Tail Risk Monotonicity in GARCH(1,1) Models

Paul Glasserman* Dan Pirjol† Qi Wu ‡

February 2023; revised August 2023

Abstract

The stationary distribution of a GARCH(1,1) process has a power law decay, under broadly applicable conditions. We study the change in the exponent of the tail decay under temporal aggregation of parameters, with the distribution of innovations held fixed. This comparison is motivated by the fact that GARCH models are often fit to the same time series at different frequencies. The resulting models are not strictly compatible so we seek more limited properties we call forecast consistency and tail consistency. Forecast consistency is satisfied through a parameter transformation. Tail consistency leads us to derive conditions under which the tail exponent increases under temporal aggregation, and these conditions cover most relevant combinations of parameters and innovation distributions. But we also prove the existence of counterexamples near the boundary of the admissible parameter region where monotonicity fails. These counterexamples include normally distributed innovations.

In memory of Tom Hurd and his longstanding interest in heavy-tailed phenomena in financial markets.

1 Introduction.

Many financial time series exhibit persistence and clustering in volatility. Once volatility becomes elevated, it tends to remain elevated for some time; high levels of volatility produce large moves in market data, and these large moves in turn fuel increased volatility. These features are captured by generalized autoregressive conditional heteroskedasticity (GARCH) models, which are widely used for this reason.

GARCH models are often fit to the same time series at different frequencies — daily, weekly, and monthly, for example. One part of a bank may be interested in forecasting exchange rates at a high frequency for trading purposes, while another part of the bank forecasts over a somewhat longer horizon for capital planning or risk management. The bank may thus end up with two GARCH models on different time scales.

Drost and Nijman (1993) showed, however, that GARCH processes are not closed under temporal aggregation. This means that if a time series is exactly described by a GARCH process at one frequency, it cannot also be exactly described by a GARCH process when observed at a lower frequency.

The analysis of Drost and Nijman (1993) applies to both stock-variable temporal aggregation and flow-variable aggregation. In the first case, the same variable is observed at different frequencies. In the second case, a variable is summed over intervals of different lengths. We focus on the simpler case of stock-variable aggregation. GARCH processes fail to be closed under both types of temporal aggregation.

In light of this fundamental limitation, what would it mean for GARCH models fit at two time scales — possibly by two groups within the same bank — to be compatible with each other? Perfect compatibility is impossible, so we propose two more limited objectives: (1) *forecast consistency*, meaning that forecasts generated from the two models using the same data should agree; and (2) *tail consistency*,

*Columbia Business School, Columbia University, 403 Uris Hall, New York, NY 10028, pg20@columbia.edu
†School of Business, Stevens Institute of Technology, Hoboken, NJ 07030, dpirjol@stevens.edu
‡City University of Hong Kong, Kowloon, Hong Kong, qi.wu@cityu.edu.hk

meaning that the stationary distributions of the two models should exhibit the same tail behavior. The first point focuses on typical values under the two models, and the second point focuses on extreme values.

In showing that GARCH processes are not closed under temporal aggregation, Drost and Nijman (1993) introduced a more general family of processes they called *weak* GARCH that is closed. They also showed how the process parameters are transformed under temporal aggregation. Building on their results, we show that forecast consistency is achieved when ordinary (strong) GARCH models are fit at different frequencies using the Drost-Nijman parameter transformation. This property argues for using this transformation to select parameters for lower-frequency models.

Our main focus is on the question of tail consistency. Under broadly applicable conditions, Mikosch and Stărică (2000) showed that the stationary distribution of a GARCH process has a power law decay. The power has a simple characterization in terms of the parameters of the process and the distribution of the innovations (the noise) driving its evolution. Under stock-variable aggregation, the stationary distribution should ideally be the same at different frequencies. As this is difficult or impossible to achieve, we investigate the more modest requirement that the power law decay be preserved at different frequencies.

We study this objective by analyzing how the exponent in the power law changes under the Drost-Nijman parameter transformation. To describe our results in more detail, write $\kappa = \kappa(\alpha, \beta, Z)$ for the exponent derived from GARCH(1,1) parameters (α, β) and innovations with the distribution of Z. Write (α_n, β_n) for the parameters derived from the Drost-Nijman parameter transformation when every nth value of the time series is observed. We study how $\kappa_n = \kappa(\alpha_n, \beta_n, Z)$ changes with n.

If κ_n increases under this transformation, then the innovations of the lower-frequency model would need to be heavier-tailed to preserve the original value of κ because a heavier-tailed Z will ordinarily lead to a smaller value of κ. In other words, for the innovations $\tilde{Z}$ of the lower-frequency model to achieve

$$\kappa_n \equiv \kappa(\alpha_n, \beta_n, Z) > \kappa(\alpha_n, \beta_n, \tilde{Z}) = \kappa(\alpha, \beta, Z) \equiv \kappa_1,$$

we need $\tilde{Z}$ to be heavier-tailed than the innovations Z of the original model. Tail consistency requires $\kappa(\alpha_n, \beta_n, \tilde{Z}) = \kappa(\alpha, \beta, Z)$.

This scenario describes the case that κ_n increases in n with Z held fixed. We will see that this covers "most" cases. In a significant portion of the (α, β) parameter space, which we characterize explicitly, both parameters decrease under the Drost-Nijman parameter transformation; within this region, the increase in κ is automatic. In the more interesting region, α first increases under temporal aggregation as β decreases, so the effect on κ depends on the distribution of Z.

Within this region, we derive a necessary and sufficient condition for κ to increase. We then establish two types of results: conditions under which κ is guaranteed to increase, and conditions leading to exceptions — cases where κ decreases. Several of our results are based on comparing innovation distributions under various convex orders.

The positive results confirm the "typical" case of κ increasing under temporal aggregation, and these results cover a wide range of parameters and distributions. The structure of exceptions is in some respects more intriguing. For example, with standard normal innovations, numerical testing finds no exceptions with $\beta < 0.993$, but we prove the existence of exceptions as $\beta \to 1$ and $\alpha \to 0$.

Although our focus is on the effect of temporal aggregation of parameters, several of our results are of independent interest in providing bounds on the tail parameter κ irrespective of temporal aggregation. Because κ describes the tail decay of the stationary distribution, it is a useful measure for risk management, and it is also important in extreme value theory for GARCH; see Bucher et al. (2020), Davison et al. (2023), Francq and Zakoian (2019), Glasserman and Wu (2018), McNeil et al. (2015), Peña et al. (2020), and Sun and Zhou (2014) for examples of applications.

The rest of this paper is organized as follows. Section 2 provides relevant background on GARCH processes and temporal aggregation. Section 3 addresses forecast consistency. Section 4 formulates the problem of tail consistency, and it then presents our general necessary and sufficient condition for κ

monotonicity. Sections 5 and 6 derive specific sufficient conditions. These conditions rely on varying levels of information about the innovations, particularly their moments. Section 7 studies asymptotics as $\beta \to 1$ and $\alpha \to 0$ leading to exceptions. Section 8 illustrates our results with empirical examples.

