notes: This table reports the finite-sample rejection rates of two-sided tests at the 5% significance level. The parameters for DGP in (31) are $(\omega, \alpha, \beta) = (-0.00127, 0.11605, 0.95)$ and $\gamma_1 = \gamma_2 = \frac{\delta}{T^{1-2/3.2}}$. The tail index is 3.2. The simulation studies are based on 10,000 replications.

We also show that the proposed methods are still robust even when the estimated tail index is not sharp. We have shown that the scaling in Section 4 depends upon the range of index parameter λ. Therefore, improper scaling might occur if we estimate the tail index inaccurately. Under the setup of [58], the tail index λ is estimated to be 3.2; thus suitable scaling should be approximately $T^{1-\frac{2}{3.2}}$. In our simulation, we experiment with several $\lambda \neq 3.2$ to see the effect of

TABLE 9. *Rejection rates under $AR(2+2)$-$GARCH(1,1)$ model with various wrong λ*

		Model averaging		Single model			Model averaging		Single model			Model averaging		Single model		
λ	T	MTMA	EW	Narrow	Middle	Full	MTMA	EW	Narrow	Middle	Full	MTMA	EW	Narrow	Middle	Full
Panel A: $\mu(\boldsymbol{\vartheta}_0)=\rho_{1,0}=0.5$ with $\boldsymbol{\vartheta}_0=(\rho_{1,0},\rho_{2,0})'=(0.5,0.3)'$							Panel B: $\mu(\boldsymbol{\vartheta}_0)=\rho_{2,0}=0.3$					Panel C: $\mu(\boldsymbol{\vartheta}_0)=\rho_{1,0}+\rho_{2,0}=0.8$				
2.8	100	0.0491	0.0487	0.0486	0.0471	0.0500	0.0569	0.0533	0.0567	0.0555	0.0566	0.0599	0.0463	0.1570	0.0632	0.0630
	400	0.0459	0.0451	0.0465	0.0460	0.0450	0.0503	0.0481	0.0562	0.0477	0.0508	0.0508	0.0465	0.1230	0.0509	0.0515
	1000	0.0484	0.0499	0.0487	0.0521	0.0510	0.0534	0.0484	0.0629	0.0513	0.0507	0.0615	0.0510	0.1205	0.0487	0.0466
3.0	100	0.0494	0.0487	0.0486	0.0471	0.0500	0.0566	0.0533	0.0567	0.0555	0.0566	0.0605	0.0463	0.1570	0.0632	0.0630
	400	0.0462	0.0451	0.0465	0.0460	0.0450	0.0504	0.0481	0.0562	0.0477	0.0508	0.0489	0.0465	0.1230	0.0509	0.0515
	1000	0.0475	0.0499	0.0487	0.0521	0.0510	0.0515	0.0484	0.0629	0.0513	0.0507	0.0556	0.0510	0.1205	0.0487	0.0466
3.4	100	0.0493	0.0487	0.0486	0.0471	0.0500	0.0564	0.0533	0.0567	0.0555	0.0566	0.0571	0.0463	0.1570	0.0632	0.0630
	400	0.0449	0.0451	0.0465	0.0460	0.0450	0.0496	0.0481	0.0562	0.0477	0.0508	0.0507	0.0465	0.1230	0.0509	0.0515
	1000	0.0479	0.0499	0.0487	0.0521	0.0510	0.0455	0.0484	0.0629	0.0513	0.0507	0.0483	0.0510	0.1205	0.0487	0.0466
3.6	100	0.0502	0.0487	0.0486	0.0471	0.0500	0.0572	0.0533	0.0567	0.0555	0.0566	0.0560	0.0463	0.1570	0.0632	0.0630
	400	0.0456	0.0451	0.0465	0.0460	0.0450	0.0496	0.0481	0.0562	0.0477	0.0508	0.0510	0.0465	0.1230	0.0509	0.0515
	1000	0.0515	0.0499	0.0487	0.0521	0.0510	0.0489	0.0484	0.0629	0.0513	0.0507	0.0475	0.0510	0.1205	0.0487	0.0466
4.0	100	0.0497	0.0487	0.0486	0.0471	0.0500	0.0577	0.0533	0.0567	0.0555	0.0566	0.0567	0.0463	0.1570	0.0632	0.0630
	400	0.0460	0.0451	0.0465	0.0460	0.0450	0.0500	0.0481	0.0562	0.0477	0.0508	0.0510	0.0465	0.1230	0.0509	0.0515
	1000	0.0500	0.0499	0.0487	0.0521	0.0510	0.0495	0.0484	0.0629	0.0513	0.0507	0.0471	0.0510	0.1205	0.0487	0.0466
4.0	100	0.0512	0.0487	0.0486	0.0471	0.0500	0.0527	0.0533	0.0567	0.0555	0.0566	0.0628	0.0463	0.1570	0.0632	0.0630
($\log T$)	400	0.0351	0.0451	0.0465	0.0460	0.0450	0.0408	0.0481	0.0562	0.0477	0.0508	0.0415	0.0465	0.1230	0.0509	0.0515
	1000	0.0486	0.0499	0.0487	0.0521	0.0510	0.0517	0.0484	0.0629	0.0513	0.0507	0.0689	0.0510	0.1205	0.0487	0.0466

Notes: This table reports the finite-sample rejection rates of two-sided tests at the 5% significance level. The parameters for DGP in (28) are $(\omega,\alpha,\beta,\delta)=(0.00000858,0.072,0.925,0.3)$ and $\gamma_1=\gamma_2=\frac{0.3}{T^{1-2/3.2}}$. The simulation studies are based on 10,000 replications. Scaling that is not specified is $T^{1-\frac{2}{\lambda}}$.

using the wrong λ. For the wrong $\lambda = 4$, we use both $\log T$ and $T^{1-\frac{2}{4}}$ as scaling in our simulations: the experimental results of the DGP (28) are displayed in Table 9.

We remind the reader that in Table 9, only MTMA and MTMS require the scaling parameter λ. Model averaging, EW, and model selection approaches perform similarly, with rejection rates near 0.05, which are better than the full model, even though we know the full model is the true DGP.

6. Conclusion

In this paper, we consider the AR-G/GARCH model and propose a robust test of a hypothesis for the MTMA and MTMS approaches over models with various lag period explanatory variables under model uncertainty. We use Monte Carlo experiments to demonstrate that the MTMA and MTMS methods work well for testing smooth functions of autocoefficients in the AR-G/GARCH models. In our method, we combine the limiting distribution of the averaging estimator with fixed weights and the Mallows-type criterion for choosing weights. The contribution in the asymptotic results of this paper is effective even when the Mallows-type criterion is replaced by any other suitable criterion. When testing the null hypothesis

$$H_0 : \mu(\boldsymbol{\vartheta}) = \mu(\boldsymbol{\vartheta}_0)$$

against $H_1 : \mu(\boldsymbol{\vartheta}) \neq \mu(\boldsymbol{\vartheta}_0)$, to simplify the choice of weights in empirical work, our simulations suggest that the EW method might be a good option.

It would be interesting to extend the same asymptotic optimality properties for the $AR(\infty)$ model in [30] with GARCH-type noise. Another possible extension is to forecast averaging by extending the result in [31] to the AR-GARCH model. As indicated by the associate editor, a general GARCH order is undoubtedly crucial for this literature; this could thus be a fruitful avenue for future research. Reference to this general framework can be found in, e.g., [61].

Appendix A: Proof of Lemma 1

Proof. The proof is adapted from [10] and [80]

First, by [29], we obtain

$$\lim_{y\to\infty} y^{\alpha} P\left[\Lambda(h_1^2) > y\right] = c_0,$$

where

$$c_0 = \frac{E\left[\left(g(z_1) + c(z_t)\Lambda(h_1^2)\right)^{\alpha} - \left(c(z_t)\Lambda(h_1^2)\right)^{\alpha}\right]}{\alpha E\left[(c(z_{t-1}))^{\alpha}\log^{+}(c(z_t))\right]}.$$

Next, we have

$$P\left[h_1 > x\right] = P\left[\Lambda(h_1^2) > \Lambda(x^2)\right] \sim c_0\left(\Lambda(x^2)\right)^{-\alpha}$$

for large x. In addition, the condition $E[(c(z_{t-1}))^{k_0}\log^+(c(z_t))] < \infty$ and $\alpha \in (0, k_0]$ imply $E|z_1|^\alpha < \infty$.

Finally, by Lemma A.2 of [10], we have

$$P[|\epsilon_1| > x] \sim c_0 E|z_1|^\alpha\left[\Lambda(x^2)\right]^{-\alpha}$$

for large x. □

Appendix B: Proof of Lemma 2

Proof. Before we prove Lemma 2, we need the following observation:

$$\begin{aligned}
\hat{\boldsymbol{\vartheta}}_m &= \left(\mathbf{X}_m'\mathbf{X}_m\right)^{-1}\mathbf{X}_m'\mathbf{Y}\\
&= \left(\mathbf{X}_m'\mathbf{X}_m\right)^{-1}\mathbf{X}_m'\left[\bar{\mathbf{Y}}_{1,p}\boldsymbol{\rho} + \bar{\mathbf{Y}}_{1-p,q}\boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m\boldsymbol{\gamma} + \bar{\mathbf{Y}}_{1-p,q}\left(\mathbf{I}_q - \boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m\right)\boldsymbol{\gamma} + \boldsymbol{\epsilon}\right]\\
&= \left(\mathbf{X}_m'\mathbf{X}_m\right)^{-1}\mathbf{X}_m'\mathbf{X}_m\boldsymbol{\vartheta}_m + \left(\mathbf{X}_m'\mathbf{X}_m\right)^{-1}\mathbf{X}_m'\bar{\mathbf{Y}}_{1-p,q}\left(\mathbf{I}_q - \boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m\right)\boldsymbol{\gamma}\\
&\quad + \left(\mathbf{X}_m'\mathbf{X}_m\right)^{-1}\mathbf{X}_m'\boldsymbol{\epsilon}\\
&= \boldsymbol{\vartheta}_m + \left(\mathbf{X}_m'\mathbf{X}_m\right)^{-1}\mathbf{S}_m'\mathbf{X}'\mathbf{X}\mathbf{S}_0\left(\mathbf{I}_q - \boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m\right)\boldsymbol{\gamma}\\
&\quad + \left(\mathbf{X}_m'\mathbf{X}_m\right)^{-1}\mathbf{S}_m'\mathbf{X}'\boldsymbol{\epsilon}.
\end{aligned}$$

The case of $\lambda \in (0, 2)$.

