Finance and Economics Discussion Series

Federal Reserve Board, Washington, D.C.
ISSN 1936-2854 (Print)
ISSN 2767-3898 (Online)

When Tails Are Heavy: The Benefits of Variance-Targeted, Non-Gaussian, Quasi-Maximum Likelihood Estimation of GARCH Models

Todd Prono

2025-075

Please cite this paper as:
Prono, Todd (2025). "When Tails Are Heavy: The Benefits of Variance-Targeted, Non-Gaussian, Quasi-Maximum Likelihood Estimation of GARCH Models," Finance and Economics Discussion Series 2025-075. Washington: Board of Governors of the Federal Reserve System, https://doi.org/10.17016/FEDS.2025.075.

When Tails Are Heavy: The Benefits of Variance-Targeted, Non-Gaussian, Quasi-Maximum Likelihood Estimation of GARCH Models[1]

Todd Prono[2]

This Version: July 2025

Abstract

In heavy-tailed cases, variance targeting the Student's-t estimator proposed in Bollerslev (1987) for the linear GARCH model is shown to be robust to density misspecification, just like the popular Quasi-Maximum Likelihood Estimator (QMLE). The resulting Variance-Targeted, Non-Gaussian, Quasi-Maximum Likelihood Estimator (VTNGQMLE) is shown to possess a stable limit, albeit one that is highly non-Gaussian, with an ill-defined variance. The rate of convergence to this non-standard limit is slow relative $\sqrt{n}$ and dependent upon unknown parameters. Fortunately, the sub-sample bootstrap is applicable, given a carefully constructed normalization. Surprisingly, both Monte Carlo experiments and empirical applications reveal VTNGQMLE to sizably outperform QMLE and other performance-enhancing (relative to QMLE) alternatives. In an empirical application, VTNGQMLE is applied to VIX (option-implied volatility of the S&P 500 Index). The resulting GARCH variance estimates are then used to forecast option-implied volatility of volatility (VVIX), thus demonstrating a link between historical volatility of VIX and risk-neutral volatility-of-volatility.

Keywords: GARCH, VIX, VVIX, heavy tails, robust estimation, variance forecasting, volatility, volatility-of-volatility. JEL codes: C13, C22, C58.

[1]The analysis and conclusions presented herein are those of the author and do not indicate concurrence by either the Federal Reserve Board or the Federal Reserve System. I owe thanks to seminar participants at the *9th International Workshop on Financial Markets and Nonlinear Dynamics* for helpful comments and discussions. I additionally owe thanks to Dong Hwan Oh for (many) detailed discussions and reviews.

[2]Federal Reserve Board; (202) 510-2398, todd.a.prono@frb.gov.

1 Introduction

The linear GARCH model of Bollerslev (1986) remains a workhorse for conditional volatility modelling in financial economics, its applications spanning portfolio formation, derivative pricing, and risk management. For a sequence $\{Y_t\}$, the most popular version of this model states

$$Y_t = \sigma_t \epsilon_t, \qquad \epsilon_t \sim i.i.d.\ D\left(0,\ 1\right), \tag{1}$$

$$\sigma_t^2 = \omega + \alpha Y_{t-1}^2 + \beta \sigma_{t-1}^2, \tag{2}$$

where D is an unknown distribution with probability density function g. The most common method for estimating (1) and (2) involves likelihood methods, which require specification of a proxy density function f, where, it is likely that $f \neq g$. This paper treats D and, therefore, g as latent. In order to increase efficiency in the GARCH model parameter estimates, ex-ante attempts are made to better match the selected f with the heavy-tailed features of data commonly modeled. The intent of these attempts, however, is not to identify g, and, therefore, achieve the Cramer-Rao lower bound. Rather, the intent is to select an f that is "closer" to g than a Gaussian density but that (like a Gaussian density) also maintains robustness in the model parameter estimates, in the (likely) case where $f \neq g$. The desired result is a non-Gaussian GARCH estimator that is robust to density misspecification and more efficient than the Gaussian alternative.

By far, the most popular choice for f is the Gaussian density, in which case, the estimator for (1) and (2) is the quasi-maximum likelihood estimator (QMLE). Explaining this popularity is the robustness of QMLE to density misspecification. Early demonstrations of QMLE as a robust estimator include Lee and Hansen (1994) as well as Lumsdaine (1996), with more recent (and more general) demonstrations including Berkes, Horváth, and Kokoszka (2003), Francq and Zakoïan (2004), and Straumann and Mikosch (2006). These more recent demonstrations also identify $E\left(\epsilon_t^4\right) < \infty$ as (close to) necessary for QMLE to be asymptotically normal.

It is well known that while robust, QMLE is not particularly efficient, especially in cases of a heavy-tailed D. Engle and Gonzalez-Rivera (1991) show, for instance, that a semi-parametric estimator for (1) and (2) bests the efficiency of QMLE by up to 50%. The evidenced wide gap between QMLE and (infeasible) full maximum likelihood estimation has encouraged a literature on GARCH estimators that aims to improve upon the efficiency of QMLE, while maintaining robustness. Examples of this literature include Francq et al. (2011a), Fan et al (2014), and Preminger and Storti (2017). Figures 5 and 11 show tail index estimates for

$\{\widehat{\epsilon}_t\}_{t=1}^T$ from daily S&P 500 (log) returns and VIX levels.[3] By the end of the, respective, data samples, point estimates no longer support $E\left(\epsilon_t^4\right) < \infty$ for S&P 500 (log) returns, while there is no evidence supporting $E\left(\epsilon_t^4\right) < \infty$ for VIX levels. From Hall and Yao (2003), when $E\left(\epsilon_t^4\right) = \infty$, QMLE has a non-Gaussian limit with a reduced rate of convergence (relative to $\sqrt{n}$). In current times, therefore, the efficiency gap for QMLE (when applied to S&P 500 returns and VIX levels, at least) appears even wider than what the literature documents.

In light of the empirical evidence in Figures 5 and 11, selecting f as the (standardized) Student's-t density of Bollerslev (1987) seems like an intuitively appealing choice. In the case where g is (very) heavy tailed but f is Gaussian, the parameter α in (2) has to, in some sense, work doubly-hard controlling for the heavy-tailed features of $\{Y_t\}_{t\in\mathbb{Z}}$ unconditionally. That is, when f is Gaussian, α is the only model parameter capable of capturing these heavy-tailed effects, where those effects source to either "reactivity" in $\left\{\sigma_t^2\right\}_{t\in\mathbb{Z}}$ to the previous period's shock or to static features of $\{\epsilon_t\}_{t\in\mathbb{Z}}$. If, instead, f is the (standardized) Student's-t density, then the additional degree-of-freedom parameter can capture the static tail features of $\{\epsilon_t\}_{t\in\mathbb{Z}}$, allowing α to focus on the dynamic features of $\left\{\sigma_t^2\right\}_{t\in\mathbb{Z}}$. The trouble with selecting f as the (standardized) Student's-t density, however, is that the resulting Non-Gaussian, Quasi-Maximum Likelihood Estimator (NGQMLE) is not robust to density misspecification (see; e.g., Newey and Steigerwald, 1997, and Fan et al., 2014). Specifically, from Fan et al. (2014, Proposition 1), bias in NGQMLE sources to under-identification of the scale of $\{\epsilon_t\}_{t\in\mathbb{Z}}$, when $f \neq g$.[4]

This paper investigates Variance-Targeted NGQMLE (VTNGQMLE) for the model of (1) and (2), where f is the (standardized) Student's-t density of Bollerslev (1987).[5] When g is (relatively) thin tailed such that $E\left(Y_t^4\right) < \infty$, VTNGQMLE is shown to be biased, just like NGQMLE, whenever $f \neq g$. In heavy-tailed cases when $E\left(Y_t^4\right) = \infty$, however, the asymptotic limit of VTNGQMLE becomes dominated by properties of the sample variance (the VT part). Explaining this dominance are different rates of convergence; specifically, the sample variance converges slower than does the likelihood function. As a result, effects from the likelihood function disappear as the sample gets large, rendering VTNGQMLE consistent, even

[3] For a regularly varying random variable, the tail index $\iota > 0$ is a moment supremum; meaning, if ϵ_t is regularly varying, then $E\left|\epsilon_t\right|^p < \infty$ if and only if $p < \iota$ (see; e.g., Resnick, 1987, for an introduction to regular variation).

[4] In the model of (1) and (2), scale of the innovations is given by ω. When f is Gaussian, ω is identified in cases where $f \neq g$. When f is non-Gaussian, identification of ω is no longer guaranteed in these same cases. Moreover, since α can be shown to depend on scale, (potential) lack of identification of ω also impacts α.

[5] See Engle and Mezrich (1996) for the initial proposal of variance-targeted estimation and Francq et al. (2011b) for an investigation into the theoretical properties of VTQMLE.

when $f \neq g$, so long as $\left\{\sigma_t^2\right\}_{t\in\mathbb{Z}}$ is mean stationary.[6,7] Consequently, in heavy-tailed cases, VTNGQMLE is robust to density misspecification, making it a member of the class of robust estimators like Francq et al. (2011a), Fan et al. (2014), and Preminger and Storti (2017).

Vaynman and Beare (2014) show that when $E\left(Y_t^4\right) = \infty$, the limit of VTQMLE is analogously dominated by properties of the sample variance. This paper (i) extends that result to a non-Gaussian likelihood, one that produces inconsistent GARCH parameter estimates in the absence of variance targeting, and (ii) considers additional (very) heavy-tailed cases that are empirically relevant. Specifically, the distributional limit of VTNGQMLE is determined in cases where $E\left(\sigma_t^4\right) = \infty$ and $E\left(\epsilon_t^4\right) = \infty$ but the distribution of $\left\{\sigma_t^2\right\}_{t\in\mathbb{Z}}$ and $\left\{\epsilon_t^2\right\}_{t\in\mathbb{Z}}$, respectively, remain in the domain of attraction of a normal law. Additionally, cases where $E\left(\sigma_t^4\right) = \infty$ and $E\left(\epsilon_t^4\right) = \infty$ but the distribution of $\left\{\sigma_t^2\right\}_{t\in\mathbb{Z}}$ and $\left\{\epsilon_t^2\right\}_{t\in\mathbb{Z}}$, respectively, is in the domain of attraction of a stable law are also considered. In this heaviest-tail case, the distributional limit of VTNGQMLE is shown to jointly depend on extremes from both $\left\{\sigma_t^2\right\}_{t\in\mathbb{Z}}$ and $\left\{\epsilon_t^2\right\}_{t\in\mathbb{Z}}$. In the case where $E\left(\sigma_t^4\right) = \infty$ but $E\left(\epsilon_t^4\right) < \infty$, in contrast, the distributional limit singularly depends on extremes from $\left\{\sigma_t^2\right\}_{t\in\mathbb{Z}}$. Consistent with the logic stated above favoring NGQMLE over QMLE (bias issues aside), simulation results, while confirming both VTQMLE and VTNGQMLE to be consistent in heavy-tailed cases and more efficient than QMLE, even when $f \neq g$, also (strongly) favor VTNGQMLE over VTQMLE, on efficiency grounds, in these same cases.

The distorting properties of the sample variance on VTQMLE are considered a cost, since these properties can prevent VTQMLE from achieving a Gaussian limit. Complicated estimators aimed at dampening the tails of $\left\{Y_t\right\}_{t\in\mathbb{Z}}$ are, thus, proposed so that VTQMLE can retain such a limit (see; e.g., Hill and Renault, 2012). This paper, in contrast, views the distorting properties of the sample variance as a benefit, since those properties enable variance-targeted estimation, generally, and VTNGQMLE, specifically, to be robust to density misspecification. Counter-balancing the non-Gaussian limit of VTNGQMLE in heavy-tailed cases are (iii) beneficial effects from the Student's-t likelihood (effects that are retained in large, though still finite, samples, owing to a relatively slow rate of convergence), and (iv) QMLE also having a non-Gaussian limit of similar, qualitative form, in these same cases (see Hall and Yao, 2003, Theorem 2.1).

Despite its non-Gaussian limit, Monte Carlo experiments reveal VTNGQMLE to perform surprisingly well against the competing robust estimators of both Fan et al. (2014) (hereafter FAN) and Preminger and Storti (2017) (hereafter LSE) . In fact, VTNGQMLE is shown to outperform both estimators in terms

[6] In this case, "large" is relative, in the sense that, owing to a slower rate of convergence, effects from the likelihood function will tend to remain, even in finite samples that are quite "large," by standard convention. This tendency is shown to be a benefit, not a cost, however.

[7] $\left\{Y_t\right\}_{t\in\mathbb{Z}}$ in (1) and (2) needs to be covariance stationary, meaning $\left\{\sigma_t^2\right\}_{t\in\mathbb{Z}}$ cannot follow an IGARCH $(1,1)$ process.

of root-mean-squared and mean-absolute-error in samples as large as $10,000$ observations. Moreover, in empirical, out-of-sample forecasting exercises using S&P 500 (log) returns and VIX levels, VTNGQMLE is shown to outperform both QMLE and FAN.[8]

Empirical applications involve forecasting the volatility of S&P 500 (log) returns and VIX levels. The former represents a standard application in financial econometrics. The latter, however, is more nuanced and leverages a characteristic unique to the S&P 500 Index. That is, for the S&P 500 Index, the following three features are directly observable: (v) the return; (vi) option-implied volatility of the return (VIX); (vii) option-implied volatility of the volatility (VVIX). Using features (vi) and (vii) robust estimation of the model in (1) and (2) on VIX demonstrates that the variance of VIX (viii) evidences rich GARCH effects and (viv) these effects are useful at forecasting option-implied volatility of volatility (VVIX), thus establishing a link between historical VIX variance and risk-neutral volatility-of-volatility.

2 Preliminaries

Define μ as a measure on a locally compact, second countable Hausdorff space E, and let $M_+(E)$ denote a collection of Radon measures on E. For $\overline{\mathbb{R}} = \mathbb{R} \cup \left\{ \begin{array}{cc} -\infty, & \infty \end{array} \right\}$, consider the bounded set $\overline{\mathbb{R}}^d \setminus \{0\}$, where bounded here means bounded away from zero. Also, $B \in B\left(\overline{\mathbb{R}}^d \setminus \{0\}\right)$ denotes a Borel σ-field defined on this bounded set. Lastly, the unit sphere is denoted by $S^{d-1} = \left\{x \in \overline{\mathbb{R}}^d : \ |x| = 1\right\}$.

For an $\mathbb{R}^d$-valued random vector $\mathbf{X}$,

Definition 1 $\mathbf{X}$ *is multivariate regularly varying with tail index* $\kappa_0 \in \left(\begin{array}{cc} 0, & \infty \end{array} \right)$ *if* $\exists$ *a sequence* $\{a_n\} \to \infty$ *and a nonnull* $\mu \in M_+\left(\overline{\mathbb{R}}^d \setminus \{0\}\right)$ *such that*

$$nP\left(a_n^{-1}\mathbf{X} \in \cdot\right) \xrightarrow{v} \mu(\cdot) \qquad \textit{as} \qquad n \to \infty,$$

where "v*" denotes "vague convergence,"*

$$\mu(sB) = s^{-\kappa_0}\mu(B),$$

$\forall\ s > 0$ *and a relatively compact* $B \in B\left(\overline{\mathbb{R}}_+^d \setminus \{0\}\right)$.

[8] In the case of VIX levels, both QMLE and FAN produce implausible estimates, while the estimates from VTNGQMLE remain "in-line" with economic rationale and empirical observation.

In addition, $\delta_{\mathbf{X}}$ denotes the Dirac measure at $\mathbf{X}$; meaning, for some set A, $\delta_{\mathbf{X}}(A) = \left\{ \begin{array}{ll} 0 & \mathbf{X} \notin A \\ 1 & \mathbf{X} \in A \end{array} \right\}$. C denotes a generic constant that can take-on different values in different places. "$\xrightarrow{d}$" denotes (weak) convergence in distribution.

3 The Model and Background Results

Under consideration is the linear GARCH $(1,1)$ model of Bollerslev (1986). Results presented herein can be extended to the general GARCH (p,q) case, where $p,\ q\ \geq\ 1$ (see; e.g., Vaynman and Beare, 2014). Focusing on the special case of $p = q = 1$, besides being the most practically relevant, also facilitates the illustration of key concepts and ideas, as well as the verification of important conditions.

For a sequence $\{Y_t\}_{t\in\mathbb{Z}}$, and a σ-algebra defined for this sequence as denoted by Ω_t,

$$Y_t = \sigma_t \epsilon_t, \qquad \epsilon_t \sim i.i.d.\ D\,(0,\ 1)\,, \tag{3}$$

$$\sigma_t^2 = \omega_0 + \alpha_0 Y_{t-1}^2 + \beta_0 \sigma_{t-1}^2, \tag{4}$$

where D is an unknown probability distribution with associated density function g. ω_0 denotes the true value of ω; ω any one of a set of possible values, and $\widehat{\omega}$ an estimate. Parallel definitions hold for all other parameter values.

ASSUMPTION 3.1.

$$\omega > 0, \qquad \alpha > 0, \qquad \beta \geq 0, \qquad \alpha + \beta < 1.$$

Under Assumption 3.1, the GARCH$(1,1)$ model being considered nests the the ARCH(1) model as a special case. Given (3), (4), and Assumption 3.1,

$$E\left(Y_t^2\right) = E\left(\sigma_t^2\right) = \frac{\omega_0}{1-\alpha_0-\beta_0} < \infty, \tag{5}$$

where, for notational convenience, $E\left(\sigma_t^2\right) = E\left(\sigma^2\right)$. As a result, (4) may be re-written as

$$\sigma_t^2 = E\left(\sigma^2\right)(1-\alpha_0-\beta_0) + \alpha_0 Y_{t-1}^2 + \beta_0 \sigma_{t-1}^2. \tag{6}$$

Let $\mathbf{X}_t = \left(\begin{array}{cc} \sigma_t^2, & Y_t^2 \end{array} \right)'$. Again owing to (3)-(4),

$$\mathbf{X}_t = \mathbf{A}_t \mathbf{X}_{t-1} + \mathbf{B}_t, \tag{7}$$

where

$$\mathbf{A}_t = \begin{pmatrix} \beta_0 & \alpha_0 \\ \beta_0 \epsilon_t^2 & \alpha_0 \epsilon_t^2 \end{pmatrix}, \qquad \mathbf{B}_t = \begin{pmatrix} \omega_0 \\ \omega_0 \epsilon_t^2 \end{pmatrix},$$

represents a stochastic recurrence equation (SRE) (see; e.g., Mikosch and Stărică, 2000, eq. 2.2). As such, Definition 1 is shown to apply (see; e.g., Mikosch and Stărică, 2000, and Basrak et al., 2002).

ASSUMPTION 3.2. *The sequence* $\left\{\epsilon_t^2\right\}_{t\in\mathbb{Z}}$ *is regularly varying, with tail index* ι_0.

Under Assumption 3.2, innovations to the GARCH$(1,1)$ model are heavy-tailed, in the sense that the (unknown) distribution for these innovations belongs to the Fréchet class, as opposed to the more commonly assumed Gumbel class.[9] Regardless of whether GARCH$(1,1)$ model innovations are heavy-tailed in a Fréchet-class sense, or (relatively) thin-tailed in a Gumbel-class sense, X_t will be regularly varying (see Mikosch, 1999, Corollary 1.4.40).

What follows in the remainder of this section is a summary of select (weakly) dependent and heavy-tailed limit theory results, upon which later sections are based. This summary draws heavily from Davis and Mikosch (1998) and Mikosch and Stărică (2000), both of which, in turn, rely on results from Davis and Hsing (1995). The intent of this summary is to introduce certain key results; not provide a comprehensive review. A detailed treatment of these results, as well as additional background information, can be found in the aforementioned works.

Lemma 1 *Given Assumptions 3.1 and 3.2, let* $\{\mathbf{X}_t\}$ *be the unique stationary solution for the SRE in (7). Then (i)* $\mathbf{X}$ *is regularly varying with tail index* $\kappa_0 \in (1,\ \iota_0)$*, and (ii)* $P\left(|\mathbf{X}| > x\right) \sim Cx^{-\kappa_0}$ *for some* $C \in (0, \infty)$.

From Definition 1, let $\{a_n\}$ satisfy

$$nP\left(|\mathbf{X}| > a_n\right) \to 1, \qquad n \to \infty \tag{8}$$

[9] See McNeil et al. (2015, Chapter 5) for definitions of the Frechét and Gumbel classes of distributions, respectively. As illustrations, the Student's-t distribution with a finite degree of freedom is a member of the Frechét class, while the Normal distribution is a member of the Gumbel class.

Given (8), Lemma 1(ii) implies that

$$a_n \sim (Cn)^{1/\kappa_0}\,, \tag{9}$$

and analogously, Assumption 3.2 implies that

$$b_n \sim (Cn)^{1/\iota_0}\,. \tag{10}$$

ASSUMPTION 3.3. *$E\left(\sigma^4\right) = \infty$, but the distribution of σ^2 remains in the domain of attraction of a normal law. In this case,*

$$H\left(a\right) = E\left(\sigma^4 \times I\left(\sigma^2 \leq a\right)\right); \qquad a_n = \inf\left\{a > 0 :\ nH\left(a\right) \leq a^2\right\}, \tag{11}$$

where H is slowly varying at ∞.

Compared to (9), Assumption 3.3 offers an alternative characterization of a_n, one that applies in the borderline case where $\kappa_0 = 2$. This condition heralds from Hall and Yao (2003), as do the following two implications; specifically,

$$\frac{a^2 P\left(\left|\sigma^2 - E\left(\sigma^2\right)\right| > a\right)}{H\left(a\right)} \to 0, \qquad a \to \infty \tag{12}$$

(see; e.g., Feller, 1996, (8.5), p. 303), and

$$\frac{aE\left(\left|\sigma^2 - E\left(\sigma^2\right)\right| \times I\left(\left|\sigma^2 - E\left(\sigma^2\right)\right| > a\right)\right)}{H\left(a\right)} \to 0, \qquad a \to \infty. \tag{13}$$

Consider the following sequence of point processes defined from the normalized process $(\mathbf{X}_t)$.

$$N_n = \sum_{t=1}^{n} \delta_{\mathbf{X}_t/a_n}, \qquad n \in \mathbb{N}, \tag{14}$$

where $\{a_n\}$ is defined in (8) and (9).

Lemma 2 *Given Lemma 1 and the sequence of point processes in (14),*

$$N_n \xrightarrow{d} N = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \delta_{P_i \mathbf{Q}_{ij}},$$

where (iii) $\sum_{i=1}^{\infty}\delta_{P_i}$ *is a Poisson process on* $\left(\begin{array}{cc} 0, & \infty \end{array}\right)$ *with absolutely continuous intensity measure*

$$v\left(dy\right)=\gamma_0\kappa_0 y^{-\kappa_0-1}dy,$$

$\kappa_0 \in \left(\begin{array}{cc} 1, & \iota_0 \end{array}\right)$, *and* $\gamma_0 \in (\begin{array}{cc} 0, & 1 \end{array}]$, *(iv)* $\sum_{j=1}^{\infty}\delta_{\mathbf{Q}_{ij}}$ *for* $i \in \mathbb{N}$ *is an i.i.d. sequence of point processes on* $\overline{\mathbb{R}}_+^2 \setminus \{0\}$ *taking values in the set*

$$\left\{\mu \in M_+\left(\overline{\mathbb{R}}_+^2 \setminus \{0\}\right)\right\} : \mu\left(\{x : |x| > 1\}\right) = 0 \qquad \textit{and} \qquad \mu\left(S\right) > 0,$$

and (v) $\sum_{i=1}^{\infty}\delta_{P_i}$ *and* $\sum_{j=1}^{\infty}\delta_{\mathbf{Q}_{ij}}$ *for* $i \in \mathbb{N}$ *are mutually independent.*

Remark 1 *From Basrak et al. (2002, Remark 2.12.), the points* $\left(\begin{array}{cc} P_i, & \mathbf{Q}_{ij} \end{array}\right)$ *correspond with the radial and spherical parts, respectively, of the limiting points* $\mathbf{X}_t/a_n$*, where the spherical part accounts for clustering behavior in the limiting point process.*

Remark 2 *Consider the sequence of point processes*

$$N_n^2 = \sum_{t=1}^{n}\delta_{\mathbf{X}_t^2/a_n^2}.$$

Given Lemma 2 and the continuous mapping theorem,

$$N_n^2 \xrightarrow{d} N^2 = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\delta_{P_i^2\mathbf{Q}_{ij}^2}.$$

In words, Lemma 2 details a convergence result from point process theory that can be used to establish the distibutional limit of the vector sequence

$$\mathbf{S}_n^k = \sum_{t=1}^{n}\mathbf{X}_t^k, \qquad k = 1, 2, \tag{15}$$

in the case where

$$E\left(\mathbf{X}_t^k\right)^l = \infty \qquad \text{for} \qquad l > 1.$$

To illustrate, consider the function $T_\varepsilon : M_P\left(\overline{\mathbb{R}}_+^2 \setminus \{0\}\right) \longrightarrow \mathbb{R}^2$ such that

$$a_n^{-k}\mathbf{S}_n^k = T_\varepsilon\left(N_n^k\right).$$

Given Lemma 2 and Remark 2,

$$T_{\varepsilon}\left(N_n^k\right) \xrightarrow{d} T_{\varepsilon}\left(N^k\right), \qquad n \to \infty.$$

In addition, given Davis and Mikosch (1998, Proposition 3.3),

$$T_{\varepsilon}\left(N^k\right) \xrightarrow{d} \mathbf{S}^k, \qquad \varepsilon \to 0,$$

where $\mathbf{S}^k$ is a vector of (κ_0/k)-stable random variables expressed in terms of the P_i's and $\mathbf{Q}_{ij}$'s in Lemma 2. The end result

$$a_n^{-k}\mathbf{S}_n^k \xrightarrow{d} \mathbf{S}^k, \tag{16}$$

is a limiting distribution for $\mathbf{S}_n^k$ with an ill-defined variance. The univariate analog to (16) was determined by Davis and Hsing (1995, Theorem 3.1). Both (16) and its univariate analog factor prominently in the limiting results developed in Section 5.

4 Estimation

For the purpose of estimating (3) and (4), assume (potentially incorrectly) that the probability density function of D is given by f, where

$$f \propto \left(1 + \frac{x^2}{\eta - 2}\right)^{\frac{-(\eta+1)}{2}}, \qquad \eta > 2,$$

in which case, the non-Gaussian likelihood based upon the standardized t_η-distribution of Bollerslev (1987) applies. In this case, let

$$\theta = \left(\begin{array}{cccc} \omega, & \alpha, & \beta, & \eta \end{array} \right) = \left(\begin{array}{cccc} \omega, & \vartheta, & \eta \end{array} \right) = \left(\begin{array}{cc} \omega, & \pi \end{array} \right).$$

Given Assumption 3.1 and its implication in (5), consider the alternative parameter vector

$$v = \left(\begin{array}{cc} s^2, & \pi \end{array} \right)$$

such that

$$\sigma_t^2(\upsilon) = \left\{ \begin{array}{ccc} c(\upsilon) + \alpha Y_{t-1}^2 + \beta \sigma_{t-1}^2(\upsilon) & \text{if} & 1 \leq t \leq n \\ s^2 & \text{if} & t \leq 1 \end{array} \right\},$$

where $c(\upsilon) = s^2(1 - \alpha - \beta)$. Then, for

$$\widehat{\sigma}_n^2 = \frac{1}{n}\sum_{t=1}^{n} Y_t^2,$$

the log likelihood function under consideration is given by

$$\log L_n\left(\begin{array}{ccc} \widehat{\sigma}_n^2, & \pi \end{array} \right) = \sum_{t=1}^{n} l_t\left(\begin{array}{ccc} \widehat{\sigma}_n^2, & \pi \end{array} \right)$$

where

$$l_t\left(\begin{array}{ccc} \widehat{\sigma}_n^2, & \pi \end{array} \right) = \log f_\eta\left(\epsilon_t \mid \Omega_{t-1}\right),$$

as defined in Bollerslev (1987, eq. 1), and $\widehat{\pi}_n$ is the solution to

$$\log L_n\left(\begin{array}{ccc} \widehat{\sigma}_n^2, & \widehat{\pi}_n \end{array} \right) = \arg\max_{\pi \in \Pi} \log L_n\left(\begin{array}{ccc} \widehat{\sigma}_n^2, & \pi \end{array} \right), \tag{17}$$

a Variance-Targeted, Non-Gaussian, Quasi-Maximum Likelihood Estimator (VTNGQMLE) for ϑ_0. The FOC from (17) is

$$0 = \frac{\partial}{\partial \pi} \log L_n\left(\widehat{v}_n\right) = \sum_{t=1}^{n} \frac{\partial l_t\left(\widehat{v}_n\right)}{\partial \pi}. \tag{18}$$

Let

$$v_0 = \left(\begin{array}{ccc} \sigma_0^2, & \vartheta_0, & \widetilde{\eta}_0 \end{array} \right),$$

where $\sigma_0^2 = E\left(\sigma^2\right)$, and $\widetilde{\eta}_0$ is interpreted as a "pseudo" truth.[10]

ASSUMPTION 4.1.

$$Q(\omega, \eta) = -\ln \omega + E\left[\ln f\left(\begin{array}{ccc} \frac{\omega_0 \epsilon_t}{\omega}, & \eta \end{array} \right)\right]$$

has a unique maximum at either ω_0 and η_0 when $f = g$ or ω_0 and $\widetilde{\eta}_0$ when $f \neq g$.

Assumption 4.1 is a generic identification condition for the scale and shape of f. It is the same as Newey and Steigerwald (1997, Assumption 2.4). When $f = g$, this assumption holds naturally. When $f \neq g$, Assumption 4.1 follows from identification of the scale and shape parameters of g.

[10] In the (likely) case where $f \neq g$, there is no "true" η. Nevertheless, in this case, $\widehat{\eta}$ converges to something, and that something is defined as a "pseudo" truth.