2 Problem Formulation

2.1 Background on GARCH.

The scalar sequence $\{X_t, t \in \mathbb{Z}\}$ is a GARCH(1,1) process if it satisfies $X_t = \sigma_t Z_t$, $t \in \mathbb{Z}$, with

$$\sigma_t^2 = \omega + \alpha X_{t-1}^2 + \beta \sigma_{t-1}^2 = \omega + \sigma_{t-1}^2(\alpha Z_{t-1}^2 + \beta), \quad t \in \mathbb{Z} \tag{1}$$

for positive constants ω, α, β, and an i.i.d. sequence of innovations $\{Z_t, t \in \mathbb{Z}\}$, with $\mathbb{E}[Z_t] = 0$ and $\mathbb{E}[Z_t^2] = 1$. Here, σ_t is called the conditional volatility of X at time t. The dynamics in (1) specify that the conditional variance σ_t^2 of X_t depends on both the lagged conditional variance σ_{t-1}^2 and on the magnitude of the lagged observation X_{t-1}^2. Thus, high levels of volatility will tend to be followed by high levels of volatility, and large moves in X_t will tend to further amplify volatility. These properties capture important features of many types of financial data; see, e.g., Bollerslev (1986), Engle (1995), and McNeil et al. (2015) for background. We introduce a set of assumptions that will be in force throughout.

Assumption 2.1. *The GARCH(1,1) parameters and innovations satisfy the following conditions:*

(A1) The parameters $\alpha, \beta > 0$ satisfy $\alpha + \beta < 1$.

(A2) The innovations Z satisfy $\mathbb{E}[Z] = 0$ and $\mathbb{E}[Z^2] = 1$.

(A3) For some $0 < r < \infty$, we have $1 < \mathbb{E}[(\alpha Z^2 + \beta)^r] < \infty$.

Assumption (A2) is a standard normalization. Assumption (A1) ensures the existence of a stationary solution to (1); see, e.g., Theorem 2 of Nelson (1990). Stationarity holds with $\alpha+\beta = 1$ if $\mathbb{E}[\log(\alpha Z^2+\beta)] < 0$, but the stronger condition in (A1) is required for temporal aggregation in Drost and Nijman (1993).

Under (A1)–(A2), the process (σ_t, X_t) admits a stationary distribution $(\sigma_\infty, X_\infty)$ for which we have the equality in distribution,

$$\sigma_\infty^2 \overset{d}{=} \omega + (\alpha Z^2 + \beta)\sigma_\infty^2. \tag{2}$$

Mikosch and Stărică (2000), see also Basrak et al. (2002), show that the stationary distribution has regularly varying tails, in the sense that

$$\mathbb{P}(\sigma_\infty > x) \sim C x^{-\kappa}, \quad x \to \infty, \tag{3}$$

and $\mathbb{P}(|X_\infty| > x) \sim C' x^{-\kappa}$, for constants C and C'. The exponent $\kappa > 0$ is determined by the distribution of $\alpha Z^2 + \beta$ through the equation

$$\mathbb{E}[(\alpha Z^2 + \beta)^{\kappa/2}] = 1. \tag{4}$$

(A3) ensures the existence of a unique solution $\kappa > 0$ to (4).

As risk measures (such as value-at-risk or expected shortfall) are often proportional to σ_t, κ provides a measure of the unconditional level of risk and is an important feature of the stationary distribution; see Glasserman and Wu (2018) for its use in setting margin requirements. Notice that GARCH processes are heavy-tailed, in the sense of (3), even when the innovations are light-tailed — for example, even if $\mathbb{E}[(\alpha Z^2 + \beta)^r]$ is finite for all $r > 0$ and (A3) holds.

We assume (A1)–(A3) in all our results. For emphasis, we sometimes refer to parameters (α, β) as admissible if they satisfy (A1), and we call Z admissible if it satisfies (A2)–(A3).

The function

$$\varphi(s) = \mathbb{E}[(\alpha Z^2 + \beta)^s], \quad s \geq 0,$$

is the moment generating function of $\log(\alpha Z^2 + \beta)$. As a moment generating function, it is log-convex on its domain, and, under (A3), its domain contains the interval $[0, r]$. We have $\varphi(0) = 1$, and, by Jensen's inequality, $\varphi'(0) = \mathbb{E}[\log(\alpha Z^2 + \beta)] \leq \log(\alpha + \beta) < 0$; at the positive root $\varphi(\kappa/2) = 1$, we therefore have $\varphi'(\kappa/2) > 0$. (A1)–(A2) imply that $\varphi(1) = \alpha + \beta < 1$, so $\kappa > 2$. With the values of α, β fixed, we refer to φ as the generating function associated with Z.

Our main objective in this paper is to investigate how κ changes under transformations of the parameters α and β resulting from temporal aggregation.

2.2 Temporal Aggregation.

As discussed in the introduction, GARCH models are often fit to the same time series at different frequencies. In the case of stock-variable aggregation, this means modeling both $\{X_t, t \in \mathbb{Z}\}$ and $\{X_{nt}, t \in \mathbb{Z}\}$, for some integer $n \geq 2$, as GARCH processes. However, Drost and Nijman (1993) showed that X_t and X_{nt} cannot both be GARCH processes: GARCH processes are not closed under temporal aggregation.

Instead, they showed that if $\{X_t, t \in \mathbb{Z}\}$ is a GARCH process then $\{X_{nt}, t \in \mathbb{Z}\}$ is a *weak* GARCH process. To explain what this means, we need additional background from Drost and Nijman (1993). Our starting point is a stationary sequence X_t with finite fourth moments, and a stationary solution σ_t^2 to the equation

$$\sigma_t^2 = \omega + \alpha X_{t-1}^2 + \beta \sigma_{t-1}^2.$$

The sequence $\{X_t, t \in \mathbb{Z}\}$ is an ordinary (or strong) GARCH process if $\{X_t/\sigma_t, t \in \mathbb{Z}\}$ are i.i.d. with mean 0 and variance 1. These i.i.d. random variables are then the Z_t appearing in (1). The sequence $\{X_t, t \in \mathbb{Z}\}$ is a *weak* GARCH process if the following two conditions are satisfied:

$$\mathbb{E}[X_t X_{t-i}^r] = 0, \quad \text{for all } i \geq 1 \text{ and } r = 0, 1, 2; \tag{5}$$

$$\mathbb{E}[(X_t^2 - \sigma_t^2) X_{t-i}^r] = 0, \quad \text{for all } i \geq 1 \text{ and } r = 0, 1, 2. \tag{6}$$

Suppose $\{X_t, t \in \mathbb{Z}\}$ is a strong GARCH process. For any integer $n \geq 2$, let $\tilde{X}_t = X_{nt}$. Drost and Nijman (1993) showed that $\{\tilde{X}_t, t \in \mathbb{Z}\}$ is a weak GARCH process satisfying

$$\tilde{\sigma}_t^2 = \omega_n + \alpha_n \tilde{X}_{t-1}^2 + \beta_n \tilde{\sigma}_{t-1}^2, \tag{7}$$

with $\tilde{\sigma}_t^2$ satisfying (6). The coefficients $(\omega_n, \alpha_n, \beta_n)$ are determined as follows

$$\omega_n = \omega \frac{1 - (\alpha + \beta)^n}{1 - (\alpha + \beta)}, \qquad \alpha_n = (\alpha + \beta)^n - \beta_n \tag{8}$$

and β_n is the solution in $(0, 1)$ of the quadratic equation

$$\frac{1 + \beta_n^2}{\beta_n} = \frac{1 + \alpha^2 \frac{1-(\alpha+\beta)^{2n-2}}{1-(\alpha+\beta)^2} + \beta^2 (\alpha + \beta)^{2n-2}}{\beta(\alpha + \beta)^{n-1}}. \tag{9}$$

Weak GARCH processes are difficult to work with and seldom used in practice. In (7), $\tilde{\sigma}_t^2$ is not the conditional variance of $\tilde{X}_t$, so weak GARCH processes are not directly applicable to volatility forecasting, which is one of the main uses of (strong) GARCH models. Even simulating a weak GARCH process presents a challenge because its driving noise is stationary but not i.i.d. (For more on weak GARCH processes, see, e.g., Alexander and Lazar (2021), Francq and Zakoian (2000), and Su and Zhu (2022).)