Denote $a_T^{(\lambda)} = \left(c_0^{(\lambda)}E|z_1|^\lambda T\right)^{1/\lambda}$, where $c_0^{(\lambda)}$ is defined in Remark 1. From [80], we have

$$\frac{1}{\left(a_T^{(\lambda)}\right)^2}\mathbf{X}'\mathbf{X} \xrightarrow{p} \Sigma^{(\lambda/2)}$$

and

$$\frac{1}{\left(a_T^{(\lambda)}\right)^2}\mathbf{X}'\boldsymbol{\epsilon} \xrightarrow{d} S_1(\lambda/2).$$

Since matrix $\mathbf{S}_m$ is non-random with elements either 0 and 1, for the m-th submodel we have

$$\frac{1}{\left(a_T^{(\lambda)}\right)^2}\mathbf{X}_m'\mathbf{X}_m \xrightarrow{p} \Sigma_m^{\lambda/2},$$

where $\Sigma_m^{(\lambda/2)} = \mathbf{S}_m'\Sigma^{(\lambda/2)}\mathbf{S}_m$ is nonsingular.

Hence, by the continuous mapping theorem, we obtain

$$\hat{\boldsymbol{\vartheta}}_m - \boldsymbol{\vartheta}_m = \left(\frac{1}{\left(a_T^{(\lambda)}\right)^2}\mathbf{X}_m'\mathbf{X}_m\right)^{-1}\left(\frac{1}{\left(a_T^{(\lambda)}\right)^2}\mathbf{S}_m'\mathbf{X}'\mathbf{X}\mathbf{S}_0\right)\left(\mathbf{I}_q - \boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m\right)\boldsymbol{\gamma}^{(\lambda)}$$

$$
\begin{aligned}
&+\left(\frac{1}{\left(a_T^{(\lambda)}\right)^2}\mathbf{X}_m'\mathbf{X}_m\right)^{-1}\mathbf{S}_m'\left(\frac{1}{\left(a_T^{(\lambda)}\right)^2}\mathbf{X}'\boldsymbol{\epsilon}\right)\\
&\xrightarrow{d}\left(\Sigma_m^{(\lambda/2)}\right)^{-1}\mathbf{S}_m'\Sigma\mathbf{S}_0\left(\mathbf{I}_q-\mathbf{\Pi}_m'\mathbf{\Pi}_m\right)\boldsymbol{\gamma}^{(\lambda)}+\left(\Sigma_m^{(\lambda/2)}\right)^{-1}\mathbf{S}_m'S_1(\lambda/2)\\
&=:\mathbf{A}_m^{(\lambda/2)}\boldsymbol{\delta}^{(\lambda)}+\mathbf{B}_m^{(\lambda/2)}S_1(\lambda/2),
\end{aligned}
$$

where $\mathbf{A}_m^{(\lambda/2)}=\left(\Sigma_m^{(\lambda/2)}\right)^{-1}\mathbf{S}_m'\Sigma\mathbf{S}_0\left(\mathbf{I}_q-\mathbf{\Pi}_m'\mathbf{\Pi}_m\right)$ and $\mathbf{B}_m^{(\lambda/2)}=\left(\Sigma_m^{(\lambda/2)}\right)^{-1}\mathbf{S}_m'$.

Note that the other cases use the same idea, with the only differences being the scaling. For the sake of completeness, we include them below.

The case of $\lambda=2$.

Thanks to [80], we have

$$\frac{1}{c_0^{(2)}T\log T}\mathbf{X}'\mathbf{X}\xrightarrow{p}\Sigma^{(1)}$$

and

$$\frac{1}{c_0^{(2)}T}\mathbf{X}'\boldsymbol{\epsilon}\xrightarrow{d}S_1(1).$$

Since matrix $\mathbf{S}_m$ is non-random with elements either 0 and 1, for the m-th submodel, we have

$$\frac{1}{c_0^{(2)}T\log T}\mathbf{X}_m'\mathbf{X}_m\xrightarrow{p}\Sigma_m^{(1)},$$

where $\Sigma_m^{(1)}=\mathbf{S}_m'\Sigma^{(1)}\mathbf{S}_m$ is nonsingular.

Similarly, by the continuous mapping theorem, we obtain

$$
\begin{aligned}
(\log T)(\hat{\boldsymbol{\vartheta}}_m-\boldsymbol{\vartheta}_m)&=\left(\frac{1}{c_0^{(2)}T\log T}\mathbf{X}_m'\mathbf{X}_m\right)^{-1}\left(\frac{1}{c_0^{(2)}T\log T}\mathbf{S}_m'\mathbf{X}'\mathbf{X}\mathbf{S}_0\right)\\
&\quad\times\left(\mathbf{I}_q-\mathbf{\Pi}_m'\mathbf{\Pi}_m\right)(\log T)\boldsymbol{\gamma}^{(2)}\\
&\quad+\left(\frac{1}{c_0^{(2)}T\log T}\mathbf{X}_m'\mathbf{X}_m\right)^{-1}\mathbf{S}_m'\left(\frac{1}{c_0^{(2)}T}\mathbf{X}'\boldsymbol{\epsilon}\right)\\
&\xrightarrow{d}\left(\Sigma_m^{(1)}\right)^{-1}\mathbf{S}_m'\Sigma_m^{(1)}\mathbf{S}_0\left(\mathbf{I}_q-\mathbf{\Pi}_m'\mathbf{\Pi}_m\right)\boldsymbol{\delta}^{(2)}+\left(\Sigma_m^{(1)}\right)^{-1}\mathbf{S}_m'S_1(1)\\
&=:\mathbf{A}_m^{(1)}\boldsymbol{\delta}^{(2)}+\mathbf{B}_m^{(1)}S_1(1),
\end{aligned}
$$

where $\mathbf{A}_m^{(1)}=\left(\Sigma_m^{(1)}\right)^{-1}\mathbf{S}_m'\Sigma_m^{(1)}\mathbf{S}_0\left(\mathbf{I}_q-\mathbf{\Pi}_m'\mathbf{\Pi}_m\right)$ and $\mathbf{B}_m^{(1)}=\left(\Sigma_m^{(1)}\right)^{-1}\mathbf{S}_m'$.

The case of $\lambda\in(2,4)$.

Again, from [80] (or [46]), we have

$$\frac{1}{T}\mathbf{X}'\mathbf{X}\xrightarrow{a.s.}\Sigma\text{ as }T\to\infty$$

and

$$T^{-2/\lambda}\mathbf{X}'\boldsymbol{\epsilon} \xrightarrow{d} S_1(\lambda/2) \text{ as } T \to \infty.$$

Since matrix $\mathbf{S}_m$ is non-random with elements either 0 and 1, for the m-th submodel we have

$$\frac{1}{T}\mathbf{X}'_m\mathbf{X}_m \xrightarrow{a.s.} \Sigma_m \text{ as } T \to \infty,$$

where $\Sigma_m = \mathbf{S}'_m\Sigma\mathbf{S}_m$ is nonsingular.

The continuous mapping theorem gives us

$$\begin{aligned} T^{1-2/\lambda}(\hat{\boldsymbol{\vartheta}}_m - \boldsymbol{\vartheta}_m) &= \left(\frac{1}{T}\mathbf{X}'_m\mathbf{X}_m\right)^{-1}\left(\frac{1}{T}\mathbf{S}'_m\mathbf{X}'\mathbf{X}\mathbf{S}_0\right)(\mathbf{I}_q - \boldsymbol{\Pi}'_m\boldsymbol{\Pi}_m)\,T^{1-2/\lambda}\boldsymbol{\gamma} \\ &\quad + \left(\frac{1}{T}\mathbf{X}'_m\mathbf{X}_m\right)^{-1}\mathbf{S}'_m\left(T^{-2/\lambda}\mathbf{X}'\boldsymbol{\epsilon}\right) \\ &\xrightarrow{d} (\Sigma_m)^{-1}\mathbf{S}'_m\Sigma\mathbf{S}_0(\mathbf{I}_q - \boldsymbol{\Pi}'_m\boldsymbol{\Pi}_m)\,\boldsymbol{\delta}^{(\lambda)} + (\Sigma_m)^{-1}\mathbf{S}'_m S_1(\lambda/2) \\ &=: \mathbf{A}_m\boldsymbol{\delta}^{(\lambda)} + \mathbf{B}_m S_1(\lambda/2), \end{aligned}$$

where $\mathbf{A}_m = (\Sigma_m)^{-1}\mathbf{S}'_m\Sigma\mathbf{S}_0(\mathbf{I}_q - \boldsymbol{\Pi}'_m\boldsymbol{\Pi}_m)$ and $\mathbf{B}_m = (\Sigma_m)^{-1}\mathbf{S}'_m$.

The case of $\lambda = 4$.

From [80], we have

$$\frac{1}{T}\mathbf{X}'\mathbf{X} \xrightarrow{a.s.} \Sigma$$

and

$$\frac{1}{\sqrt{T\log T}}\mathbf{X}'\boldsymbol{\epsilon} \xrightarrow{d} \mathbf{N}(\mathbf{0}, \Omega),$$

where Ω is defined in Proposition 2. Since matrix $\mathbf{S}_m$ is non-random with elements either 0 and 1, for the m-th submodel we have

$$\frac{1}{T}\mathbf{X}'_m\mathbf{X}_m \xrightarrow{a.s.} \Sigma_m.$$

Similarly, by the continuous mapping theorem, we obtain

$$\begin{aligned} \frac{\sqrt{T}}{\log T}(\hat{\boldsymbol{\vartheta}}_m - \boldsymbol{\vartheta}_m) &= \left(\frac{1}{T}\mathbf{X}'_m\mathbf{X}_m\right)^{-1}\left(\frac{1}{T}\mathbf{S}'_m\mathbf{X}'\mathbf{X}\mathbf{S}_0\right)(\mathbf{I}_q - \boldsymbol{\Pi}'_m\boldsymbol{\Pi}_m)\,\frac{\sqrt{T}}{\log T}\boldsymbol{\gamma}^{(\lambda)} \\ &\quad + \left(\frac{1}{T}\mathbf{X}'_m\mathbf{X}_m\right)^{-1}\mathbf{S}'_m\left(\frac{1}{\sqrt{T\log T}}\mathbf{X}'\boldsymbol{\epsilon}\right) \\ &\xrightarrow{d} (\Sigma_m)^{-1}\mathbf{S}'_m\Sigma\mathbf{S}_0(\mathbf{I}_q - \boldsymbol{\Pi}'_m\boldsymbol{\Pi}_m)\,\boldsymbol{\delta}^{(4)} + (\Sigma_m)^{-1}\mathbf{S}'_m\mathbf{N}(\mathbf{0}, \Omega) \\ &=: \mathbf{N}(\mathbf{A}_m\boldsymbol{\delta}^{(4)}, (\Sigma_m)^{-1}\Omega_m((\Sigma_m)^{-1})'), \end{aligned}$$

where $\mathbf{A}_m = (\Sigma_m)^{-1}\mathbf{S}'_m\Sigma\mathbf{S}_0(\mathbf{I}_q - \boldsymbol{\Pi}'_m\boldsymbol{\Pi}_m)$ and $\Omega_m = \mathbf{S}'_m\Omega\mathbf{S}_m$. The proof is complete. □