Taking a first-order Taylor Expansion of (18) around $\upsilon = \upsilon_0$ produces

$$\begin{aligned} 0 &= \sum_{t=1}^{n} \left\{ \frac{\partial l_t(\upsilon_0)}{\partial \pi} + \frac{\partial^2 l_t(\upsilon_i)}{\partial \pi \partial \upsilon'} \left(\widehat{\upsilon}_n - \upsilon_0 \right) \right\} \\ &= \sum_{t=1}^{n} \left\{ \frac{\partial l_t(\upsilon_0)}{\partial \pi} + \frac{\partial^2 l_t(\upsilon_i)}{\partial \pi \partial \sigma^2} \left(\widehat{\sigma}_n^2 - \sigma_0^2 \right) + \frac{\partial^2 l_t(\upsilon_i)}{\partial \pi \partial \vartheta'} \left(\widehat{\vartheta}_n - \vartheta_0 \right) + \frac{\partial^2 l_t(\upsilon_i)}{\partial \pi \partial \eta} \left(\widehat{\eta}_n - \widetilde{\eta}_0 \right) \right\} \end{aligned} \tag{19}$$

where υ_i lies on the line segment between $\widehat{\upsilon}_n$ and υ_0. Letting

$$J_n = \frac{1}{n}\sum_{t=1}^{n} \frac{\partial^2 l_t(\upsilon_i)}{\partial \pi \partial \vartheta'}, \quad K_n = \frac{1}{n}\sum_{t=1}^{n} \frac{\partial^2 l_t(\upsilon_i)}{\partial \pi \partial \sigma^2}, \quad M_n = \frac{1}{n}\sum_{t=1}^{n} \frac{\partial^2 l_t(\upsilon_i)}{\partial \pi \partial \eta}, \quad Z_n = \frac{1}{n}\sum_{t=1}^{n} \frac{\partial l_t(\upsilon_0)}{\partial \pi},$$

(19) becomes

$$0 = nZ_n + J_n n \left(\widehat{\vartheta}_n - \vartheta_0 \right) + K_n n \left(\widehat{\sigma}_n^2 - \sigma_0^2 \right) + M_n n \left(\widehat{\eta}_n - \widetilde{\eta}_0 \right). \tag{20}$$

ASSUMPTION 4.2.

$$J_n \xrightarrow{a.s.} J \equiv E\left(\frac{\partial^2 l_t(\upsilon_0)}{\partial \pi \partial \vartheta'} \right), \quad K_n \xrightarrow{a.s.} K \equiv E\left(\frac{\partial^2 l_t(\upsilon_0)}{\partial \pi \partial \sigma^2} \right), \quad M_n \xrightarrow{a.s.} M \equiv E\left(\frac{\partial^2 l_t(\upsilon_0)}{\partial \pi \partial \eta} \right).$$

When $\eta = \infty$, Assumption 4.2 follows from Vaynman and Beare (2014, Lemma 1). Given Assumption 4.2, rearranging (20) and substituting population moments for sample moments produces

$$n a_n^{-1} \left(\widehat{\vartheta}_n - \vartheta_0 \right) = -J^{-1} \left\{ K n a_n^{-1} \left(\widehat{\sigma}_n^2 - \sigma_0^2 \right) + M n a_n^{-1} \left(\widehat{\eta}_n - \widetilde{\eta}_0 \right) + n a_n^{-1} Z_n \right\}. \tag{21}$$

5 Asymptotics

This section considers the large-sample implications of (21) in six cases ranging from (relatively) thin-tailed to (very) heavy-tailed. To preview the results, the large-sample properties of VTNGQMLE look very different depending on whether a thin-tailed or thick-tailed case applies.

5.1 Case 1: $\kappa_0 > 2$; $\widetilde{\eta}_0 = \infty$.

In this case, the likelihood used in estimation is Gaussian, so the second term on the right-hand-side of (21) drops out. In addition, $a_n^{-1} = n^{-1/2}$ so that

$$\sqrt{n} \left(\widehat{\sigma}_n^2 - \sigma_0^2 \right) \xrightarrow{d} N\left(0, V_{\sigma^2} \right), \tag{22}$$

by a CLT for weakly dependent data, and

$$\sqrt{n}Z_n \xrightarrow{d} N\left(0, V_Z\right), \tag{23}$$

by, for instance, Hall and Yao (2003, Theorem 2.1(a)). Moreover,

$$\sqrt{n}\left(\widehat{\vartheta}_n - \vartheta_0\right) \xrightarrow{d} N\left(0, V_\vartheta\right), \tag{24}$$

by Francq et al. (2011, Theorem 1.1); in which case, given (22) and (23), V_ϑ is seen to depend upon both V_{σ^2} and V_Z. Moreover, given Francq et al. (2011, Corollary 2), the VTQMLE cannot be asymptotically more efficient than QMLE.

5.1.1 Case 2: $\kappa_0 > 2$; $\widetilde{\eta}_0 \in \left(\ 2, \quad \infty\ \right)$.

In this case, all three terms on the right-hand-side of (21) matter. $a_n^{-1} = n^{-1/2}$ continues to hold, as do both (22) and (23), except that the latter now follows from Fan et al. (2014, Theorem 2). Given Assumption 4.1, we can posit that

$$\sqrt{n}\left(\widehat{\eta}_n - \widetilde{\eta}_0\right) \xrightarrow{d} N\left(0, V_{\widetilde{\eta}}\right). \tag{25}$$

In (25), Fan et al. (2014, Section 5.3) establishes the rate of convergence as $\sqrt{n}$ and the limit as Gaussian, both so long as $\iota_0 > 1$.

We can further posit that

$$\sqrt{n}\left(\left(\widehat{\vartheta}_n + \widehat{C}_n\right) - \vartheta_0\right) \xrightarrow{d} N\left(0, \dot{V}_\vartheta\right), \tag{26}$$

using (22) and results from Fan et al. (2014, Theorem 2). In (26), however, and in contrast to (24),$\widehat{C}_n \xrightarrow{p} C \neq 0$, thus rendering $\widehat{\vartheta}_n$ from (17) generally inconsistent. The presence of a non-asymptotically-vanishing $\widehat{C}_n$ follows from Fan et al. (2014, Proposition 1). Specifically, for our chosen f, VTNGQMLE fails to identify the scale of the true model innovations, whenever $f \neq g$ (see, additionally, Newey and Steigerwald, 1997). This issue of under-identification impacts $\widehat{\vartheta}_n$, generally, because it impacts $\widehat{\alpha}_n$, specifically. Consequently, $\widehat{\vartheta}_n$ from VTNGQMLE is inconsistent because the scale of the GARCH$(1,1)$ model's innovations is not identified.

It is well known that NGQMLE is inconsistent whenever $f \neq g$. It turns out that in this case, VT-NGQMLE inherits this same undesirable property.

5.1.2 Case 3: $\kappa_0 = 2$; $\widetilde{\eta}_0 \in \left(\ 2, \ \ \infty \ \right)$.

In this case, apply Assumption 3.3. In addition, let

$$\begin{aligned} U_n &= a_n^{-1} \sum_{t=1}^{n} \left(Y_t^2 - E\left(Y^2\right) \right) \\ &= a_n^{-1} \sum_{t=1}^{n} \left(\left(\sigma_t^2 + W_t \right) - E\left(Y^2\right) \right); \qquad W_t = \left(\epsilon_t^2 - 1 \right) \times \sigma_t^2 \\ &= a_n^{-1} \sum_{t=1}^{n} W_t + a_n^{-1} \sum_{t=1}^{n} \left(\sigma_t^2 - E\left(\sigma^2\right) \right) \\ &= I(a) + II(a) \end{aligned} \tag{27}$$

Following notation from Mikosch and Stărică (2000), for a random variable X, let

$$\gamma_{n,X}\left(h\right) = \frac{1}{n} \sum_{t=1}^{n-h} X_t X_{t+h}.$$

Given (6),

$$\begin{aligned} II(a) &= a_n^{-1} \sum_{t=1}^{n} \alpha_0 \left(Y_{t-1}^2 - E\left(Y^2\right) \right) + \beta_0 \left(\sigma_{t-1}^2 - E\left(\sigma^2\right) \right) \\ &= \alpha_0 n a_n^{-1} \left(\gamma_{n,Y}\left(0\right) - E\left(Y^2\right) \right) + \beta_0 n a_n^{-1} \left(\gamma_{n,\sigma}\left(0\right) - E\left(\sigma^2\right) \right) \\ &= \left(\frac{\alpha_0}{1 - \beta_0} \right) n a_n^{-1} \left(\gamma_{n,Y}\left(0\right) - E\left(Y^2\right) \right) \end{aligned} \tag{28}$$

Plugging (28) back into (27) produces

$$U_n = \left(\frac{1 - \beta_0}{1 - \alpha_0 - \beta_0} \right) a_n^{-1} \sum_{t=1}^{n} W_t \tag{29}$$

Theorem 3 *Given Assumptions 3.1 and 3.3,*

$$n a_n^{-1} \left(\widehat{\sigma}_n^2 - \sigma_0^2 \right) \xrightarrow{d} N\left(0, \ \left(\frac{1 - \beta_0}{1 - \alpha_0 - \beta_0} \right)^2 V_{\overline{W}} \right) \tag{30}$$

where a_n is given by (11), and $V_{\overline{W}}$ is defined in (65) of the Appendix.

Proof. Unless otherwise stated, all proofs appear in Appendix A. ■

Motivated by results in Hall and Yao (2003), Theorem 3 establishes $\widehat{\sigma}_n^2$ as asymptotically normal in the borderline case where $\kappa_0 = 2$. In this case, (23) continues to hold as in Case 2, as does (25). Moreover,

given Assumption 3.3,

$$n^{1/2}a_n^{-1} = O(1),$$

in which case, from (21),

$$na_n^{-1}\left(\widehat{\vartheta}_n - \vartheta_0\right) = -J^{-1}\left\{Kna_n^{-1}\left(\widehat{\sigma}_n^2 - \sigma_0^2\right) + M\sqrt{n}\left(\widehat{\eta}_n - \widetilde{\eta}_0\right) + \sqrt{n}Z_n\right\}.$$

As a result, VTNGQMLE remains an inconsistent estimator of $\widehat{\vartheta}_n$, owing to the (asymptotic) effects of Z_n.

5.2 Case 4: $\kappa_0 \in \left(\ 1,\ \ 2\ \right)$; $\iota_0 > 2$.

Given Figure 5, this case has been empirically relevant for SPX (log) returns, at least, in the past

Theorem 4 *Given Lemma 1, (9), Assumption 4.2, and (21), if $E\left(Y_t^4\right) = \infty$; σ_t^2 is in the domain of attraction of a κ_0-stable law, and $E\left(\epsilon_t^4\right) < \infty$, then*

$$na_n^{-1}\left(\widehat{\vartheta}_n - \vartheta_0\right) = -J^{-1}Kna_n^{-1}\left(\widehat{\sigma}_n^2 - \sigma_0^2\right) + o_p(1). \tag{31}$$

When $E\left(Y_t^4\right) = \infty$ because $\kappa_0 < 2$, and $E\left(\epsilon_t^4\right) < \infty$ (or, equivalently, $\iota_0 > 2$), the distributional limit of $\widehat{\vartheta}_n$ becomes dominated by the limit of $\widehat{\sigma}_n^2$ (an analogous result is reported in Vaynman and Beare, 2014, for VTQMLE). This dominance sources to a slower rate of convergence for $\widehat{\sigma}_n^2$ compared to either $\widehat{\eta}_n$ or the score of the likelihood function. Since $\widehat{\sigma}_n^2 \xrightarrow{p} \sigma_0^2$, an effect of this dominance is that VTNGQMLE becomes a consistent estimator for $\widehat{\vartheta}_n$. As such, VTNGQMLE is a robust estimator like QMLE and the multi-step estimators aimed at improving QMLE, like Fan et al. (2014) and Preminger and Storti (2017).

Theorem 5 *Given Lemma 1 and (9), if $E\left(Y_t^4\right) = \infty$; σ_t^2 is in the domain of attraction of a κ_0-stable law, and $E\left(\epsilon_t^4\right) < \infty$, then*

$$na_n^{-1}\left(\widehat{\sigma}_n^2 - \sigma_0^2\right) \xrightarrow{d} \left(\frac{1-\beta_0}{1-\alpha_0-\beta_0}\right) U_{\sigma^2}, \tag{32}$$

where U_{σ^2} is the κ_0-stable random variable given in (72).

Remark 3 *The method of proof behind Theorem 5 borrows from both Davis and Mikosch (1998) and Mikosch and Stărică (2000). Theorem 5 is also closely related to Vaynman and Beare (2014, Theorem 4).*

Remark 4 *The limit in (32) relates to the P_i's and $\mathbf{Q}_{ij}$'s in Lemma 2.*

Established in Theorem 5 is a stable limit for $\widehat{\sigma}_n^2$ that is highly non-Gaussian. Convergence to this non-Gaussian limit is also slower than the usual $\sqrt{n}$ rate and dependent upon the tails of $\left\{\sigma_t^2\right\}$. In this case, while both $\left\{\sigma_t^2\right\}$ and $\left\{\epsilon_t^2\right\}$ are allowed to be heavy tailed, only the tail properties of the former impact the limit in (32).[11]

Given (31) and (32), establishing the limit of $\widehat{\vartheta}_n$ requires a straight-forward application of Slutsky's Theorem to produce

$$na_n^{-1}\left(\widehat{\vartheta}_n - \vartheta_0\right) \xrightarrow{d} -\left(\frac{1-\beta_0}{1-\alpha_0-\beta_0}\right) J^{-1} K U_{\sigma^2}. \tag{33}$$

5.3 Case 5: $\kappa_0 \in \left(\ 1,\ \ \iota_0\ \right)$; $\iota_0 = 2$.

This second borderline case was first introduced and studied in Hall and Yao (2003). Given Figure 6, this case appears to be empirically relevant for SPX (log) returns in contemporaneous times. For analyzing this case, the following Condition is important.

ASSUMPTION 5.1. *$E\left(\epsilon^4\right) = \infty$, but the distribution of ϵ^2 remains in the domain of attraction of a normal law. In this case,*

$$H(b) = E\left(\epsilon^4 \times I\left(\epsilon^2 \leq b\right)\right); \qquad b_n = \inf\left\{b > 0 :\ nH(b) \leq b^2\right\},$$

where H is slowly varying at ∞.

Remark 5 *Assumption 5.1 parallels Assumption 3.3 but for ϵ^2 and is identical to Hall and Yao (2003, eq. 2.8). It controls the rate of tail decay in the distribution of ϵ^2.*

In this borderline case, (31) continues to hold, in which case, the asymptotic properties of $\widehat{\vartheta}_n$ remain dominated by those of $\widehat{\sigma}_n^2$. Establishing the stable limit of $\widehat{\sigma}_n^2$, however, becomes more complicated compared to Theorem 5, since it is now the case that $E\left(\epsilon^4\right) = \infty$. Nevertheless, with the aid of Assumption 5.1 and its associated implications (see 12 and 13, appropriately modified for ϵ^2), (32) continues to hold, as established by the following Theorem.

Theorem 6 *Given Lemma 1, (9), and Assumption 5.1, if $E\left(Y_t^4\right) = \infty$, and σ_t^2 is in the domain of attraction of a κ_0-stable law, then (32) continues to hold.*

[11] Just as the impact of the likelihood function vanishes in (31), the impact of extremes in $\left\{\epsilon_t^2\right\}$ vanish in determining the limit of $\widehat{\sigma}_n^2$. The notation for the limiting variable U_{σ^2} emphasizes this singular impact.

Under Theorem 6, despite ϵ^2 being a heavier-tailed process compared to Case 4, its tail properties continue to exercise no effect on the asymptotic limit of $\widehat{\sigma}_n^2$. Moving from Case 1 to Case 2, the rate of convergence changes but the distributional limit remains (generally) the same. Moving from Case 4 to Case 5, in contrast, both the rate of convergence and the distributional limit remain unaltered. Explaining this difference between borderline cases are the dual results that (1) the rate of convergence implied by $\{b_n\}$ is faster than the rate of convergence implied by $\{a_n\}$, causing any effects related to the former to vanish, and (2) extremes in σ_t^2 continue to dominate the asymptotic behavior of $\sum_{t=1}^{n} Y_t^2$.

An immediate consequence of (31) and (32) continuing to hold is that (33) also continues to hold.

5.4 Case 6: $\kappa_0 \in \left(1, \quad \iota_0 \right)$; $\iota_0 < 2$.

There is strong empirical evidence supporting this very heavy-tailed case as being relevant for VIX (see Figures 11 and 12). In addition, empirical relevance of this case even for SPX (log) returns, in contemporaneous times, cannot be dismissed (see Figures 5 and 6).

Theorem 7 *Given Lemma 1, (9), (10), (20), and Assumption 4.2, if $E\left(Y_t^4\right) = \infty$; σ_t^2 is in the domain of attraction of a κ_0-stable law; $E\left(\epsilon_t^4\right) = \infty$, and ϵ^2 is in the domain of attraction of a ι_0-stable law, then*

$$na_n^{-1}b_n^{-1}\left(\widehat{\vartheta}_n - \vartheta_0\right) = -J^{-1}Kna_n^{-1}b_n^{-1}\left(\widehat{\sigma}_n^2 - \sigma_0^2\right) + o_p\left(1\right). \tag{34}$$

In this case, the analog to (27) is

$$U_n = a_n^{-1}b_n^{-1}\sum_{t=1}^{n}\left(Y_t^2 - E\left(Y^2\right)\right),$$

in which case,

$$U_n = \left(\frac{1-\beta_0}{1-\alpha_0-\beta_0}\right)a_n^{-1}b_n^{-1}\sum_{t=1}^{n} W_t, \tag{35}$$

following the steps outlined in (27)–(29). Given (34), analysis of $a_n^{-1}b_n^{-1}\sum_{t=1}^{n} W_t$ then determines the asymptotic limit of $\widehat{\vartheta}_n$.

Theorem 8 *Given Lemma 1, (9) and (10), if $E\left(Y_t^4\right) = \infty$; σ_t^2 is in the domain of attraction of a κ_0-stable law; $E\left(\epsilon_t^4\right) = \infty$, and ϵ^2 is in the domain of attraction of a ι_0-stable law, then*

$$na_n^{-1}b_n^{-1}\left(\widehat{\sigma}_n^2 - \sigma_0^2\right) \xrightarrow{d} \left(\frac{1-\beta_0}{1-\alpha_0-\beta_0}\right)U_{\epsilon^2,\,\sigma^2}, \tag{36}$$

where $U_{\epsilon^2,\,\sigma^2}$ is the κ_0-stable random variable given in (86).

Under Theorem 8, and for the first time, tail properties of both $\left\{\sigma_t^2\right\}$ and $\left\{\epsilon_t^2\right\}$ matter in determining the asymptotic limit in (36).

Remark 6 *In the proof of Theorem 8, when establishing the asymptotic variance of certain sums as negligible, it appears insufficient to rely solely on the normalizing constants $\{a_n\}$, since doing so implies explosive (as opposed to dampened) behaviour in the affected sums, as n grows large. Joint reliance on the normalizing constants $\{a_n\}$ and $\{b_n\}$, however, enables the variance of these affected sums to smoothly vanish. Moreover, the asymptotic limit of the remaining sum is seen to depend on both $\{a_n\}$ and $\{b_n\}$, as opposed to just $\{a_n\}$ alone.*

Remark 7 *The limits in (32) and (36) are not the same, but they are similar in a qualitative sense (see; e.g., Davis and Mikosch, 1998, Remark 3.2). That is, $U_{\epsilon^2,\,\sigma^2}$ can be expressed in terms of quantities that are qualitatively similar to the P_i's and $\mathbf{Q}_{ij}$'s in Lemma 2.*

The limit of VTNGQMLE in (36) appears (qualitatively) similar to the limit of QMLE, as determined by Hall and Yao (2003, Theorem 2.1(c)). Under Case 6, consequently, it is unclear which estimator (VTNGQLME or QMLE) dominates the other, on efficiency grounds. A similar statement appears to hold true when comparing VTNGQMLE to the multi-step estimator of Preminger and Storti (2017) that assumes $E\left(Y_t^2\right) < \infty$ (hereafter LSE), since $\sqrt{n}$ asymptotic normality of this estimator also depends on $E\left(\epsilon_t^4\right) < \infty$, just as in the QMLE case.[12] The multi-step estimator of Fan et al. (2014), on the other hand, (hereafter FAN) should be more efficient (asymptotically) than VTNGQMLE, since the former should be $\sqrt{n}$ asymptotically normal, so long as $\iota_0 > 1$.

Lastly, given (34) and (36), Slutsky's Theorem establishes

$$na_n^{-1}b_n^{-1}\left(\widehat{\vartheta}_n - \vartheta_0\right) \xrightarrow{d} -\left(\frac{1-\beta_0}{1-\alpha_0-\beta_0}\right) J^{-1}KU_{\epsilon^2,\,\sigma^2}. \tag{37}$$

6 Bootstrap Inference

Troubles with the results in (33) and (37) are twofold:

[12] In the case of QMLE, $E\left(\epsilon_t^4\right) < \infty$ is necessary for $\sqrt{n}$ asymptotic normality. Still in the case of QMLE, Hall and Yao (2003) show that when $E\left(\epsilon_t^4\right) = \infty$, the asympotic limit is α-stable with a slower rate of convergence that depends upon the tail properties of $\left\{\epsilon_t^2\right\}$. Owing to this result, it seems reasonable to conclude that when $E\left(\epsilon_t^4\right) = \infty$, LSE, too, would have an α-stable limit and a rate of convergence slower than $\sqrt{n}$. This conclusion, however, has not been formally established.

1. the precise form of the distributional limits is awkward, rendering how to construct asymptotic confidence bands unclear;

2. the rate of convergence depends upon distributional characteristics that are unknown.

Owing to these twin troubles, it is equally unclear the practical relevance of (33) and (37). To help dispel these troubles, consider

$$\widehat{\tau}_n^2 = n^{-1}\sum_{t=1}^{n} Y_t^4 - \left(n^{-1}\sum_{t=1}^{n} Y_t^2 \right)^2, \tag{38}$$

which is analogous to a finite-sample variance for $\widehat{\sigma}_n^2$, if the true variance were well defined.

Theorem 9 *Under Case 6 and the assumptions of Theorem 8,*

$$n a_n^{-2} b_n^{-2} \widehat{\tau}_n^2 \xrightarrow{d} U_{\epsilon^4,\, \sigma^4}, \tag{39}$$

where $U_{\epsilon^4,\, \sigma^4}$ is a $(\kappa_0/2)$-stable random variable determined by the extremes of both $\left\{\epsilon_t^2\right\}$ and $\left\{\sigma_t^2\right\}$ (see the proof of Theorem 9 for additional details).

Remark 8 *Hall and Yao (2003) consider a statistic analogous to (38) that is based on ϵ_t, as opposed to Y_t. Denote the Hall and Yao (2003) statistic $\widehat{\tau}_n^2(\epsilon_t)$, so that the statistic in (38) can be denoted $\widehat{\tau}_n^2(Y_t)$. Because $\{\epsilon_t\}$ is i.i.d., an appropriately scaled version of $\widehat{\tau}_n^2(\epsilon_t)$ can be shown to have a stable limit using results from Feller (1971) and Lepage et al. (1981), even when $E\left(\epsilon_t^4\right) = \infty$. Complicating an analgous demonstration for $\widehat{\tau}_n^2(Y_t)$, in the case where $E\left(Y_t^4\right) = \infty$, is dependence in $\{Y_t\}$ that sources to $\left\{\sigma_t^2\right\}$. Theorem (9) establishes a stable distributional limit for $\widehat{\tau}_n^2(Y_t)$ by relying upon the convergence results summarized in Section 3.*

Corollary 10 *Under Case 4 and the assumptions of Theorem 5,*

$$n a_n^{-2} \widehat{\tau}_n^2 \xrightarrow{d} U_{\sigma^4}, \tag{40}$$

where U_{σ^4} is a $(\kappa_0/2)$-stable random variable determined by the extremes of $\left\{\sigma_t^2\right\}$.

Proof. The general method of proof follows the same arguments in the proof of Theorem 5 immediately below (71) and through to the end. Establishing

$$T_\varepsilon\left(N^2\right) \xrightarrow{d} U_{\sigma^4}, \qquad \varepsilon \to 0$$

for an appropriately defined $T_{\varepsilon}(\cdot)$ and the N^2 in Remark 2 follows from Davis and Hsing (1995, Theorem 3.1(i)). ■

Theorems (8) and (9) demonstrate that individually $na_n^{-1}b_n^{-1}\left(\widehat{\sigma}_n^2-\sigma_0^2\right)$ and $na_n^{-2}b_n^{-2}\widehat{\tau}_n^2$ have proper limiting distributions. The following theorem and corollary establish that these (weak) marginal convergence results are also joint.

Theorem 11 *Under Case 6 and the assumptions of Theorem 8,*

$$\left(\begin{array}{cc} na_n^{-1}b_n^{-1}\left(\widehat{\sigma}_n^2-\sigma_0^2\right), & na_n^{-2}b_n^{-2}\widehat{\tau}_n^2 \end{array}\right) \xrightarrow{d} \left(\begin{array}{cc} U_{\epsilon^2,\,\sigma^2}, & U_{\epsilon^4,\,\sigma^4} \end{array}\right), \tag{41}$$

where the, respective, marginal limits are those from (36) and (39).

Corollary 12 *Under Case 4 and the assumptions of Theorem 5,*

$$\left(\begin{array}{cc} na_n^{-1}\left(\widehat{\sigma}_n^2-\sigma_0^2\right), & na_n^{-2}\widehat{\tau}_n^2 \end{array}\right) \xrightarrow{d} \left(\begin{array}{cc} U_{\,\sigma^2}, & U_{\,\sigma^4} \end{array}\right), \tag{42}$$

where the, respective, marginal limits are those from (32) and (40).

Proof. The method of proof follows that of Theorem (11) (see Appendix A). Alternatively, let

$$\left(\begin{array}{cc} na_n^{-1}\left(\widehat{\sigma}_n^2-\sigma_0^2\right), & na_n^{-2}\widehat{\tau}_n^2 \end{array}\right) = \left(\begin{array}{cc} \ddot{U}_{n,\varepsilon}, & V_{n,\varepsilon} \end{array}\right),$$

and note that $\ddot{U}_{n,\varepsilon}$ is a special case of $U_{n,\varepsilon}$ in Vaynman and Beare (2014, eq. 35). Then (42) can be established by following the steps outlined in Vaynman and Beare (2014, proof of Theorem 4). ■

With the aid of the continuous mapping theorem and Slutsky's Theorems, from (41) follows that

$$\sqrt{n}\left(\begin{array}{c} \widehat{\vartheta}_n-\vartheta_0 \\ \widehat{\tau}_n \end{array}\right) \xrightarrow{d} -\left(\frac{1-\beta_0}{1-\alpha_0-\beta_0}\right) J^{-1}K\left(\begin{array}{c} U_{\epsilon^2,\,\sigma^2} \\ U_{\epsilon^4,\,\sigma^4}^{1/2} \end{array}\right). \tag{43}$$

The power of (43) is that the left-hand-side has a proper limiting distribution, and the rate of convergence is known. Moreover, given (42), it is evident that the left-hand-side of (43) has a proper limiting distribution under Cases 3–6. In fact, the left-hand-side of (43) has a proper limiting distribution under all of the cases considered in Section 5. In Cases 1 and 2, the result is trivial, since $\widehat{\tau}_n$ has a degenerate limit, with the standard $\sqrt{n}$ rate of convergence. In Case 3, $\widehat{\tau}_n$ maintains a degenerate limit; however, the rate of convergence to that degenerate limit is now unknown. Fortunately, the rate of convergence to a non-degenerate limit for

$\left(\widehat{\vartheta}_n - \vartheta_0\right)$ is also unknown and happens to depend upon the same latent factor in such a precise way that the effect of this (common) latent factor cancels out. In Cases 4–6, the limit of $\widehat{\tau}_n$ becomes non-degenerate, but the rate of convergence to that non-degenerate limit remains precisely aligned with the rate of convergence in $\left(\widehat{\vartheta}_n - \vartheta_0\right)$ in such a way that the normalized statistic $\left(\frac{\widehat{\vartheta}_n - \vartheta_0}{\widehat{\tau}_n}\right)$ converges at the standard $\sqrt{n}$ rate. Owing to this result, $\left(\frac{\widehat{\vartheta}_n - \vartheta_0}{\widehat{\tau}_n}\right)$ can be bootstrapped using the re-sampling scheme described in Hall and Yao (2003, Section 3.2), which then approximates the limiting result in (43), as demonstrated by Hall and Yao (2003, Theorem 3.2).

7 Monte Carlo Experiments

The sequence $\{\epsilon_t\}$ is drawn from the skewed student's-t density of Hansen (1994). This density has two parameters, λ and η, with the former governing skewness, the latter governing the tails, and up to the ηth moment of the distribution being well defined. Values for these parameters are

$$\lambda_0 = \left(\begin{array}{cccc} 0.00, & 0.40, & 0.80, & 0.99 \end{array} \right); \qquad \eta_0 = \left(\begin{array}{cccc} 8.5, & 4.5, & 4.0, & 3.5 \end{array} \right).$$

As λ increases, so, too, does skewness, while as η decreases, tail thickness increases. $\eta_0 = 8.5$ is a (relatively) thin-tailed case, while the remaining values for η_0 correspond with heavy-tailed cases. When $\eta_0 = 4.5$, QMLE is asymptotically normal (AN). When $\eta_0 = 4.0$, AN of QMLE is preserved, but with a convergence rate slower than $\sqrt{n}$, while when $\eta_0 = 3.5$, QMLE is no longer AN, instead converging to a limit that appears qualitatively similar to the one discovered for VTNGQMLE.

Non-zero skewness levels are considered for two reasons. First, $\lambda \neq 0$ introduces a density misspecification, since f is symmetric. Second, non-zero skewness is an empirical feature of both SPX log returns and VIX levels, especially, in recent times (see Figure 16), where the former tends to be negative and the latter strongly positive.[13] Unreported results indicate no material differences between positive and negative values for λ_0; consequently, only results for positive values are reported.[14]

Across the different parameterizations of the innovation density, the different GARCH$(1,1)$ model parameters are given in Table 1. The estimators under study are QMLE, NGQMLE, VTNGQMLE, FAN, and LSE. NGQLME is the Student's-t estimator of Bollerslev (1987), while FAN and LSE are the estimators of Fan et al. (2014) and Preminger and Storti (2017), respectively. Away from the case $\lambda = 0$, NGQMLE is not robust, while FAN, LSE, and QMLE are all robust estimators. Samples sizes for the simulations range

[13] The positive skewness in VIX levels is "natural," in the sense that VIX > 0 because it is a volatility.

[14] Simulation results using negative values of λ_0 are available upon request.

from $500 - 100,000$, with all simulations conducted over $10,000$ trials.[15] Summary statistics for the simulations include mean bias and inter-decile range. Also reported are ratios of the root-mean-squared- and mean-absolute-error (each measured with respect to the true parameter value) divided by the corresponding measure for QMLE. Termed "efficiency ratios," values less than one indicate improved efficiency of the given estimator over QMLE. Figures 17–24 depict results for Specification III.

7.1 Bias

For $\widehat{\alpha}_n$, as λ_0 increases, NGQMLE displays a growing bias that can be quite severe, particularly in large samples (see Figure 17). Bias in VTNGQMLE (relative to NGQMLE), in contrast, behaves quite differently, tending to decrease (rather sharply) as the sample size increases. At very large samples ($> 10,000$), VTNGQMLE appears to retain a small amount of bias. Large sample results in Section 5 are based on a first-order approximation to the score of the likelihood function. This residual bias, then, is consistent with higher-order effects. Since this residual bias is (1) orders-of-magnitude smaller than the bias affecting NGQMLE and (2) additionally materially smaller than the bias displayed by VTQMLE, the latter being a consistent estimator (see Francq et al., 2011, Theorem 1.1), any retained bias in VTNGQMLE, and the higher-order terms causing it, appears to be only of secondary importance. Consequently, simulation results confirm VTNGQMLE to be a consistent estimator, comparable to FAN, LSE, and QMLE.