In light of these limitations of weak GARCH processes, it is not uncommon to see ordinary (strong) GARCH models fit to the same time series at different frequencies. We know that the resulting models are inherently incompatible. What conditions can we impose to make them in some sense consistent with each other? We propose two conditions:

(1) *Forecast consistency:* Forecasts from the high-frequency and low-frequency models should agree when both are provided with low-frequency data;

(2) *Tail consistency:* The high-frequency and low-frequency models should have the same power law exponent κ.

The first condition focuses on typical values; the second focuses on extremes. Condition (2) is appropriate because X_{nt} has the same stationary distribution as X_t, so models of the data at two frequencies should ideally have stationary distributions with equally heavy tails. We develop these conditions in the next two sections.

We limit our analysis to stock-variable aggregation, which compares X_{nt} at different values of n, as opposed to flow-variable aggregation, which compares partial sums of X_t of different lengths. We have extended some of our results to flow-variable aggregation in unpublished work. The Drost and Nijman (1993) parameter transformation is considerably more complicated for flow-variable aggregation; moreover, it depends on the kurtosis of the innovations and is therefore not distribution-free.

Market returns are usually treated as flow variables; market prices and balance sheet data observed at specific points in time are more naturally viewed as stock variables than flow variables. A problem of stock-variable aggregation arises, for example, in comparing publicly released quarterly balance sheet data with confidential data observed at higher frequencies. Stock-variable aggregation is also referred to as systematic sampling; see, e.g., Mamingi (2017), Su and Zhu (2022), Teles (2023), and the many references therein.

3 Forecast Consistency.

We now formulate the idea of forecast consistency precisely, and we show that it is satisfied by the Drost-Nijman parameters; indeed, related consistency considerations motivated the parameter transformation they introduced. We start with the high-frequency model $\{X_t, t \in \mathbb{Z}\}$ in (1), where the innovations Z_t are i.i.d. We consider the problem of forecasting future values of X_t^2. Because $X_t^2 = \sigma_t^2 Z_t^2$, with Z_t independent of σ_t and independent of all $\{X_s, \sigma_s, Z_s, s \leq t-1\}$, forecasting X_t^2 is equivalent to forecasting σ_t^2.

Define a low-frequency strong GARCH process $\{\bar{X}_t, t \in n\mathbb{Z}\}$ indexed by $t \in \{0, \pm n, \pm 2n, \dots\}$ by setting

$$\bar{X}_{t+n} = \bar{\sigma}_{t+n}\bar{Z}_{t+n}, \quad \bar{\sigma}_{t+n}^2 = \bar{\omega} + \bar{\alpha}\bar{X}_t^2 + \bar{\beta}\bar{\sigma}_t^2, \tag{10}$$

for some parameters $(\bar{\omega}, \bar{\alpha}, \bar{\beta})$ and i.i.d. innovations $\{\bar{Z}_t, t \in n\mathbb{Z}\}$ satisfying (A1)–(A2). We have not made any assumptions about how the high-frequency and low-frequency models are related.

We consider forecasts based on *best linear predictors* or, equivalently, linear projections. The best linear predictor of a square-integrable random variable Y given a finite or infinite sequence $\epsilon_1, \epsilon_2, \dots$ of square-integrable random variables, is a linear combination $\sum_{i\geq 0} a_i\epsilon_i$, with $\epsilon_0 = 1$, satisfying

$$\mathbb{E}[(Y - \sum_{i\geq 0} a_i\epsilon_i)\epsilon_j] = 0, \quad j = 0, 1, \dots.$$

In other words, the forecast error $Y - \sum_{i\geq 0} a_i\epsilon_i$ should be uncorrelated with all the ϵ_j. We denote the best linear predictor by $[Y|\epsilon_1, \epsilon_2, \dots]$. To indicate its value at specific outcomes e_i of the ϵ_i, we write $[Y|\epsilon_1 = e_1, \epsilon_2 = e_2, \dots]$.

Proposition 3.1. *Suppose that the X_t in the high-frequency model (1) and the $\bar{X}_t$ in the low-frequency model (10) have finite fourth moments. Suppose that in the low-frequency model (10) we have $(\bar{\omega}, \bar{\alpha}, \bar{\beta}) = (\omega_n, \alpha_n, \beta_n)$, as defined in (8)–(9). Then, for any $k \geq 1$, we have the forecast consistency property*

$$\begin{aligned}&[\bar{X}_{t+kn}^2|\bar{X}_t^2 = v_t, \bar{X}_{t-n}^2 = v_{t-n}, \bar{X}_{t-2n}^2 = v_{t-2n}, \dots] \\ &\quad = [X_{t+kn}^2|X_t^2 = v_t, X_{t-n}^2 = v_{t-n}, X_{t-2n}^2 = v_{t-2n}, \dots],\end{aligned} \tag{11}$$

for any $v_t, v_{t-n}, v_{t-2n}, \ldots$.

This result says that with the parameters in (8)–(9), the low-frequency model on the left will produce the same forecast as the high-frequency model on the right, when the two models are fed the same history. In practice, we expect the high-frequency model to have access to a richer history, which would lead to a different forecast. The point of (11) is that a richer history is the *only* reason the two models make different forecasts. When they are both limited to the low-frequency data, they will make the same forecasts, even though the low-frequency model is a (strong) GARCH process, whereas the temporally aggregated high-frequency model is not. This is an important and attractive consistency property in fitting models on two time scales.

Proof. Bollerslev (1988) observed that in a GARCH process the squared values follow an ARMA process; in the case of (10), this becomes

$$\bar{X}_{t+n}^2 = \bar{\omega} + (\bar{\alpha} + \bar{\beta})\bar{X}_t^2 + \bar{\eta}_{t+n} - \bar{\beta}\bar{\eta}_t, \quad t \in n\mathbb{Z}, \tag{12}$$

with $\bar{\eta}_t = \bar{X}_t^2 - \bar{\sigma}_t^2$. This recursion is a direct consequence of (10). It defines an ARMA(1,1) process indexed by $t \in \{0, \pm n, \pm 2n, \ldots, \}$ because the $\bar{\eta}_t$ are stationary with mean zero and are uncorrelated with each other. If we start from the high-frequency model (1) and consider the subsequence X_{nt}^2, $t \in \mathbb{Z}$, Drost and Nijman (1993) showed that this sequence also follows an ARMA(1,1) process,

$$X_{t+n}^2 = \omega_n + (\alpha_n + \beta_n)X_t^2 + \eta_{t+n} - \beta_n\eta_t, \quad t \in n\mathbb{Z}, \tag{13}$$

with parameters $(\omega_n, \alpha_n, \beta_n)$ as defined in (8)–(9); indeed, the equations in (8)–(9) are derived precisely to arrive at (13). In (13), the η_t, $t \in n\mathbb{Z}$, are again stationary with mean zero, finite variance, and uncorrelated with each other, so (13) does indeed define an ARMA(1,1) process. An expression for η_t is derived in Drost and Nijman (1993).