Appendix C: Proof of Lemma 3

Proof. Define $\boldsymbol{\gamma}_{m^c}^{(\lambda)} = \{\gamma_j^{(\lambda)} : \gamma_j^{(\lambda)} \notin \boldsymbol{\gamma}_m^{(\lambda)}, \text{for } j = 1, \dots, q\}$, where $\boldsymbol{\gamma}_{m^c}^{(\lambda)}$ is the set of parameters $\gamma_j^{(\lambda)}$ that are not included in submodel m. We write $\mu(\boldsymbol{\vartheta}) := \mu(\boldsymbol{\rho}, \boldsymbol{\gamma}_m^{(\lambda)}, \boldsymbol{\gamma}_{m^c}^{(\lambda)})$ and $\mu(\boldsymbol{\vartheta}_m) := \mu(\boldsymbol{\rho}, \boldsymbol{\gamma}_m^{(\lambda)}, \mathbf{0})$.

The case of $\lambda \in (0, 2)$.

Note that $\boldsymbol{\gamma}^{(\lambda)}$ does not depend on the sample size T. We apply Taylor's theorem on $\mu(\boldsymbol{\vartheta})$ at $(\boldsymbol{\rho}, \boldsymbol{\gamma}_m^{(\lambda)}, \mathbf{0})$ to obtain

$$\begin{aligned}
\mu(\boldsymbol{\rho}, \boldsymbol{\gamma}_m^{(\lambda)}, \boldsymbol{\gamma}_{m^c}^{(\lambda)}) &= \mu(\boldsymbol{\rho}, \boldsymbol{\gamma}_m^{(\lambda)}, \mathbf{0}) + \mathbf{D}'_{\boldsymbol{\gamma}_{m^c}} \boldsymbol{\gamma}_{m^c}^{(\lambda)} + h(\boldsymbol{\rho}, \boldsymbol{\gamma}_m^{(\lambda)}, \boldsymbol{\gamma}_{m^c}^{(\lambda)}) |\boldsymbol{\Pi}_{m^c} \boldsymbol{\gamma}^{(\lambda)}| \\
&= \mu(\boldsymbol{\rho}, \boldsymbol{\gamma}_m^{(\lambda)}, \mathbf{0}) + \mathbf{D}'_{\boldsymbol{\gamma}} \left(\mathbf{I}_q - \boldsymbol{\Pi}'_m \boldsymbol{\Pi}_m\right) \boldsymbol{\gamma}^{(\lambda)} + h(\boldsymbol{\vartheta}) |\boldsymbol{\Pi}_{m^c} \boldsymbol{\gamma}^{(\lambda)}|.
\end{aligned}$$

Hence

$$\mu(\boldsymbol{\vartheta}) - \mu(\boldsymbol{\vartheta}_m) = \mathbf{D}'_{\boldsymbol{\gamma}} \left(\mathbf{I}_q - \boldsymbol{\Pi}'_m \boldsymbol{\Pi}_m\right) \boldsymbol{\gamma}^{(\lambda)} + h(\boldsymbol{\vartheta}) |\boldsymbol{\Pi}_{m^c} \boldsymbol{\gamma}^{(\lambda)}|.$$

Next, by Assumption 2 and Lemma 2, together with the delta method, we obtain

$$\begin{aligned}
\mu(\hat{\boldsymbol{\vartheta}}_m) - \mu(\boldsymbol{\vartheta}) &= \left(\mu(\hat{\boldsymbol{\vartheta}}_m) - \mu(\boldsymbol{\vartheta}_m)\right) - \left(\mu(\boldsymbol{\vartheta}) - \mu(\boldsymbol{\vartheta}_m)\right) \\
&\xrightarrow{d} \mathbf{D}'_{\boldsymbol{\vartheta}_m} \left[\mathbf{A}_m^{(\lambda/2)} \boldsymbol{\gamma}^{(\lambda)} + \mathbf{B}_m^{(\lambda/2)} S_1(\lambda/2)\right] \\
&\quad - \mathbf{D}'_{\boldsymbol{\gamma}} \left(\mathbf{I}_q - \boldsymbol{\Pi}'_m \boldsymbol{\Pi}_m\right) \boldsymbol{\gamma}^{(\lambda)} - h(\boldsymbol{\vartheta}) |\boldsymbol{\Pi}_{m^c} \boldsymbol{\gamma}^{(\lambda)}| \\
&= \left[\mathbf{D}'_{\boldsymbol{\vartheta}} \mathbf{S}_m \left(\mathbf{S}'_m \Sigma^{(\lambda/2)} \mathbf{S}_m\right)^{-1} \mathbf{S}'_m \Sigma^{(\lambda/2)} \mathbf{S}_0 - \mathbf{D}'_{\boldsymbol{\vartheta}} \mathbf{S}_0\right] \\
&\quad \times \left(\mathbf{I}_q - \boldsymbol{\Pi}'_m \boldsymbol{\Pi}_m\right) \boldsymbol{\gamma}^{(\lambda)} \\
&\quad - h(\boldsymbol{\vartheta}) |\boldsymbol{\Pi}_{m^c} \boldsymbol{\gamma}| + \mathbf{D}'_{\boldsymbol{\vartheta}} \mathbf{S}_m \left(\mathbf{S}'_m \Sigma^{(\lambda/2)} \mathbf{S}_m\right)^{-1} \mathbf{S}'_m S_1(\lambda/2) \\
&=: \mathbf{D}'_{\boldsymbol{\vartheta}} \mathbf{C}_m^{(\lambda/2)} \boldsymbol{\gamma}^{(\lambda)} - h(\boldsymbol{\vartheta}) |\boldsymbol{\Pi}_{m^c} \boldsymbol{\gamma}^{(\lambda)}| + \mathbf{D}'_{\boldsymbol{\vartheta}} \mathbf{V}_m^{(\lambda/2)} S_1(\lambda/2),
\end{aligned}$$

where $\mathbf{C}_m^{(\lambda/2)} = \left(\mathbf{V}_m \Sigma^{(\lambda/2)} - \mathbf{I}_{p+q}\right) \mathbf{S}_0 \left(\mathbf{I}_q - \boldsymbol{\Pi}'_m \boldsymbol{\Pi}_m\right)$ with $\mathbf{V}_m^{(\lambda/2)} = \mathbf{S}_m \left(\mathbf{S}'_m \Sigma^{(\lambda/2)} \mathbf{S}_m\right)^{-1} \mathbf{S}'_m$. Again, other cases are similar to the case of $(0, 2)$ and proofs are given below for completeness.

The case of $\lambda = 2$.

In this case, $\boldsymbol{\gamma}^{(2)} = O\left(1/\log T\right)$. Again, by Taylor's theorem of $\mu(\boldsymbol{\vartheta})$ at $(\boldsymbol{\rho}, \boldsymbol{\gamma}_m^{(2)}, \mathbf{0})$, we have

$$\begin{aligned}
\mu(\boldsymbol{\vartheta}) &= \mu(\boldsymbol{\rho}, \boldsymbol{\gamma}_m^{(2)}, \mathbf{0}) + \mathbf{D}'_{\boldsymbol{\gamma}_{m^c}^{(2)}} \boldsymbol{\gamma}_{m^c}^{(2)} + O\left(1/(\log T)^2\right) \\
&= \mu(\boldsymbol{\rho}, \boldsymbol{\gamma}_m^{(2)}, \mathbf{0}) + \mathbf{D}'_{\boldsymbol{\gamma}} \left(\mathbf{I}_q - \boldsymbol{\Pi}'_m \boldsymbol{\Pi}_m\right) \boldsymbol{\gamma}^{(2)} + O\left(1/(\log T)^2\right),
\end{aligned}$$

so

$$\mu(\boldsymbol{\vartheta}) - \mu(\boldsymbol{\vartheta}_m) = \mathbf{D}'_{\boldsymbol{\gamma}} \left(\mathbf{I}_q - \boldsymbol{\Pi}'_m \boldsymbol{\Pi}_m\right) \boldsymbol{\gamma}^{(2)} + O\left(1/(\log T)^2\right).$$

Next, by Assumption 2, Lemma 2 and the delta method, we obtain

$$
\begin{aligned}
(\log T)(\mu(\hat{\boldsymbol{\vartheta}}_m) - \mu(\boldsymbol{\vartheta})) &= (\log T)\left(\mu(\hat{\boldsymbol{\vartheta}}_m) - \mu(\boldsymbol{\vartheta}_m)\right) - (\log T)\left(\mu(\boldsymbol{\vartheta}) - \mu(\boldsymbol{\vartheta}_m)\right) \\
&\xrightarrow{d} \mathbf{D}'_{\boldsymbol{\vartheta}_m}\left[\mathbf{A}_m^{(1)}\boldsymbol{\delta}^{(2)} + \mathbf{B}_m^{(1)}S_1(1)\right] - \mathbf{D}'_{\boldsymbol{\gamma}}\left(\mathbf{I}_q - \boldsymbol{\Pi}'_m\boldsymbol{\Pi}_m\right)\boldsymbol{\delta}^{(2)} \\
&= \left[\mathbf{D}'_{\boldsymbol{\vartheta}}\mathbf{S}_m\left(\mathbf{S}'_m\Sigma^{(1)}\mathbf{S}_m\right)^{-1}\mathbf{S}'_m\Sigma^{(1)}\mathbf{S}_0 - \mathbf{D}'_{\boldsymbol{\vartheta}}\mathbf{S}_0\right] \\
&\quad \times \left(\mathbf{I}_q - \boldsymbol{\Pi}'_m\boldsymbol{\Pi}_m\right)\boldsymbol{\delta}^{(2)} \\
&\quad + \mathbf{D}'_{\boldsymbol{\vartheta}}\mathbf{S}_m\left(\mathbf{S}'_m\Sigma^{(1)}\mathbf{S}_m\right)^{-1}\mathbf{S}'_m S_1(1) \\
&=: \mathbf{D}'_{\boldsymbol{\vartheta}}\mathbf{C}_m^{(1)}\boldsymbol{\delta}^{(2)} + \mathbf{D}'_{\boldsymbol{\vartheta}}\mathbf{V}_m^{(1)}S_1(1),
\end{aligned}
$$

where $\mathbf{V}_m^{(1)} = \mathbf{S}_m\left(\mathbf{S}'_m\Sigma^{(1)}\mathbf{S}_m\right)^{-1}\mathbf{S}'_m$ and $\mathbf{C}_m^{(1)} = \left(\mathbf{V}_m\Sigma^{(1)} - \mathbf{I}_{p+q}\right)\mathbf{S}_0(\mathbf{I}_q - \boldsymbol{\Pi}'_m\boldsymbol{\Pi}_m)$.