For $\widehat{\beta}_n$, bias in NGQMLE decreases sharply with the sample size, indicating NGQMLE to be a consistent estimator for β_0 (see Figure 18). In fact, for $\widehat{\beta}_n$, NGQMLE tends to display the least bias of all the estimators being studied and under all the simulation designs considered. This result confirms the theoretical prediction in Fan et al. (2014) that bias in NGQMLE sources to under-identification of scale. $\widehat{\beta}_n$ is unaffected by scale, in which case, NGQMLE is a robust estimator for β_0. VTNGQMLE of $\widehat{\beta}_n$ tends to be close to NGQMLE in terms of bias and, consequently, tends to display among the least bias of the estimators being studied, except under very large sample sizes. Parallel to $\widehat{\alpha}_n$, any retained bias in $\widehat{\beta}_n$ from VTNGQMLE sources to higher-order effects, which, owing to results that follow, are of (decidedly) second-order importance.

7.2 Dispersion

For $\widehat{\alpha}_n$, except in the largest samples, NGQMLE and VTNGQMLE tend to be noticeably less disperse than the other estimators. Under all simulation designs considered, the rate of convergence for FAN and

[15] The first 200 observations within each trial are dropped in order to avoid initialization effects. Very large samples are considered because of the slow convergence rates identified under Cases 4 and 6 (see, also, Hall and Yao, 2003, Theorem 2.1(b)–(c), for QMLE).

NGQMLE should be $\sqrt{n}$. The rate of convergence for VTNGQMLE, however, should always be less than $\sqrt{n}$ and should be the slowest in the case where $\eta_0 = 3.5$. Consistent with these predictions, the rate of reduction in dispersion appears muted for VTNGQMLE compared to both FAN and NGQMLE (see Figure 19). Moreover, the difference between rates of reduction in dispersion appears most apparent in the case where $\eta_0 = 3.5$.

For $\widehat{\beta}_n$, NGQMLE and VTNGQMLE are consistently the least disperse estimators, followed by FAN and LSE (see Figure 20). QMLE and VTQMLE are the bottom-two, in terms of dispersion, and appear (effectively) indistinguishable.

7.3 Efficiency

For $\widehat{\alpha}_n$, as λ_0 increases, the bias in NGQMLE grows in importance and eventually dominates both efficiency ratios, causing NGQMLE to become the least efficient estimator (see Figures 21 and 23). This dominance, however, takes a surprisingly long while to set in, only severely and adversely impacting the very largest sample sizes. For empirically-relevant sample sizes (i.e., $T \in \left[\begin{array}{cc} 500, & 2,500 \end{array} \right]$), NGQMLE beats all other estimators except VTNGQMLE in terms of RMSE (see Figure 21), with a similar result holding for MAE (see Figure 23). The source of this outperformance appears to be (despite the materially higher bias) the material reduction in dispersion that the Student's-t likelihood affords (specifically, estimation of a degrees-of-freedom parameter) relative to the competing estimators. Moreover, except in the largest samples considered ($\geq 50,000$), NGQMLE sizably outperforms QMLE, in terms of both RMSE and MAE. Consequently, as a practical matter, in heavy-tailed cases, and despite the presence of material bias, NGQMLE appears preferable to QMLE.

For $\widehat{\alpha}_n$ and $\lambda_0 > 0$, VTNGQMLE consistently beats NGQMLE (see Figures 21 and 23). Moreover, and surprisingly, in these same cases, VTNGQMLE consistently beats both FAN and LSE in samples as large as $T = 2,500$, generally. In the heaviest-tailed case of $\eta_0 = 3.5$, specifically, VTNGQMLE beats FAN and LSE in samples as large as $T = 10,000$. Consequently, gains in VTNGQMLE over FAN and LSE appear to be finite-sample phenomena; however, in heavy-tailed cases, these gains (1) are rather sizable and (2) extend into finite samples that are (very) common to empirical applications.

Still for $\widehat{\alpha}_n$, in the (relatively) thin-tailed case of $\eta_0 = 8.5$, VTNGQMLE is never more efficient (under either efficiency ratio) than QMLE in the largest sample size, making gains in VTNGQMLE over QMLE also a finite-sample phenomena; albeit, one that similarly persists into surprisingly large samples. In cases where $\eta_0 \leq 4.5$, however, material efficiency gains in VTNGQMLE over QMLE begin appearing even in

the largest sample size, when $\lambda_0 > 0$. When $\eta_0 = 4.0$, VTNGQMLE is more efficient than QMLE across all sample sizes considered, with this tendency preserved in the heaviest-tailed case of $\eta_0 = 3.5$. From Hall and Yao (2003, Theorem 2.1(b)–(c)), the large sample properties of QMLE change when $\eta_0 \leq 4.0$. In these same cases, the relative large sample properties between VTNGQMLE and QMLE appear to change as well.

For $\widehat{\alpha}_n$ using RMSE, Francq et al. (2011, p. 630) reports that when the true ARCH (1) innovations are heavy-tailed, VTQMLE "performs remarkably well and even outperforms QMLE." The sample size upon which this result is based is $T = 500$. Results in Figures 21 and 23 show that this outperformance of VTQMLE over QMLE (1) extends to the GARCH $(1,1)$ case and (2) covers sample sizes much larger than $T = 500$. For instance, when $\eta_0 \leq 4.0$, VTQMLE bests QMLE (in terms of either RMSE or MAE) in samples as large as $T = 10,000$. However, across all samples considered, VTNGQMLE always bests VTQMLE (and by large amounts), and in the heaviest-tailed cases of $\eta_0 \leq 4.0$, VTQMLE tends not to outperform QMLE in the largest sample size, while VTNGQMLE does.

For $\widehat{\beta}_n$ overall, VTNGQMLE and NGQMLE tend to be the most efficient estimators, with NGQMLE performing the best across all specifications considered (see Figures 22 and 24). Only at samples larger than $T = 10,000$ does there appear any appreciable difference between NGQMLE and VTNGQMLE, with that difference favoring NGQMLE. In these same (very) large sample cases ($\geq 50,000$), both FAN and LSE outperform VTNGQMLE but neither outperforms NGQMLE. Consequently, not only is $\widehat{\beta}_n$ from NGQMLE robust to density misspecification (as further explored in the next section), NGQMLE is the best estimator for β_0 out of all the estimators considered.

Consequently, in the family of robust GARCH estimators, VTNGQMLE appears tough to beat. Compared to both FAN and LSE, VTNGQMLE is also the simplest to implement, requiring the fewest computational steps.

8 Explaining the Results

Let

$$\overline{\sigma}_t^2 = \frac{\sigma_t^2}{\omega_0}. \tag{44}$$

Conditional on (44), the model of (3) and (4) can be re-cast as

$$Y_t = \sqrt{\omega_0}\overline{\sigma}_t \epsilon_t, \tag{45}$$

$$\overline{\sigma}_t^2 = 1 + \left(\frac{\alpha_0}{\omega_0}\right) Y_{t-1}^2 + \beta_0 \overline{\sigma}_{t-1}^2. \tag{46}$$

From (45), the constant parameter ω_0 can be seen as the scale of the model's innovations. From (46), "reactivity" of the conditional variance to the previous period's innovation is seen to depend on scale.[16] Also from (46), the portion of the previous period's conditional variance affecting the current period's conditional variance is seen to be invariant to scale. Consequently, difficulties in estimating ω_0 have the potential to adversely impact the estimation of α_0, while such difficulties should not impact the estimation of β_0. Conversely, improvements in the estimation of ω_0 have the potential to benefit the estimation of α_0, as shown; e.g., by Fan et al. (2014).

Consider, next, the following generalization to the model of (3),

$$\begin{aligned} Y_t &= \eta_{f,0} \sigma_t \epsilon_t, \\ &= \ddot{\sigma}_t \epsilon_t \end{aligned} \tag{47}$$

where

$$\eta_{f,0} = \underset{\eta_f > 0}{\arg\max} E\left[-\log \eta_f + \log f\left(\frac{\epsilon}{\eta_f}\right)\right], \tag{48}$$

with the expectation is taken under g, and

$$\begin{aligned} \ddot{\sigma}_t^2 &= \left(\eta_{f,0}^2 \omega_0\right) + \left(\eta_{f,0}^2 \alpha_0\right) Y_{t-1}^2 + \beta_0 \ddot{\sigma}_{t-1}^2 \\ &= \overline{\omega}_0 + \overline{\alpha}_0 Y_{t-1}^2 + \beta_0 \ddot{\sigma}_{t-1}^2. \end{aligned} \tag{49}$$

(47) and (48) herald from Fan et al. (2014, eq. 6), where $\eta_{f,0}$ acts as a scale adjustment parameter.[17] The model of (47) and (48) compliments the finding from Newey and Steigerwald (1997) that GARCH-style models require additional parameters for correcting discrepancies between f and g, so as to ensure identification of NGQMLE.[18] When either $f = g$ or $f \propto e^{\frac{-x^2}{2}}$, $\eta_{f,0} = 1$ (see Fan et al., 2014, Proposition 1), in which case, the baseline model of (3) and (4) applies. As a result, no adjustment factor is necessary for the scale estimate from QMLE. However, when $f \neq g$ and $f \propto e^{\frac{-x^2}{2}}$ does not hold (as is the case here), $\eta_{f,0} \neq 1$. In this case, owing to (49), incorrectly assuming that $\eta_{f,0} = 1$ results in a biased estimate of scale. Moreover, the same bias impacting scale will also (and equally) impact "reactivity."

Consider estimation of (47) and (49) ignoring the presence of $\eta_{f,0}$ and, therefore, implicitly assuming

[16] That is, GARCH "reactivity" is the ARCH parameter normalized by the (unconditional) scale of the model's innovations.

[17] Specifically, $\eta_{f,0}$ measures the "distance" between f and g.

[18] Those discrepancies relate to location and scale.

$\eta_{f,0} = 1$.[19] The following two cases are considered: $E\left(Y_t^4\right) < \infty$ (the thin-tailed case); $E\left(Y_t^4\right) = \infty$ (the heavy-tailed case). Let $\widehat{\overline{\omega}}$ denote the NGQMLE estimate of $\overline{\omega}_0$ and $\widehat{\overline{\omega}}^{VT}$ the VTNGQMLE estimate, with parallel definitions holding for other parameters in (49).

ASSUMPTION 8.1 *Under both the thin- and heavy-tailed cases,*

$$p\lim\left(\widehat{\overline{\omega}}_n\right) = \overline{\omega}_0; \quad p\lim\left(\widehat{\overline{\alpha}}_n\right) = \overline{\alpha}_0; \quad p\lim\left(\widehat{\beta}_n\right) = \beta_0. \tag{50}$$

Under Assumption 8.1, neither ω_0 nor α_0 are identified, owing to the distorting presence of $\eta_{f,0}$. What are identified, however, are $\overline{\omega}_0$ and $\overline{\alpha}_0$, which can be interpreted as reduced-form parameters. Consequently, NGQMLE (minus any scale correction) consistently estimates the reduced-form conditional variance in (49). The mistake, then, is treating $\widehat{\overline{\omega}}_n$ and $\widehat{\overline{\alpha}}_n$ as structural estimates. When $f \neq g$ and $f \propto e^{\frac{-x^2}{2}}$ does not hold, NGQMLE under-identifies the (structural) GARCH model.

Monte Carlo results support Assumption 8.1. In (50), bias in $\widehat{\overline{\omega}}_n$ and $\widehat{\overline{\alpha}}_n$ as estimates of ω_0 and α_0, respectively, is precisely the same, as it stems from the same distorting property introduced by $\eta_{f,0}$. This prediction is confirmed by comparing Figures 17 and 25. In large samples, the size and sign of the bias in NGQMLE estimates for ω_0 and α_0, respectively, are identical across all simulation designs considered for which $f \neq g$. In addition, no (asymptotic) bias is detected for the NGQMLE estimates for β_0.

Under both the thin- and heavy-tailed cases, $\widehat{\overline{\omega}}_n^{VT}$ is given by

$$\widehat{\overline{\omega}}_n^{VT} = \widehat{\eta}_{f,n}^2 \widehat{\omega}_n = \widehat{\eta}_{f,n}^2 \left\{\widehat{\sigma}_n^2 \left(1 - \widehat{\overline{\alpha}}_n^{VT} - \widehat{\beta}_n\right)\right\},$$

in which case, the scale of the GARCH model innovations is, in turn, a scaled version of the unconditional variance of Y_t, where the scaling coefficients are the parameters governing short-term, conditional variance dynamics. Under the thin-tailed case, the probability limit of $\widehat{\overline{\omega}}_n^{VT}$ is

$$\begin{aligned} p\lim\left(\widehat{\overline{\omega}}_n^{VT}\right) &= \eta_{f,0}^2 \left\{\sigma_0^2 \left(1 - \overline{\alpha}_0 - \beta_0\right)\right\} \\ &= \eta_{f,0}^2 \omega_0 - \alpha_0 \sigma_0^2 \eta_{f,0}^2 \left(\eta_{f,0}^2 - 1\right) \\ &= p\lim\left(\widehat{\overline{\omega}}_n\right) - \alpha_0 \sigma_0^2 \eta_{f,0}^2 \left(\eta_{f,0}^2 - 1\right), \end{aligned} \tag{51}$$

where the second equality follows from $p\lim\left(\widehat{\eta}_{f,n}^2\right) = \eta_{f,0}^2$ by Fan et al. (2014, Theorem 1) and $p\lim\left(\widehat{\overline{\alpha}}_n^{VT}\right) =$

[19] Fan et al. (2014), in contrast, accounts for the presence of $\eta_{f,0}$ by estimating (48) in a preliminary step, using $\{\widehat{\epsilon}\}$ from QMLE. Unlike in Newey and Steigerwald (1997), it is not possible to jointly estimate $\eta_{f,0}$ along with the other GARCH parameters, since $\eta_{f,0}$ is not separately identified.

$\overline{\alpha}_0$ by the discussion under Case 2. In this case, $\widehat{\overline{\omega}}_n^{VT}$ is also a biased estimator of scale, but the form of the bias differs from that affecting $\widehat{\overline{\omega}}_n$.

Under the heavy-tailed case, the probability limit of $\widehat{\overline{\omega}}_n^{VT}$ becomes

$$\begin{aligned} p\lim\left(\widehat{\overline{\omega}}_n^{VT}\right) &= \eta_{f,0}^2\left\{\sigma_0^2\left(1-\alpha_0-\beta_0\right)\right\} \\ &= \eta_{f,0}^2\omega_0 \\ &= p\lim\left(\widehat{\overline{\omega}}_n\right), \end{aligned} \tag{52}$$

where the first equality follows from $p\lim\left(\widehat{\overline{\alpha}}_n^{VT}\right)=\alpha_0$ by Theorem 4 or 7. In this case, $\widehat{\overline{\omega}}_n^{VT}$ remains a biased estimator of scale, but the form of the bias is now precisely the same as that affecting $\widehat{\overline{\omega}}_n$.

Monte Carlo results support the prediction in (52). Biases in $\widehat{\overline{\omega}}_n^{VT}$ and $\widehat{\overline{\omega}}_n$ tend to be right on top of each other, in cases where $f \neq g$ (see Figure 25). Differences in these biases tend to be minor and, consequently, source to $p\lim\left(\widehat{\overline{\alpha}}_n^{VT}\right)=\overline{\alpha}_0$ being true to a first-order approximation.

Section 5 discovers that VTNGQMLE is a biased estimator for α_0 in the thin-tailed case (including the borderline case of Assumption 3.3) but a consistent estimator in the heavy-tailed case. (51) and (52) reveal that in neither case is VTNGQMLE a consistent estimator of scale. Fan et al. (2014) shows that accounting for the scale correction parameter in (47) and, therefore, solving the identification problem evident in (49), results in a more efficient estimator than QMLE, consistent with the argument put forth in Section 1. That is, in heavy-tailed cases, a heavy-tailed likelihood can distinguish heavy-tailed effects that are static in nature from heavy-tailed effects that arise due to short-run fluctuations in the conditional variance. A Gaussian likelihood, in contrast, cannot make this distinction. Holding heavy-tailed likelihoods back, however, is their inability to identify (and consistently estimate) scale, when those likelihoods depart from the truth. Fan et al. (2014) removes this impediment and shows that the resulting efficiency gains (relative to QMLE) can be substantial. Section 7 shows that VTNGQMLE can be a materially more efficient estimator for α_0 than FAN, LSE, and QMLE. But VTNGQMLE does not solve the identification problem associated with scale, whenever $f \neq g$. So what is going on? Results presented in Section 7 hint at an answer.

From Section 7, NGQMLE is shown (consistent with popular belief) to be a biased estimator away from the true innovation density. Despite being a biased estimator for α_0, NGQMLE is also shown (contrary to popular belief) to perform surprisingly well against robust alternatives in finite samples of surprisingly large sizes. The reason behind this surprisingly strong performance is the reduction in dispersion afforded by NGQMLE (see Figure 17). Analogous reductions in dispersion tend also to be afforded to the NGQMLE estimates of ω_0 (see Figure 26), where these reductions (relative to robust alternatives) tend to grow as η_0

shrinks. Moreover, these reductions in dispersion tend to be so great as to overwhelm the presence of bias, resulting in estimates of ω_0 that tend to be more efficient (in terms of either RMSE or MAE) than robust alternatives in finite samples as large as $T = 2,500$ (see Figures 27 and 28). From (52) and confirmed in Figure 25, NGQMLE and VTNGQMLE for ω_0 are closely linked in terms of bias. It turns out, NGQMLE and VTNGQMLE for ω_0 are also closely linked in terms of dispersion (see Figure 26) as well as in terms of efficiency (see Figures 27 and 28). Consequently, VTNGQMLE tends to produce a more efficient estimate of scale, in heavy-tailed cases, than does either FAN, LSE, or QMLE in sample sizes as large as $T = 2,500$. In these same cases, the outperformance of the VTNGQMLE estimates of α_0 over those from FAN, LSE, or QMLE is at its highest (compare Figures 21 and 27 as well as Figures 23 and 28). As a result, efficiency gains in VTNGQMLE over FAN, LSE, and QMLE can be attributed to the same factor identified in Fan et al. (2014): improvements in the estimate of scale for the GARCH model innovations. In contrast to Fan et al. (2014), however, in the case of VTNGQMLE, this improvement is being afforded by a biased estimator.

Further departing from Fan et al. (2014), improvements in the estimation of ω_0 do not appear to be the only factor contributing to the outperformance of VTNGQMLE over FAN, LSE, and QMLE: the very act of variance targeting itself appears to be a second contributing factor. Why? VTQMLE estimates of ω_0 are never more efficient than the QMLE alternatives in any of the simulation designs considered (see Figures 27 and 28). Despite this fact, VTNGQMLE estimates of α_0 nonetheless deliver sizable efficiency gains over QMLE alternatives in heavy-tailed cases (as also reported in Francq et al., 2011), where these gains tend to increase as the tails grow thicker. Explaining the difference in gains earned using VTNGQMLE over VTQMLE links to the former's improvements in estimating scale over the latter's. Both estimators, however, also appear to enjoy a boost afforded from the very act of VT. Supporting the existence of this shared boost is the fact that efficiency gains in the VTNGQMLE estimates of α_0 tend to persist into larger sample sizes even after the efficiency gains in the VTNGQMLE estimates of ω_0 have disappeared (again, compare Figures 21 and 27 as well as Figures 23 and 28).

9 S&P 500 Index-Related Volatility Estimation and Forecasting

The S&P 500 Index is a unique financial instrument in that the following three quantities are each directly observed daily: (i) the return on the index; (ii) option-implied volatility on the index (VIX); (iii) option-implied volatility on VIX, or option-implied volatility of volatility (VVIX). Using a historical time series of (i), it is standard to apply the model of (3) and (4) for the purpose of forecasting return variance and comparing out-of-sample results against the realized (return) variance (see; e.g., Andersen and Bollerslev,

1998). Following this convention, Section 9.0.1 compares $\widehat{\vartheta}_n$ from QMLE, VTNGQMLE, and FAN, using the results for each, respective, estimator to generate out-of-sample volatility forecasts 1-, 5-, 10-, and 21-days-ahead to determine which estimator produces the best forecasts at each horizon. Aiding these forecast comparisons are the loss functions RMSE and QLIKE, since both are "robust" in the sense discussed by Patton (2011).

Applying the model of (3) and (4) to the historical time series of (ii), though less standard, produces estimates of the historical volatility of VIX. Given that (iii) is directly observable, it is feasible to ask whether estimates of the historical volatility of VIX are useful in forecasting VVIX. In this sense, VVIX acts analogously as the realized (return) variance in that both (potentially) can be used to compare the efficacy of competing volatility forecasts, despite the fact that VVIX is not an unbiased measure (or proxy) of the realized volatility of VIX.[20] Section 9.1 proposes a forecasting model for VVIX that takes GARCH volatility of VIX forecasts as inputs.[21] 1-day-ahead VVIX forecasts are then constructed using 1-day-ahead volatility forecasts of VIX from QMLE, VTNGQMLE, and FAN, and the performance of these competing VVIX forecasts are compared using the RMSE and QLIKE loss functions.

9.0.1 Returns

Figures 1 and 2 depict rolling window estimates of $\widehat{\alpha}_n$ from daily S&P 500 returns, first over a lengthy period beginning 12/27/1999, and then over a shortened period immediately following the worst (financially speaking) of the COVID crisis. In the GARCH $(1,1)$ model, the parameter α_0 measures the "reactivity" of return variance to the previous period's return shock. Evident in Figure 1, return variance has become an increasingly reactive process through time, and in a statistically significant way. Evident in Figure 2, when reactivity is at its highest, $\widehat{\alpha}_n$ from QMLE is the largest, followed by FAN and then by VTNGQMLE. Oh and Patton (2024) document a tendency for QMLE-based GARCH volatility forecasts to "overshoot" their target (the realized return variance) following a large return shock. Figure 1 suggests this tendency to be the most acute in recent times. VTNGQMLE is the least impacted by this tendency (compared to both QMLE and FAN), however, making VTNGQMLE (in some sense) comparable to the local maximum likelihood estimator of Oh and Patton (2024).

In the GARCH $(1,1)$ model, $\phi_0 = \alpha_0 + \beta_0$ measures persistence in the variance process. Evident in

[20] The realized volatility of VIX is determined under the historical measure, while VVIX is determined under the risk-neutral measure. Consequently, the latter contains a variance-of-the-variance risk premium not present in the former (see; e.g., Huang et al., 2019).

[21] This model provides reduced-form scale corrections for the variance-of-variance risk premium, and so can be interpreted as internalizing the bias in VVIX as a proxy for the (latent) realized volatility of VIX.

Figure 3, variance persistence has been on the decline in recent years and in a statistically significant way. This decline is the most acute under QMLE. Evident in Figure 4, VTNGQMLE and FAN, in contrast, both indicate more modest declines in variance persistence. A variance process that is more reactive and less persistent is harder to forecast. The fact that both VTNGQMLE and FAN dampen these trends, (potentially) foreshadow their tendency to produce more stable and, thus, more reliable variance forecasts.

Figure 5 depicts $2 \times \iota_0$ estimates from the Hill (1975) estimator, together with one-sided 95% confidence bands, constructed using the standard error estimator in Hill (2010).[22] In the middle of the sample, these estimates provide no evidence in favor of

$$H_0 : \iota_0 < 2, \tag{53}$$

thus indicating QMLE to be $\sqrt{n}$ asymptotically normal. Under Cases 4–6, in contrast, VTNGQMLE has a non-Gaussian limit, to which convergence is slower than $\sqrt{n}$. Collectively, these results imply that, in the middle of the sample (when GARCH volatility was relatively less reactive and relatively more persistent), QMLE performed better than VTNGQMLE. Towards the end of the sample, however, there is now evidence favoring (53); in fact, evident in Figure 5, $\widehat{\iota}_n < 2$. From Hall and Yao (2003, Theorem 2.1), when $\iota_0 < 2$, QMLE also has a non-Gaussian limit, with a slower rate of convergence compared to $\sqrt{n}$. At the end of the sample, therefore, it is less apparent that QMLE should outperform VTNGQMLE.

Table 2 summarizes out-of-sample comparisons of the GARCH volatility forecasts produced by QMLE, VTNGQMLE, and FAN, respectively, using the RMSE and QLIKE loss functions and the standard "RV5" proxy for the latent variance. Comparisons are conducted over two forecast evaluation samples, one beginning on 5/1/2020 and one on 1/3/2022 (the approximate date where the difference between $\widehat{\alpha}_n$ from QMLE and VTNGQMLE gaps out and remains wide through to the end of the sample; see, Figure 2). Over these samples, k-period-ahead forecasts are generated each day, where $k \in \left(\begin{array}{cccc} 1, & 5, & 10, & 21 \end{array} \right)$. By RMSE, QMLE is the best; although, VTNGQMLE is fairly close behind. FAN, interestingly, tends to noticeably lag both QMLE and VTNGQMLE. By QLIKE, however, a different story emerges. In this case, VTNGQMLE is the consistent winner, while FAN continues to lag behind.

9.1 VIX

It is standard convention to model SPX log returns, since the underlying index levels appear (at least) to be well approximated as an $I(1)$ process. VIX levels, on the other hand (precisely because they measure volatilities), should be both strictly stationary and ergodic (see; e.g., Nelson, 1990, and Lumsdaine, 1996).

[22] Specifically, depicted in Figure 5 are tail index estimates for $\{|\widehat{\epsilon}_t|\}_{t=1}^{T}$ as determined using VTNGQMLE.

Consequently, it should be not only feasible, but also preferable, to extract the conditional variance of VIX directly from VIX levels, rather than from VIX log returns, with one important caveat. When modeling daily SPX log returns, it is also standard to ignore the conditional mean, since it is small and doing so exercises (very) little impact on $\widehat{\vartheta}_n$. The VIX series, however (again, because it is a series of volatilities), clusters, and the degree of this clustering indicates that conditional mean dynamics are important. Andersen et al. (2003) study the realized return variance series and find it to display long-memory properties. The model of Corsi (2009) uses lower-frequency covariates as proxies for long-memory properties. Motivated by these results, consider the following extension of the model in (3) and (4).

$$\rho(L)(1-L)^{d_0} Y_t = \theta(L)\epsilon_t, \tag{54}$$

$$\epsilon_t = \sigma_t \eta_t, \qquad \eta_t \sim i.i.d.\ D(0,\ 1), \tag{55}$$

$$\sigma_t^2 = \omega_0 + \alpha_0 \epsilon_{t-1}^2 + \beta_0 \sigma_{t-1}^2, \tag{56}$$

where $\rho(L) = \left(1 - \rho_{Y,0}L\right)$; $\theta(L) = \left(1 - \theta_{\epsilon,0}L\right)$, and L is the lag operator. (54) is an ARFIMA $\left(\begin{array}{ccc} 1, & d, & 1 \end{array}\right)$ model for $\{Y_t\}_{t\in\mathbb{Z}}$, where $d \in \left(\begin{array}{ccc} 0, & 0.50 \end{array}\right)$ governs long-memory dynamics. The estimator for (54) is full maximum likelihood (see Sowell, 1992). Using this estimator, $\{\widehat{\epsilon}_t\}_{t=1}^T$ is obtained and from which (55) and (56) are estimated in a second step.

Table 3 summarizes estimation results of (54) on a lengthy VIX sample (see the Notes to Table 3 for additional details). The estimate $\widehat{d}$ is inside of, but near, it's upper bound, indicating the VIX series to be a covariance stationary and (strongly) long-memory process. As a benchmark, parameter estimates including the constraint $d = 0$ are also summarized in Table 3, where this constraint forces the conditional mean of VIX to display only short-memory properties. Notice that $\widehat{\rho}_Y$ is significantly different in the two cases, with $\widehat{\rho}_Y$ being much closer to 1 in the case where $d = 0$, compared to the case where d is (jointly) estimated. When $d = 0$, $\widehat{\rho}_Y$ is forced to perform "double-duty," controlling for both short- and long-run dynamics. When d is freely estimated, on the other hand, $\widehat{\rho}_Y$ only governs short-run dynamics, while $\widehat{d}$ determines long-run dynamics. In the case of VIX, at least, allowing for long-run dynamics results in less persistent short-run dynamics.

Figures 7 and 8 depict rolling window estimates of $\widehat{\alpha}_n$ from VIX (see the Notes to Figures7–10 for additional details). Analogous to the case for return variance in Figures 1 and 2, VIX variance "reactivity" has been increasing through time. The level of VIX variance "reactivity," however, is higher than that of return variance "reactivity," and in a statistically significant way (compare Figure 8 against Figure 2).

Differences in $\widehat{\alpha}_n$ between QMLE, VTNGQMLE, and FAN also appear accentuated in the VIX variance case, compared to the return variance case. Specifically, evident in Figure 8, $\widehat{\alpha}_n$ from VTNGQMLE appears materially muted compared to either QMLE or FAN. Since heightened values of $\widehat{\alpha}_n$ tend to be associated GARCH volatility forecast "overshoot," VTNGQMLE appears (far) less prone to this difficulty than either QMLE or FAN.

Figures 9 and 10 track persistence in VIX variance through time. In contrast to the case for return variance (see Figures 3 and 4), where all three estimators tend to indicate a declining trend in persistence, in the case of VIX variance, only VTNGQMLE signals a declining trend; QMLE and FAN both imply increasing trends, occurring at the end of the sample. What's more, towards the end of the sample, $\widehat{\phi}_n > 1$ for both QMLE and FAN, indicating the variance of VIX to be either "integrated" or even explosive, while for VTNGQMLE, $\widehat{\phi}_n$ remains (comfortably) inside of the unit boundary. VVIX (owing to it being observable) appears to be, not only mean stationary, but also covariance stationary.[23] It seems counterintuitive, then, for implied vol-of-vol to appear covariance stationary, while historical vol-of-VIX appears (under QMLE and FAN, at least) either "integrated" or explosive. Regardless, for QMLE, VTNGQMLE, and FAN, material differences between in-sample estimates foreshadow accentuated differences between out-of-sample volatility forecasts, compared to the return variance case.

Figures 11 and 12 depict rolling $2 \times \iota_0$ estimates for GARCH $(1,1)$ model innovations to VIX (see the Notes to Figures 11 and 12 for additional details). Consider the one-side null of $\iota_0 \geq 2$. The full sample offers (very) little support for this null (see Figure 11), and emerging from the COVID crisis, there is no support for this null (see Figure 12). Monte Carlo results under Case 6 evidence material efficiency gains of VTNGQMLE over both QMLE and FAN. Figures 11 and 12 support Case 6 as being empirically relevant for the variance of VIX. Additionally, notice that if $\iota_0 < 2$ (as is strongly supported by Figures 11 and 12), then QMLE has a non-Gaussian limit and a convergence rate slower than $\sqrt{n}$ (see Hall and Yao, 2003, Theorem 2.1), comparable to the findings for VTNGQMLE in Section 5.