We now see from (12) that the linear projection of $\bar{X}_{t+kn}^2$ onto $1, \bar{X}_t^2, \bar{X}_{t-n}^2, \ldots$ is the best linear predictor for an ARMA(1,1) process. The best linear predictor for an ARMA(1,1) process can be found in, for example, equation (4.2.39) of Hamilton (1994); its coefficients are completely determined by the ARMA parameters $\bar{\omega}$, $(\bar{\alpha} + \bar{\beta})$, and $\bar{\beta}$ in (12). Similarly, we see from (13) that the linear projection of X_{t+kn}^2 onto $1, X_t^2, X_{t-n}^2, \ldots$ is the best linear predictor of an ARMA(1,1) process with parameters ω_n, $(\alpha_n + \beta_n)$, and β_n. Thus, if we take $(\bar{\omega}, \bar{\alpha}, \bar{\beta}) = (\omega_n, \alpha_n, \beta_n)$, then the coefficients of the best linear predictors for the two ARMA models coincide, and (11) follows. In other words, once we match the coefficients in the ARMA processes (12) and (13), the coefficients in their best linear predictors will match, even though the processes are driven by different noise sequences $\bar{\eta}_t$ and η_t. The noise sequences do not affect the forecasts once we specify past values $v_t, v_{t-n}, \ldots$ in (11). □

On the left side of (11), we could replace the best linear predictor with the conditional expectation — the two coincide for a (strong) GARCH process. But we could not make such a substitution on the right side of (11); the conditional expectation for the temporally aggregated process need not be linear in the temporally aggregated history.

4 Tail Consistency and Monotonicity

4.1 A Motivating Numerical Experiment.

Proposition 3.1 provides a guarantee of forecast consistency using the Drost-Nijman temporally aggregated parameters for the lower-frequency model, so from now on we will assume that lower-frequency models use these parameters. But suppose we want to go beyond a point forecast and simulate the future evolution of X_t to estimate risk measures, for example. Indeed, the ability to make this type of distributional forecast is an important advantage of working with a strong GARCH model rather than a weak GARCH model. Doing so requires specifying a distribution for the i.i.d. innovations in the low-frequency GARCH model (10). Ideally, we would like to be able to choose the innovations so that the stationary distributions under the high- and low-frequency models coincide; but that is too much to ask for, given that GARCH models are not closed under temporal aggregation. Instead, we consider the weaker objective of selecting $\tilde{Z}$ so that the correct tail decay is reproduced

$$\kappa(\alpha_n, \beta_n, \tilde{Z}) = \kappa(\alpha, \beta, Z). \tag{14}$$

In selecting an innovation distribution to use for simulation, there are at least two approaches one could follow: use a standard choice, such as a normal distribution, or estimate the empirical distribution of the innovations. If we take Z to be normal, we can ask whether (14) holds if $\tilde{Z}$ is also normal; our analysis in later sections addresses this type of question.[1] For additional motivation, we consider an example using empirical distributions from simulated data.

Once the parameters ω, α, β are fixed (or estimated) from observations of X_t, estimation of the empirical distribution of the innovations works as follows. Using the recursion $\sigma_{t+1}^2 = \omega + \alpha X_t^2 + \beta\sigma_t^2$, with some initial value σ_0^2, one can evaluate the sequence of σ_t^2 values. From these, one can extract the innovations $Z_t = X_t/\sigma_t$ to estimate their distribution. This process can be applied to data observed at different frequencies. We are interested in whether it provides the consistency in (14).

We simulate a high-frequency model with $\omega = 0.02$, $\alpha = 0.1$, $\beta = 0.8$, and innovations drawn from Z_5, with $Z_\nu = \sqrt{\frac{\nu-2}{\nu}}t_\nu$ the Student-t distribution with ν degrees of freedom, rescaled to unit variance. Our high-frequency data is thus drawn exactly from a GARCH(1,1) model. We simulate 10^5 observations and we aggregate the data at multiples of $n = \{2, 4, 5\}$. At each level of aggregation, we use the Drost-Nijman parameters $(\omega_n, \alpha_n, \beta_n)$ to extract the "implied" empirical distribution of the innovations $Z^{(n)}$. (We do not find statistically significant serial correlation in the implied innovations, even though they are not theoretically independent for $n > 1$.)

We next test whether tail consistency holds. More precisely we are asking whether the equality $\kappa(\alpha_n, \beta_n, Z^{(n)}) = \kappa(\alpha, \beta, Z_5)$ holds, with $Z^{(n)}$ the empirical distribution extracted at aggregation level n. As a test of this equality we use the empirical distributions of the innovations to compute the expectation[2] using a simulated time series with N observations

$$\hat{\varphi}(s, \alpha_n, \beta_n) = \frac{1}{N}\sum_{i=1}^{N}(\alpha_n Z_i^{(n)2} + \beta_n)^s \tag{15}$$

at the point $s = \frac{1}{2}\kappa(\alpha, \beta, Z_5)$. Tail consistency of the empirical distribution is achieved if this quantity approaches 1. For this simulation we use 100 samples of 10^5 observations each. The estimates for this quantity are shown in the fourth column of Table 1. The results for $n > 1$ are consistently less than 1 which indicate that the tail exponent of the empirical distribution is larger than $\kappa(\alpha, \beta, Z_5)$. These estimates indicate that using the empirical distributions $Z^{(n)}$ does not produce tail consistency.

What is less clear is why. This comparison reflects changes in the parameters α_n, β_n together with possible changes in the innovations $Z^{(n)}$. To isolate these effects, we focus on the question of how $\kappa_n = \kappa(\alpha_n, \beta_n, Z)$ changes with n with Z fixed. The fifth column of Table 1 (calculated using numerical integration) shows what happens when we fix the innovations at the original distribution, Z_5. Holding this distribution fixed, κ_n becomes progressively larger under temporal aggregation.

The last two columns in the table show that we can offset this effect by adjusting the degrees-of-freedom parameter ν_n (for the innovations Z_{ν_n}) downward with increasing n. With the appropriate adjustment, we can achieve tail consistency, as shown in the last column. In decreasing ν_n, we are

[1] The widely followed V-lab website `https://vlab.stern.nyu.edu/` makes GARCH estimates publicly available for many financial time series using quasi-maximum likelihood assuming normal innovations.

[2] This is similar to the method used by Mikosch and Stărică (2000) to compute the tail exponent of the empirical distribution.

Table 1: Comparison of κ values using Student-t distribution with five degrees of freedom, Z_5, and Student-t distributions with ν_n degrees of freedom Z_{ν_n}. The fourth column shows $\hat{\varphi}(\frac{1}{2}\kappa(\alpha,\beta,Z_5),\alpha_n,\beta_n)$ computed using the empirical distribution $Z^{(n)}$ determined as explained in text. Tail consistency requires that this quantity is equal to 1; values smaller than 1 indicate a higher tail exponent under the empirical distribution than $\kappa(\alpha,\beta,Z_5)$. The adjustment in ν_n achieves tail consistency.

n	α_n	β_n	$\hat{\varphi}(\frac{1}{2}\kappa,\alpha_n,\beta_n)$	$\kappa(\alpha_n,\beta_n,Z_5)$	ν_n	$\kappa(\alpha_n,\beta_n,Z_{\nu_n})$
1	0.1	0.8	0.93 ± 0.08	4.58	5	4.58
2	0.1043	0.7057	0.79 ± 0.27	4.75	4.816	4.58
4	0.0935	0.5625	0.53 ± 0.20	4.89	4.686	4.58
5	0.0862	0.5043	0.46 ± 0.30	4.92	4.656	4.58

using progressively heavier-tailed innovations at lower frequencies; doing so offsets a general tendency for $\kappa(\alpha_n,\beta_n,Z)$ to increase with n, for fixed Z.