The case of $\lambda \in (2,4)$.

Recall that $\boldsymbol{\gamma}^{(\lambda)} = O\left(T^{\frac{2}{\lambda}-1}\right)$. We apply Taylor's Theorem on the function $\mu(\boldsymbol{\vartheta})$ at $(\boldsymbol{\rho}, \boldsymbol{\gamma}_m^{(\lambda)}, \mathbf{0})$ to obtain

$$
\begin{aligned}
\mu(\boldsymbol{\vartheta}) &= \mu(\boldsymbol{\rho}, \boldsymbol{\gamma}_m^{(\lambda)}, \mathbf{0}) + \mathbf{D}'_{\boldsymbol{\gamma}_{m^c}}\boldsymbol{\gamma}_{m^c}^{(\lambda)} + O(T^{\frac{4}{\lambda}-2}) \\
&= \mu(\boldsymbol{\rho}, \boldsymbol{\gamma}_m^{(\lambda)}, \mathbf{0}) + \mathbf{D}'_{\boldsymbol{\gamma}}\left(\mathbf{I}_q - \boldsymbol{\Pi}'_m\boldsymbol{\Pi}_m\right)\boldsymbol{\gamma}^{(\lambda)} + O(T^{\frac{4}{\lambda}-2}).
\end{aligned}
$$

Therefore,

$$
\mu(\boldsymbol{\vartheta}) - \mu(\boldsymbol{\vartheta}_m) = \mathbf{D}'_{\boldsymbol{\gamma}}\left(\mathbf{I}_q - \boldsymbol{\Pi}'_m\boldsymbol{\Pi}_m\right)\boldsymbol{\gamma}^{(\lambda)} + O(T^{4/\lambda-2}).
$$

Next, by Assumption 1 together with the delta method, we have

$$
\begin{aligned}
T^{1-2/\lambda}\left(\mu(\hat{\boldsymbol{\vartheta}}_m) - \mu(\boldsymbol{\vartheta})\right) &= T^{1-2/\lambda}\left(\mu(\hat{\boldsymbol{\vartheta}}_m) - \mu(\boldsymbol{\vartheta}_m)\right) - T^{1-2/\lambda}\left(\mu(\boldsymbol{\vartheta}) - \mu(\boldsymbol{\vartheta}_m)\right) \\
&\xrightarrow{d} \mathbf{D}'_{\boldsymbol{\vartheta}_m}\left[\mathbf{A}_m\boldsymbol{\delta}^{(\lambda)} + \mathbf{B}_m S_1(\lambda/2)\right] - \mathbf{D}'_{\boldsymbol{\gamma}}\left(\mathbf{I}_q - \boldsymbol{\Pi}'_m\boldsymbol{\Pi}_m\right)\boldsymbol{\delta}^{(\lambda)} \\
&= \left[\mathbf{D}'_{\boldsymbol{\vartheta}}\mathbf{S}_m\left(\mathbf{S}'_m\Sigma\mathbf{S}_m\right)^{-1}\mathbf{S}'_m\Sigma\mathbf{S}_0 - \mathbf{D}'_{\boldsymbol{\vartheta}}\mathbf{S}_0\right] \\
&\quad \times \left(\mathbf{I}_q - \boldsymbol{\Pi}'_m\boldsymbol{\Pi}_m\right)\boldsymbol{\delta}^{(\lambda)} \\
&\quad + \mathbf{D}'_{\boldsymbol{\vartheta}}\mathbf{S}_m\left(\mathbf{S}'_m\Sigma\mathbf{S}_m\right)^{-1}\mathbf{S}'_m S_1(\lambda/2) \\
&=: \mathbf{D}'_{\boldsymbol{\vartheta}}\mathbf{C}_m\boldsymbol{\delta}^{(\lambda)} + \mathbf{D}'_{\boldsymbol{\vartheta}}\mathbf{V}_m S_1(\lambda/2),
\end{aligned}
$$

where $\mathbf{C}_m = \left(\mathbf{V}_m\Sigma - \mathbf{I}_{p+q}\right)\mathbf{S}_0\left(\mathbf{I}_q - \boldsymbol{\Pi}'_m\boldsymbol{\Pi}_m\right)$ and $\mathbf{V}_m = \mathbf{S}_m\left(\mathbf{S}'_m\Sigma\mathbf{S}_m\right)^{-1}\mathbf{S}'_m$.

The case of $\lambda = 4$.

Recall that $\boldsymbol{\gamma}^{(4)} = O\left(\sqrt{\log T/T}\right)$. Applying Taylor's theorem on $\mu(\boldsymbol{\vartheta})$ at $(\boldsymbol{\rho}, \boldsymbol{\gamma}_m^{(4)}, \mathbf{0})$ gives

$$
\begin{aligned}
\mu(\boldsymbol{\vartheta}) &= \mu(\boldsymbol{\rho}, \boldsymbol{\gamma}_m^{(4)}, \mathbf{0}) + \mathbf{D}'_{\boldsymbol{\gamma}_{m^c}}\boldsymbol{\gamma}_{m^c}^{(4)} + O(\log T/T) \\
&= \mu(\boldsymbol{\rho}, \boldsymbol{\gamma}_m^{(4)}, \mathbf{0}) + \mathbf{D}'_{\boldsymbol{\gamma}}\left(\mathbf{I}_q - \boldsymbol{\Pi}'_m\boldsymbol{\Pi}_m\right)\boldsymbol{\gamma}^{(4)} + O(\log T/T).
\end{aligned}
$$

Namely,

$$\mu(\boldsymbol{\vartheta}) - \mu(\boldsymbol{\vartheta}_m) = \mathbf{D}_{\boldsymbol{\gamma}}' \left(\mathbf{I}_q - \mathbf{\Pi}_m'\mathbf{\Pi}_m\right) \boldsymbol{\gamma}^{(4)} + O(\log T/T).$$

Next, Assumption 1 and the delta method give

$$\begin{aligned}
\sqrt{T/\log T}\left(\mu(\hat{\boldsymbol{\vartheta}}_m) - \mu(\boldsymbol{\vartheta})\right) &= \sqrt{T/\log T}\left(\mu(\hat{\boldsymbol{\vartheta}}_m) - \mu(\boldsymbol{\vartheta}_m)\right) \\
&\quad - \sqrt{T/\log T}\left(\mu(\boldsymbol{\vartheta}) - \mu(\boldsymbol{\vartheta}_m)\right) \\
&\xrightarrow{d} \mathbf{D}_{\boldsymbol{\vartheta}_m}' \left[\mathbf{A}_m \boldsymbol{\delta}^{(4)} + \mathbf{B}_m \mathbf{N}(\mathbf{0}, \Omega)\right] \\
&\quad - \mathbf{D}_{\boldsymbol{\gamma}}' \left(\mathbf{I}_q - \mathbf{\Pi}_m'\mathbf{\Pi}_m\right) \boldsymbol{\delta}^{(4)} \\
&= \left[\mathbf{D}_{\boldsymbol{\vartheta}}'\mathbf{S}_m \left(\mathbf{S}_m'\Sigma\mathbf{S}_m\right)^{-1} \mathbf{S}_m'\Sigma\mathbf{S}_0 - \mathbf{D}_{\boldsymbol{\vartheta}}'\mathbf{S}_0\right] \\
&\quad \times \left(\mathbf{I}_q - \mathbf{\Pi}_m'\mathbf{\Pi}_m\right) \boldsymbol{\delta}^{(4)} \\
&\quad + \mathbf{D}_{\boldsymbol{\vartheta}}'\mathbf{S}_m \left(\mathbf{S}_m'\Sigma\mathbf{S}_m\right)^{-1} \mathbf{S}_m'\mathbf{N}(\mathbf{0}, \Omega) \\
&= N(\mathbf{D}_{\boldsymbol{\vartheta}}'\mathbf{C}_m\boldsymbol{\delta}^{(4)}, \mathbf{D}_{\boldsymbol{\vartheta}}'\mathbf{V}_m\Omega\mathbf{V}_m'\mathbf{D}_{\boldsymbol{\vartheta}}).
\end{aligned}$$

The proof is complete. □

Appendix D: Proof of Lemma 4

Proof. Note that a weighted average of the stable (or normal) distributions is again a stable (or normal respectively) distribution; see Remark 4.

The case of $\lambda \in (0, 2)$.