Let $\widehat{\sigma}_{t \mid t-1}$ denote the out-of-sample GARCH volatility forecast for VIX from (56). Figure 13 compares $\widehat{\sigma}_{t \mid t-1}$ from VTNGQMLE to VVIX on date t. Visually, out-of-sample GARCH volatility forecasts for VIX display similar dynamics compared to VVIX. These visual similarities are confirmed by a correlation coefficient of 0.53 between $\left\{\widehat{\sigma}_{t \mid t-1}\right\}_{t=1}^{T}$ and $\{VVIX_t\}_{t=1}^{T}$ for the full forecast evaluation sample (see the Notes to Figures 13–15 for additional details). Also visually apparent in Figure 13 is that the two series are not on the same scale. This visual dissimilarity should not be that surprising, since VVIX is anticipated to

[23] Estimating (54) on VVIX produces $\widehat{d} < 0.50$ and $\widehat{\rho}_{Y,0} < 1$. These results are not reported herein but are available upon request.

contain a vol-of-vol risk premium that should not be present in the historical volatility of VIX (see; e.g., Huang et al., 2019). This scale difference needs to be addressed, however, if $\left\{\widehat{\sigma}_{t \mid t-1}\right\}_{t=1}^{T}$ is to serve as a forecasting instrument for $\{VVIX_t\}_{t=1}^{T}$. Towards that end, consider the following model for adjusting the scale of $\sigma_{t \mid t-1}$ for the purpose of forecasting a target variable U_t. Let

$$V_t = \frac{U_t}{\sigma_t}, \tag{57}$$

where σ_t is given in (56).

$$V_t = \zeta_0 + \rho_{V,0} V_{t-1} + \theta_{\upsilon,0} \nu_{t-1} + \nu_t, \tag{58}$$

where $\{\nu_t\}_{t \in \mathbb{Z}^*}$ are i.i.d. innovations and $\mathbb{Z} \in \mathbb{Z}^*$, so that

$$V_{t \mid t-1} = \zeta_0 + \rho_{V,0} V_{t-1} + \theta_{\upsilon,0} \nu_{t-1}$$

and

$$U_{t \mid t-1} = V_{t \mid t-1} \times \sigma_{t \mid t-1}. \tag{59}$$

In the current application, $U_t = VVIX_t$. Dynamics in (58) are limited to being short-memory. Rolling estimates of $\rho_{V,0}$ (not reported here, but available upon request) are all comfortably inside of the unit boundary, indicating that short-run dynamics (at least as a proxy), are not a bad fit; especially, since only short-run forecasts are being made.

(57)–(58) control for the vol-of-vol risk premium in VVIX, allowing that risk premium to exercise both constant and time-varying effects on scale. This time-varying scale factor is then forecast out-of-sample, and the resulting out-of-sample forecast is combined with an out-of-sample GARCH volatility of VIX forecast to produce the forecast of VVIX in (59). The complete model of (54)–(59), then, produces a forecast of VVIX that uses the GARCH volatility of VIX as its principle input.

The dynamic scale factor model of (57)–(59) additionally, however, has a more general interpretation. Consider U_t as an observable proxy for the true (and latent) volatility that $\sigma_{t \mid t-1}$ is intended to forecast. For illustrative purposes, suppose $\sigma_{t \mid t-1}$ is the GARCH return volatility from the previous section, so that a good candidate for U_t is the realized return volatility.[24] In this case, $\zeta_0 = 1$ and $\rho_{V,0} = 0$ in (58), since the realized return volatility is an unbiased estimator for the true (and latent) return volatility (see; e.g., Barndorff-Nielsen and Shephard, 2004). Consequently, $\sigma_{t \mid t-1}$ can be used as an unadjusted forecast

[24] In other words, $U_t = \sqrt{RV5_t}$.

for the realized return volatility, consistent with standard practice. With this illustrative example in mind, consider VVIX, not as the target variable of interest directly, but rather as an observable proxy for the true (and latent) volatility of VIX. Owing to the presence of a vol-of-vol risk premium, VVIX can be anticipated to be a biased proxy for the volatility of VIX. The model of (57)–(59), then, can be seen as correcting for this bias, thus allowing VVIX to be used as the predicted variable in an evaluation of the GARCH volatility of VIX forecasts, where that evaluation looks to examine the efficacy of the GARCH volatility of VIX forecasts as predictive instruments for the true (and latent) volatility of VIX.

Figure 14 shows the results of applying the model in (57)–(59) to adjust (or correct) the GARCH volatility of VIX forecasts in Figure 13. As is evident, the predicted variable (VVIX) and the out-of-sample forecasts are now on the same scale. Moreover, the full-sample correlation between the adjusted forecasts and VVIX is, essentially, the same (0.51 versus 0.53), indicating that adjusting the forecasts does (practically) nothing to alter the predictive power of the GARCH volatility of VIX forecasts. Consequently, Figure 14 evidences that GARCH volatility of VIX is, in fact, useful at forecasting implied volatility-of-volatility (VVIX).

Also evidenced in Figure 14 is a tendency for the forecasts to "overshoot" their target. Perhaps this tendency shouldn't be too surprising, given the heightened levels of GARCH variance "reactivity" observed across different estimators (see Figure 7). To help mitigate this tendency, the following strategy (motivated by the "averaged-forecasting" approach used in De Nard et al., 2021) is adopted.[25] For any date t, it is possible to generate two forecasts, $\widehat{\sigma}_{t \mid t-1}$ and $\widehat{\sigma}_{t \mid t-2}$. The single point forecast for date t is then given by $\frac{\widehat{\sigma}_{t \mid t-1}+\widehat{\sigma}_{t \mid t-2}}{2}$, and this average forecast is substituted for $\widehat{\sigma}_{t \mid t-1}$ in (59). The result of performing this substitution is evidenced in Figure 15. The effect is a fairly apparent reduction in forecast variability, generally, and, more importantly, forecast extremes, specifically. Interestingly, the correlation between these adjusted average forecasts and VVIX increases to 0.65 (from 0.51). Forecast evaluations performed using the QMLE, VTNGQMLE, and FAN estimators all substitute $\frac{\sigma_{t \mid t-1}+\sigma_{t \mid t-2}}{2}$ for $\sigma_{t \mid t-1}$ in (59).

Table 2 also summarizes out-of-sample comparisons of the GARCH volatility forecasts produced by QMLE, VTNGQMLE, and FAN, respectively, using the RMSE and QLIKE loss functions and VVIX as the predicted variable.[26] Comparisons are conducted over the same two forecast evaluation samples used

[25]Consider the set of daily forecasts $\left\{\sigma_{t+h \mid t}\right\}_{h=1}^{H=21}$. The "averaged forecast" from this set is given by $H^{-1}\sum_{h=1}^{H=21}\sigma_{t+h \mid t}$. This "averaged forecast" is a proxy for the monthly volatility forecast. By analogy, the desired forecast here is a daily forecast. That daily forecast is proxied by an "averaged forecast" taken over a near neighborhood behind the desired forecast date. That is, the "averaged forecast" is $H^{-1}\sum_{h=1}^{H=2}\sigma_{t \mid t-h}$.

[26]Following the discussion above, VVIX can be interpreted either as the target variable being forecasted of a (biased) proxy for the true (and latent) volatility of VIX.

in evaluating the SPX return volatility estimates. Over these samples, only 1-day-ahead forecasts are considered. By RMSE, VTNGQMLE is now the clear winner, with FAN second and QMLE a close third. By QLIKE, the rankings remain unaltered. These results are consistent with the conclusions drawn from the parameter estimates depicted in Figures 7–10. These results further bolster the strong performance of VTNGQMLE in the Monte Carlo experiments.

Table 4 compares the "average forecasting" method for generating 1-day-ahead VVIX forecasts to the standard method (both of which are described above). For comparison purposes, out-of-sample results from the "average forecasting" method are depicted in Figure 15, while results from the standard method are depicted in Figure 14. Consistent with these figures, the "average forecasting" method beats the standard method in terms of RSME. Somewhat surprising, the "average forecasting" method also beats the standard method in terms of QLIKE.

10 Conclusion

Motivated by the NGQMLE of Bollerlsev (1987), this paper considers the VTNGQMLE, determining its limiting properties, studying its finite-sample properties, and applying it in a series of empirical investigations into volatility forecasting. In heavy-tailed cases, VTNGQMLE is shown to be a robust estimator, like QMLE, FAN, and LSE. In these same cases, when the likelihood function is misspecified, VTNGQMLE is shown to perform (surprisingly) well, both in simulation and empirically. In fact, VTNGQMLE is shown to be very hard to beat, both by the popular QMLE and by alternative (robust) estimators aimed at improving the QMLE result.

Explaining the popularity of QMLE is it being robust and $\sqrt{n}$ asymptotically normal, under fairly general conditions. Previous works demonstrate that QMLE loses its Gaussian limit when the model errors become (very) heavy-tailed (see; e.g., Hall and Yao, 2003). In earlier years, this case, while theoretically interesting, did not appear empirically relevant. In recent times, however, this case has become empirically relevant. Moreover, in this case, both QMLE and multi-step estimators aimed at producing more efficient (relative to QMLE) estimates perform (relatively) poorly. VTNGQMLE, in contrast, performs markedly better; in part, because of its reliance upon a heavy-tailed (though misspecified) likelihood function that removes some emphasis from the ARCH parameter as the single model parameter responsible for capturing heavy-tailed features in the unconditional distribution of the random variable being modeled. In recent times, therefore, VTNGQMLE appears to deserve serious consideration over QMLE and competing estimators because VTNGQMLE (i) is comparable in complexity relative to QMLE but (ii) delivers sizably

improved results.

References

[1] Andersen, T.G. & T. Bollerslev (1998) Answering the skeptics: yes, standard volatility models do provide accurate forecasts. *International Economic Review* 39(4), 885-905.

[2] Andersen, T.G., T. Bollerslev, F.X Diebold & P. Labys (2003) Modeling and forecasting realized volatility. *Econometrica* 71(2), 579-625.

[3] Barndorff-Nielsen, O.E. & N. Shephard (2004) Econometric analysis of realized covariation: high frequency based covariance, regression, and correlation in financial economics. *Econometrica* 72(3), 885-925.

[4] Basrak, B., R.A. Davis & T. Mokosch (2002) Regular variation of GARCH processes. Stochastic Processes and their Applications 99(1), 95-115.

[5] Berkes, I., L. Horváth & P. Kokoszka (2003) GARCH processes: structure and estimation. *Bernoulli* 9, 201-227.

[6] Bollerslev, T. (1986) Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31, 307-327.

[7] Bollerslev, T. (1987) A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return. *The Review of Economics and Statistics* 69(3), 542-547.

[8] Corsi, F. (2009) A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics* 7(2), 174-196.

[9] Davis, R.A., & T. Hsing (1995) Point process and partial sum convergence for weakly-dependent random variables with infinite variance. *The Annals of Probaility* 23(2), 879-917.

[10] De Nard, G., O. Ledoit & M. Wolf (2021) Factor models for portfolio selection in large dimensions: the good, the better, and the ugly. *Journal of Financial Econometrics* 19(2), 236-257.

[11] Engle, R.F., & G. Gonzalez-Rivera (1991) Semiparametric ARCH models. *Journal of Business and Economic Statistics* 9, 345-359.

[12] Engle, R.F., & J. Mezrich (1996) GARCH for groups. *Risk* 9, 36-40.

[13] Feller. W. (1971) *An introduction to probability theory and its applications*. New York: Wiley.

[14] Fan, J., L. Qi and D. Xiu (2014) Quasi-maximum likelihood estimation of garch models with heavy-tailed likelihoods. *Journal of Business and Economic Statistics* 32, 178-191.

[15] Francq, C., G. Lepage & J.M. Zakoïan (2011a), Two-stage non gaussian qml estimation of garch models and testing the efficiency of the gaussian qmle. *Journal of Econometrics* 165, 246-257.

[16] Francq, C., L. Horváth & J-M. Zakoïan (2011b) Merits and drawbacks of variance targeting in GARCH Models. *Journal of Financial Econometrics* 9(4), 619-656.

[17] Francq, C., & J.-M. Zakoïan (2004) Maximum likelihood estimation of pure garch and arma-garch processes. *Bernoulli* 10, 605-637.

[18] Hansen, B.E. (1994) Autoregressive conditional density estimation. *International Economic Review* 35, 705-730.

[19] Hall, P. & Q. Yao (2003) Inference in arch and garch models with heavy-tailed errors. *Econometrica* 71, 285-317.

[20] Hill, B.M. (1975) A simple general approach to inference about the tail of a distribution. *Annals of Statistics* 5, 1163-1174.

[21] Hill, J.B. (2010) On tail index estimation for dependent, heterogeneous data." *Econometric Theory* 26(5): 1398-1436.

[22] Hill, J.B. & E. Renault (2012) Variance targeting for heavy tailed time series. Unpublished manuscript.

[23] Huang, D., C. Schlag, I. Shaliastovich, and J. Thimme (2019) Volatility-of-volatility risk. *Journal of Financial and Quantitative Analysis* 54(6), 2423-2452.

[24] LePage, R., M. Woodroofe & J. Zinn (1981) Convergence to a stable distribution via order statistics. *The Annals of Probability* 9(4), 624-632.

[25] Lee, S.W. & B.E. Hansen (1994) Asymptotic theory for the garch(1,1) quasi-maximum likelihood estimator. *Econometric Theory* 10, 29-52.

[26] Lumsdaine, R.L. (1996) Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH (1,1) and covariance stationary GARCH (1,1) models. *Econometrica* 64, 575-596.

[27] McNeil, A.J., R. Frey & P. Embrechts (2015) *Quantitative Risk Management: Concepts Techniques, and Tools, Revised Edition*. Princeton University Press: Princeton, NJ.

[28] Nelson, D.B. (1990) Stationarity and persistence in the GARCH (1,1) model, *Econometric Theory* 6(3), 318-334.

[29] Mikosch, T. (1999) Regular variation, subexponentiality and their applications in probability theory. Lecture notes for the workshop "Heavy Tails and Queques," EURANDOM, Eindhoven, Netherlands.

[30] Mikosch, T. & C. Stărică (2000) Limit theory for the sample autocorrelations and extremes of a garch(1,1) process. *The Annals of Statistics* 28, 1427-1451.

[31] Mikosch, T. & D. Straumann (2006) Quasi-maximum-likelihood estimation in conditionally heteroskedastic time series: a stochastic recurrence equations approach. *The Annals of Statistics* 34, 2449-2495.

[32] Newey, W.K. & D.G. Steigerwald (1997) Asymptotic bias for quasi-maximum-likelihood estimators in conditional heteroskedasticity models. *Econometrica* 65(3), 587-599.

[33] Oh, D.H. & A.J. Patton (2024) Better the devil you know: improved forecasts from imperfect models, *Journal of Econometrics* 242(1), 105767.

[34] Patton, A.J. (2011) Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics* 160(1), 246-256.

[35] Resnick, S.I. (1987) *Extreme Values, Regular Variation, and Point Processes*. New York: Springer-Verlag.

[36] Sowell, F. (1992) Maximum likelihood estimation of stationary univariate fractionally integrated time series models. *Journal of Econometrics* 53, 165-188.

[37] Straumann, D. & T. Mikosch (2006) Quasi-maximum likelihood estimation in conditionally heteroskedastic time series: a stochastic recurrence approach. *The Annals of Statistics* 34(5), 2449-2495.

[38] Vaynman, I. & B.K. Beare (2014) Stable limit theory for the variance targeting estimator, in Y. Chang, T.B. Fomby & J.Y. Park (eds), *Essays in Honor of Peter C.B. Phillips*, vol. 33 of *Advances in Econometrics:* Emerald Group Publishing Limited, chapter 24, 639-672.

11 Appendix A (Proofs)

Proof of Theorem 3. Let

$$I_{tn} = I\left(\left|\sigma_t^2 - E\left(\sigma^2\right)\right| > a_n\right), \qquad J_{tn} = 1 - I_{tn} = I\left(\left|\sigma_t^2 - E\left(\sigma^2\right)\right| \leq a_n\right).$$

Then

$$\begin{aligned} a_n^{-1}\sum_{t=1}^{n} W_t &= a_n^{-1}\sum_{t=1}^{n} W_t \times I_{tn} + a_n^{-1}\sum_{t=1}^{n} W_t \times J_{tn} \\ &= I(b) + II(b) \end{aligned}$$

$$\begin{aligned} I(b) &= a_n^{-1}\sum_{t=1}^{n}\left\{\left(\epsilon_t^2 - 1\right) \times \sigma_t^2 \times I_{tn} - \left(\epsilon_t^2 - 1\right) \times E\left(\sigma^2\right) \times I_{tn} + \left(\epsilon_t^2 - 1\right) \times E\left(\sigma^2\right) \times I_{tn}\right\} \\ &= a_n^{-1}\sum_{t=1}^{n}\left(\epsilon_t^2 - 1\right) \times \left(\sigma_t^2 - E\left(\sigma^2\right)\right) \times I_{tn} + a_n^{-1}\sum_{t=1}^{n}\left(\epsilon_t^2 - 1\right) \times E\left(\sigma^2\right) \times I_{tn} \\ &= I(c) + II(c) \end{aligned}$$

By Markov's Inequality,

$$\begin{aligned} P\left(|II(c)| > C\right) &\leq C^{-1}E\left(\left|a_n^{-1}\sum_{t=1}^{n}\left(\epsilon_t^2 - 1\right) \times E\left(\sigma^2\right) \times I_{tn}\right|\right) \\ &\leq C^{-1}na_n^{-1}E\left(\left|\epsilon_t^2 - 1\right|\right) \times E\left(\sigma^2\right) \times E\left(I\left(\left|\sigma_t^2 - E\left(\sigma^2\right)\right| > a_n\right)\right) \\ &\leq Ca_n^{-1}nP\left(\left|\sigma_t^2 - E\left(\sigma^2\right)\right| > a_n\right) \\ &\leq Ca_n^{-1}\left(\frac{a_n^2 P\left(\left|\sigma_t^2 - E\left(\sigma^2\right)\right| > a_n\right)}{H\left(a_n\right)}\right) \\ &\longrightarrow 0 \end{aligned}$$

as $n \to \infty$, where the fourth inequality follows from (11), and the (weak) convergence result follows

from (12). Next, and also by Markov's Inequality,

$$\begin{aligned} P\left(|I(c)| > C\right) &\leq C^{-1}E\left(\left|a_n^{-1}\sum_{t=1}^{n}\left(\epsilon_t^2-1\right)\times\left(\sigma_t^2-E\left(\sigma^2\right)\right)\times I_{tn}\right|\right) \\ &\leq Cna_n^{-1}E\left(\left|\left(\epsilon_t^2-1\right)\times\left(\sigma_t^2-E\left(\sigma^2\right)\right)\times I_{tn}\right|\right) \\ &\leq Ca_n^{-1}nE\left(\left|\sigma_t^2-E\left(\sigma^2\right)\right|\times I\left(\left|\sigma_t^2-E\left(\sigma^2\right)\right|>a_n\right)\right) \\ &\leq C\frac{a_nE\left(\left|\sigma_t^2-E\left(\sigma^2\right)\right|\times I\left(\left|\sigma_t^2-E\left(\sigma^2\right)\right|>a_n\right)\right)}{H\left(a_n\right)} \\ &\longrightarrow 0 \end{aligned}$$

as $n\to\infty$, where the fourth inequality follows from (11), and the (weak) convergence result follows from (13). Consequently,

$$a_n^{-1}\sum_{t=1}^{n}W_t = a_n^{-1}\sum_{t=1}^{n}W_t\times J_{tn}+o_p\left(1\right),$$

and

$$\begin{aligned} Var\left(a_n^{-1}\sum_{t=1}^{n}W_t\times J_{tn}\right) &= na_n^{-2}E\left(\left(\epsilon_t^2-1\right)^2\times\sigma_t^4\times I\left(\left|\sigma_t^2-E\left(\sigma^2\right)\right|\leq a_n\right)\right) \\ &= Ca_n^{-2}nE\left(\sigma_t^4\times I\left(\left|\sigma_t^2-E\left(\sigma^2\right)\right|\leq a_n\right)\right) \\ &\leq C\times\left(\frac{E\left(\sigma_t^4\times I\left(\left|\sigma_t^2-E\left(\sigma^2\right)\right|\leq a_n\right)\right)}{E\left(\sigma_t^4\times I\left(\sigma_t^2\leq a_n\right)\right)}\right), \end{aligned} \tag{60}$$

where the inequality follows from (11). For sufficiently large n,

$$I\left(\left|\sigma_t^2-E\left(\sigma^2\right)\right|\leq a_n\right)\geq I\left(\sigma_t^2\leq a_n\right) \tag{61}$$

Given (61), because $E\left(\sigma^2\right)$ does not depend on n, $\exists$ a C such that

$$I\left(\left|\sigma_t^2-E\left(\sigma^2\right)\right|\leq a_n\right)=I\left(\sigma_t^2\leq a_n+C\right) \tag{62}$$

Given (62),

$$\frac{E\left(\sigma_t^4 \times I\left(\left|\sigma_t^2 - E\left(\sigma^2\right)\right| \leq a_n\right)\right)}{E\left(\sigma_t^4 \times I\left(\sigma_t^2 \leq a_n\right)\right)} = \frac{\int_0^{a_n+C} \sigma^4 f\left(\sigma^2\right) d\sigma^2}{\int_0^{a_n} \sigma^4 f\left(\sigma^2\right) d\sigma^2} = 1 + \frac{\int_{a_n}^{a_n+C} \sigma^4 f\left(\sigma^2\right) d\sigma^2}{\int_0^{a_n} \sigma^4 f\left(\sigma^2\right) d\sigma^2} \tag{63}$$

Given (63),

$$\lim_{n\to\infty} Var\left(a_n^{-1}\sum_{t=1}^{n} W_t \times J_{tn}\right) = C < \infty. \tag{64}$$

Given (64), in turn, it is possible to apply a CLT to $a_n^{-1}\sum_{t=1}^{n} W_t$ following analogous arguments given by Hall and Yao (2003, p. 306-307). Let the result of this application be

$$a_n^{-1}\sum_{t=1}^{n} W_t \xrightarrow{d} N\left(0,\ V_{\overline{W}}\right). \tag{65}$$

(30) then follows from Slutsky's Theorem.■

Remark 9 *An alternative way of establishing that (60) is bounded is to note that, given Assumption 3.3,* $\left(\frac{E\left(\sigma_t^4 \times I\left(\left|\sigma_t^2 - E\left(\sigma^2\right)\right| \leq a_n\right)\right)}{E\left(\sigma_t^4 \times I\left(\sigma_t^2 \leq a_n\right)\right)}\right)$ *is a ratio of slowly varying functions that, as such, has a finite limit.*

Proof of Theorem 4. For $i = 1, 2$, let

$$X\left(i\right) = \left\{ \begin{array}{cc} \widehat{\eta}_n - \widetilde{\eta}_0 & \text{if } i = 1 \\ Z_n & \text{if } i = 2 \end{array} \right\}. \tag{66}$$

Under Case 2, it is established that $\sqrt{n}X\left(i\right)$ converges to a stable limit, so long as $\iota_0 > 1$. Consider

then

$$\begin{aligned} na_n^{-1}X(i) &= \left(n^{1/2}a_n^{-1}\right) \times \left(n^{1/2}X(i)\right) \\ &= \left(n^{1/2}\left(Cn^{1/\kappa_0}\right)^{-1}\right) \times \left(n^{1/2}X(i)\right) \\ &= C \times \left(n^{\frac{\kappa_0-2}{2\kappa_0}}\right) \times \left(n^{1/2}X(i)\right) \\ &= C \times o_p(1) \times O_p(1) \\ &= o_p(1). \end{aligned}$$

where the third equality follows, since $n^{1/2}X(i)$ converges to a stable limit, and $n^{\frac{\kappa_0-2}{2\kappa_0}} \longrightarrow 0$ as $n \to \infty$, since $\kappa_0 < 2$.■

Remark 10 *$n^{1/2}$ is increasing at a faster rate than $n^{\frac{\kappa_0-2}{2\kappa_0}}$ is decreasing. Consequently, $n^{1/2}X(i)$ reaches its stable limit first and then is driven towards zero by $n^{\frac{\kappa_0-2}{2\kappa_0}}$.*

Proof of Theorem 5. Starting from (29), for a $\varepsilon > 0$,

$$\begin{aligned} a_n^{-1}\sum_{t=1}^{n} W_t &= a_n^{-1}\sum_{t=1}^{n} W_t \times I\left(\sigma_t^2 > a_n\varepsilon\right) + a_n^{-1}\sum_{t=1}^{n} W_t \times I\left(\sigma_t^2 \leq a_n\varepsilon\right) \\ &= I(d) + II(d) \end{aligned} \tag{67}$$

$$\begin{aligned} Var\left(II(d)\right) &= na_n^{-2}Var\left(\left(\epsilon_t^2 - 1\right) \times \sigma_t^2 \times I\left(\sigma_t^2 \leq a_n\varepsilon\right)\right) \\ &= na_n^{-2}E\left(\left(\epsilon_t^2 - 1\right)^2 \times \sigma_t^4 \times I\left(\sigma_t^2 \leq a_n\varepsilon\right)\right) \\ &= Cna_n^{-2}E\left(\sigma_t^4 \times I\left(\sigma_t^2 \leq a_n\varepsilon\right)\right), \end{aligned} \tag{68}$$

where necessary for the third equality is $E\left(\left(\epsilon_t^2 - 1\right)^2\right) < \infty$, which is established by $E\left(\epsilon_t^4\right) < \infty$. Since given Theorem 1, σ_t^2 is regularly varying with tail index κ_0, for a function L that is slowly

varying at ∞,

$$\begin{aligned} E\left(\sigma_t^4 \times I\left(\sigma_t^2 \leq a_n\varepsilon\right)\right) &= \int_0^{a_n\varepsilon} \left(\sigma^2\right)^2 f\left(\left(\sigma^2\right)\right) d\sigma^2 \\ &\sim C\left(-\kappa_0\right) \int_0^{a_n\varepsilon} \left(\sigma^2\right)^{2-\kappa_0-1} L\left(\sigma^2\right) d\sigma^2 \\ &\sim C\left(-\kappa_0\right)\left(2-\kappa_0\right)^{-1} \left\{\left(\sigma^2\right)^{2-\kappa_0} L\left(\sigma^2\right) \mid_0^{a_n\varepsilon}\right\} \\ &\sim C\left(-\kappa_0\right)\left(2-\kappa_0\right)^{-1} \left(a_n\varepsilon\right)^2 \left(a_n\varepsilon\right)^{-\kappa_0} L\left(a_n\varepsilon\right) \\ &\sim C a_n^2 \varepsilon^2 P\left(\sigma^2 > a_n\varepsilon\right) \end{aligned} \tag{69}$$

where the first $\sim$ follows from Mikosch (1999, Theorem 1.2.9), the second $\sim$ from Mikosch (1999, Theorem 1.2.6), and the last $\sim$ from Lemma 1. Putting (69) and (68) together,

$$Var\left(II(d)\right) \sim C\varepsilon^2 n P\left(\sigma^2 > a_n\varepsilon\right),$$

in which case,

$$\lim_{n\to\infty} Var\left(II(d)\right) \sim C \times \varepsilon^{2-\kappa_0}$$

by Definition 1, and further

$$\lim_{n\to\infty}\lim_{\varepsilon\to 0} Var\left(II(d)\right) \sim 0,$$

since $\kappa_0 \in (1,2)$. As a result,

$$a_n^{-1}\sum_{t=1}^{n} W_t = a_n^{-1}\sum_{t=1}^{n} W_t \times I\left(\sigma_t^2 > a_n\varepsilon\right) + o_p\left(1\right). \tag{70}$$

Let

$$U_{n,\varepsilon} = a_n^{-1}\sum_{t=1}^{n} W_t \times I\left(\sigma_t^2 > a_n\varepsilon\right), \tag{71}$$

and define the function $T_\varepsilon :\ M_P\left(\overline{\mathbb{R}}_+^2 \setminus \{0\}\right) \longrightarrow \mathbb{R}$ as

$$T_\varepsilon\left(\sum_{i=1}^{\infty}\delta_{\mathbf{x}_i}\right) = \sum_{i=1}^{\infty}\left(X_{i,2} - X_{i,1}\right) \times I\left(X_{i,1} > \varepsilon\right),$$

such that, given (14), $U_{n,\varepsilon} = T_\varepsilon\left(N_n\right)$. From Lemma 2, $N_n \xrightarrow{d} N$ as $n \to \infty$, in which case,

$T_\varepsilon(N_n) \xrightarrow{d} T_\varepsilon(N)$ as $n \to \infty$ by the continuous mapping theorem. Lastly,

$$T_\varepsilon(N) \xrightarrow{d} U_{\sigma^2}, \qquad \varepsilon \to 0, \tag{72}$$

by Davis and Hsing (1995, Theorem 3.1(ii)), where U_{σ^2} is a κ_0-stable random variable that can be expressed in terms of the P_i's and $\mathbf{Q}_{ij}$'s in Lemma 2. Given (29), (72) then results in

$$U_n \xrightarrow{d} \left(\frac{1-\beta_0}{1-\alpha_0-\beta_0}\right) U_{\sigma^2}.$$

■

Remark 11 *(71) is a special case of Vaynman and Beare (2014, eq. 35). Consequently, the result in (72) also follows from the proof of Vaynman and Beare (2014, Theorem 4), starting from equaton 35 and proceeding to the end.*

Proof of Theorem 6. Given (67),

$$\begin{aligned} II(d) &= a_n^{-1}\sum_{t=1}^{n}\left(\epsilon_t^2-1\right)\times I\left(\epsilon_t^2 > b_n\varepsilon\right)\times\sigma_t^2\times I\left(\sigma_t^2\le a_n\varepsilon\right) \\ &\quad + a_n^{-1}\sum_{t=1}^{n}\left(\epsilon_t^2-1\right)\times I\left(\epsilon_t^2 \le b_n\varepsilon\right)\times\sigma_t^2\times I\left(\sigma_t^2\le a_n\varepsilon\right) \\ &= I(e)+II(e). \end{aligned}$$

$$\begin{aligned} Var\left(II(e)\right) &= a_n^{-2}n\times Var\left(\left(\epsilon_t^2-1\right)\times I\left(\epsilon_t^2\le b_n\varepsilon\right)\times\sigma_t^2\times I\left(\sigma_t^2\le a_n\varepsilon\right)\right) \\ &= E\left(\left(\epsilon_t^2-1\right)^2\times I\left(\epsilon_t^2\le b_n\varepsilon\right)\right)\times a_n^{-2}n\times E\left(\sigma_t^4\times I\left(\sigma_t^2\le a_n\varepsilon\right)\right) \\ &= \left\{E\left(\epsilon_t^4\times I\left(\epsilon_t^2\le b_n\varepsilon\right)\right)+C\right\}\times a_n^{-2}n\times E\left(\sigma_t^4\times I\left(\sigma_t^2\le a_n\varepsilon\right)\right) \\ &\le \left\{b_n^2n^{-1}+C\right\}\times a_n^{-2}n\times E\left(\sigma_t^4\times I\left(\sigma_t^2\le a_n\varepsilon\right)\right) \\ &\le C\times a_n^{-2}n\times E\left(\sigma_t^4\times I\left(\sigma_t^2\le a_n\varepsilon\right)\right) \\ &\longrightarrow 0, \end{aligned}$$

as $n\to\infty$ and $\varepsilon\to 0$, where both inequalities follow from Assumption 5.1, and convergence sources

to (69) and the results that follow.