This general tendency — and exceptions to it — are the focus of the rest of the paper. If we know that $\kappa(\alpha_n,\beta_n,Z)$ is increasing in n, then we know at least qualitatively that we need to replace Z with a heavier-tailed distribution at larger n in order to achieve tail consistency. Within a parametric family of innovation distributions, one can search for the appropriate adjustment, as we did with ν_n in Table 1; but, more generally, we would like to find conditions under which we know that the distributional adjustment needs to move toward heavier-tailed or lighter-tailed innovations.

One way to make precise the idea that $\tilde{Z}$ is heavier-tailed than Z is to impose the convex order $Z^2 \leq_{cx} \tilde{Z}^2$, meaning that $\mathbb{E}[g(Z^2)] \leq \mathbb{E}[g(\tilde{Z}^2)]$, for all convex functions g for which the expectations exist. As the function $x \mapsto (\alpha x+\beta)^s$ is convex for all $s \geq 1$, for all admissible α,β, $Z^2 \leq_{cx} \tilde{Z}^2$ implies that $\kappa(\alpha,\beta,Z) \geq \kappa(\alpha,\beta,\tilde{Z})$. We end this section by applying this idea to the Student-t distribution and confirming that varying the parameter ν has the offsetting effect suggested by Table 1.

Proposition 4.1. *For any admissible α,β, $\kappa(\alpha,\beta,Z_\nu)$ is increasing in $\nu \in (2,\infty)$, where Z_ν is the scaled Student-t distribution, scaled so that $\mathbb{E}[Z_\nu^2]=1$.*

4.2 Parameter Derivatives Under Temporal Aggregation.

Expressions (8)–(9) remain well-defined if we extend them to real values of $n \geq 1$. From any initial point (α_1,β_1), (8)–(9) then yield a trajectory of values (α_n,β_n), $n \geq 1$, with n now varying continuously. These trajectories are illustrated in Figure 4.1 in the (β,α) plane. (See also Figure 1 of Drost and Nijman (1993).) As suggested by the figure, in region A, both parameters decrease whereas in region B, α_n increases as β_n decreases. (Of the (β_n,α_n) values in Table 1, the case $n=1$ is in region B and the others are in region A.) The change in κ under temporal aggregation will depend on the changes in (α_n,β_n), so we need to characterize these regions explicitly, which we do in the following result. Recalling that we have extended n to real values, we use α_1' and β_1' to denote derivatives with respect to n evaluated at $n=1$.

Proposition 4.2. *For all admissible (α,β), we have $\beta_1' < 0$. With*

$$\alpha_c(\beta) = -\frac{\beta(2-\beta^2) - \sqrt{\beta^2(2-\beta^2)^2+(1-\beta^2)^3}}{1-\beta^2}, \tag{16}$$

we have $\alpha_1' > 0$ if $\alpha > \alpha_c(\beta)$, and $\alpha_1' < 0$ if $\alpha < \alpha_c(\beta)$.

The threshold $\alpha_c(\cdot)$ defines the boundary between regions A and B in Figure 4.1, so the proposition confirms the pattern in the figure.

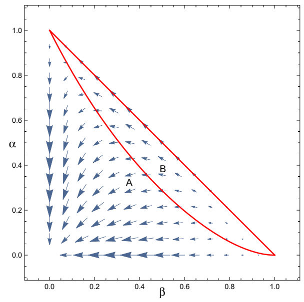

Figure 4.1: Regions A,B in the (β, α) parameter space. In the interior of region B, α increases under temporal aggregation; β decreases in both regions.

Proof. Differentiation in (8)–(9) yields

$$\alpha_1'(\alpha, \beta) = \frac{\log(\alpha+\beta)}{[1-(\alpha+\beta)^2][1-\beta^2]}\alpha f_\alpha(\alpha, \beta) \tag{17}$$

$$\beta_1'(\alpha, \beta) = \frac{\log(\alpha+\beta)}{[1-(\alpha+\beta)^2][1-\beta^2]}\beta f_\beta(\alpha, \beta) \tag{18}$$

with

$$\begin{aligned} f_\alpha(\alpha, \beta) &= 1 - \alpha^2 - 4\alpha\beta - 2\beta^2 + \beta^2(\alpha+\beta)^2 \\ f_\beta(\alpha, \beta) &= 1 + \alpha^2 - 2\alpha\beta - 2\beta^2 + \beta^2(\alpha+\beta)^2 \,. \\ &= [1 - \beta(\alpha+\beta)]^2 + \alpha^2 > 0 \,. \end{aligned}$$

The sign of β_1' follows from the last inequality.

The function f_α is quadratic in α. By writing it as

$$f_\alpha(\alpha, \beta) = -\alpha^2(1-\beta^2) - 2\beta\alpha(2-\beta^2) + (1-\beta^2)^2,$$

we see that the equation $f_\alpha = 0$ has the real root α_c in $[0, 1]$. For $\alpha > \alpha_c$ the function f_α is negative, and for $\alpha < \alpha_c$ it is positive. □

A trajectory starting from (β, α) will be in region A after n steps if $\alpha_1'(\beta_n, \alpha_n) < 0$. This condition can alternatively be formulated using the sign of the derivative α_n' evaluated at the initial point (β, α) through the following lemma.

Lemma 4.1. *We have*

$$n\alpha_n'(\alpha, \beta) = \alpha_1'(\alpha_n, \beta_n)\,, \quad n\beta_n'(\alpha, \beta) = \beta_1'(\alpha_n, \beta_n)\,. \tag{19}$$

Proof. The composition rule under double aggregation gives $\alpha_{mn}(\alpha, \beta) = \alpha_n(\alpha_m, \beta_m)$. Taking a derivative with respect to m we obtain $\frac{d}{dm}\alpha_{mn}(\alpha, \beta) = n\alpha_{mn}'(\alpha, \beta) = \alpha_m'(\alpha_n, \beta_n)$. Taking $m = 1$ gives the first relation (19). The second relation follows analogously. □

4.3 Monotonicity Condition.

Holding the distribution of Z fixed, the mapping $n \mapsto (\alpha_n, \beta_n)$ induces a mapping $n \mapsto \kappa_n := \kappa(\alpha_n, \beta_n)$. For parameters (α_n, β_n) in region A, where both parameters decrease, κ_n must increase to satisfy (4). Within region B, the effect on κ_n is not obvious. With Z fixed, we will say that κ *monotonicity holds* at

(α, β) if $\kappa_1' \geq 0$, where κ_1' denotes the derivative with respect to the temporal aggregation parameter n, evaluated at $n = 1$ and at parameter values (α, β). We want to identify regions of the parameter space at which $\kappa_1' \geq 0$, for a fixed Z or a family of distributions. As long as a trajectory of (α_n, β_n) evolves within the region where $\kappa_1' \geq 0$, κ_n continues to increase. The trajectory starting from any admissible (α, β) eventually enters region A in Figure 4.1 and thus has $\kappa_n' \geq 0$, for all sufficiently large n. In the following, φ denotes the generating function associated with Z.