Recall the fact that the asymptotic distribution of the averaging estimator is a stable distribution if each submodel estimator has a stable distribution. An easy calculation gives

$$\begin{aligned}
\hat{\mu}(\mathbf{w}) - \mu(\boldsymbol{\vartheta}) &= \sum_{m=1}^{M} w_m \left(\mu(\hat{\boldsymbol{\vartheta}}_m) - \mu(\boldsymbol{\vartheta})\right) \\
&\xrightarrow{d} \sum_{m=1}^{M} w_m \left[\mathbf{D}_{\boldsymbol{\vartheta}}'\mathbf{C}_m^{(\lambda/2)}\boldsymbol{\gamma}^{(\lambda)} - h(\boldsymbol{\vartheta})|\mathbf{\Pi}_{m^c}\boldsymbol{\gamma}^{(\lambda)}| + \mathbf{D}_{\boldsymbol{\vartheta}}'\mathbf{V}_m^{(\lambda/2)} S_1(\lambda/2)\right] \\
&= \mathbf{D}_{\boldsymbol{\vartheta}}'\mathbf{C}^{(\lambda/2)}(\mathbf{w})\boldsymbol{\gamma}^{(\lambda)} - h(\boldsymbol{\vartheta}) \sum_{m=1}^{M} w_m |\mathbf{\Pi}_{m^c}\boldsymbol{\gamma}^{(\lambda)}| \\
&\quad + \mathbf{D}_{\boldsymbol{\vartheta}}'\mathbf{V}^{(\lambda/2)}(\mathbf{w}) S_1(\lambda/2),
\end{aligned}$$

where $\mathbf{C}^{(\lambda/2)}(\mathbf{w}) = \sum_{m=1}^{M} w_m \mathbf{C}_m^{(\lambda/2)}$ and $\mathbf{V}^{(\lambda/2)}(\mathbf{w}) = \sum_{m=1}^{M} w_m \mathbf{V}_m^{(\lambda/2)}$. Constants $\mathbf{D}_{\boldsymbol{\vartheta}}'\mathbf{C}^{(\lambda/2)}(\mathbf{w})\boldsymbol{\gamma}^{(\lambda)} - h(\boldsymbol{\vartheta})\sum_{m=1}^{M} w_m |\mathbf{\Pi}_{m^c}\boldsymbol{\gamma}^{(\lambda)}|$ and $\mathbf{D}_{\boldsymbol{\vartheta}}'\mathbf{V}^{(\lambda/2)}(\mathbf{w})$ represent location and scale, respectively.

The case of $\lambda = 2$.

Similar to the case of $\lambda \in (0, 2)$ except for the scaling, we obtain

$$\begin{aligned}\log T\left(\hat{\mu}(\mathbf{w}) - \mu(\boldsymbol{\vartheta})\right) &= \sum_{m=1}^{M} w_m \log T\left(\mu(\hat{\boldsymbol{\vartheta}}_m) - \mu(\boldsymbol{\vartheta})\right) \\ &\xrightarrow{d} \sum_{m=1}^{M} w_m \left[\mathbf{D}_{\boldsymbol{\vartheta}}'\mathbf{C}_m^{(1)}\boldsymbol{\delta}^{(2)} + \mathbf{D}_{\boldsymbol{\vartheta}}'\mathbf{V}_m^{(1)}S_1(1)\right] \\ &= \mathbf{D}_{\boldsymbol{\vartheta}}'\mathbf{C}^{(1)}(\mathbf{w})\boldsymbol{\delta}^{(2)} + \mathbf{D}_{\boldsymbol{\vartheta}}'\mathbf{V}^{(1)}(\mathbf{w})S_1(1),\end{aligned}$$

where $\mathbf{C}^{(1)}(\mathbf{w}) = \sum_{m=1}^{M} w_m \mathbf{C}_m^{(1)}$ and $\mathbf{V}^{(1)}(\mathbf{w}) = \sum_{m=1}^{M} w_m \mathbf{V}_m^{(1)}$.

The case of $\lambda \in (2, 4)$.

Again, we have the following:

$$\begin{aligned}T^{1-2/\lambda}\left(\hat{\mu}(\mathbf{w}) - \mu(\boldsymbol{\vartheta})\right) &= \sum_{m=1}^{M} w_m T^{1-2/\lambda}\left(\mu(\hat{\boldsymbol{\vartheta}}_m) - \mu(\boldsymbol{\vartheta})\right) \\ &\xrightarrow{d} \sum_{m=1}^{M} w_m \left[\mathbf{D}_{\boldsymbol{\vartheta}}'\mathbf{C}_m\boldsymbol{\delta}^{(\lambda)} + \mathbf{D}_{\boldsymbol{\vartheta}}'\mathbf{V}_m S_1(\lambda/2)\right] \\ &= \mathbf{D}_{\boldsymbol{\vartheta}}'\mathbf{C}(\mathbf{w})\boldsymbol{\delta}^{(\lambda)} + \mathbf{D}_{\boldsymbol{\vartheta}}'\mathbf{V}(\mathbf{w})S_1(\lambda/2),\end{aligned}$$

where $\mathbf{C}(\mathbf{w}) = \sum_{m=1}^{M} w_m \mathbf{C}_m$ and $\mathbf{V}(\mathbf{w}) = \sum_{m=1}^{M} w_m \mathbf{V}_m$.

The case of $\lambda = 4$.

The following is true:

$$\begin{aligned}\sqrt{T/\log T}\left(\hat{\mu}(\mathbf{w}) - \mu(\boldsymbol{\vartheta})\right) &= \sum_{m=1}^{M} w_m \sqrt{T/\log T}\left(\mu(\hat{\boldsymbol{\vartheta}}_m) - \mu(\boldsymbol{\vartheta})\right) \\ &\xrightarrow{d} \sum_{m=1}^{M} w_m N(\mathbf{D}_{\boldsymbol{\vartheta}}'\mathbf{C}_m\boldsymbol{\delta}^{(4)}, \mathbf{D}_{\boldsymbol{\vartheta}}'\mathbf{V}_m\Omega\mathbf{V}_m\mathbf{D}_{\boldsymbol{\vartheta}}) \\ &=: \sum_{m=1}^{M} w_m R_m.\end{aligned}$$

Similarly, the asymptotic distribution of the averaging estimator is a weighted average of normal distributions, which is still a normal distribution.

Simple calculation gives us the mean and variance. The mean is

$$\begin{aligned}E\left[\sum_{m=1}^{M} w_m R_m\right] &= \sum_{m=1}^{M} w_m E\left[R_m\right] = \sum_{m=1}^{M} w_m \mathbf{D}_{\boldsymbol{\vartheta}}'\mathbf{C}_m\boldsymbol{\delta}^{(4)} \\ &= \mathbf{D}_{\boldsymbol{\vartheta}}'\sum_{m=1}^{M} w_m \mathbf{C}_m\boldsymbol{\delta}^{(4)} =: \mathbf{D}_{\boldsymbol{\vartheta}}'\mathbf{C}(\mathbf{w})\boldsymbol{\delta}^{(4)},\end{aligned}$$

where $\mathbf{C}(\mathbf{w}) = \sum_{m=1}^{M} w_m \mathbf{C}_m$, and the covariance between any two submodels is

$$\begin{aligned}
\mathrm{Cov}\,(R_m, R_k) &= E\left[(R_m - E[R_m]) \cdot (R_k - E[R_k])\right] \\
&= E\left[\mathbf{D}'_{\boldsymbol{\vartheta}} \mathbf{V}_m \mathbf{R} (\mathbf{D}'_{\boldsymbol{\vartheta}} \mathbf{V}_k \mathbf{R})^{-1}\right] \\
&= \mathbf{D}'_{\boldsymbol{\vartheta}} \mathbf{V}_m E\left[\mathbf{R}\mathbf{R}'\right] \mathbf{V}'_k \mathbf{D}_{\boldsymbol{\vartheta}} \\
&= \mathbf{D}'_{\boldsymbol{\vartheta}} \mathbf{V}_m \Omega \mathbf{V}'_k \mathbf{D}_{\boldsymbol{\vartheta}},
\end{aligned}$$

where $\mathbf{R} \overset{d}{=} \mathbf{N}(\mathbf{0}, \Omega)$. Therefore, the variance is

$$\begin{aligned}
\mathrm{Var}\left(\sum_{m=1}^{M} w_m R_m\right) &= \sum_{m=1}^{M} w_m^2 \mathrm{Var}\,(R_m) + 2\sum\sum_{m \neq k} w_m w_k \mathrm{Cov}\,(R_m, R_k) \\
&= \sum_{m=1}^{M} w_m^2 \mathbf{D}'_{\boldsymbol{\vartheta}} \mathbf{V}_m \Omega \mathbf{V}_m \mathbf{D}_{\boldsymbol{\vartheta}} + 2\sum\sum_{m \neq k} w_m w_k \mathbf{D}'_{\boldsymbol{\vartheta}} \mathbf{V}_m \Omega \mathbf{V}_k \mathbf{D}_{\boldsymbol{\vartheta}} \\
&=: V(\mathbf{w}),
\end{aligned}$$

which completes the proof. □

Appendix E: Proof of Remark 8

In this section, we show two equalities of Remark 8:

$$\hat{\boldsymbol{\epsilon}}'_m \hat{\boldsymbol{\epsilon}} = \hat{\boldsymbol{\epsilon}}' \hat{\boldsymbol{\epsilon}}$$

and

$$\hat{\boldsymbol{\epsilon}}'_m \hat{\boldsymbol{\epsilon}}_l - \hat{\boldsymbol{\epsilon}}' \hat{\boldsymbol{\epsilon}} = (\hat{\boldsymbol{\epsilon}}_m - \hat{\boldsymbol{\epsilon}})'(\hat{\boldsymbol{\epsilon}}_l - \hat{\boldsymbol{\epsilon}}).$$

Proof. We observe

$$\begin{aligned}
\hat{\boldsymbol{\epsilon}}'_m \hat{\boldsymbol{\epsilon}} &= (\mathbf{Y} - \mathbf{X}_m \hat{\boldsymbol{\vartheta}}_m)'(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\vartheta}}) \\
&= \mathbf{Y}'\mathbf{Y} - \mathbf{Y}'\mathbf{X}_m \hat{\boldsymbol{\vartheta}}_m - \mathbf{Y}'\mathbf{X}\hat{\boldsymbol{\vartheta}} + \hat{\boldsymbol{\vartheta}}'_m \mathbf{X}'_m \mathbf{X} \hat{\boldsymbol{\vartheta}},
\end{aligned}$$

and

$$\begin{aligned}
\hat{\boldsymbol{\vartheta}}'_m \mathbf{X}'_m \mathbf{X} \hat{\boldsymbol{\vartheta}} &= \mathbf{Y}'\mathbf{X}_m \left(\mathbf{X}'_m \mathbf{X}_m\right)^{-1} \mathbf{X}'_m \mathbf{X} \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{X}'\mathbf{Y} \\
&= \mathbf{Y}'\mathbf{X}_m \left(\mathbf{X}'_m \mathbf{X}_m\right)^{-1} \mathbf{S}'_m \left[\mathbf{X}'\mathbf{X} \left(\mathbf{X}'\mathbf{X}\right)^{-1}\right] \mathbf{X}'\mathbf{Y} \\
&= \mathbf{Y}'\mathbf{X}_m \left[\left(\mathbf{X}'_m \mathbf{X}_m\right)^{-1} \mathbf{S}'_m \mathbf{X}'\mathbf{Y}\right] \\
&= \mathbf{Y}'\mathbf{X}_m \hat{\boldsymbol{\vartheta}}_m,
\end{aligned}$$

and similarly,

$$\hat{\boldsymbol{\vartheta}}' \mathbf{X}'\mathbf{X} \hat{\boldsymbol{\vartheta}} = \mathbf{Y}'\mathbf{X}\hat{\boldsymbol{\vartheta}}.$$