$$\begin{aligned}
Var\left(I\left(e\right)\right) &= E\left(\left(\epsilon_t^2-1\right)^2\times I\left(\epsilon_t^2>b_n\varepsilon\right)\right)\times a_n^{-2}n\times E\left(\sigma_t^4\times I\left(\sigma_t^2\leq a_n\varepsilon\right)\right)\\
&= \left\{E\left(\epsilon_t^4\times I\left(\epsilon_t^2>b_n\varepsilon\right)\right)+C\right\}\times a_n^{-2}n\times E\left(\sigma_t^4\times I\left(\sigma_t^2\leq a_n\varepsilon\right)\right)\\
&= E\left(\epsilon_t^4\times I\left(\epsilon_t^2>b_n\varepsilon\right)\right)\times a_n^{-2}n\times E\left(\sigma_t^4\times I\left(\sigma_t^2\leq a_n\varepsilon\right)\right)\\
&\quad +C\times a_n^{-2}n\times E\left(\sigma_t^4\times I\left(\sigma_t^2\leq a_n\varepsilon\right)\right)\\
&= \left(b_n^2n^{-1}\right)\times\left(b_n^{-2}n\right)E\left(\epsilon_t^4\times I\left(\epsilon_t^2>b_n\varepsilon\right)\right)\times a_n^{-2}n\times E\left(\sigma_t^4\times I\left(\sigma_t^2\leq a_n\varepsilon\right)\right)\\
&\quad +C\times a_n^{-2}n\times E\left(\sigma_t^4\times I\left(\sigma_t^2\leq a_n\varepsilon\right)\right)\\
&\leq \left(b_n^2n^{-1}\right)\times\left(\frac{E\left(\epsilon_t^4\times I\left(\epsilon_t^2>b_n\varepsilon\right)\right)}{H\left(b_n\right)}\right)\times a_n^{-2}n\times E\left(\sigma_t^4\times I\left(\sigma_t^2\leq a_n\varepsilon\right)\right)\\
&\quad +C\times a_n^{-2}n\times E\left(\sigma_t^4\times I\left(\sigma_t^2\leq a_n\varepsilon\right)\right)\\
&\leq \left(b_nn^{-1}\right)\times\left(\frac{b_nE\left(\epsilon_t^4\times I\left(\epsilon_t^2>b_n\varepsilon\right)\right)}{H\left(b_n\right)}\right)\times a_n^{-2}n\times E\left(\sigma_t^4\times I\left(\sigma_t^2\leq a_n\varepsilon\right)\right)\\
&\quad +C\times a_n^{-2}n\times E\left(\sigma_t^4\times I\left(\sigma_t^2\leq a_n\varepsilon\right)\right),
\end{aligned}$$

following from both Assumption 5.1 and (13) adapted for ϵ^2. Noting that $b_nn^{-1}=o(1)$,

$$\begin{aligned}
\lim_{n\to\infty}Var\left(I\left(e\right)\right) &\leq o\left(1\right)\times\varepsilon^{2-\kappa_0}+C\times\varepsilon^{2-\kappa_0}\\
&\longrightarrow 0,
\end{aligned}$$

as $\varepsilon\to 0$, (see, again, (69) and the results that follow). Consequently, (70) continues to hold and, from which, the result in (33) follows (see the proof of Theorem 5), since the limiting random variable continues to be determined by (only) the extremes of σ_t^2, despite $E\left(\epsilon_t^2\right)=\infty$.■

Proof of Theorem 7. Given (66),

$$\begin{aligned}
na_n^{-1}b_n^{-1}X\left(i\right) &= \left(n^{1/2}a_n^{-1}\right)\times\left(b_n^{-1}\right)\times\left(n^{1/2}X\left(i\right)\right)\\
&= C\times\left(n^{\frac{\kappa_0-2}{2\kappa_0}}\right)\times\left(n^{\frac{-1}{\iota_0}}\right)\times\left(n^{1/2}X\left(i\right)\right)\\
&= C\times o\left(1\right)\times o\left(1\right)\times O_p\left(1\right)\\
&= o_p\left(1\right),
\end{aligned}$$

where the third equality follows since $\kappa_0<2$, and $n^{1/2}X\left(i\right)$ converges to a stable limit (see the proof

of Theorem 4).■

Proof of Theorem 8. Starting from (35), the analog to (67) is

$$\begin{aligned} a_n^{-1}b_n^{-1}\sum_{t=1}^{n}W_t &= a_n^{-1}b_n^{-1}\sum_{t=1}^{n}W_t \times I\left(\sigma_t^2 > a_n\varepsilon\right) + a_n^{-1}b_n^{-1}\sum_{t=1}^{n}W_t \times I\left(\sigma_t^2 \leq a_n\varepsilon\right) \\ &= I(d) + II(d), \end{aligned}$$

preserving the same notation from the proof of Theorem 5. Consider

$$\begin{aligned} I(d) &= a_n^{-1}b_n^{-1}\sum_{t=1}^{n}W_t \times I\left(\epsilon_t^2 > b_n\varepsilon\right) \times I\left(\sigma_t^2 > a_n\varepsilon\right) \\ &\quad + a_n^{-1}b_n^{-1}\sum_{t=1}^{n}W_t \times I\left(\epsilon_t^2 \leq b_n\varepsilon\right) \times I\left(\sigma_t^2 > a_n\varepsilon\right) \\ &= I\left(di\right) + I\left(dii\right), \end{aligned}$$

and

$$\begin{aligned} I(d) &= a_n^{-1}b_n^{-1}\sum_{t=1}^{n}W_t \times I\left(\epsilon_t^2 > b_n\varepsilon\right) \times I\left(\sigma_t^2 \leq a_n\varepsilon\right) \\ &\quad + a_n^{-1}b_n^{-1}\sum_{t=1}^{n}W_t \times I\left(\epsilon_t^2 \leq b_n\varepsilon\right) \times I\left(\sigma_t^2 \leq a_n\varepsilon\right) \\ &= II\left(di\right) + II\left(dii\right). \end{aligned}$$

$$\begin{aligned} Var\left(II\left(dii\right)\right) &= a_n^{-2}b_n^{-2}nVar\left(\left(\epsilon_t^2 - 1\right) \times I\left(\epsilon_t^2 \leq b_n\varepsilon\right) \times \sigma_t^2 \times I\left(\sigma_t^2 \leq a_n\varepsilon\right)\right) \\ &= \left(n^{-1}\right) \times \left\{\left(b_n^{-2}n\right) \times E\left(\left(\epsilon_t^2 - 1\right)^2 \times I\left(\epsilon_t^2 \leq b_n\varepsilon\right)\right)\right\} \times \left\{\left(a_n^{-2}n\right) \times E\left(\sigma_t^4 \times I\left(\sigma_t^2 \leq a_n\varepsilon\right)\right)\right\} \\ &= \left(n^{-1}\right) \times II\left(diii\right) \times II\left(d\,iv\right), \end{aligned}$$

where

$$\begin{aligned} II\left(diii\right) &= \left(b_n^{-2}n\right) \times \left\{E\left(\epsilon_t^4 \times I\left(\epsilon_t^2 \leq b_n\varepsilon\right)\right) - 2E\left(\epsilon_t^2 \times I\left(\epsilon_t^2 \leq b_n\varepsilon\right)\right) + E\left(I\left(\epsilon_t^2 \leq b_n\varepsilon\right)\right)\right\} \\ &= \left(b_n^{-2}n\right) \times \left\{E\left(\epsilon_t^4 \times I\left(\epsilon_t^2 \leq b_n\varepsilon\right)\right) + C\right\} \\ &= \left(b_n^{-2}n\right) \times E\left(\epsilon_t^4 \times I\left(\epsilon_t^2 \leq b_n\varepsilon\right)\right) + o\left(1\right), \end{aligned}$$

since

$$\begin{aligned} b_n^{-2} n &= \left(C n^{\frac{1}{\iota_0}} \right)^{-1} n \\ &= C \times \left(n^{\frac{\iota_0 = 2}{\iota_0}} \right) \\ &\longrightarrow 0 \end{aligned} \tag{73}$$

as $n \to \infty$, given $\iota_0 < 2$.

$$\begin{aligned} E\left(\epsilon_t^4 \times I\left(\epsilon_t^2 \leq b_n \varepsilon\right)\right) &= \int_0^{b_n \varepsilon} \left(\epsilon^2\right)^2 f\left(\epsilon^2\right) d\epsilon^2 \\ &\sim C \times (-\iota_0) \times \int_0^{b_n \varepsilon} \left(\epsilon^2\right)^{2-\iota_0-1} L\left(\epsilon^2\right) d\epsilon^2 \\ &\sim C \times (-\iota_0) \times (2-\iota_0)^{-1} \times \left\{ \left(\epsilon^2\right)^{2-\iota_0} L\left(\epsilon^2\right) \Big|_0^{b_n \varepsilon} \right\} \\ &\sim C \times \left(b_n \varepsilon\right)^2 \left(b_n \varepsilon\right)^{-\iota_0} L\left(b_n \varepsilon\right) \\ &\sim C \times \left(b_n \varepsilon\right)^2 \times P\left(\epsilon^2 > b_n \varepsilon\right), \end{aligned} \tag{74}$$

where the first $\sim$ follows from Mikosch (1999, Theorem 1.2.9), the second from Mikosch (1999, Theorem 1.2.6), and the final from Lemma 1. Consequently,

$$II\left(diii\right) = C \times \left(\varepsilon\right)^2 \times nP\left(\epsilon^2 > b_n \varepsilon\right),$$

in which case,

$$\lim_{n \to \infty} II\left(diii\right) = C \times \left(\varepsilon\right)^{2-\iota_0},$$

which, in turn, implies that

$$\lim_{n \to \infty} \lim_{\varepsilon \to 0} II\left(diii\right) \sim 0,$$

since $\iota_0 < 2$, and

$$\lim_{n \to \infty} \lim_{\varepsilon \to 0} II\left(d\,iv\right) \sim 0,$$

given (68) and the arguments that (immediately) follow. As a result,

$$\lim_{n \to \infty} \lim_{\varepsilon \to 0} Var\left(II\left(di\right)\right) \sim 0. \tag{75}$$

Next,

$$\begin{aligned}
II\left(di\right) &= a_n^{-1}b_n^{-1}\sum_{t=1}^{n}\left\{\epsilon_t^2 \times I\left(\epsilon_t^2 > b_n\varepsilon\right) - I\left(\epsilon_t^2 > b_n\varepsilon\right)\right\} \times \sigma_t^2 \times I\left(\sigma_t^2 \leq a_n\varepsilon\right) \\
&= \left(n^{-\frac{1}{2}}b_n^{-1}\right)\sum_{t=1}^{n}\left\{\epsilon_t^2 \times I\left(\epsilon_t^2 > b_n\varepsilon\right) - I\left(\epsilon_t^2 > b_n\varepsilon\right)\right\} \times \left\{a_n^{-1} \times n^{\frac{1}{2}} \times \sigma_t^2 \times I\left(\sigma_t^2 \leq a_n\varepsilon\right)\right\} \\
&\leq C \times \left(n^{-\frac{1}{2}}b_n^{-1}\right)\sum_{t=1}^{n}\left\{\epsilon_t^2 \times I\left(\epsilon_t^2 > b_n\varepsilon\right) - I\left(\epsilon_t^2 > b_n\varepsilon\right)\right\} \\
&\leq C \times \left\{\left(n^{-\frac{1}{2}}b_n^{-1}\right)\sum_{t=1}^{n}\epsilon_t^2 \times I\left(\epsilon_t^2 > b_n\varepsilon\right) - b_n^{-1}\left\{n^{-\frac{1}{2}}\sum_{t=1}^{n}\epsilon_t^2 \times I\left(\epsilon_t^2 > b_n\varepsilon\right)\right\}\right\} \\
&\leq C \times \left\{\left(n^{-\frac{1}{2}}b_n^{-1}\right)\sum_{t=1}^{n}\epsilon_t^2 \times I\left(\epsilon_t^2 > b_n\varepsilon\right) - o\left(1\right) \times O_p\left(1\right)\right\} \\
&\leq C \times n^{-\frac{1}{2}} \times \left\{b_n^{-1}\sum_{t=1}^{n}\epsilon_t^2 \times I\left(\epsilon_t^2 > b_n\varepsilon\right)\right\} + o_p\left(1\right)
\end{aligned}$$

where the first inequality follows from (68) and the results that (immediately) follow, and the third inequality follows from a central limit theorem for i.i.d. data. Since (1) $\{\epsilon_t\}_{t\in\mathbb{Z}}$ is i.i.d., and (2) (8) holds for ϵ and the normalizing constants b_n,

$$\lim_{n\to\infty}\lim_{\varepsilon\to 0} b_n^{-1}\sum_{t=1}^{n}\epsilon_t^2 \times I\left(\epsilon_t^2 > b_n\varepsilon\right) = Z_{\epsilon^2}, \tag{76}$$

where Z_{ϵ^2} follows a ι_0-stable law (see; e.g., LePage et al., 1981, Theorem 1). Given (1) and (2), sufficient for (76) is that the distribution of ϵ has a balanced tail (see, Feller, 1971), as defined in Davis and Hsing (1995, eq. 1.2)). Given (76),

$$II\left(di\right) \leq C \times o\left(1\right) \times O_p\left(1\right) + o_p\left(1\right) \leq o_p\left(1\right),$$

in which case, given (75), $II\left(d\right) \leq o_p\left(1\right)$. Consequently, the asymptotic limit of $a_n^{-1}b_n^{-1}\sum_{t=1}^{n}W_t$ is

determined by $I\left(di\right)$ and $I\left(dii\right)$.

$$\begin{aligned} I\left(dii\right) &= a_n^{-1}b_n^{-1}\sum_{t=1}^{n}\left(\epsilon_t^2-1\right)\times I\left(\epsilon_t^2\leq b_n\varepsilon\right)\times\sigma_t^2\times I\left(\sigma_t^2>a_n\varepsilon\right) \\ &= a_n^{-1}b_n^{-1}\sum_{t=1}^{n}\left\{\epsilon_t^2\times I\left(\epsilon_t^2\leq b_n\varepsilon\right)-I\left(\epsilon_t^2\leq b_n\varepsilon\right)\right\}\times\sigma_t^2\times I\left(\sigma_t^2>a_n\varepsilon\right) \\ &= a_n^{-1}b_n^{-1}\sum_{t=1}^{n}\epsilon_t^2\times I\left(\epsilon_t^2\leq b_n\varepsilon\right)\times\sigma_t^2\times I\left(\sigma_t^2>a_n\varepsilon\right) \\ &\quad -a_n^{-1}b_n^{-1}\sum_{t=1}^{n}\sigma_t^2\times I\left(\sigma_t^2>a_n\varepsilon\right)\times I\left(\epsilon_t^2\leq b_n\varepsilon\right) \\ &= I\left(e\right)+II\left(e\right). \end{aligned}$$

$$\begin{aligned} II\left(e\right) &\leq \left\{a_n^{-1}\sum_{t=1}^{n}\sigma_t^2\times I\left(\sigma_t^2>a_n\varepsilon\right)\right\}\times\left\{b_n^{-1}\sum_{t=1}^{n}I\left(\epsilon_t^2\leq b_n\varepsilon\right)\right\} \\ &\leq \left\{a_n^{-1}\sum_{t=1}^{n}\sigma_t^2\times I\left(\sigma_t^2>a_n\varepsilon\right)\right\}\times\left(b_n^{-1}n^{\frac{1}{2}}\right)\left\{n^{-\frac{1}{2}}\sum_{t=1}^{n}I\left(\epsilon_t^2\leq b_n\varepsilon\right)\right\} \\ &\leq \left\{a_n^{-1}\sum_{t=1}^{n}\sigma_t^2\times I\left(\sigma_t^2>a_n\varepsilon\right)\right\}\times o\left(1\right)\times O_p\left(1\right), \end{aligned}$$

where the final inequality follows from (73) and a central limit theorem for i.i.d. data. Given Lemma 1 and (8),

$$\lim_{n\to\infty}\lim_{\varepsilon\to 0}a_n^{-1}\sum_{t=1}^{n}\sigma_t^2\times I\left(\sigma_t^2>a_n\varepsilon\right)=Z_{\sigma^2}, \tag{77}$$

where Z_{σ^2} follows a κ_0-stable law (see Davis and Hsing, 1995, Theorem 3.1.ii). Consequently,

$$II\left(e\right)\leq O_p\left(1\right)\times o_p\left(1\right)\leq o_p\left(1\right). \tag{78}$$

Next,

$$\begin{aligned} I\left(e\right) &= \left(n^{-\frac{1}{2}}\right)\times\left\{a_n^{-1}\sum_{t=1}^{n}\left\{\left(b_n^{-1}n^{\frac{1}{2}}\right)\times\epsilon_t^2\times I\left(\epsilon_t^2\leq b_n\varepsilon\right)\right\}\times\left\{\sigma_t^2\times I\left(\sigma_t^2>a_n\varepsilon\right)\right\}\right\} \\ &\leq C\times\left(n^{-\frac{1}{2}}\right)\times\left\{a_n^{-1}\sum_{t=1}^{n}\left\{\sigma_t^2\times I\left(\sigma_t^2>a_n\varepsilon\right)\right\}\right\} \\ &\leq C\times o\left(1\right)\times O_p\left(1\right), \end{aligned} \tag{79}$$

where the first inequality follows from (74) and the results that (immediately) follow, and the third

inequality follows from (77). Combining (78) and (79) implies that $I(dii) \leq o_p(1)$. What then remains to consider is

$$\begin{aligned} I(di) &= a_n^{-1}b_n^{-1}\sum_{t=1}^{n}\left(\epsilon_t^2-1\right)\times I\left(\epsilon_t^2>b_n\varepsilon\right)\times\sigma_t^2\times I\left(\sigma_t^2>a_n\varepsilon\right) \\ &= a_n^{-1}b_n^{-1}\sum_{t=1}^{n}\epsilon_t^2\times I\left(\epsilon_t^2>b_n\varepsilon\right)\times\sigma_t^2\times I\left(\sigma_t^2>a_n\varepsilon\right) \\ &\quad -a_n^{-1}b_n^{-1}\sum_{t=1}^{n}\sigma_t^2\times I\left(\sigma_t^2>a_n\varepsilon\right)\times I\left(\epsilon_t^2>b_n\varepsilon\right) \\ &= I(f)+II(f). \end{aligned}$$

First,

$$\begin{aligned} II(f) &\leq \left\{a_n^{-1}\sum_{t=1}^{n}\sigma_t^2\times I\left(\sigma_t^2>a_n\varepsilon\right)\right\}\times\left\{b_n^{-1}\sum_{t=1}^{n}I\left(\epsilon_t^2>b_n\varepsilon\right)\right\} \\ &\leq \left\{a_n^{-1}\sum_{t=1}^{n}\sigma_t^2\times I\left(\sigma_t^2>a_n\varepsilon\right)\right\}\times\left(n^{\frac{1}{2}}b_n^{-1}\right)\times\left\{n^{-\frac{1}{2}}\sum_{t=1}^{n}I\left(\epsilon_t^2>b_n\varepsilon\right)\right\} \\ &\leq I(g)\times o(1)\times II(g), \end{aligned}$$

where the third inequality follows from (73). Moreover, since $I(g)\xrightarrow{w}Z_{\sigma^2}$ (see 77), and $II(g)\xrightarrow{d}N(0,V)$ given a (standard) central limit theorem for i.i.d. data, $II(f)\leq o_p(1)$.

Finally, consider

$$\begin{aligned} I(f) &= a_n^{-1}b_n^{-1}\sum_{t=1}^{n}\left\{\epsilon_t^2\times I\left(\epsilon_t^2>b_n\varepsilon\right)\right\}\times\left\{\sigma_t^2\times I\left(\sigma_t^2>a_n\varepsilon\right)\right\} \\ &= \widetilde{U}_{n,\varepsilon}. \end{aligned} \tag{80}$$

Let $Y_t=\epsilon_t^2$ and $Z_t=\sigma_t^2$, noting that Y_t and Z_t are independent. Further, let $\widetilde{\mathbf{X}}_t=\left(\begin{array}{cc} Y_t, & Z_t \end{array}\right)$, with the associated polar coordinates $\left(\begin{array}{cc} \left|\widetilde{\mathbf{X}}_t\right|, & \theta\left(\widetilde{\mathbf{X}}_t\right) \end{array}\right)$, where $A\in B\left(\left[\begin{array}{cc} 0, & 2\pi \end{array}\right)^{d-1}\right)$, the Borel

subsets of $[\ 0,\quad 2\pi\ )^{d-1}$, for $1 \leq d \leq 2$. For a $r > 0$, consider

$$\begin{aligned}
& \frac{P\left(\ \left|\widetilde{\mathbf{X}}_t\right| > ur,\quad \theta\left(\widetilde{\mathbf{X}}_t\right) \in A\ \right)}{P\left(\left|\widetilde{\mathbf{X}}_t\right| > r\right)} \\
= & \frac{P\left(\ |Y_t| > ur,\quad |Z_t| > ur,\quad \theta\left(X_t\right) \in A\ \right)}{P\left(\ |Y_t| > r,\quad |Z_t| > r\ \right)} \\
= & \left\{\frac{P\left(|Y_t| > ur\right)}{P\left(|Y_t| > r\right)}\right\} \times \left\{\frac{P\left(|Z_t| > ur\right)}{P\left(|Z_t| > r\right)}\right\},
\end{aligned}$$

in which case,

$$\lim_{u\to\infty} \frac{P\left(\ \left|\widetilde{\mathbf{X}}_t\right| > ur,\quad \theta\left(\widetilde{\mathbf{X}}_t\right) \in A\ \right)}{P\left(\left|\widetilde{\mathbf{X}}_t\right| > r\right)} = \left\{r^{-\iota_0}\right\} \times \left\{r^{-\kappa_0}\right\} < \infty \tag{81}$$

$\forall\ r$, given Assumption 3.2 and Lemma 1. Given (81), in turn, Davis and Mikosch (1988, eq. 2.1) is satisfied for $\widetilde{\mathbf{X}}_t$ (see also Resnick, 1986). Moreover, given Carrasco and Chen (2002, Corollary 6), $\left\{\widetilde{\mathbf{X}}_t\right\}$ is strong mixing, in which case, Davis and Mikosch (1998, eq 2.3) is also satisfied. Next, note that time-dependence in $\left\{\widetilde{\mathbf{X}}_t\right\}$ drives entirely from $\{Z_t\}$, where

$$\begin{aligned}
\sigma_t^2 &= \omega + \alpha Y_{t-1}^2 + \beta\sigma_{t-1}^2 \\
&= \omega + \sigma_{t-1}^2\left(\alpha\epsilon_{t-1}^2 + \beta\right) \\
&= \sigma_{t-1}^2 A_t + B_t \\
&= \prod_{i=1}^{t} A_i\sigma_0^2 + \sum_{i=1}^{t}\prod_{j=i+1}^{t} A_j B_i \\
&= \mathbf{I}_{t,1}\sigma_0^2 + \mathbf{I}_{t,2},
\end{aligned} \tag{82}$$

where the final equality follows from recursive substitution. From Mikosch and Stărică (2000), (82) is a valid stochastic recurrence equation, for which

$$\begin{aligned}
P\left(\sigma_t^2 > a_n y \mid \sigma_0^2 > a_n y\right) &\leq P\left(\mathbf{I}_{t,1}\sigma_0^2 > a_n y \mid \sigma_0^2 > \frac{a_n y}{2}\right) + P\left(\mathbf{I}_{t,2} > \frac{a_n y}{2} \mid \sigma_0^2 > a_n y\right) \\
&\leq I\left(h\right) + II\left(h\right).
\end{aligned}$$

Using Markov's inequality,

$$\begin{aligned} I(h) &= \frac{P\left(\ \mathbf{I}_{t,1}\sigma_0^2 > \frac{a_n y}{2}\ ,\ \ \sigma_0^2 > a_n y\ \right)}{P\left(\sigma_0^2 > a_n y\right)} \\ &= \frac{P\left(\mathbf{I}_{t,1}\sigma_0^2 \times I\left(\sigma_0^2 > a_n y\right) > \frac{a_n y}{2}\right)}{P\left(\sigma_0^2 > a_n y\right)} \\ &\leq \frac{2 \times (a_n y)^{-1} \times E\left(\mathbf{I}_{t,1}\sigma_0^2 \times I\left(\sigma_0^2 > a_n y\right)\right)}{P\left(\sigma_0^2 > a_n y\right)} \\ &\leq \frac{2 \times (a_n y)^{-1} \times E\left(\mathbf{I}_{t,1}\right) \times E\left(\sigma_0^2 \times I\left(\sigma_0^2 > a_n y\right)\right)}{P\left(\sigma_0^2 > a_n y\right)}, \end{aligned}$$

where

$$E\left(\mathbf{I}_{t,1}\right) = E\left(\prod_{i=1}^{t} A_i\right) = \prod_{i=1}^{t} E\left(A_i\right) = E\left(A\right)^t = b^t,$$

$b < 1$, and

$$\begin{aligned} E\left(\sigma_0^2 \times I\left(\sigma_0^2 > a_n y\right)\right) &= \int_{a_n y}^{\infty} \sigma_0^2 f\left(\sigma_0^2\right) d\sigma_0^2 \\ &\sim C \times (-\kappa_0) \int_{a_n y}^{\infty} \left(\sigma_0^2\right)^{-\kappa_0} L\left(\sigma_0^2\right) d\sigma_0^2 \\ &\sim C \times (-\kappa_0)\left(-\kappa_0 + 1\right)^{-1} (a_n y)^{-\kappa_0 + 1} L\left(a_n y\right) \\ &\sim C \times (a_n y) \times (a_n y)^{-\kappa_0} L\left(a_n y\right) \\ &\sim C \times (a_n y) \times P\left(\sigma_0^2 > a_n y\right), \end{aligned}$$

with the first $\sim$ following from Mikosch (1999, Theorem 1.2.9), and the third $\sim$ following from Mikosch (199, Theorem 1.2.6(b)). As a result,

$$\begin{aligned} I(h) &\leq \frac{2 \times (a_n y)^{-1} \times b^t \times C \times (a_n y) \times P\left(\sigma_0^2 > a_n y\right)}{P\left(\sigma_0^2 > a_n y\right)} \\ &\leq C \times b^t. \end{aligned} \tag{83}$$

Next, by independence and Markov's inequality,

$$\begin{aligned} II\left(h\right) &= P\left(\mathbf{I}_{t,2} > \frac{a_n y}{2}\right) \\ &= P\left(\sum_{i=1}^{t}\prod_{j=i+1}^{t} A_j B_i > \frac{a_n y}{2}\right) \\ &\leq P\left(\sum_{i=1}^{\infty}\prod_{j=i+1}^{\infty} A_j B_i > \frac{a_n y}{2}\right) \\ &\leq \left(\frac{a_n y}{2}\right)^{-1} \times E\left(\sum_{i=1}^{\infty}\prod_{j=i+1}^{\infty} A_j B_i\right) \\ &\leq \left(\frac{a_n y}{2}\right)^{-1} \times E\sum_{i=1}^{\infty}\prod_{j=i+1}^{\infty} E\left(A_j\right) B_i \\ &\leq C \times a_n^1, \end{aligned} \tag{84}$$

since $E\left(A\right) < 1$. For a sequence of positive integers $\{r_n\}$, where $r_n \to \infty$ and $\frac{n}{r_n} \to \infty$ as $n \to \infty$, putting together the results in (83) and (84) produces

$$\begin{aligned} &\lim_{k\to\infty}\lim_{n\to\infty}\sup P\left(\bigvee_{k\leq|t|\leq r_n} \sigma_t^2 > a_n y \mid \sigma_0^2 > a_n y\right) \\ \leq\ &\lim_{k\to\infty}\lim_{n\to\infty}\sup 2\left(m+1\right)\sum_{t=k}^{r_n+m} P\left(\sigma_t^2 > a_n y \mid \sigma_0^2 > a_n y\right) \\ \leq\ &\lim_{k\to\infty} C \times \sum_{t=k}^{\infty} b^t \\ \leq\ &0, \end{aligned}$$

thus establishing that Davis and Mikosch (1998, eq 2.10) holds for $\widetilde{\mathbf{X}}_t$, which, in turn, establishes that γ in Davis and Mikosch (1998, eq 2.11) exists. Suppose that $\gamma \neq 0$.

Consider the function $\widetilde{T}_\varepsilon :\ M_P\left(\overline{\mathbb{R}}_+^2 \setminus \{0\}\right) \longrightarrow \mathbb{R}$ as

$$\widetilde{T}_\varepsilon\left(\sum_{i=1}^{\infty}\delta_{Y_i Z_i}\right) = \sum_{i=1}^{\infty}\left\{Z_i \times I\left(Z_i > \varepsilon\right)\right\}\left\{Y_i \times I\left(Y_i > \varepsilon\right)\right\}.$$

Further let

$$\widetilde{N}_n = \sum_{t=1}^{n}\delta_{a_n^{-1}b_n^{-1}Y_t Z_t}. \tag{85}$$

Given (80),

$$\widetilde{U}_{n,\varepsilon} = \widetilde{T}_{\varepsilon}\left(\widetilde{N}_n\right).$$

From Davis and Mikosch (1998, Proposition 3.1, Remark 3.2), $\widetilde{N}_n \xrightarrow{d} \widetilde{N}$ as $n \to \infty$, and then from the continuous mapping theorem, $\widetilde{T}_{\varepsilon}\left(\widetilde{N}_n\right) \xrightarrow{d} \widetilde{T}_{\varepsilon}\left(\widetilde{N}\right)$ as $n \to \infty$. Lastly, from Davis and Hsing (1995, Theorem 3.1(ii)),

$$\widetilde{T}_{\varepsilon}\left(\widetilde{N}\right) \xrightarrow{d} U_{\epsilon^2,\,\sigma^2}, \qquad \varepsilon \to 0, \tag{86}$$

where $U_{\epsilon^2,\,\sigma^2}$ is a κ_0-stable random variable that can be expressed in terms of quantities qualitatively similar to the P_i's and $\mathbf{Q}_{ij}$'s in Lemma 2, in which case,

$$\widetilde{U}_{n,\varepsilon} \xrightarrow{d} U_{\epsilon^2,\,\sigma^2}.$$

■

Proof of Theorem 9. A univariate analog to (15) is

$$S_n^i = \sum_{t=1}^{n} Y_t^i \qquad \text{for} \qquad i = 2, 4.$$

Given this univariate analog,

$$\begin{aligned}
na_n^{-2}b_n^{-2}\widehat{\tau}_n^2 &= \left(a_n^{-2}b_n^{-2}\right) \times S_n^4 - \left(na_n^{-2}b_n^{-2}\right) \times \left(n^{-1}S_n^2\right)^2 \\
&= \left(a_n^{-2}b_n^{-2}\right) \times S_n^4 - \left(na_n^{-2}b_n^{-2}\right) \times \left(n^{-1}a_nb_n\right)^2 \times \left(\left(a_n^{-1}b_n^{-1}\right) \times S_n^2\right)^2 \\
&= \left(a_n^{-2}b_n^{-2}\right) \times S_n^4 - \left(n^{-1}\right) \times \left(\left(a_n^{-1}b_n^{-1}\right) \times S_n^2\right)^2 \\
&= \left(a_n^{-2}b_n^{-2}\right) \times S_n^4 - o\left(1\right) \times O_p\left(1\right) \\
&= \left(a_n^{-2}b_n^{-2}\right) \times S_n^4 + o_p\left(1\right),
\end{aligned}$$

where the fourth equality follows from the proof of Theorem (8). Consequently, the asymptotic limit of $\widehat{\tau}_n^2$ is determined by S_n^4.