Proposition 4.3. *Under the admissibility conditions (A1)–(A3), κ monotonicity holds at (α, β) if and only if*

$$\varphi\left(\frac{\kappa}{2} - 1\right) \geq \delta(\alpha, \beta) := -\frac{f_\alpha(\alpha, \beta)}{2\alpha\beta(\alpha + \beta)}. \tag{20}$$

Proof. As $\varphi(s)$ depends smoothly on s, α, and β, and α_n and β_n depend smoothly on n, we may differentiate κ_n with respect to n, applying the chain rule to get

$$\kappa_1' = \alpha_1' \partial_\alpha \kappa(\alpha, \beta) + \beta_1' \partial_\beta \kappa(\alpha, \beta). \tag{21}$$

The partial derivatives of κ can be evaluated by differentiating the equation $\varphi(\kappa/2) = 1$ with respect to α, β and rearranging to get

$$\partial_\alpha \kappa(\alpha, \beta) = -\frac{\kappa(\alpha, \beta)}{\varphi'(\kappa/2)} \mathbb{E}[Z^2(\alpha Z^2 + \beta)^{\kappa/2-1}] < 0$$
$$\partial_\beta \kappa(\alpha, \beta) = -\frac{\kappa(\alpha, \beta)}{\varphi'(\kappa/2)} \mathbb{E}[(\alpha Z^2 + \beta)^{\kappa/2-1}] < 0\,;$$

we noted in Section 2.1 that $\varphi'(\kappa/2) > 0$. Substituting into (21) gives

$$\kappa_1' = -\frac{\kappa(\alpha, \beta)}{\varphi'(\kappa/2)} \mathbb{E}[\{\alpha_1' Z^2 + \beta_1'\}(\alpha Z^2 + \beta)^{\frac{\kappa}{2}-1}].$$

It follows that $\kappa_1' \geq 0$ if and only if

$$\mathbb{E}[(\alpha_1' Z^2 + \beta_1')(\alpha Z^2 + \beta)^{\kappa/2-1}] \leq 0\,. \tag{22}$$

If $\alpha_1' \leq 0$, this inequality is automatic, and so is (20) because $-f_\alpha$ has the same sign as α_1', and $\varphi(s) > 0$, for all s. If $\alpha_1' > 0$, substituting $Z^2 = \frac{1}{\alpha}((\alpha Z^2 + \beta) - \beta)$ and multiplying by α shows that (22) is equivalent to

$$\alpha_1' + (\alpha\beta_1' - \beta\alpha_1')\varphi(\kappa/2 - 1) \leq 0\,.$$

As $\alpha\beta_1' - \beta\alpha_1' < 0$, this is equivalent to a lower bound on $\varphi(\kappa/2 - 1)$. Substituting the relations (17)–(18) for α_1', β_1' gives the bound (20). □

The condition in (20) is automatically satisfied in region A, where $\delta(\alpha, \beta) \leq 0$ because $f_\alpha(\alpha, \beta) \geq 0$. Our focus throughout the rest of the paper is on understanding when (20) holds within region B. We focus on the derivative κ_1' for tractability. Derivatives at larger values of n can be reduced to derivatives at $n = 1$ through Lemma 4.1.

The lower bound $\delta(\alpha, \beta)$ satisfies the inequality

$$\delta(\alpha, \beta) \leq \alpha + \beta\,, \quad \alpha + \beta \leq 1\,. \tag{23}$$

This can be seen by writing (23) as

$$(\alpha + \beta)^2[(\alpha + \beta)^2 - 1] \geq (1 + \alpha^2)[(\alpha + \beta)^2 - 1]\,,$$

and noting that $(\alpha + \beta)^2 \leq 1 + \alpha^2$. Because $\varphi(1) = \alpha + \beta$ and φ is convex, a simple condition for κ monotonicity is $\varphi'(1) > 0$ and $\kappa > 4$, as this implies $\varphi(\kappa/2 - 1) > \varphi(1) \geq \delta(\alpha, \beta)$ through (23).

5 Moment-Free Bounds.

Conditions for κ monotonicity in region B cannot be formulated purely in terms of (α, β), as the first part of our next result shows; we need to add conditions on κ or Z. In this section, we keep Z general and restrict κ; in the subsequent two sections, we add moment information about Z and weaken the conditions we need on κ. In the following, let $\delta_+(\alpha, \beta)$ denote $\max\{0, \delta(\alpha, \beta)\}$, and interpret $\log 0$ as $-\infty$.

Theorem 5.1. *(i) For any (α, β) with $\delta(\alpha, \beta) > 0$, there exists an admissible innovation distribution for which κ monotonicity fails at (α, β).*

(ii) If $\kappa > 4$, then

$$(\alpha+\beta)^{\frac{\kappa}{2}-1} \leq \varphi\left(\frac{\kappa}{2}-1\right) \leq (\alpha+\beta)^{1/(\frac{\kappa}{2}-1)}; \tag{24}$$

κ monotonicity is guaranteed if $\kappa \leq 2 + 2\log\delta_+(\alpha, \beta)/\log(\alpha+\beta)$.

(iii) If $2 < \kappa \leq 4$, then

$$(\alpha+\beta)^{1/(\frac{\kappa}{2}-1)} \leq \varphi\left(\frac{\kappa}{2}-1\right) \leq (\alpha+\beta)^{\frac{\kappa}{2}-1}, \tag{25}$$

and

$$\beta^{\frac{\kappa}{2}-1} \leq \varphi\left(\frac{\kappa}{2}-1\right); \tag{26}$$

κ monotonicity is guaranteed if $\kappa \geq 2 + 2\log(\alpha+\beta)/\log\delta_+(\alpha, \beta)$ or $\kappa \leq 2 + 2\log\delta_+(\alpha, \beta)/\log\beta$. In particular, κ monotonicity is guaranteed if

$$\log\delta_+(\alpha, \beta) \leq -\sqrt{\log\beta\log(\alpha+\beta)}. \tag{27}$$

The sufficient condition in (27) is stated purely in terms of α and β, so it is particularly easy to check. It does not contradict the statement in (i) because (27) implies κ monotonicity only for $\kappa \in (2, 4]$. Anywhere (27) holds, the violation guaranteed by (i) must have $\kappa > 4$.

We noted in Section 2.2 that Drost and Nijman (1993) assume a finite fourth moment in defining a weak GARCH process, as required in (6). We assumed finite fourth moments in Proposition 3.1 to ensure the existence of a best linear predictor. This condition would entail $\kappa > 4$, but a finite fourth moment is not needed for the parameter transformation (8)–(9), so our question of κ monotonicity remains well-posed for $\kappa \in (2, 4]$.

5.1 An Extremal Case.

As already suggested by Theorem 5.1, the case $\kappa \in (2, 4)$ is often qualitatively different from $\kappa > 4$ because in the first case $\kappa/2 - 1 < 1 < \kappa/2$, and monotonicity depends on $\varphi(\kappa/2 - 1)$. In this section, we show that among all distributions with a given $\kappa \in (2, 4)$, κ monotonicity is guaranteed if it holds for a certain two-point distribution. This result provides a simple sufficient condition for monotonicity for $\kappa \in (2, 4)$ at a fixed (α, β): if κ monotonicity holds for this extremal two-point distribution, it holds for all innovations that yield the same κ at (α, β).