So, we have

$$\hat{\boldsymbol{\epsilon}}_m'\hat{\boldsymbol{\epsilon}} = \mathbf{Y}'\mathbf{Y} - \mathbf{Y}'\mathbf{X}\hat{\boldsymbol{\vartheta}} = \mathbf{Y}'\mathbf{Y} - \mathbf{Y}'\mathbf{X}\hat{\boldsymbol{\vartheta}} - \mathbf{Y}'\mathbf{X}\hat{\boldsymbol{\vartheta}} + \hat{\boldsymbol{\vartheta}}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\vartheta}} = \hat{\boldsymbol{\epsilon}}'\hat{\boldsymbol{\epsilon}}.$$

Therefore, for each $m, l < M$, we have

$$\hat{\boldsymbol{\epsilon}}_m'\hat{\boldsymbol{\epsilon}}_l - \hat{\boldsymbol{\epsilon}}'\hat{\boldsymbol{\epsilon}} = (\hat{\boldsymbol{\epsilon}}_m - \hat{\boldsymbol{\epsilon}})'(\hat{\boldsymbol{\epsilon}}_l - \hat{\boldsymbol{\epsilon}}).$$ □

Appendix F: Proof of Theorem 1

Proof. Minimizing $\mathbf{w}'\boldsymbol{\zeta}_T^{(\lambda)}\mathbf{w}$ over $\mathbf{w}$ is a convex minimization problem due to its quadratic form; hence it has a unique minimizer $\hat{\mathbf{w}}^{(\lambda)}$. The minimizer $\hat{\mathbf{w}}^{(\lambda)}$ converges in distribution to the minimizer $\mathbf{w}^{*(\lambda)}$ of $\mathbf{w}'\boldsymbol{\zeta}^{*}{}_T^{(\lambda)}\mathbf{w}$ by Theorem 2.7 of [45].

The case of $\lambda \in (0,2)$.

Note that $\zeta_{m,l}^{(\lambda)} = \left(a_T^{(\lambda)}\right)^{-2}(\hat{\boldsymbol{\epsilon}}_m - \hat{\boldsymbol{\epsilon}})'(\hat{\boldsymbol{\epsilon}}_l - \hat{\boldsymbol{\epsilon}})$. Therefore, we obtain

$$\zeta_{m,l}^{(\lambda)} = (\mathbf{S}_m\hat{\boldsymbol{\vartheta}}_m - \hat{\boldsymbol{\vartheta}})'\left[\left(a_T^{(\lambda)}\right)^{-2}\mathbf{X}'\mathbf{X}\right](\mathbf{S}_l\hat{\boldsymbol{\vartheta}}_l - \hat{\boldsymbol{\vartheta}}).$$

Thanks to Proposition 2 and Lemma 2, we have

$$\begin{aligned}
(\mathbf{S}_m\hat{\boldsymbol{\vartheta}}_m - \hat{\boldsymbol{\vartheta}}) &= \mathbf{S}_m(\hat{\boldsymbol{\vartheta}}_m - \boldsymbol{\vartheta}_m) + (\mathbf{S}_m\boldsymbol{\vartheta}_m - \boldsymbol{\vartheta}) - (\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}) \\
&\xrightarrow{d} \mathbf{S}_m\mathbf{A}_m^{(\lambda/2)}\boldsymbol{\gamma}^{(\lambda)} + \mathbf{S}_m\mathbf{B}_m^{(\lambda/2)}S_1(\lambda/2) - \mathbf{S}_0\left(\mathbf{I}_q - \boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m\right)\boldsymbol{\gamma}^{(\lambda)} \\
&\quad - \left(\Sigma^{(\lambda/2)}\right)^{-1}S_1(\lambda/2) \\
&= \left(\mathbf{S}_m\left(\Sigma_m^{(\lambda/2)}\right)^{-1}\mathbf{S}_m'\Sigma^{(\lambda/2)} - \mathbf{I}_{p+q}\right)\mathbf{S}_0\left(\mathbf{I}_q - \boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m\right)\boldsymbol{\gamma}^{(\lambda)} \\
&\quad + \left(\mathbf{S}_m\left(\Sigma_m^{(\lambda/2)}\right)^{-1}\mathbf{S}_m' - \left(\Sigma^{(\lambda/2)}\right)^{-1}\right)S_1(\lambda/2) \\
&= \mathbf{C}_m^{(\lambda/2)}\boldsymbol{\gamma}^{(\lambda)} + \left(\mathbf{V}_m^{(\lambda/2)} - \left(\Sigma^{(\lambda/2)}\right)^{-1}\right)S_1(\lambda/2) \\
&\stackrel{d}{=} \Gamma_m^{(\lambda)}.
\end{aligned}$$

On the other hand, we have

$$\left(a_T^{(\lambda)}\right)^{-2}\mathbf{X}'\mathbf{X} \to \Sigma^{(\lambda/2)}. \tag{F.1}$$

By Slutsky's theorem, we obtain

$$\zeta_{m,l}^{(\lambda)} \xrightarrow{d} \left(\Gamma_m^{(\lambda)}\right)'\Sigma^{(\lambda/2)}\Gamma_l^{(\lambda)}.$$

Finally, since each weight and submodel estimator both can be expressed in terms of the same stable random vector, there is joint convergence in distribution of all $\mu(\hat{\boldsymbol{\vartheta}}_m)$ and $\hat{w}_m^{(\lambda)}$, where $\hat{w}_m^{(\lambda)}$ is the m-th element of $\hat{\mathbf{w}}^{(\lambda)}$. Thus, the second result holds due to (F.1) and Lemma 4.

The case of $\lambda = 2$.

Note that $\zeta_{m,l}^{(2)} = \frac{\log T}{c_0^{(2)} T}(\hat{\boldsymbol{\epsilon}}_m - \hat{\boldsymbol{\epsilon}})'(\hat{\boldsymbol{\epsilon}}_l - \hat{\boldsymbol{\epsilon}})$. Therefore, we obtain

$$\zeta_{m,l}^{(2)} = \log T(\mathbf{S}_m \hat{\boldsymbol{\vartheta}}_m - \hat{\boldsymbol{\vartheta}})' \left[\frac{1}{c_0^{(2)} T \log T} \mathbf{X}'\mathbf{X} \right] \log T(\mathbf{S}_l \hat{\boldsymbol{\vartheta}}_l - \hat{\boldsymbol{\vartheta}}).$$

From Proposition 2 and Lemma 2, we have

$$\begin{aligned}
\log T(\mathbf{S}_m \hat{\boldsymbol{\vartheta}}_m - \hat{\boldsymbol{\vartheta}}) &= \mathbf{S}_m \log T(\hat{\boldsymbol{\vartheta}}_m - \boldsymbol{\vartheta}_m) + \log T(\mathbf{S}_m \boldsymbol{\vartheta}_m - \boldsymbol{\vartheta}) - \log T(\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}) \\
&\xrightarrow{d} \mathbf{S}_m \mathbf{A}_m^{(1)} \boldsymbol{\delta}^{(2)} + \mathbf{S}_m \mathbf{B}_m^{(1)} S_1(1) - \mathbf{S}_0 \left(\mathbf{I}_q - \boldsymbol{\Pi}_m' \boldsymbol{\Pi}_m \right) \boldsymbol{\delta}^{(2)} \\
&\quad - \left(\Sigma^{(1)} \right)^{-1} S_1(1) \\
&= \left(\mathbf{S}_m \left(\Sigma_m^{(1)} \right)^{-1} \mathbf{S}_m' \Sigma^{(1)} - \mathbf{I}_{p+q} \right) \mathbf{S}_0 \left(\mathbf{I}_q - \boldsymbol{\Pi}_m' \boldsymbol{\Pi}_m \right) \boldsymbol{\delta}^{(2)} \\
&\quad + \left(\mathbf{S}_m \left(\Sigma_m^{(1)} \right)^{-1} \mathbf{S}_m' - \left(\Sigma^{(1)} \right)^{-1} \right) S_1(1) \\
&= \mathbf{C}_m^{(1)} \boldsymbol{\delta}^{(2)} + \left(\mathbf{V}_m^{(1)} - \left(\Sigma^{(1)} \right)^{-1} \right) S_1(1) \\
&= \Gamma_m^{(2)},
\end{aligned}$$

in which we use the fact that $\mathbf{S}_0 \boldsymbol{\Pi}_m' = \mathbf{S}_m \left(\mathbf{0}_{p \times q_m}', \mathbf{I}_{q_m} \right)'$. On the other hand, we have

$$\frac{1}{c_0^{(2)} T \log T} \mathbf{X}'\mathbf{X} \to \Sigma^{(1)}.$$

Therefore,

$$\zeta_{m,l}^{(2)} \xrightarrow{d} \left(\Gamma_m^{(2)} \right)' \Sigma^{(1)} \Gamma_l^{(2)} \tag{F.2}$$

by Slutsky's theorem. Finally, since each weight and submodel estimator both can be expressed in terms of the same stable random vector, there is joint convergence in distribution of all $\mu(\hat{\boldsymbol{\vartheta}}_m)$ and $\hat{w}_m^{(\lambda)}$, where $\hat{w}_m^{(\lambda)}$ is the m-th element of $\hat{\mathbf{w}}^{(\lambda)}$. Thus, the second result holds by combining (F.2) and Lemma 4.

The case of $\lambda \in (2, 4)$.