Also from the proof of Theorem 8, $\widetilde{\mathbf{X}}_t$ is regularly varying with tail index κ_0. By Mikosch (1999, Proposition 1.5.14), $\widetilde{\mathbf{X}}_t^2 = \begin{pmatrix} Y_t^2, & Z_t^2 \end{pmatrix}$ is regularly varying with tail index $\kappa_0/2$.

Let $\widetilde{N}_n^2 = \sum_{t=1}^{n} \delta_{a_n^{-2}b_n^{-2}Y_t^2 Z_t^2}$. Given $\widetilde{N}_n \xrightarrow{d} \widetilde{N}$ as $n \to \infty$, from the proof of Theorem 8, $\widetilde{N}_n^2 \xrightarrow{d} \widetilde{N}^2$

as $n \to \infty$, given Remark 2. Since (1) $\widetilde{\mathbf{X}}_t^2$ is regularly varying and (2) $\widetilde{N}_n^2 \xrightarrow{d} \widetilde{N}^2$ as $n \to \infty$, Davis and Hsing (1995, Theorem 3.1(i)) can be applied to establish

$$\left(a_n^{-2} b_n^{-2}\right) \times S_n^4 \xrightarrow{d} U_{\epsilon^4,\, \sigma^4},$$

where $U_{\epsilon^4,\, \sigma^4}$ is a $(\kappa_0/2)$-stable random variable that can be expressed in terms of the limiting points for Y_t^2/b_n^2 and Z_t^2/a_n^2.■

Proof of Theorem 11. Recalling the definition of $\widetilde{U}_{n,\varepsilon}$ from (80), let

$$\widetilde{U}_{n,\varepsilon}^2 = a_n^{-2} b_n^{-2} \sum_{t=1}^{n} \left\{\epsilon_t^4 \times I\left(\epsilon_t^2 > b_n \varepsilon\right)\right\} \times \left\{\sigma_t^4 \times I\left(\sigma_t^2 > a_n \varepsilon\right)\right\},$$

and define the function $\widetilde{T}_\varepsilon :\ M_P\left(\overline{\mathbb{R}}_+^2 \setminus \{0\}\right) \longrightarrow \mathbb{R}^2$ as

$$\widetilde{T}_\varepsilon\left(\sum_{i=1}^{\infty} \delta_{Y_i Z_i}\right) = \left(\ \sum_{i=1}^{\infty} \left\{Z_i \times I\left(Z_i > \varepsilon\right)\right\}\left\{Y_i \times I\left(Y_i > \varepsilon\right)\right\},\ \ \sum_{i=1}^{\infty} \left\{Z_i^2 \times I\left(Z_i > \varepsilon\right)\right\}\left\{Y_i^2 \times I\left(Y_i > \varepsilon\right)\right\}\ \right).$$

Further recall the definition of $\widetilde{N}_n$ from (85). Then

$$\left(\ \widetilde{U}_{n,\varepsilon},\ \ \widetilde{U}_{n,\varepsilon}^2\ \right) = \widetilde{T}_\varepsilon\left(\widetilde{N}_n\right).$$

The conditions requisite for

$$\widetilde{N}_n \xrightarrow{d} \widetilde{N}, \qquad n \to \infty$$

are established in the proof of Theorem 8. That

$$\widetilde{T}_\varepsilon\left(\widetilde{N}_n\right) \xrightarrow{d} \widetilde{T}_\varepsilon\left(\widetilde{N}\right), \qquad n \to \infty$$

then follows from the continuous mapping theorem. Lastly,

$$\widetilde{T}_\varepsilon\left(\widetilde{N}_n\right) \xrightarrow{d} \left(\ U_{\epsilon^2,\, \sigma^2},\ \ U_{\epsilon^4,\, \sigma^4}\ \right), \qquad \varepsilon \to 0$$

results from Davis and Mikosch (1998, Proposition 3.3), where the marginal limits are described at the end of the proofs to Theorems (8) and (9), respectively.■

12 Appendix B (Tables)

Table 1: Parameter Configurations

Specification	θ_0		
	ω_0	α_0	β_0
I	0.05	0.05	0.90
II	0.05	0.10	0.85
III	0.05	0.20	0.75
IV	0.05	0.30	0.65

Notes to Table 1. Different GARCH$(1,1)$ parameter values considered in the Monte Carlo simulations.

Table 2: Out-of-Sample Forecast Comparisons

Asset	Eval. Sample (Beg.)	Loss Function	Estimator	Forecast Horizon			
				1-Step	5-Steps	10-Steps	21-Steps
SPX	5/1/2020	RMSE	QMLE	6.687	7.152	7.367	7.487
			VTNGQMLE	6.820	7.338	7.593	7.713
			FAN	7.057	7.718	8.143	8.611
		QLIKE	QMLE	3.463	3.481	3.492	3.502
			VTNGQMLE	3.463	3.480	3.491	3.501
			FAN	3.468	3.488	3.503	3.522
	1/3/2022	RMSE	QMLE	6.476	6.790	7.074	7.213
			VTNGQMLE	6.550	6.902	7.206	7.263
			FAN	6.873	7.414	7.937	8.449
		QLIKE	QMLE	3.496	3.510	3.524	3.538
			VTNGQMLE	3.494	3.506	3.519	3.532
			FAN	3.500	3.516	3.535	3.559
VIX	5/1/2020	RMSE	QMLE	34.026			
			VTNGQMLE	26.301			
			FAN	31.441			
		QLIKE	QMLE	5.636			
			VTNGQMLE	5.628			
			FAN	5.633			
	1/3/2022	RMSE	QMLE	27.446			
			VTNGQMLE	18.747			
			FAN	24.307			
		QLIKE	QMLE	5.557			
			VTNGQMLE	5.548			
			FAN	5.554			

Notes to Table 2. Daily SPX, VIX, and VVIX levels source to Bloomberg, L.P. The GARCH $(1,1)$ model of (3) and (4) is estimated on SPX log returns using a fixed 10-year look-back window beginning on 4/30/2020 and rolling through the end of the sample on 10/29/2024. For $k \in \{1, 5, 10, 21\}$, on each day of the sample, k-period-ahead GARCH volatility forecasts are constructed using QMLE, VTNGQMLE, and FAN. RMSE and QLIKE loss functions evaluate the efficacy of these GARCH forecasts, using the standard "RV5" proxy for the latent variance. The ARFIMA $(1, d, 1)$ model of (54) is estimated on VIX levels using a fixed 20-year look-back window beginning on 4/30/2020 and rolling through the end of the data sample on 10/29/2024. The GARCH $(1,1)$ model of (55) and (56) is then estimated on the daily ARFIMA $(1, d, 1)$ model innovations using QMLE, VTNGQMLE, and FAN on a fixed 10-year look-back window beginning on 4/30/2020 and rolling through the end of the data sample on 10/29/2024, so as to produce the out-of-sample forecasts $\widehat{\sigma}_{t \mid t-1}$ and $\widehat{\sigma}_{t \mid t-2}$ on each date t of the sample. For each estimator, the dynamic scale factor model of (57)–(59) is estimated on a fixed 10-year look-back window beginning on 4/30/2020 and rolling through the end of the data sample on 10/29/2024, so as to produce $\widehat{U}_{t \mid t-1} = \widehat{VVIX}_{t \mid t-1}$ for each date t in the sample, where $\frac{\sigma_{t \mid t-1} + \sigma_{t \mid t-2}}{2}$ replaces $\sigma_{t \mid t-1}$ in (59). $\left\{\widehat{VVIX}_{t \mid t-1}\right\}_{t=1}^{T}$ formed using the competing estimators are compared using the RMSE and QLIKE loss functions.

Table 3: ARFIMA $\left(\ 1,\quad d,\quad 1\ \right)$ Estimates for VIX

model	para.	est.	stderror	95% C.I.	
ARFIMA $\left(\ 1,\quad d,\quad 1\ \right)$	ρ_Y	0.830	0.045	0.742	0.918
	θ	-0.494	0.065	-0.622	-0.366
	d	0.497	0.002	0.494	0.500
ARFIMA $\left(\ 1,\quad 0,\quad 1\ \right)$	ρ_Y	0.984	0.005	0.975	0.994
	θ	-0.138	0.032	-0.202	-0.075
	d	0.000			

Notes to Table 3. Daily VIX levels source to Bloomberg, L.P. The ARFIMA $\left(\ 1,\quad d,\quad 1\ \right)$ model is fit to these VIX levels over a lengthy sample beginning 1/2/1990 and running through 10/13/2022. Reported standard errors are robust in the Huber-White "sandwich" estimator sense.

Table 4: Out-of-Sample VVIX Forecast Construction Comparisons

Eval. Sample (Beg.)	Forecast	Loss Function RMSE	QLIKE
5/1/2020	1-Step	43.469	5.634
	Avg.	26.301	5.628
1/3/2022	1-Step	26.801	5.552
	Avg.	18.747	5.548

Notes to Table 4. Daily VIX levels source to Bloomberg, L.P. The ARFIMA $(1,\ d,\ 1)$ model of (54) is estimated on VIX levels using a fixed 20-year look-back window beginning on 4/30/2020 and rolling through the end of the data sample on 10/29/2024. The GARCH $(1,1)$ model of (55) and (56) is estimated on the daily ARFIMA $(1,\ d,\ 1)$ model innovations using VTNGQMLE on a fixed 10-year look-back window beginning on 4/30/2020 and rolling through the end of the data sample on 10/29/2024, so as to produce the out-of-sample forecasts $\widehat{\sigma}_{t\ |\ t-1}$ and $\widehat{\sigma}_{t\ |\ t-2}$ on each date t of the sample. The dynamic scale factor model of (57)–(59) is estimated on a fixed 10-year look-back window beginning on 4/30/2020 and rolling through the end of the data sample on 10/29/2024, so as to produce for each date t in the sample $\widehat{U}_{t\ |\ t-1} = \widehat{VVIX}_{t\ |\ t-1}$ as denoted by "1-Step", and $\widehat{U}_{t\ |\ t-1} = \widehat{VVIX}_{t\ |\ t-1}$ as denoted by "Avg." , where, in this case, $\frac{\sigma_{t\ |\ t-1}+\sigma_{t\ |\ t-2}}{2}$ replaces $\sigma_{t\ |\ t-1}$ in (59). $\left\{\widehat{VVIX}_{t\ |\ t-1}\right\}_{t=1}^{T}$ formed as "1-Step" and "Avg." are then compared using the RMSE and QLIKE loss functions.

13 Appendix C (Figures)

Notes to Figures 1–4. Daily SPX index levels source to Bloomberg, L.P. The GARCH $(1, 1)$ model of (3) and (4) is estimated on daily log returns constructed from these index levels using a fixed 10-year look-back window beginning on 12/23/1999 and rolling through the end of the data sample on 10/29/2024. The GARCH estimators being compared are QMLE, VTNGQMLE, and FAN. For these estimators, the figures plot $\widehat{\alpha}$ and $\widehat{\phi} = \widehat{\alpha} + \widehat{\beta}$ as solid lines. Also plotted as dashed lines are 2-sided, 95% confidence bands for the QMLE estimates, constructed using Huber-White "sandwich" standard error estimates.

Notes to Figures 5–6. Daily SPX index levels source to Bloomberg, L.P. The GARCH $(1, 1)$ model of (3) and (4) is estimated on daily log returns constructed from these index levels using VTQMLE and a fixed 10-year look-back window beginning on 12/23/1999 and rolling through the end of the data sample on 10/29/2024.On each day beginning 12/23/1999, GARCH $(1, 1)$ model innovations for the fixed 10-year look-back window are estimated, to which the tail index estimator of Hill (1975) is applied using a threshold of 5% of the largest innovations in absolute value.[27] One-sided, 95% confidence bands (dashed lines) for the Hill (1975) estimates (solid lines) are also constructed using the robust standard error estimator developed in Hill (2010).

Notes to Figures 7–10. Daily VIX levels source to Bloomberg, L.P. These VIX levels are mean-filtered using an ARFIMA $(1,\ d,\ 1)$ model, where $d \in (0,\ 0.5)$, estimated using the full maximum likelihood estimator of Sowell (1992) on a fixed 20-year look-back window beginning on 12/31/2009 and rolling through the end of the data sample on 10/29/2024. The GARCH $(1, 1)$ model of (3) and (4) is then estimated on the daily ARFIMA $(1,\ d,\ 1)$ model innovations using a fixed 10-year look-back window beginning on 12/31/2009 and rolling through the end of the data sample on 10/29/2024. The GARCH estimators being compared are QMLE, VTNGQMLE, and FAN. For these estimators, the figures plot $\widehat{\alpha}$ and $\widehat{\phi} = \widehat{\alpha} + \widehat{\beta}$ as solid lines.

Notes to Figures 11–12. Daily VIX levels source to Bloomberg, L.P. These VIX levels are mean-filtered using an ARFIMA $(1,\ d,\ 1)$ model, where $d \in (0,\ 0.5)$, estimated using the full maximum likelihood estimator of Sowell (1992) on a fixed 20-year look-back window beginning on 12/31/2009 and rolling through the end of the data sample on 10/29/2024. The GARCH $(1, 1)$ model of (55) and (56) is then estimated on the daily ARFIMA $(1,\ d,\ 1)$ model innovations using VTNGQMLE and a fixed 10-year look-back window beginning on 12/31/2009 and rolling through the end of the data sample on 10/29/2024. On each day beginning 12/31/2009, GARCH $(1, 1)$ model innovations for the fixed 10-year look-back window are estimated, to which the tail index estimator of Hill (1975) is applied using a threshold of 5% of the largest innovations in absolute value. One-sided, 95% confidence bands (dashed lines) for the Hill (1975) estimates (solid lines) are also constructed using the robust standard error estimator developed in Hill (2010).

Notes to Figures 13–15. Daily VIX levels source to Bloomberg, L.P. These VIX levels are mean-filtered using an ARFIMA $(1,\ d,\ 1)$ model, where $d \in (0,\ 0.5)$, estimated using the full maximum likelihood estimator of Sowell (1992) on a fixed 20-year look-back window beginning on 05/01/2020 and rolling through the end of the data sample on 10/29/2024. The GARCH $(1, 1)$ model of (55) and (56) is then estimated on the daily ARFIMA $(1,\ d,\ 1)$ model innovations using VTNGQMLE and a fixed 10-year look-back window beginning on 05/01/2020 and rolling through the end of the data sample on 10/29/2024. On each day beginning 04/30/2020, a 1-step-ahead and 2-steps-ahead out-of-sample GARCH volatility forecast is made, so that for each date t in the forecast-evaluation sample, there are two out-of-sample, GARCH volatility forecasts, $\widehat{\sigma}_{t\ |\ t-1}$ and $\widehat{\sigma}_{t\ |\ t-2}$. Figure 13 compares $\widehat{\sigma}_{t\ |\ t-1}$ against the actual VVIX value on date t. Figure 14 depicts the out-of-sample VVIX forecast constructed using $\widehat{\sigma}_{t\ |\ t-1}$ and $\widehat{V}_{t\ |\ t-1}$, comparing that forecast against the actual VVIX value on date t. Figure 15 depicts the out-of-sample VVIX forecast constructed from $\frac{\widehat{\sigma}_{t\ |\ t-1}+\widehat{\sigma}_{t\ |\ t-2}}{2}$ and $\widehat{V}_{t\ |\ t-1}$, comparing that forecast against the actual VVIX value on date t. In Figures 14 and 15, MAX is the largest VVIX value ever observed.

Notes to Figure 16. Daily SPX index and VIX levels source to Bloomberg, L.P. For SPX log returns, skewness estimates on a fixed 10-year look-back window beginning on 05/01/2020 and rolling through the end of the data sample on 10/29/2024 are shown. For VIX levels, skewness estimates for the innovations to an ARFIMA $(1,\ d,\ 1)$ model fit to a fixed 10-year look-back window beginning on 05/01/2020 and rolling through the end of the data sample on 10/29/2024 are shown.

[27] This 5% threshold means that the 126 largest daily innovations (in terms of absolute value) are input into the Hill (1975) estimator.

Figure 1: SPX $\widehat{\alpha}_n$ (Full Sample)

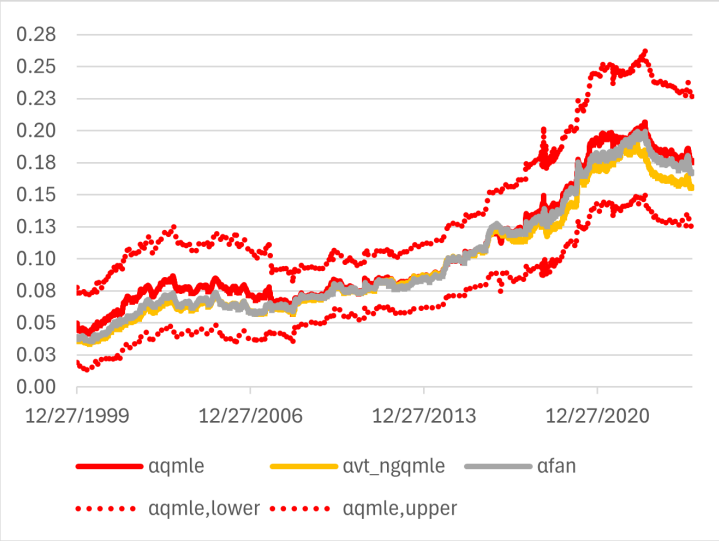


Figure 2: SPX $\widehat{\alpha}_n$ (Post COVID)

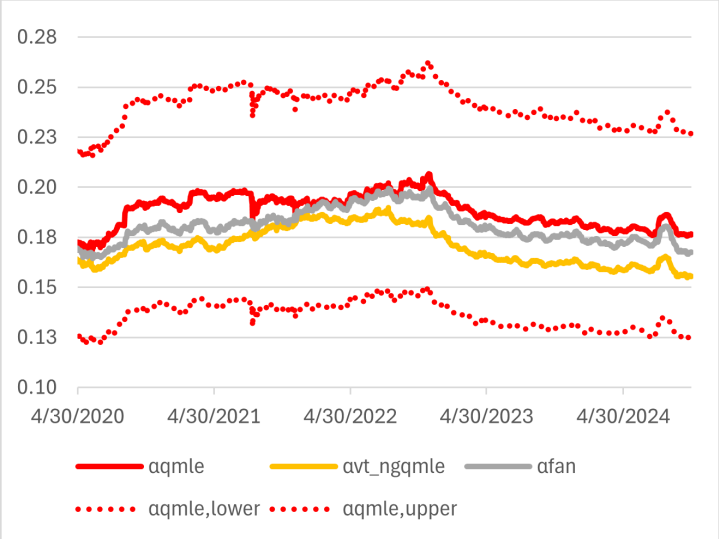


Figure 3: SPX $\widehat{\phi}_n$ (Full Sample)

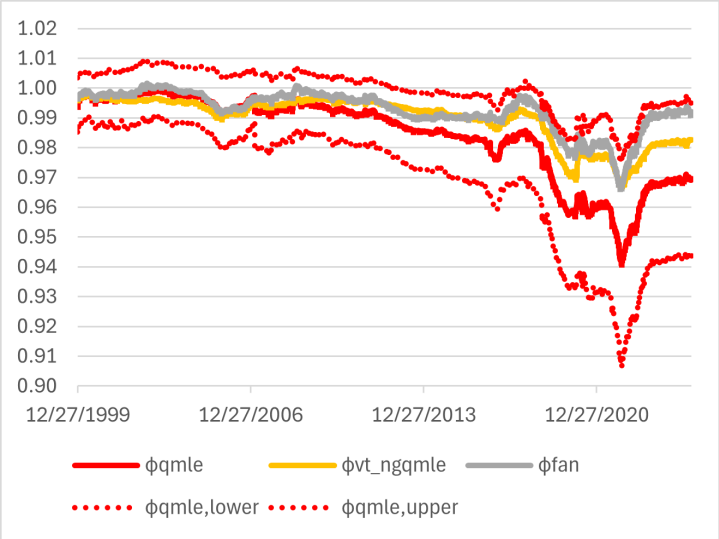


Figure 4: SPX $\widehat{\phi}_n$ (Post COVID)

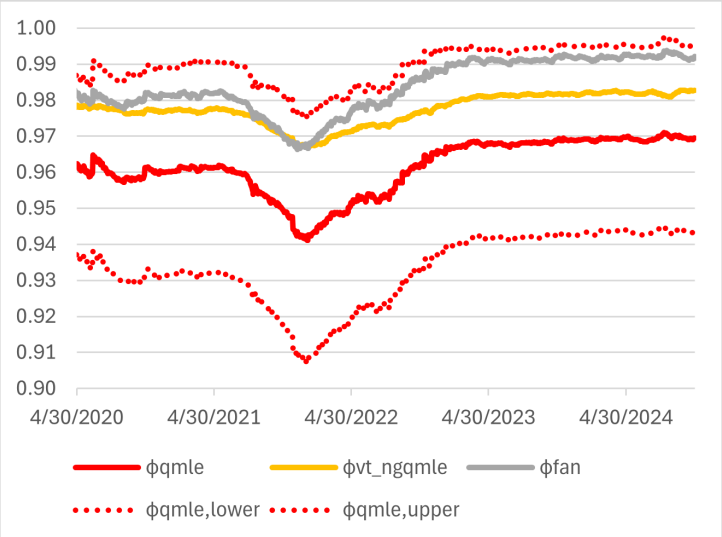


Figure 5: SPX $\widehat{\iota}_n$ (Full Sample)

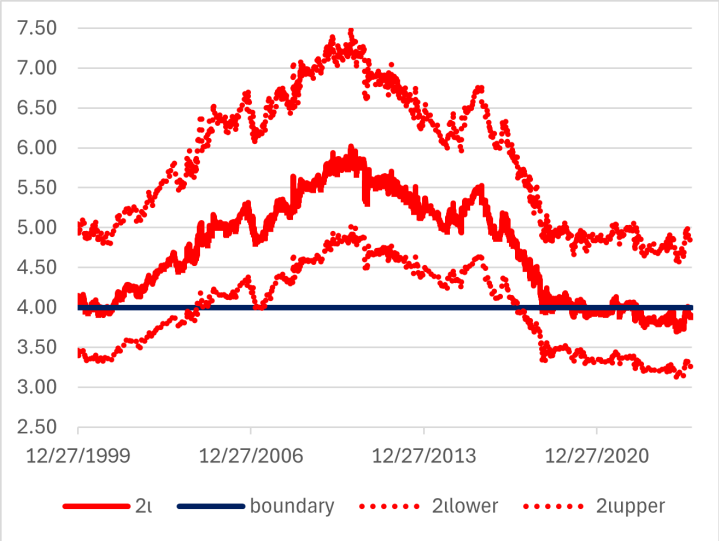


Figure 6: SPX $\widehat{\iota}_n$ (Post COVID)

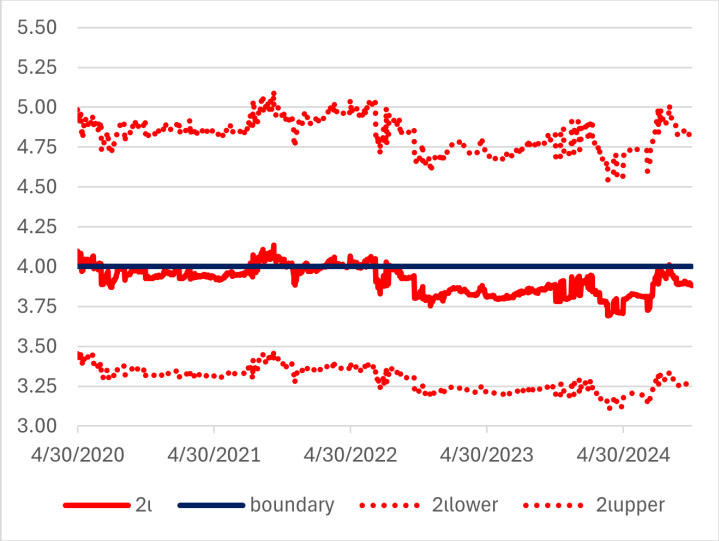


Figure 7: VIX $\widehat{\alpha}_n$ (Full Sample)

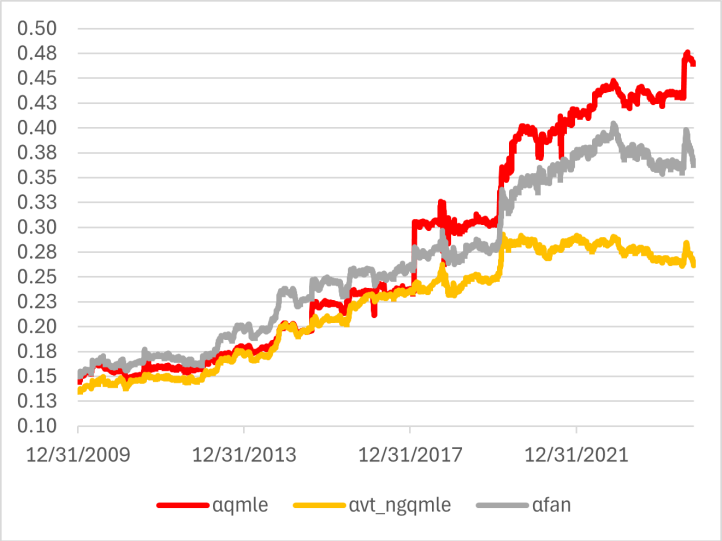


Figure 8: VIX $\widehat{\alpha}_n$ (Post COVID)

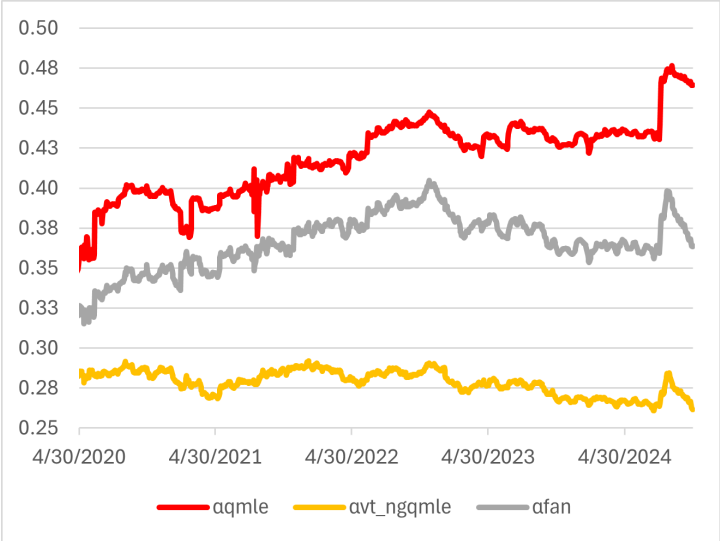


Figure 9: VIX $\widehat{\phi}_n$ (Full Sample)

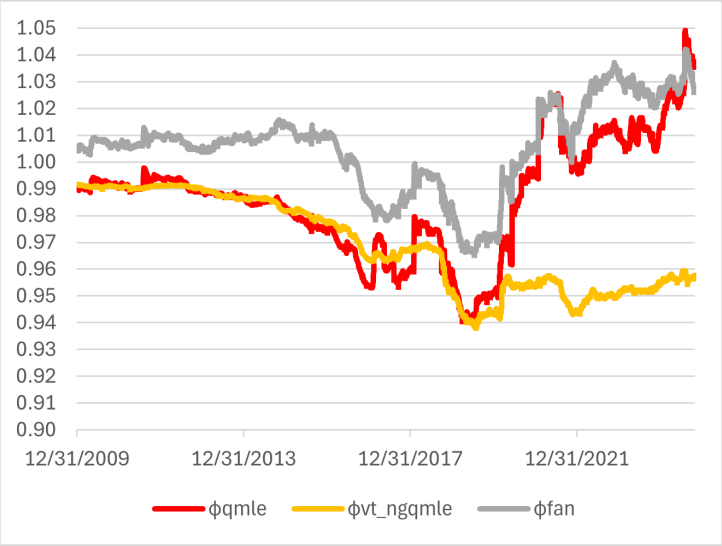


Figure 10: VIX $\widehat{\phi}_n$ (Post COVID)

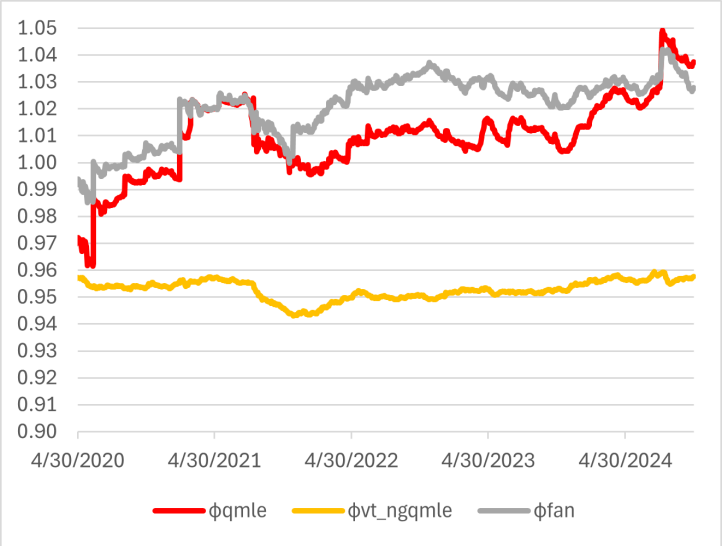


Figure 11: VIX $\widehat{\iota}_n$ (Full Sample)