For any $x \geq 1$, let B_x denote a random variable supported in $\{0, x\}$ with mass $1/x$ at x, and thus with $\mathbb{E}[B_x] = 1$. The associated generating function is given by

$$\varphi_{B_x}(s) = \beta^s + \frac{1}{x}[(\alpha x+\beta)^s - \beta^s], \quad s \geq 0. \tag{28}$$

We think of B_x as a candidate distribution for the squared innovations. The innovations themselves could have the symmetric distribution on $\{-\sqrt{x}, 0, \sqrt{x}\}$ with mass $(x-1)/x$ at zero. Condition (A3) requires $x > 1$.

Theorem 5.2. *Fix any admissible (α, β). For each $\kappa \in (2,4)$, there is an $x > 1$ for which $\varphi_{B_x}(\kappa/2) = 1$. If κ monotonicity holds for this B_x, then it holds for all admissible innovations with the same κ.*

We can make the condition more explicit as follows:

Corollary 5.1. *Suppose $\kappa \in (2,4)$ and $\beta^{\kappa/2-1} < \delta(\alpha,\beta)$. Define*

$$x^* = \frac{1 - \beta\delta(\alpha,\beta) - \alpha\beta^{\kappa/2-1}}{\alpha[\delta(\alpha,\beta) - \beta^{\kappa/2-1}]}.$$

If $\varphi_{B_{x^*}}(\kappa/2) \geq 1$, then κ monotonicity holds for all innovations with the same κ. If $\kappa \in (2,4)$ and $\beta^{\kappa/2-1} \geq \delta(\alpha,\beta)$, κ monotonicity holds without further conditions.

This corollary provides a simple check for monotonicity for $\kappa \in (2,4)$ at any (α, β). In extensive numerical testing over a grid of (α, β) pairs and values of κ in $(2,4)$, we have not found any instance in which the conditions in the corollary fail to hold. This numerical evidence thus suggests that κ monotonicity holds whenever $\kappa \in (2,4)$, regardless of the innovation distribution. Recall that Theorem 5.1 already tightly constrained the possibility of exceptions for κ values in this range.

5.2 Unimodal Innovations.

The violations of κ monotonicity constructed in Theorem 5.1(i) relied on distributions of Z^2 that put nearly all mass on 1, which would be unnatural in applications. The range of guaranteed κ monotonicity is much wider if we restrict attention to unimodal innovations with mode zero. The innovations commonly used in practice satisfy these conditions. As a preview of our results, Figure 5.1 illustrates the region of (β, α) pairs for which κ monotonicity holds for this class of innovations when $\kappa \geq 4$.

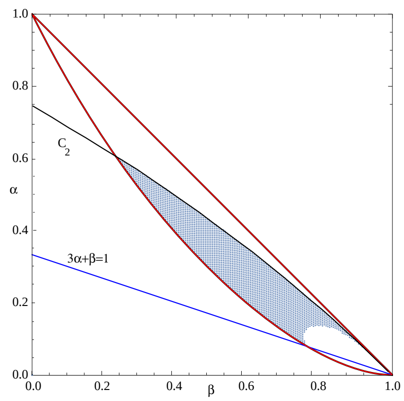

Figure 5.1: In the shaded region, κ monotonicity holds for any unimodal admissible innovation distribution with mode zero and $\kappa \geq 4$.

The precise unimodality condition we impose is that the cumulative distribution function of Z be convex on $(-\infty, 0)$ and concave on $(0, \infty)$. Our bounds are based on the special case $U \sim U[-\sqrt{3}, \sqrt{3}]$, the uniform distribution on the interval $[-\sqrt{3}, \sqrt{3}]$. The admissibility condition (A3) requires $\mathbb{P}(\alpha U^2 + \beta > 1) > 0$ and thus $3\alpha + \beta > 1$, so the results in this section are limited to parameters satisfying this condition. Without this condition, the equation $\mathbb{E}[(\alpha U^2 + \beta)^s] = 1$ would fail to have a solution $s > 0$. For this choice of innovation, we have an explicit expression for the function φ.

Lemma 5.1. *With $U \sim U[-\sqrt{3}, \sqrt{3}]$, we have*

$$\varphi_U(s) = \mathbb{E}[(\alpha U^2 + \beta)^s] = \beta^s {}_2F_1\left(\frac{1}{2}, -s, \frac{3}{2}; -\frac{3\alpha}{\beta}\right),$$

where ${}_2F_1$ denotes the Gauss hypergeometric function.

Proof. By definition,

$$\varphi_U(s) = \frac{1}{2\sqrt{3}} \int_{-\sqrt{3}}^{\sqrt{3}} (\alpha u^2 + \beta)^s \, du = \beta^s \int_0^1 (1 + 3\alpha u^2/\beta)^s \, du.$$

From 15.3.1 in Abramowitz and Stegun (1972), we have the representation

$${}_2F_1(\frac{1}{2}, -s, \frac{3}{2}; -x) = \frac{1}{2} \int_0^1 t^{-1/2}(1 + tx)^s dt = \int_0^1 (1 + u^2 x)^s du,$$

from which the result follows. □

The uniform case provides a powerful bound. We show in Appendix B.2 that for any unimodal innovation distribution, $\varphi(s) \geq \varphi_U(s)$, for all $s \geq 1$. A condition on φ_U thus leads to κ monotonicity for a broad class of unimodal innovations. We write κ_U for the strictly positive root of $\varphi_U(\kappa_U/2) = 1$. We need $3\alpha + \beta > 1$ for the existence of κ_U, and the quadratic parameter constraint in the following result is equivalent to $\kappa_U \geq 4$. The quadratic constraint is indicated by the curve C_2 in Figure 5.1.

Theorem 5.3. *Suppose* $3\alpha + \beta \geq 1$, $9\alpha^2/5 + 2\alpha\beta + \beta^2 \leq 1$, *and*

$$\min_{s \in [1, \kappa_U/2 - 1]} \varphi_U(s) \geq \delta(\alpha, \beta). \tag{29}$$

Then κ *monotonicity holds at* (α, β) *for any admissible innovation distribution that is unimodal with mode zero and has* $\kappa \geq 4$.

Figure 5.1 shows that condition (29) holds for a large proportion of (β, α) pairs. Theorem 5.3 improves on Theorem 5.1 in the sense that κ monotonicity is guaranteed in the shaded region regardless of the value of κ. Above the C_2 curve, we have $\kappa < 4$ for unimodal innovations. As we noted at the end of Section 5.1, numerical testing based on the condition in Corollary 5.1 suggests that κ monotonicity always holds with $\kappa < 4$. The main question left open by Figure 5.1 (and Theorem 5.3) is whether we have κ monotonicity with unimodal innovations when β is close to 1.

6 Convexity Bounds.

Using the first three moments of Z^2 we can derive further conditions for κ monotonicity. These results, derived in Section 6.1, use a generalization of the standard convex stochastic order. In Section 6.2, we derive conditions that apply when we have sufficiently many moments of Z^2 to provide integer bounds on κ.

6.1 s-Convexity Bounds.

The s-convex stochastic orderings, introduced by Denuit et al. (1998), are generalizations of the ordinary convex stochastic order; see also Section 3.A.5 of Shaked and Shanthikumar (2007). For any $s = 1, 2, \ldots$, let $\mathcal{U}_{s-cx}$ denote the class of functions $g : [0, \infty) \to \mathbb{R}$ whose derivative of order s exists and satsifies $g^{(s)}(x) \geq 0$, for all $x \in [0, \infty)$. For non-negative random variables X and Y, we write $X \preceq_{s-cx} Y$ if $\mathbb{E}[g(X)] \leq \mathbb{E}[g(Y)]$, for all $g \in \mathcal{U}_{s-cx}$ for which the expectations exist. The 1-convex order is the ordinary stochastic order, and the 2-convex order is the usual convex order. We will apply the cases $s = 3, 4$. For other results on stochastic comparisons of GARCH processes, see Bellini et al. (2014).