Note that $\zeta_{m,l}^{(\lambda)} = T^{1-4/\lambda}(\hat{\boldsymbol{\epsilon}}_m - \hat{\boldsymbol{\epsilon}})'(\hat{\boldsymbol{\epsilon}}_l - \hat{\boldsymbol{\epsilon}})$. Therefore, we obtain

$$\zeta_{m,l}^{(\lambda)} = T^{1-2/\lambda}(\mathbf{S}_m \hat{\boldsymbol{\vartheta}}_m - \hat{\boldsymbol{\vartheta}})' \left[T^{-1} \mathbf{X}'\mathbf{X} \right] T^{1-2/\lambda}(\mathbf{S}_l \hat{\boldsymbol{\vartheta}}_l - \hat{\boldsymbol{\vartheta}}).$$

Thanks to Proposition 2 and Lemma 2, we have

$$\begin{aligned}
T^{1-2/\lambda}(\mathbf{S}_m \hat{\boldsymbol{\vartheta}}_m - \hat{\boldsymbol{\vartheta}}) &= \mathbf{S}_m T^{1-2/\lambda}(\hat{\boldsymbol{\vartheta}}_m - \boldsymbol{\vartheta}_m) \\
&\quad + T^{1-2/\lambda}(\mathbf{S}_m \boldsymbol{\vartheta}_m - \boldsymbol{\vartheta}) - T^{1-2/\lambda}(\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}) \\
&\xrightarrow{d} \mathbf{S}_m \mathbf{A}_m \boldsymbol{\delta} + \mathbf{S}_m \mathbf{B}_m S_1(\lambda/2) - \mathbf{S}_0 \left(\mathbf{I}_q - \boldsymbol{\Pi}_m' \boldsymbol{\Pi}_m \right) \boldsymbol{\delta}^{(\lambda)}
\end{aligned}$$

$$
\begin{aligned}
&- \Sigma^{-1} S_1(\lambda/2) \\
&= \left(\mathbf{S}_m \Sigma_m^{-1} \mathbf{S}_m' \Sigma - \mathbf{I}_{p+q}\right) \mathbf{S}_0 \left(\mathbf{I}_q - \mathbf{\Pi}_m' \mathbf{\Pi}_m\right) \boldsymbol{\delta}^{(\lambda)} \\
&\quad + \left(\mathbf{S}_m \Sigma_m^{-1} \mathbf{S}_m' - \Sigma^{-1}\right) S_1(\lambda/2) \\
&= \mathbf{C}_m \boldsymbol{\delta}^{(\lambda)} + \left(\mathbf{V}_m - \Sigma^{-1}\right) S_1(\lambda/2) \\
&= \Gamma_m^{(\lambda)}.
\end{aligned}
$$

We also have

$$T^{-1}\mathbf{X}'\mathbf{X} \to \Sigma.$$

By Slutsky's theorem, we obtain

$$\zeta_{m,l}^{(\lambda)} \xrightarrow{d} \left(\Gamma_m^{(\lambda)}\right)' \Sigma \Gamma_l^{(\lambda)}. \tag{F.3}$$

Finally, since the each weight and submodel estimator both can be expressed in terms of the same stable random vector, there is joint convergence in distribution of all $\mu(\hat{\boldsymbol{\vartheta}}_m)$ and $\hat{w}_m^{(\lambda)}$, where $\hat{w}_m^{(\lambda)}$ is the m-th element of $\hat{\mathbf{w}}^{(\lambda)}$. Thus, the second result holds by combining (F.3) and Lemma 4.

The case of $\lambda = 4$.

Note that $\zeta_{m,l}^{(4)} = \frac{1}{\log T}(\hat{\boldsymbol{\epsilon}}_m - \hat{\boldsymbol{\epsilon}})'(\hat{\boldsymbol{\epsilon}}_l - \hat{\boldsymbol{\epsilon}})$. Therefore, we obtain

$$\zeta_{m,l}^{(4)} = \sqrt{T/\log T}(\mathbf{S}_m \hat{\boldsymbol{\vartheta}}_m - \hat{\boldsymbol{\vartheta}})' \left[T^{-1}\mathbf{X}'\mathbf{X}\right] \sqrt{T/\log T}(\mathbf{S}_l \hat{\boldsymbol{\vartheta}}_l - \hat{\boldsymbol{\vartheta}}).$$

By 2 and Lemma 2, we have

$$
\begin{aligned}
\sqrt{T/\log T}(\mathbf{S}_m \hat{\boldsymbol{\vartheta}}_m - \hat{\boldsymbol{\vartheta}}) &= \mathbf{S}_m \sqrt{T/\log T}(\hat{\boldsymbol{\vartheta}}_m - \boldsymbol{\vartheta}_m) + \sqrt{T/\log T}(\mathbf{S}_m \boldsymbol{\vartheta}_m - \boldsymbol{\vartheta}) \\
&\quad - \sqrt{T/\log T}(\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}) \\
&\xrightarrow{d} \mathbf{S}_m \mathbf{A}_m \boldsymbol{\delta} + \mathbf{S}_m \mathbf{B}_m \mathbf{N}(\mathbf{0}, \Omega) - \mathbf{S}_0 \left(\mathbf{I}_q - \mathbf{\Pi}_m' \mathbf{\Pi}_m\right) \boldsymbol{\delta}^{(4)} \\
&\quad - \Sigma^{-1} \mathbf{N}(\mathbf{0}, \Omega) \\
&= \left(\mathbf{S}_m \Sigma_m^{-1} \mathbf{S}_m' \Sigma - \mathbf{I}_{p+q}\right) \mathbf{S}_0 \left(\mathbf{I}_q - \mathbf{\Pi}_m' \mathbf{\Pi}_m\right) \boldsymbol{\delta}^{(4)} \\
&\quad + \left(\mathbf{S}_m \Sigma_m^{-1} \mathbf{S}_m' - \Sigma^{-1}\right) \mathbf{N}(\mathbf{0}, \Omega) \\
&= \mathbf{N}(\mathbf{C}_m \boldsymbol{\delta}^{(4)}, \left(\mathbf{V}_m - \Sigma^{-1}\right) \Omega \left(\mathbf{V}_m - \Sigma^{-1}\right)') \\
&= \Gamma_m^{(4)}.
\end{aligned}
$$

On the other hand, we have

$$T^{-1}\mathbf{X}'\mathbf{X} \to \Sigma.$$

By Slutsky's theorem, we obtain

$$\zeta_{m,l}^{(4)} \xrightarrow{d} \left(\Gamma_m^{(4)}\right)' \Sigma \Gamma_l^{(4)}. \tag{F.4}$$

Finally, since the each weight and submodel estimator both can be expressed in terms of the same normal random vector, there is joint convergence in distribution of all $\mu(\hat{\boldsymbol{\vartheta}}_m)$ and $\hat{w}_m^{(\lambda)}$, where $\hat{w}_m^{(\lambda)}$ is the m-th element of $\hat{\mathbf{w}}^{(\lambda)}$. Hence, the second result holds due to the combination of (F.4) and Lemma 4. □

Acknowledgments

We are thankful for the invaluable suggestions of the associate editor and editors of *Electronic Journal of Statistics*.

Funding

Shang-Yuan Shiu appreciates the financial support provided by grants MOST 109-2115-M-008-006-MY2 and MOST 111-2115-M-008-005-MY2 of the National Science and Technology Council, Taiwan. Hsin-Chieh Wong appreciates the financial support partially provided by grants NSTC 111-2118-M-305-004-MY2 of the National Science and Technology Council, Taiwan and 112-NTPU_ORDA-F-003 of National Taipei University, Taiwan. Hsin-Chieh Wong would also like to thank the University Social Responsibility Program (2023–2024 program), "Promoting a Sustainable Low-Carbon Economic Living Circle in Sanxia-Yingge Area" by the Ministry of Education sponsorship to National Taipei University, for funds supporting this research.

References

[1] Aue, A., Berkes, I., and Horváth, L. (2006). "Strong approximation for the sums of squares of augmented GARCH sequences." *Bernoulli* **12**, 583–608. MR2248229

[2] Aue, A. and Horváth, L. (2007). "A limit theorem for mildly explosive autoregression with stable errors." *Econometric Theory* **23**, 201–220. MR2344396

[3] Basrak, B., Davis, R. A., and Mikosch, T. (2002). "Regular variation of GARCH processes." *Stochastic Processes and their Applications* **99**, 95–115. MR1894253

[4] Bollerslev, T. (1986). "Generalized autoregressive conditional heteroskedasticity." *Journal of Econometrics* **31**, 307–327. MR0853051

[5] Brockwell, P. J. and Davis, R. A. (1991) "Time Series: Theory and Methods." Second Edition. Springer-Verlag. MR1093459

[6] Campbell, J. Y., Lo, A. W., and McKinley, A. C. (1997). "The Econometrics of Financial Markets." Princeton University Press, Princeton.

[7] Cavaliere, G., Georgiev, I., and Taylor, A. M. R. (2018). "Unit root inference for non-stationary linear processes driven by infinite variance innovations." *Econometric Theory* **34**, 302–348. MR3783999

[8] Chan, N. H. (1990). "Inference for near-integrated time series with infinite variance." *Journal of the American Statistical Association* **85**, 1069–1074. MR1134503

[9] Chan, N. H. and Tran, L. T. (1989). "On the first-order autoregressive process with infinite variance." *Econometric Theory* **5**, 354–362. MR1029881

[10] Chan, N. H. and Zhang, R. M. (2010). "Inference for unit-root models with infinite variance GARCH errors." *Statistica Sinica* **20**, 1363–1393. MR2777329

[11] Cheng, X. and Hansen, B. E. (2015). "Forecasting with factor-augmented regression: A frequentist model averaging approach." *Journal of Econometrics* **186**, 280–293. MR3343787

[12] Cheng, T. C. F., Ing, C. K., and Yu, S. H. (2015). "Toward optimal model averaging in regression models with time series errors." *Journal of Econometrics* **189**, 321–334. MR3414903

[13] Claeskens, G. and Hjort, H. L. (2008). "Model Selection and Model Averaging." Cambridge University Press, Cambridge. MR2431297

[14] Davis, R. A. (1996). "Gauss–Newton and M-estimation for ARMA processes with infinite variance." *Stochastic Processes and their Applications* **63**, 75–95. MR1411191

[15] Davis, R. A., Knight, K., and Liu, J. (1992). "M-estimation for autoregressions with infinite variance." *Stochastic Processes and their Applications* **40**, 145–180. MR1145464

[16] Davis, R. A. and Mikosch, T. (1998). "The sample autocorrelations of heavy-tailed processes with applications to ARCH." *Annals of Statistics* **26**, 2049–2080. MR1673289

[17] Duan, J. C. (1997). "Augmented $GARCH(p, q)$ process and its diffusion limit." *Journal of Econometrics* **79**, 97–127. MR1457699

[18] DuMouche, W. H. (1983). "Estimating the stable index c in order to measure tail thickness." *Annals of Statistics* **11**, 1019–1031. MR0865342

[19] Elliott, G., Gargano, A., and Timmermann, A. (2013). "Complete subset regressions." *Journal of Econometrics* **177**, 357–373. MR3118566

[20] Embrechts, P., C. Klüppelberg, and T. Mikosch 1997. "Modelling Extremal Events for Insurance and Finance." Springer, Berlin. MR1458613

[21] Engle, R. F. (1982). "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation." *Econometrica* **50**, 987–1007. MR0666121

[22] Engle, R. F. (1990). "Discussion: Stock market volatility and the crash of 1987." *Review of Financial Studies* **3**, 103–106.