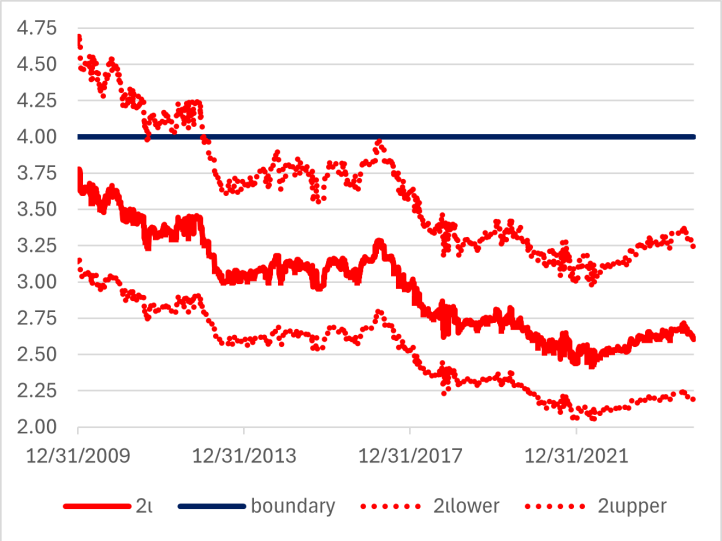


Figure 12: VIX $\widehat{\iota}_n$ (Post COVID)

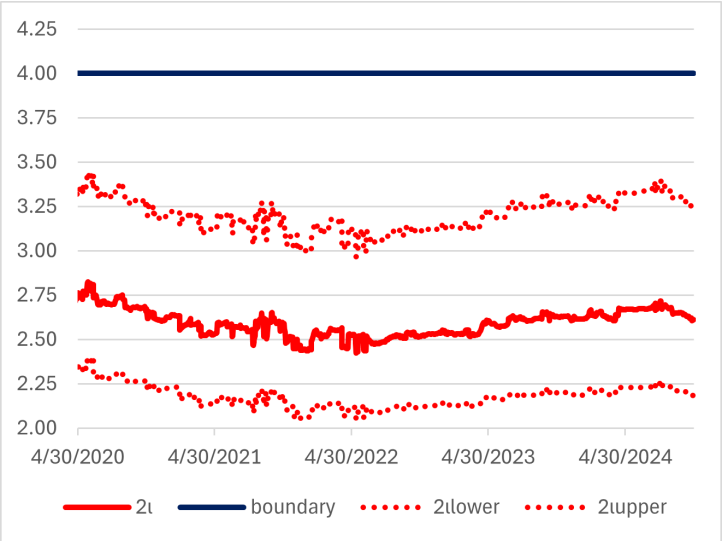


Figure 13: VIX 1-Step GARCH Vol Forecasts

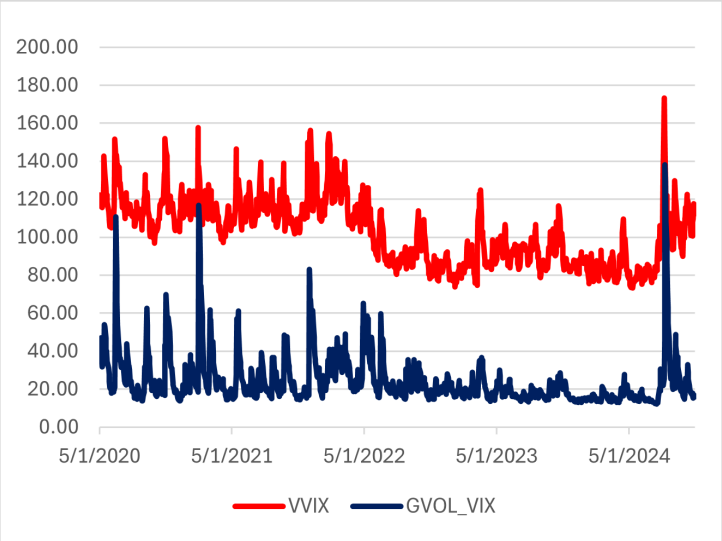


Figure 14: VVIX 1-Step GARCH Vol Forecasts

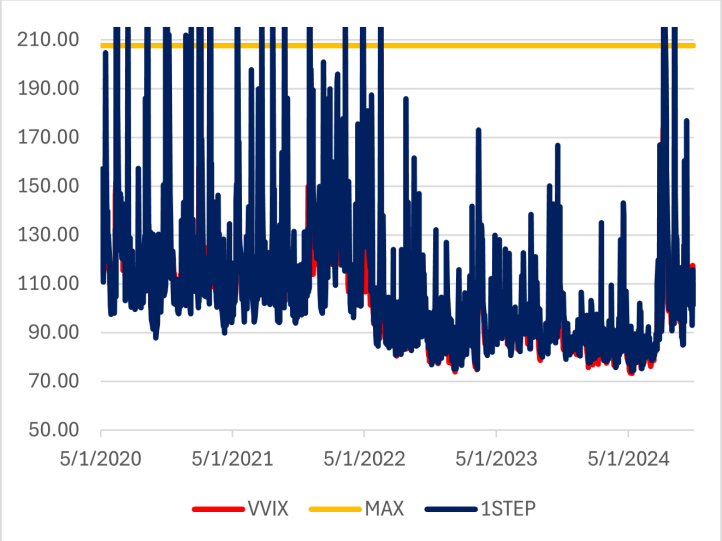


Figure 15: VVIX Average GARCH Vol Forecasts

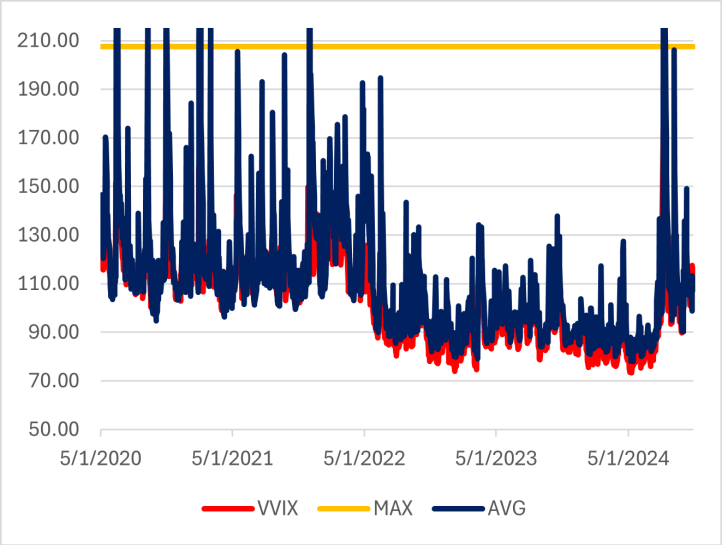


Figure 16: SPX and VIX Skewness

FIGURE 17: MEAN BIAS COMPS (ALPHA)

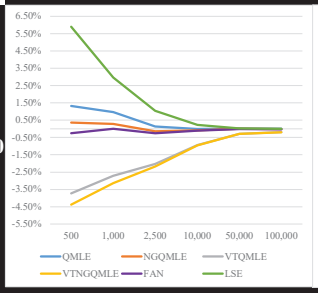
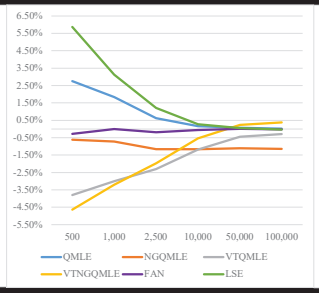
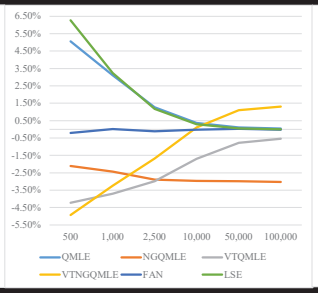
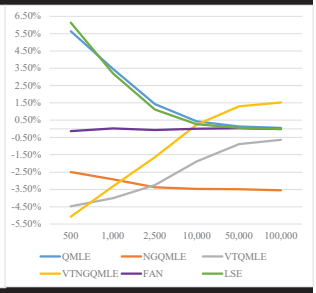
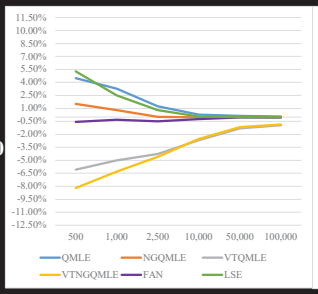
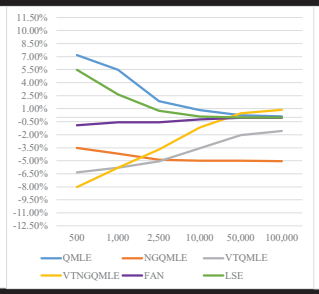
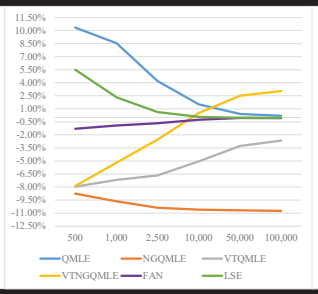
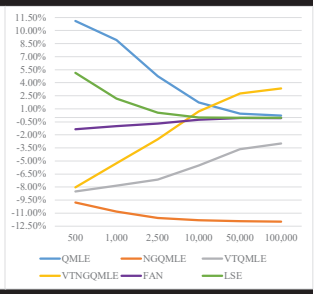
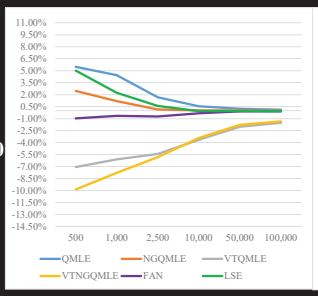
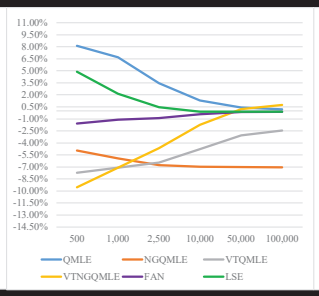
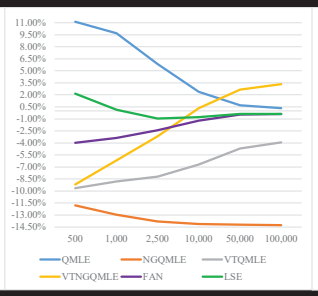
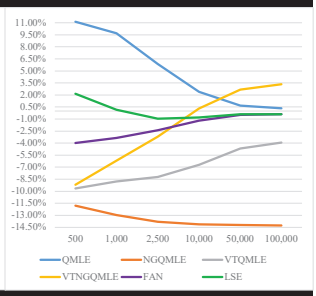
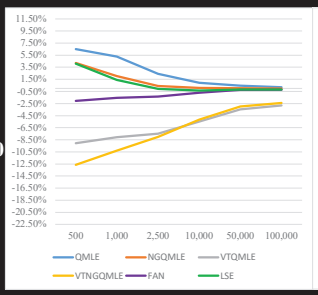
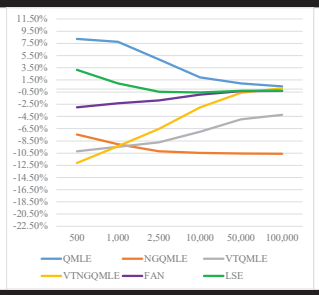
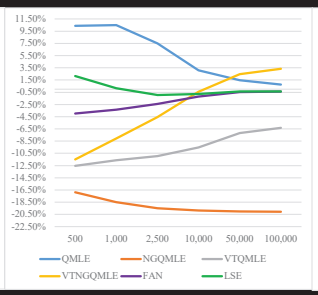
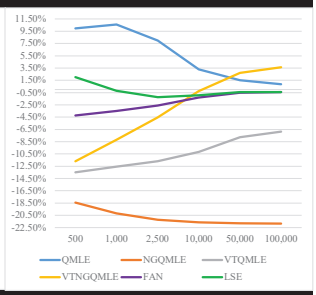
	λ = 0.00	λ = 0.40	λ = 0.80	λ = 0.99
η = 8.50				
η = 4.50				
η = 4.00				
η = 3.50				

FIGURE 18: MEAN BIAS COMPS (BETA)

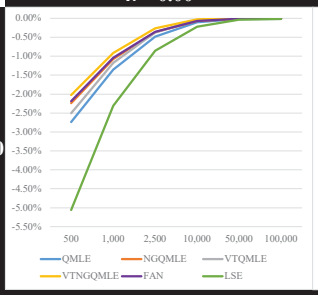
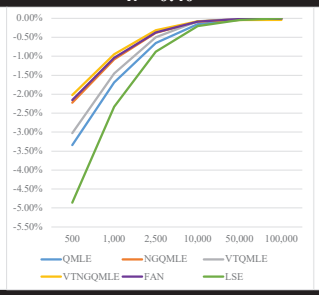
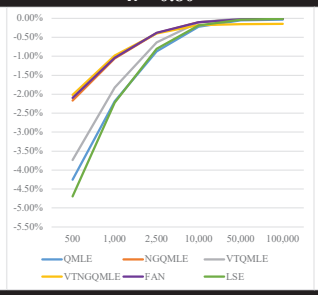
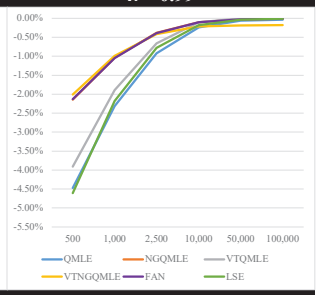
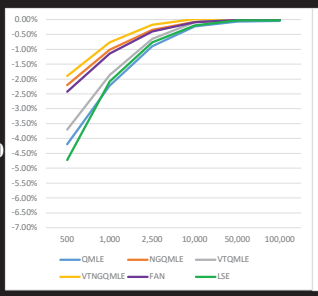
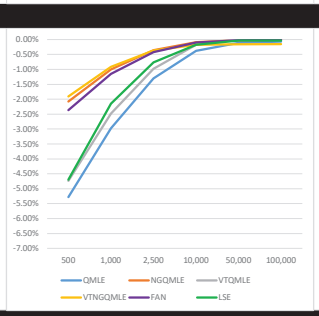
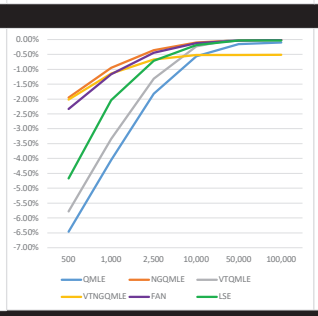
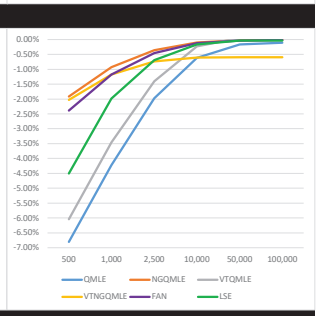
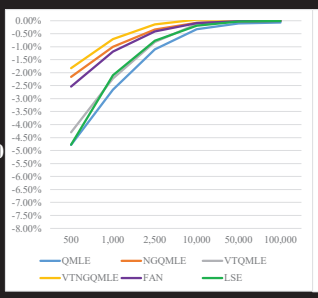
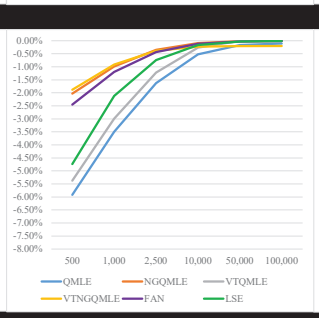
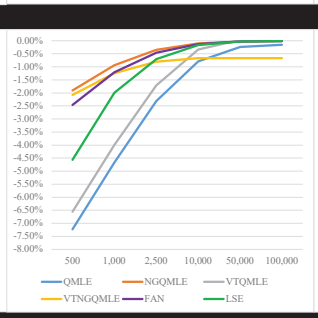
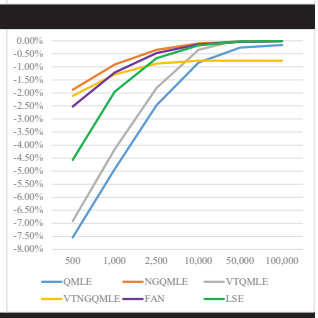
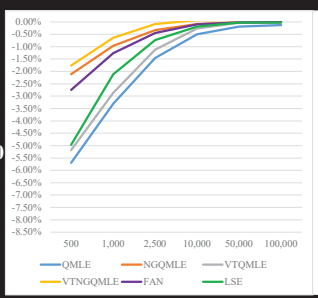
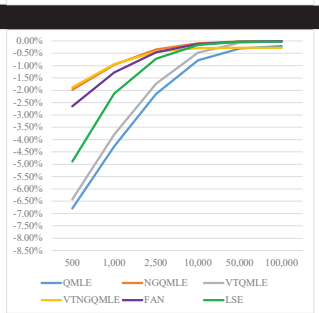
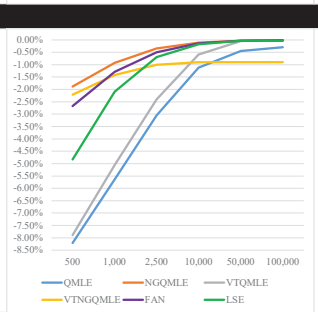
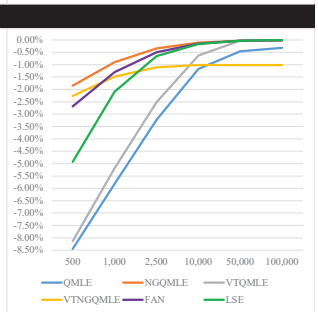
	λ = 0.00	λ = 0.40	λ = 0.80	λ = 0.99
η = 8.50				
η = 4.50				
η = 4.00				
η = 3.50				

FIGURE 19: DISPERSION COMPS (ALPHA)

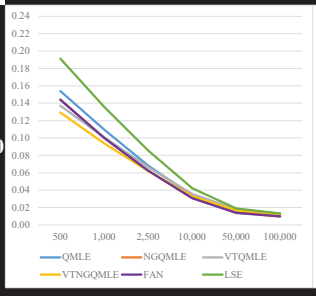
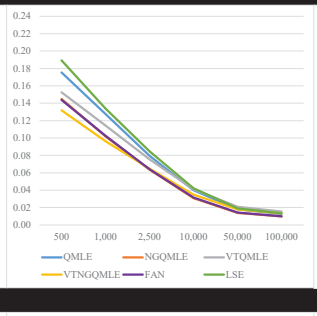
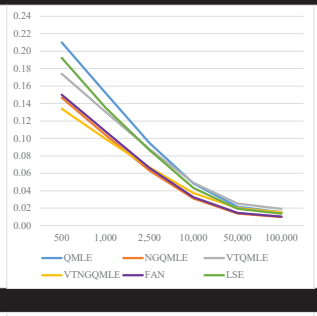
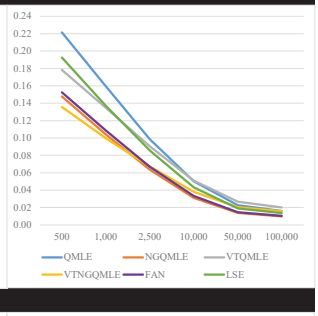
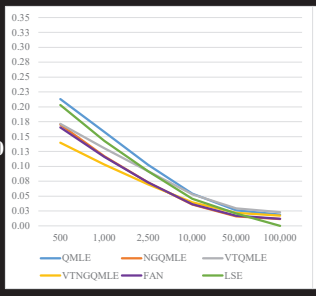
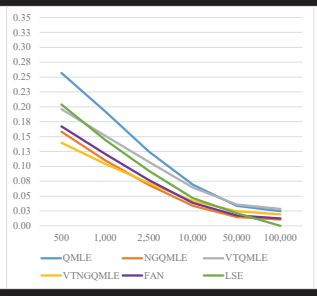
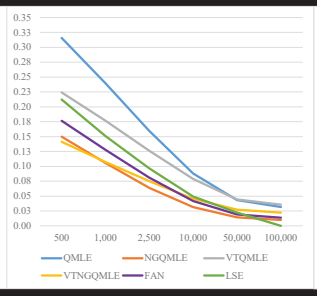
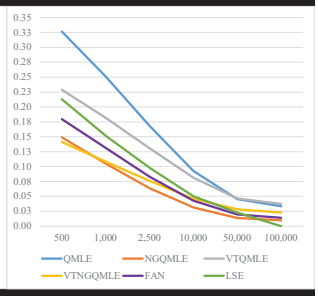
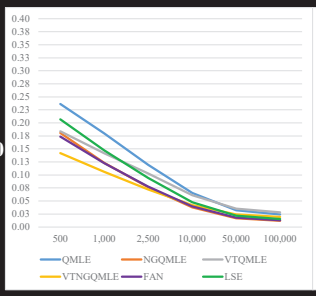
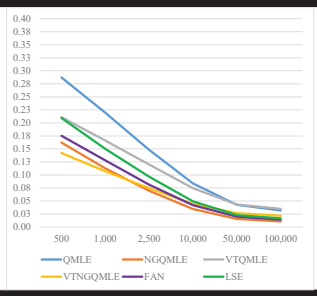
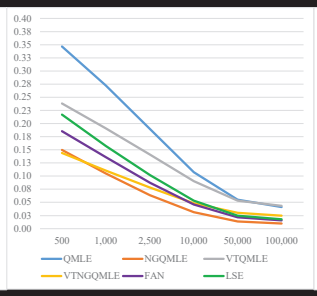
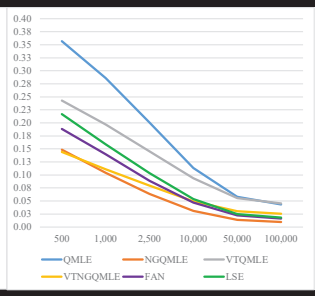
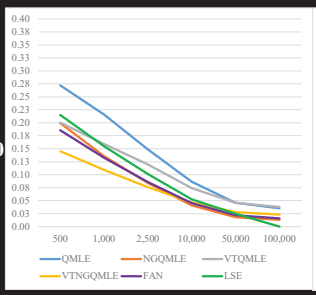
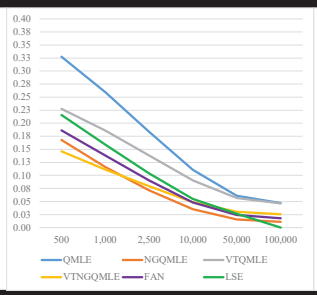
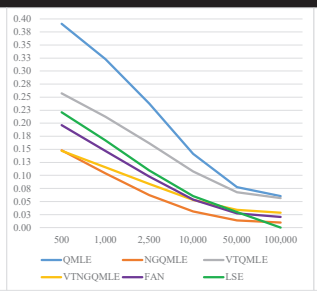
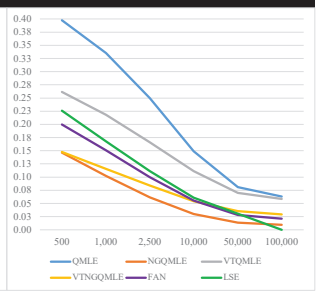
	λ = 0.00	λ = 0.40	λ = 0.80	λ = 0.99
η = 8.50				
η = 4.50				
η = 4.00				
η = 3.50				

FIGURE 20: DISPERSION COMPS (BETA)

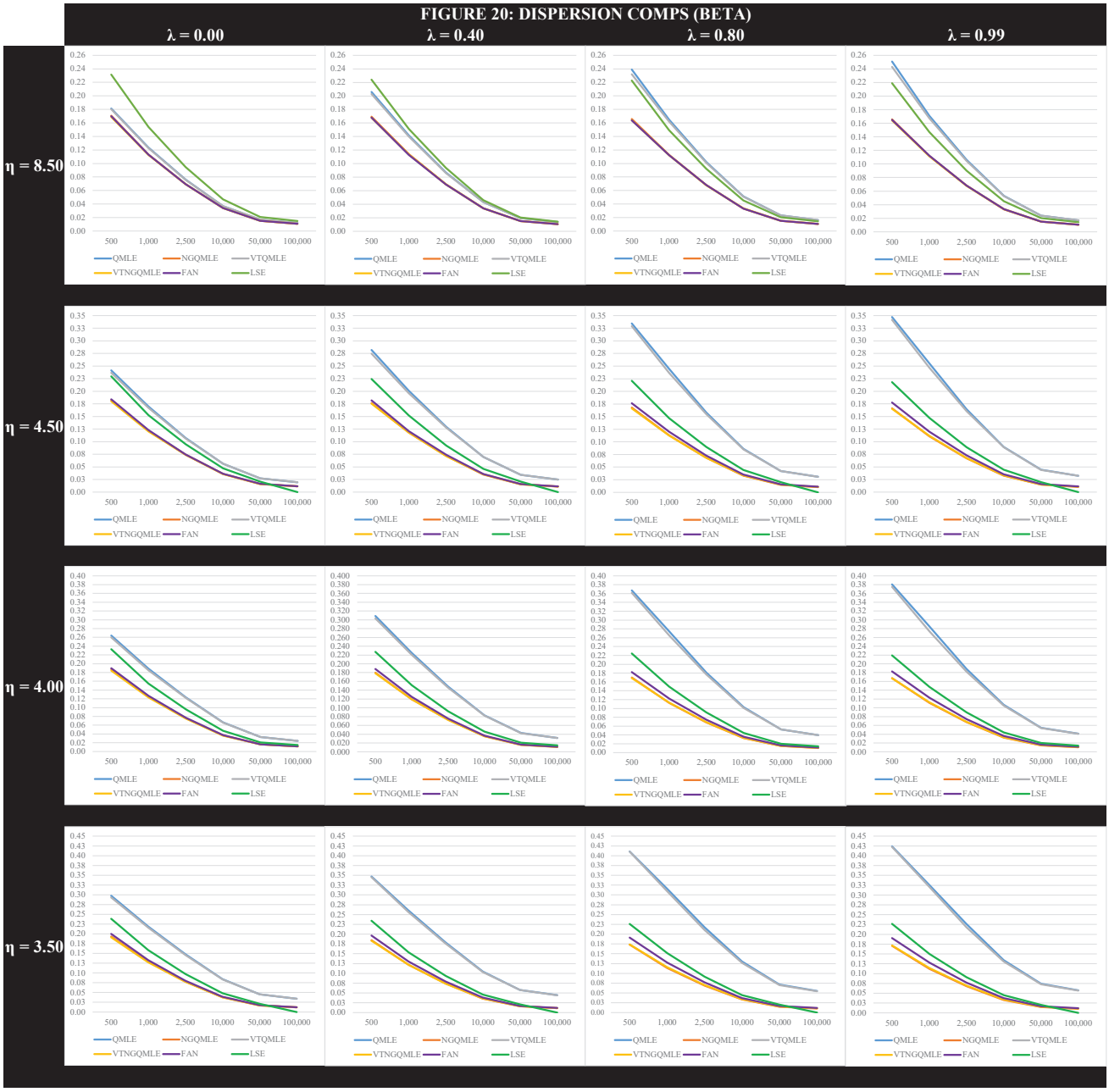


FIGURE 21: ROOT-MEAN-SQUARED-ERROR COMPS (ALPHA)

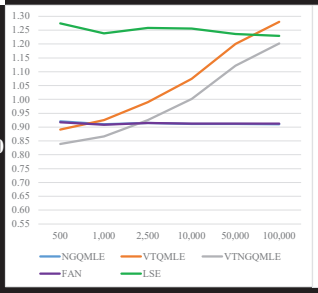
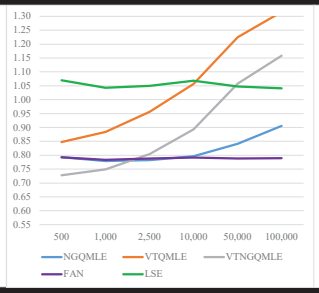
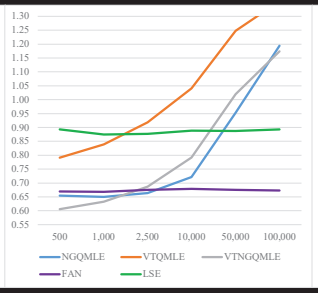
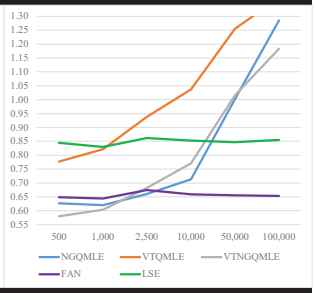
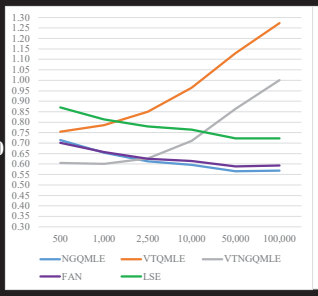
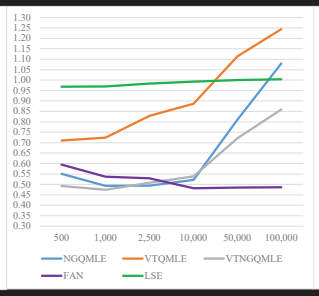
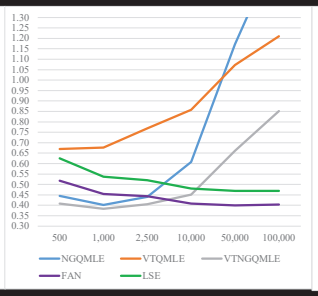
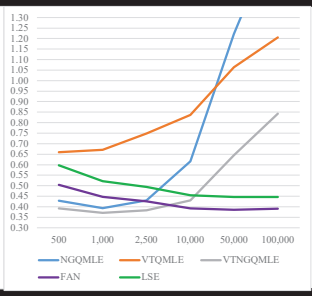
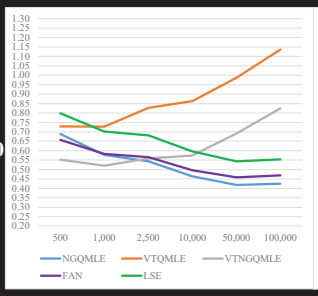
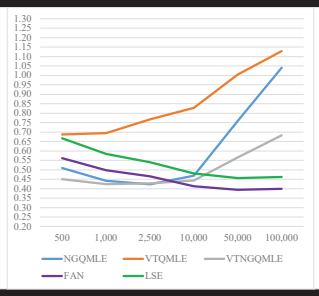
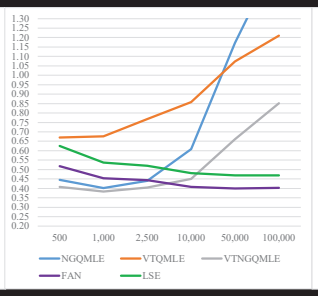
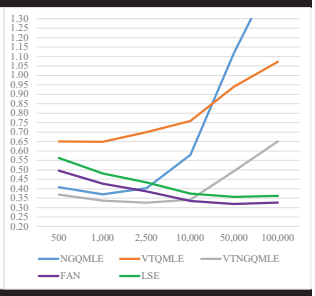
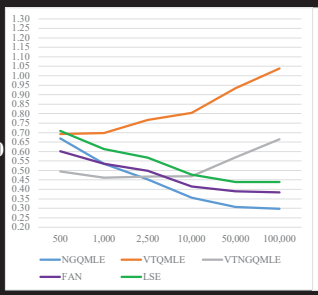
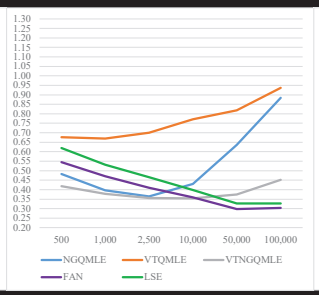
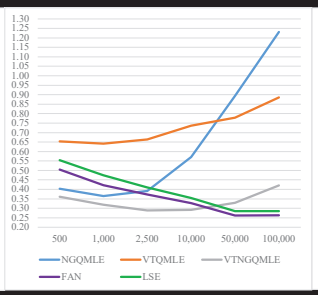
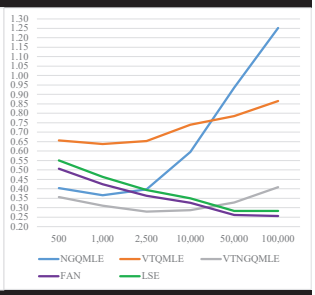
	λ = 0.00	λ = 0.40	λ = 0.80	λ = 0.99
η = 8.50				
η = 4.50				
η = 4.00				
η = 3.50				

FIGURE 22: ROOT-MEAN-SQUARED-ERROR COMPS (BETA)

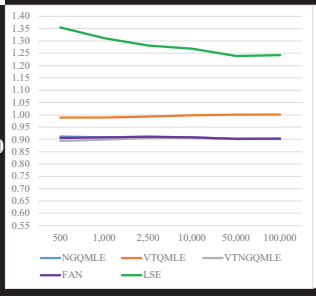
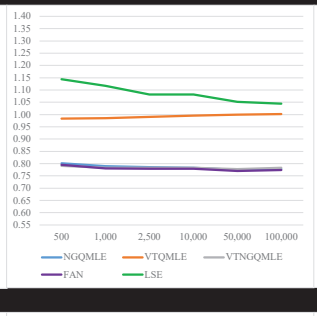
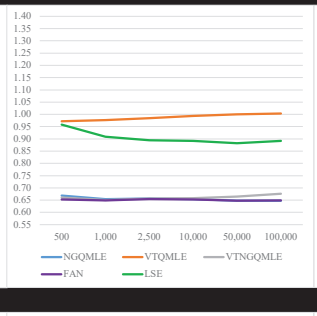
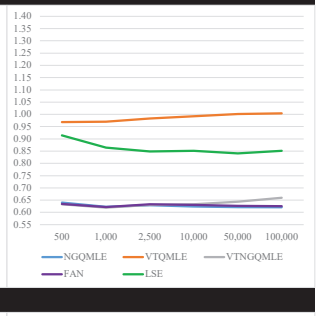
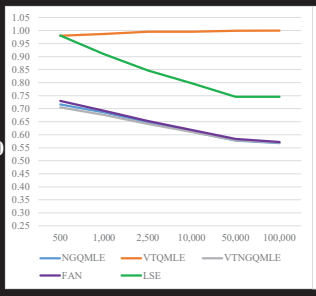
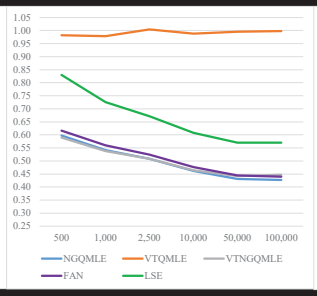
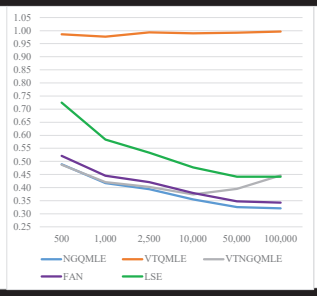
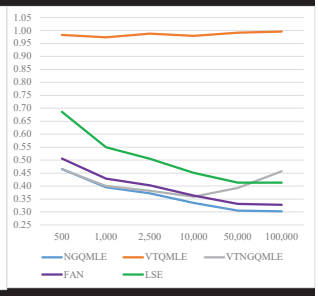
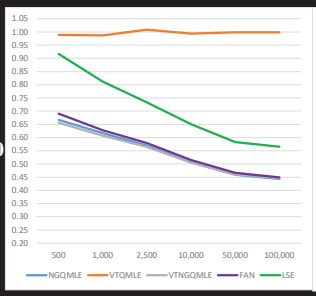
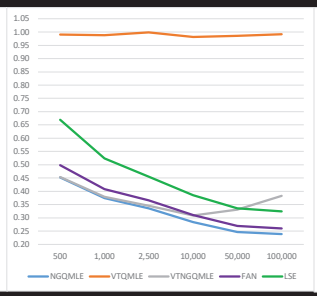
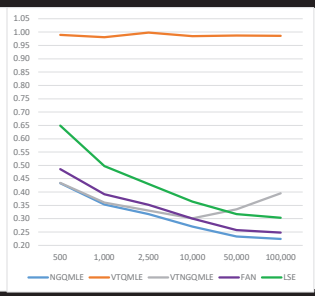
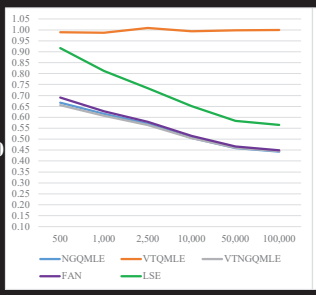
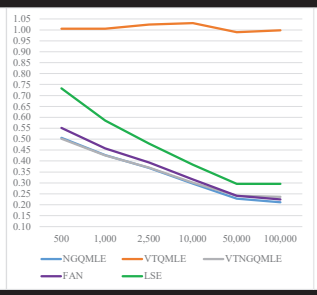
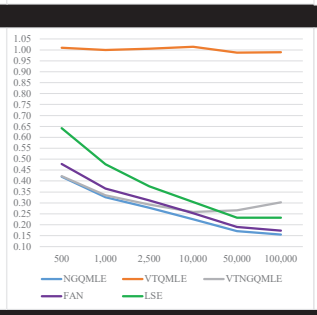
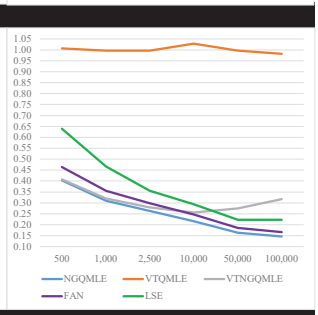
	λ = 0.00	λ = 0.40	λ = 0.80	λ = 0.99
η = 8.50				
η = 4.50				
η = 4.00				
η = 3.50				

FIGURE 23: MEAN-ABSOLUTE-ERROR COMPS (ALPHA)

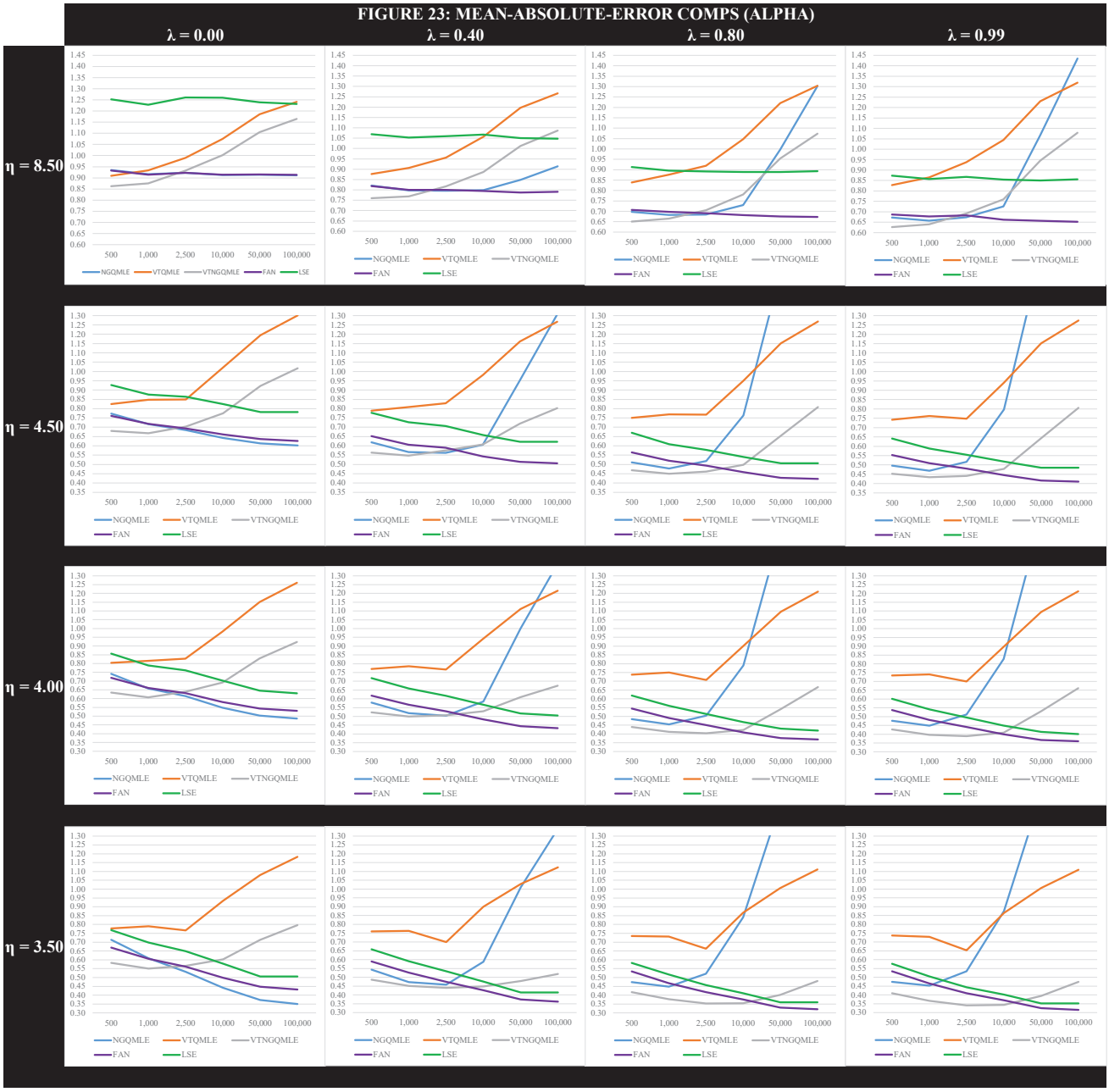


FIGURE 24: MEAN-ABSOLUTE-ERROR COMPS (BETA)

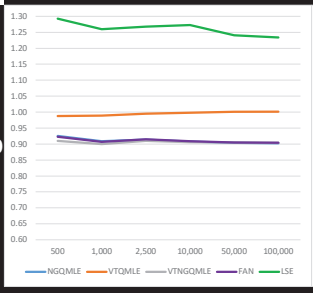
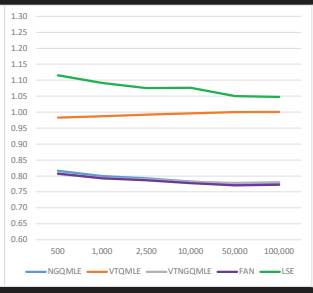
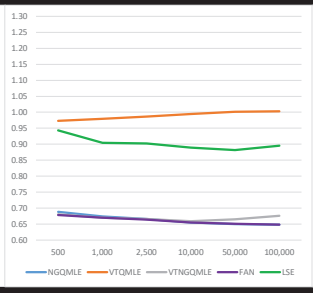
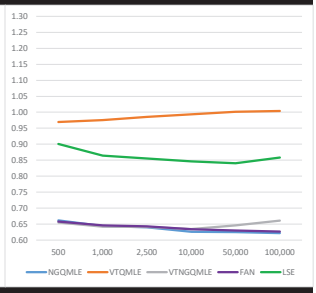
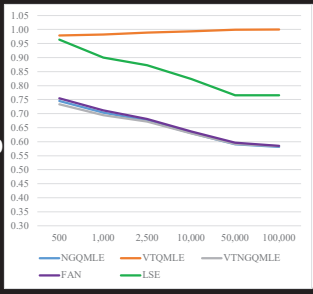
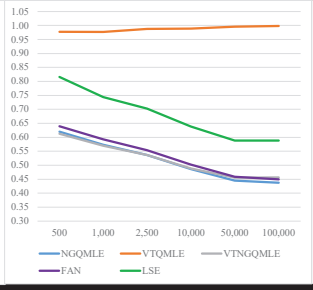
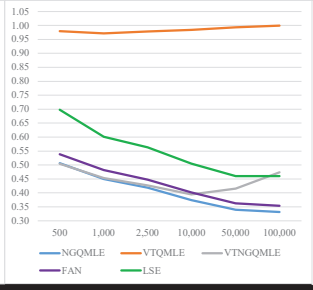
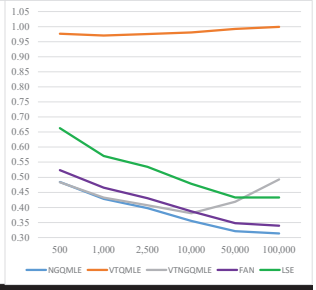
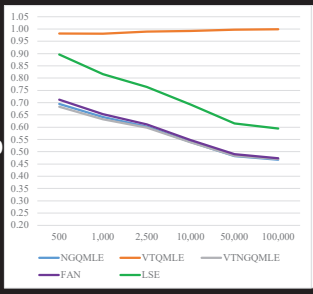
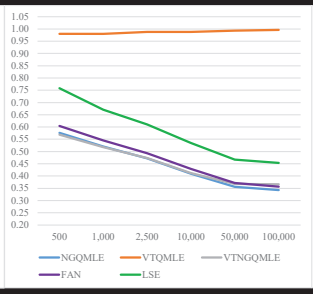
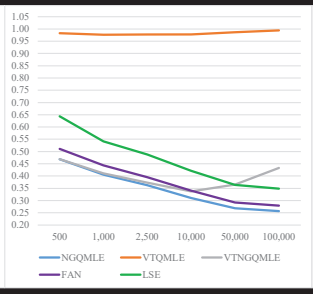
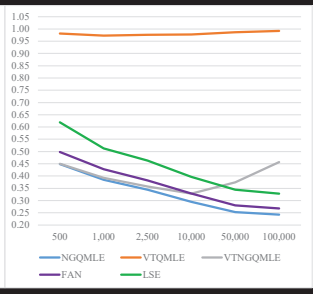
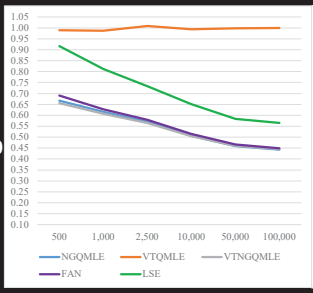
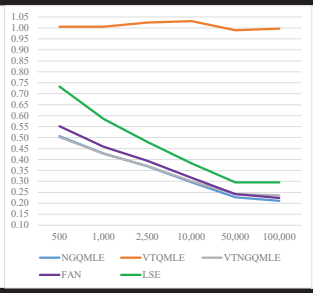
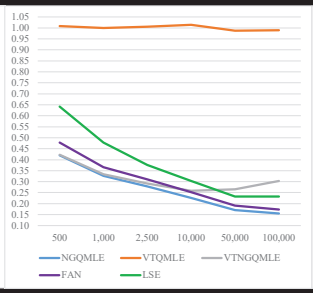
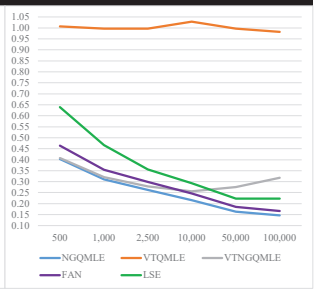
	λ = 0.00	λ = 0.40	λ = 0.80	λ = 0.99
η = 8.50				
η = 4.50				
η = 4.00				
η = 3.50				

FIGURE 25: MEAN BIAS COMPS (OMEGA)

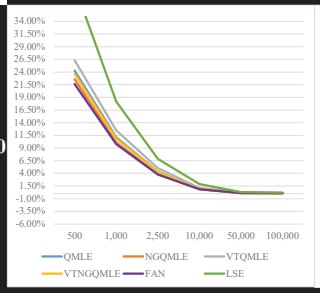
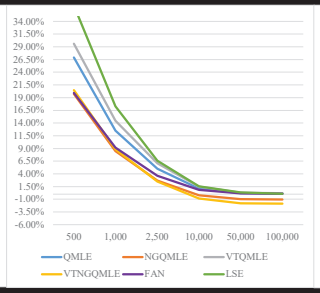
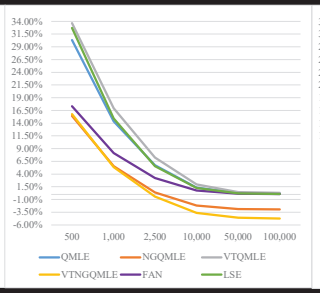
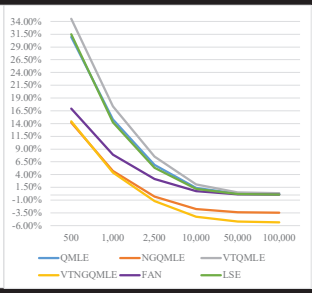
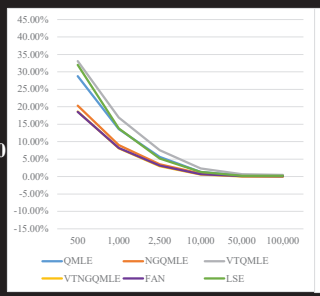
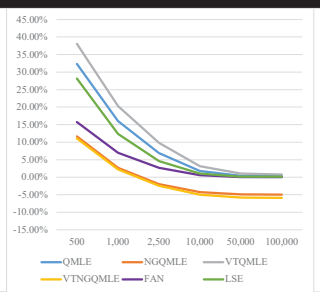
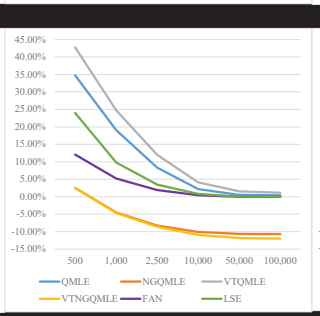
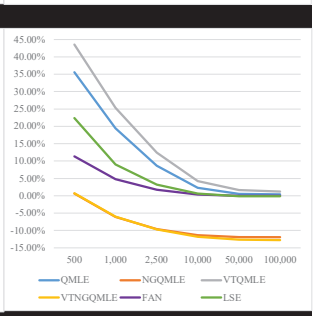
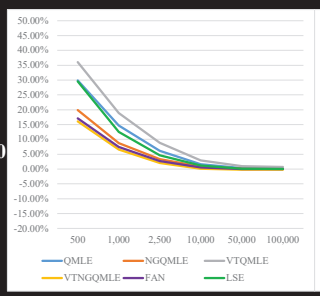
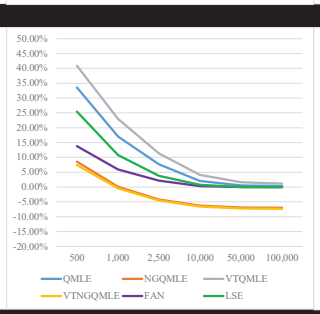
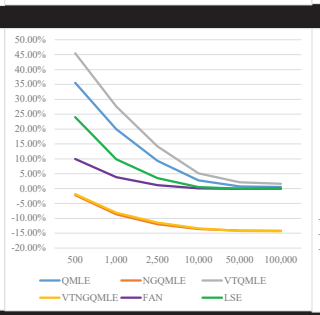
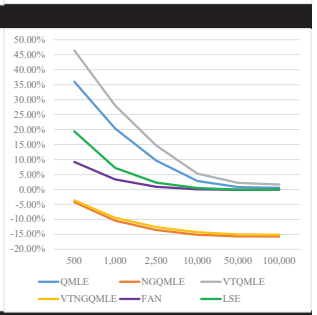
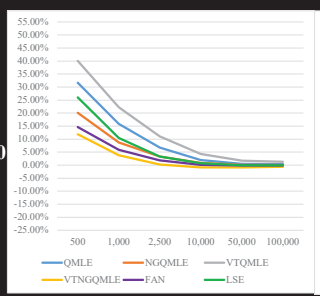
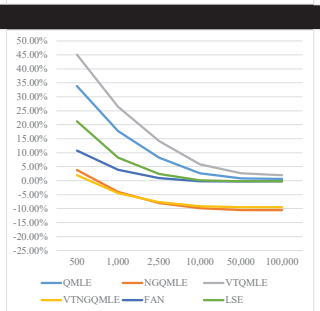
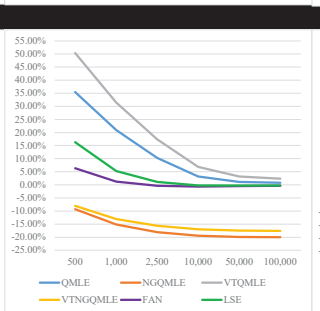
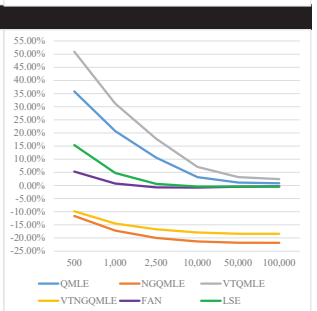
	λ = 0.00	λ = 0.40	λ = 0.80	λ = 0.99
η = 8.50				
η = 4.50				
η = 4.00				
η = 3.50				

FIGURE 26: DISPERSION COMPS (OMEGA)

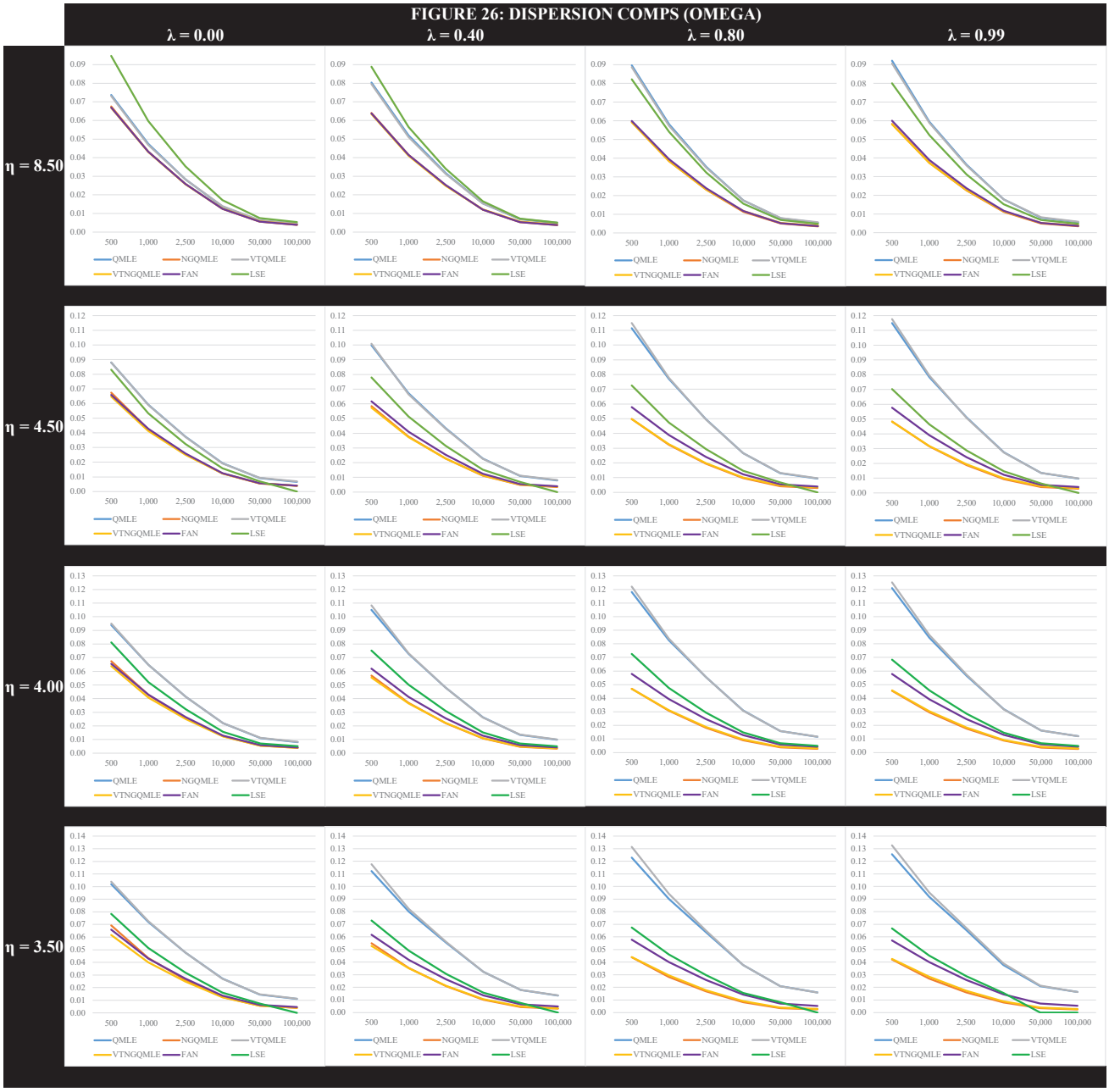


FIGURE 27: ROOT-MEAN-SQUARED-ERROR COMPS (OMEGA)

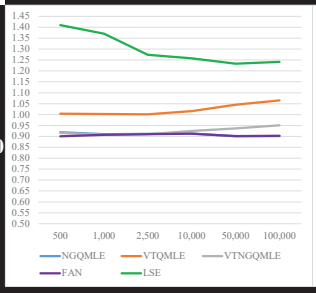
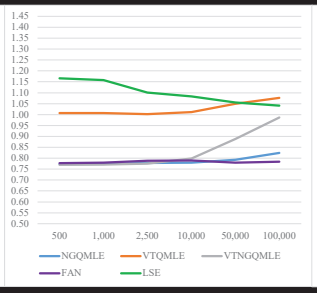
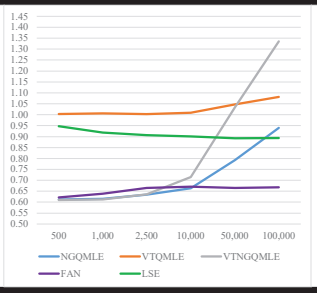
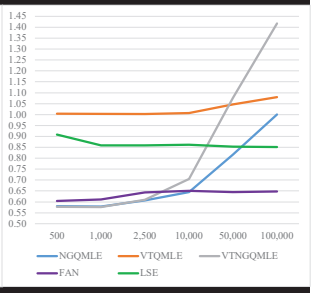
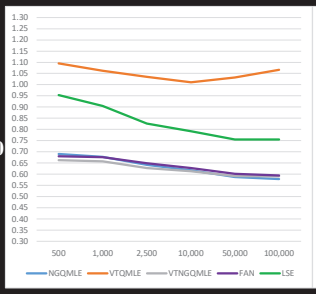
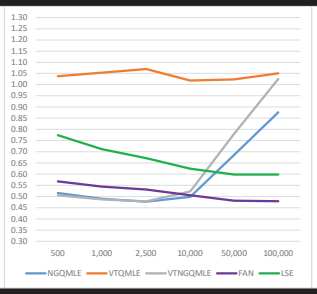
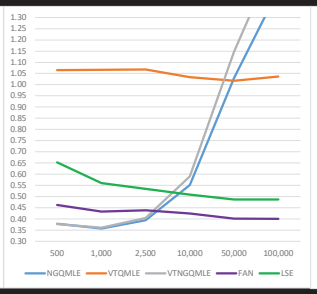
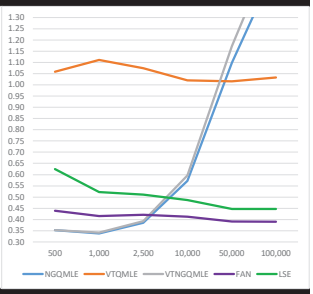
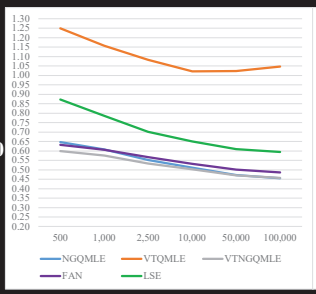
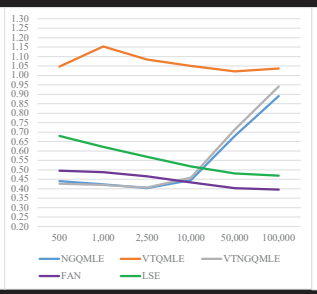
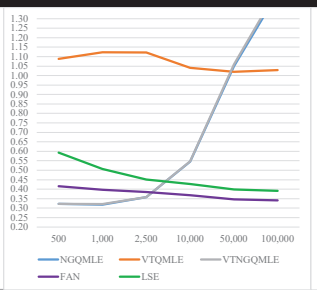
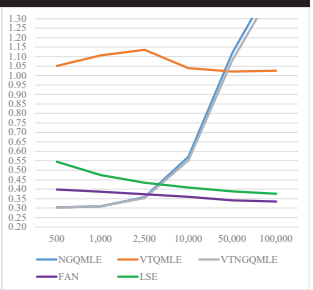
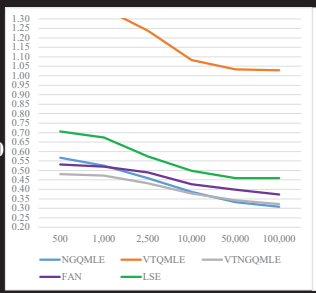
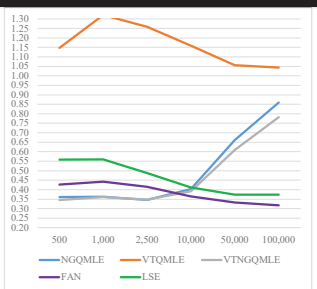
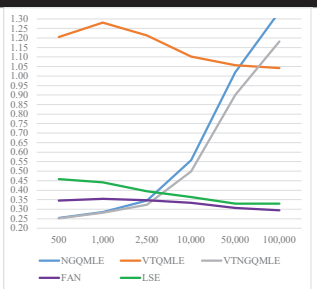
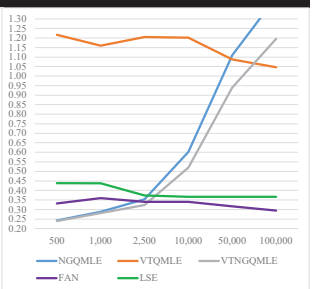
	λ = 0.00	λ = 0.40	λ = 0.80	λ = 0.99
η = 8.50				
η = 4.50				
η = 4.00				
η = 3.50				

FIGURE 28: MEAN-ABSOLUTE-ERROR COMPS (OMEGA)