[23] Fama, E. (1965). "The behavior of stock market prices." *Journal of Business* **38**, 34–105.

[24] Fornari, F. and Mele, A. (1997). "Sign- and volatility-switching ARCH models: Theory and applications to international stock markets." *Journal of Applied Econometrics* **12**, 1779–1801.

[25] Franses, P. H. and Van Dijk, R. (1996). "Forecasting stock market volatility using (non-linear) GARCH models." *Journal of Forecasting* **15**, 229–235.

[26] Georgiev, I., Rodrigues, P. M. M., and Taylor, A. M. R. (2017). "Unit root inference and heavy-tailed innovations." *Journal of Time Series Analysis* **38**, 733–768. MR3689443

[27] Geweke, J. (1986). "Modeling persistence of conditional variances: a comment." *Econometric Reviews* **5**, 57–61. MR2364358

[28] Glosten, L. R., Jagannathan, R., and Runkle, D. (1993). "On the relation between the expected value and the volatility of the nominal excess return on stocks." *Journal of Finance* **48**, 1779–1801.

[29] Goldie, C. M. (1991). "Implicit renewal theory and tails of solutions of

random equations." *Annals of Applied Probability* **1**, 126–166. MR1097468

[30] Hansen, B. E. (2007). "Least squares model averaging." *Econometrica* **75**, 1175–1186. MR2333497

[31] Hansen, B. E. (2008). "Least-squares forecast averaging." *Journal of Econometrics* **146**, 342–350. MR2465178

[32] Hansen, B. E. (2009). "Averaging estimators for regressions with a possible structural break." *Econometric Theory* **25**, 1498–1514. MR2557572

[33] Hansen, B. E. (2010). "Averaging estimators for autoregressions with a near unit root." *Journal of Econometrics* **158**, 142–155. MR2671270

[34] Hansen, B. E. (2014). "Model averaging, asymptotic risk, and regressor groups." *Quantitative Economics* **5**, 495–530. MR3286414

[35] Hansen, B. E. and Racine, J. S. (2012). "Jackknife model averaging." *Journal of Econometrics* **167**, 38–46. MR2885437

[36] He, C. and Terasvirta, T. (1999). "Properties of moments of a family of GARCH processes." *Journal of Econometrics* **92**, 173–192. MR1706988

[37] Hirschberg, J. and Lye, J. (2010)."A geometric comparison of the Delta and Fieller Confidence intervals." *The American Statistician* **64**, 234–241. MR2757169

[38] Hirschberg, J. and Lye, J. (2017)."Inverting the indirect—The ellipse and the boomerang: Visualizing the confidence intervals of the structural coefficient from two-stage least squares." *Journal of Econometrics* **199**, 173–183. MR3681024

[39] Hjort, N. L. and Claeskens, G. (2003a). "Frequentist model average estimators." *Journal of American Statistical Association* **98**, 879–899. MR2041481

[40] Hjort, N. L. and Claeskens, G. (2003b). "Rejoinder to the focused information criterion and frequentist model average estimators." *Journal of American Statistical Association* **98**, 938–945. MR2041482

[41] Hoeting, J. A., Madigan, D., Raftery, A. E., and Volinsky, C. T. (1999). "Bayesian model averaging: A tutorial." *Statistical Science* **14**, 382–417. MR1765176

[42] Huang, H., Leng, X., Liu, X., and Peng, L. (2020). "Unified inference for an AR process regardless of finite or infinite variance GARCH errors." *Journal of Financial Econometrics* **18**, 425–470.

[43] Ibragimov, R. and Müller, U. K. (2010). "t-statistic based correlation and heterogeneity robust inference." *Journal of Business & Economic Statistics* **28**, 453–468. MR2732358

[44] Ing, C. K. and Wei, C. Z. (2005). "Order selection for same-realization predictions in autoregressive processes." *Annals of Statistics* **33**, 2423–2474. MR2211091

[45] Kim, J. and Pollard, D. (1990). "Cube root asymptotics." *Annals of Statistics* **18**, 191–219. MR1041391

[46] Lange, T. (2011). "Tail behavior and OLS estimator in AR-GARCH models." *Statistica Sinica* **21**, 1191–1200. MR2827520

[47] Leeb, H. and Pötscher, B. (2005). "Model selection and inference: Facts and fiction." *Econometric Theory* **21**, 21–59. MR2153856

[48] Li, K. C. (1987). "Asymptotic optimality for C_p, C_L, cross-validation and

generalized cross-validation: Discrete index set." *Annals of Statistics* **15**, 958–975. MR0902239
[49] Liu, C. A. (2015). "Distribution theory of the least squares averaging estimator." *Journal of Econometrics* **186**, 142–159. MR3321531
[50] Leung, G. and Barron, A. (2006). "Information theory and mixing least-squares regressions." *IEEE Transactions Information Theory* **52**, 3396–3410. MR2242356
[51] Mallows, C. L. (1973). "Some comments on C_p." *Technometrics* **15**, 661–675.
[52] Malmsten, H. (2004). "Evaluating exponential GARCH models." SSE/EFI Working Paper Series in Economics and Finance 564, Stockholm School of Economics.
[53] Mandelbrot, B. (1969). "Long-run linearity, locally Gaussian process, H-spectra, and infinite variances." *International Economic Review* **10**, 82–111.
[54] McCullough, H. (1978). "Continuous-time processes with stable increments." *Journal of Business* **51**, 601–619.
[55] Mikosch, T. 2003. "Modeling dependence and tails of financial time series." In: Finkenstädt, B., Rootzén, H. (Eds.), Extreme Value Theory and Applications. Chapman and Hall, Chichester, pp. 185–286.
[56] Mikosch, T., Gadrich, T., Klüppelberg, C., and Adler, R. J. (1995). "Parameter estimation for ARMA models with infinite variance innovations." *Annals of Statistics* **23**, 305–326. MR1331670
[57] Mikosch, T. and Stărică, C. (2000). "Limit theory for the sample autocorrelations and extremes of a $GARCH(1,1)$ process." *Annals of Statistics* **28**, 1427–1451. MR1805791
[58] Mikosch, T. and Straumann, D. (2002). "Whittle estimation in a heavy-tailed $GARCH(1,1)$ model." *Stochastic Processes and their Applications* **100**, 187–222. MR1919613
[59] Mikosch, T. and de Vries, C. G. (2013). "Heavy tails of OLS." *Journal of Econometrics* **172**, 205–221. MR3010612
[60] Nelson, D. B. (1991). "Conditional heteroskedasticity in asset returns: A new approach." *Econometrica* **59**, 347–370. MR1097532
[61] Noguchi, K., Aue, A, and Burman, P. (2016). "Exploratory analysis and modeling of stock returns." *Journal of Computational and Graphical Statistics* **25**, 363–381. MR3499685
[62] Pedersen, R. S. (2020). "Robust inference in conditionally heteroskedastic autoregressions." *Econometric Reviews* **39**, 244–259. MR4045797
[63] Pesaran, M. H. (1997). "The role of economic theory in modelling the long run." *Economic Journal* **107**, 178–191.
[64] Pötscher, B. (2006). "The distribution of model averaging estimators and an impossibility result regarding its estimation." *Lecture Notes-Monograph Series* **52**, 113–129. MR2427842
[65] Raftery, A. E., Madigan, D., and Hoeting, J. A. (1997). "Bayesian model averaging for regression models." *Journal of the American Statistical Association* **92**, 179–191. MR1436107
[66] Raftery, A. E. and Zheng, Y. (2003). "Discussion: performance of Bayesian

model averaging." *Journal of the American Statistical Association* **98**, 931–938.

[67] Reed, W. R. and Zhu, M. (2017). "On estimating long-run effects in models with lagged dependent variables." *Economic Modelling* **64**, 302–311.

[68] Resnick, S. (1987). "Extreme Values, Regular Variation, and Point Processes." Applied Probability, Springer. MR0827332

[69] Ruppert, D. (2004)."Statistics and Finance: An Introduction." Springer, New York. MR2247523

[70] Samorodnitsky, G. and Taqqu, M. (1994). "Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance, Stochastic Modeling." New York, NY: Chapman & Hall. MR1280932

[71] Schwert, G. W. (1989). "Why does stock market volatility change over time?" *Journal of Finance* **45**, 1129–1155.

[72] Sentana, E. (1995). "Quadratic ARCH models." *Review of Economic Studies* **62**, 639–661.

[73] Sohn, S. Y. and Lim, M. (2007). "Hierarchical forecasting based on AR-GARCH model in a coherent structure." *European Journal of Operational Research* **176**, 1033–1040.

[74] Taylor, S. (1986). "Modelling Financial Time Series." Wiley.

[75] Wan, A. T. K., Zhang, X., and Zou, G. (2010). "Least squares model averaging by Mallows criterion." *Journal of Econometrics* **156**, 277–283. MR2609932

[76] Wong, H. C., Chung, M. H., Fuh, C. D., and Pang, T. (2022). "Robust tests of stock return predictability under heavy-tailed innovations." *Applied Mathematics-A Journal of Chinese Universities*, to appear.

[77] Wong, H. C. and Tsay, W. J. (2022). "Mallows model averaging estimator for the MIDAS model with Almon polynomial weight." *Statistica Sinica* **32**, 1811–1833. MR4478179

[78] Zakoian, J. M. (1994). "Threshold heteroskedastic models." *Journal of Economic Dynamics and Control* **18**, 931–955. MR1462740

[79] Zhu, K. and Ling, S. (2011). "Global self-weighted and local quasi-maximum exponential likelihood estimators for ARMA-GARCH/IGARCH models." *Annals of Statistics* **39**, 2131–2163. MR2893864

[80] Zhang, R. and Ling, S. (2015). "Asymptotic inference for AR models with heavy-tailed G-GARCH noises." *Econometric Theory* **31**, 880–890. MR3377272

[81] Zhang, X., Wan, A. T. K., and Zou, G. (2013). "Model averaging by jackknife criterion in models with dependent data." *Journal of Econometrics* **174**, 82–94. MR3045021