Let $\mathcal{B}_s(\mu_1, \ldots, \mu_{s-1})$ denote the set of random variables on $[0, \infty)$ with moments $\mu_1, \ldots, \mu_{s-1}$. Denuit et al. (1998) characterize the minimal distributions within these sets in the s-convex order. In other words, for each s they define a random variable X_s such that $X_s \preceq_{s-cx} Y$, for all non-negative random variables Y. (The maximal counterparts are defined only on bounded intervals $[a, b]$; as noted in Denuit et al. (1998), the minimal distributions hold on $[0, \infty)$.) We will use these results to bound Z^2, and as we require $\mathbb{E}[Z^2] = 1$, we take $\mu_1 = 1$.

For $s = 3$, the minimal distribution is supported on $\{0, \mu_2\}$ with mass $1/\mu_2$ at μ_2. For $s = 4$, the minimal distribution is supported on points $\{y_1, y_2\}$, $y_2 > 1 > y_1$ defined by

$$y_{2/1} = \frac{\mu_3 - \mu_2 \pm \sqrt{(\mu_3 - \mu_2)^2 - 4(\mu_2 - 1)(\mu_3 - \mu_2^2)}}{2(\mu_2 - 1)}$$

with mass $p = (y_2 - 1)/(y_2 - y_1)$ on y_1. We will write $\phi_s(u)$ for $\mathbb{E}[(\alpha X_s + \beta)^u]$

To illustrate, consider distributions matching the first few moments of Z^2 when $Z \sim N(0, 1)$, meaning $\mu_1 = 1$, $\mu_2 = 3$, and $\mu_3 = 15$. Then

$$\phi_3(u) = \mathbb{E}[(\alpha X_3 + \beta)^u] = (1 - \frac{1}{\mu_2})\beta^u + \frac{1}{\mu_2}(\alpha\mu_2 + \beta)^u = \frac{2}{3}\beta^u + \frac{1}{3}(3\alpha + \beta)^u\,,$$

and

$$\phi_4(u) = \mathbb{E}[(\alpha X_4 + \beta)^u] = p(\alpha y_1 + \beta)^u + (1 - p)(\alpha y_2 + \beta)^u\,,$$

where $y_1 = 3 - \sqrt{6}$, $y_2 = 3 + \sqrt{6}$, $p = \frac{1}{2}(1 + \sqrt{\frac{2}{3}})$.

Because the functions ϕ_3 and ϕ_4 are based on minimal distributions, they provide lower bounds on the generating function φ associated with any Z for which Z^2 has the same first three or four moments. The bound $\varphi(u) \geq \phi_3(u)$ holds where x^u is 3-convex (i.e., $u \notin (1, 2)$), and the bound $\varphi(u) \geq \phi_4(u)$ holds where x^u is 4-convex (i.e., $u \notin (0, 1) \cup (2, 3)$). These properties lead to the following result, which allows us to verify κ monotonicity based on the first few moments of the squared innovations.

Theorem 6.1. *Let X_s be s-convex minimal elements of $\mathcal{B}_s(\mu_1, \ldots, \mu_{s-1})$, $s = 3, 4$, and let ϕ_s be the associated generating functions. For any κ satisfying*

$$\kappa/2 \in [1, 2] \cup [3, \infty) \quad \text{and} \quad \phi_3\left(\frac{\kappa}{2} - 1\right) \geq \delta(\alpha, \beta), \tag{30}$$

κ monotonicity holds under any admissible innovation Z with $Z^2 \in \mathcal{B}_3(\mu_1, \mu_2)$ satisfying $\varphi(\kappa/2) = 1$. If

$$\kappa/2 \in [2, 3] \cup [4, \infty) \quad \text{and} \quad \phi_4\left(\frac{\kappa}{2} - 1\right) \geq \delta(\alpha, \beta), \tag{31}$$

then κ monotonicity holds under any admissible innovation Z with $Z^2 \in \mathcal{B}_4(\mu_1, \mu_2, \mu_3)$ satisfying $\varphi(\kappa/2) = 1$.

Figure 6.1 shows the (β, α) pairs where κ monotonicity is guaranteed by Theorem 6.1, when the moments are chosen to match the first three moments of Z^2, with $Z \sim N(0, 1)$. The two labeled curves show the set of (β, α) pairs for which $\varphi(2) = 1$ ($\kappa = 4$) and $\varphi(3) = 1$ ($\kappa = 6$), respectively, under normal innovations. So, for example, at (β, α) pairs in the shaded region above the $\kappa = 4$ curve, we have $\kappa < 4$ (because $\varphi(2) > 1$), and κ monotonicity is guaranteed for normal innovations by (30). In the shaded region between the two curves, κ monotonicity is guaranteed for normal innovations by (31). The figure indicates that κ monotonicity is guaranteed for normal innovations throughout nearly all of region B, but we will show in Section 7 that there are indeed exceptions in the lower right corner.

With $\alpha\mu_2 + \beta > 1$, X_3 is an admissible choice of Z^2, and with $\alpha y_2 + \beta > 1$, X_4 is admissible. Because these random variables are minimal in the corresponding convex orders, the conditions in Theorem 6.1 are in a sense the best possible conditions, given the moments μ_2 and μ_3.

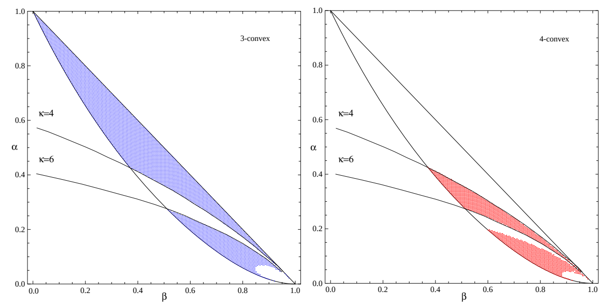

Figure 6.1: Points (β, α) where κ monotonicity is guaranteed by (30) (left panel) and by (31) (right panel) for innovations matching the first three moments of Z^2, $Z \sim N(0,1)$.

6.2 s-Convexity with Moment Bounds.

Theorem 6.1 provides monotonicity guarantees for specific values of κ and thus requires knowledge of κ. In this section, we derive conditions based on bounds on κ. Some of our results use bounds on κ based on integer moments of Z^2. Knowing the moments of Z^2 is equivalent to knowing the values of $\varphi(n)$ for $n \in \mathbb{N}$, because

$$\varphi(n) = \sum_{k=0}^{n} \binom{n}{k} \alpha^k \beta^{n-k} \mathbb{E}[Z^{2k}].$$

In some cases, $\varphi(n)$ can be calculated recursively, without evaluating $\mathbb{E}[Z^{2k}]$:

Proposition 6.1. *With standard normal Z and any $s \geq 1$,*

$$\varphi(s+1) = [\alpha(1+2s) + \beta]\varphi(s) - 2\alpha\beta s \varphi(s-1)\,.$$

For double exponential innovations,

$$\varphi(s+1) = \alpha(1+s)(1+2s)\varphi(s) - 2\alpha\beta s(s+1)\varphi(s-1) + \beta^{s+1}\,.$$

Proof. Write $\varphi(s+1)$ as $\alpha\mathbb{E}[Zf(Z)] + \beta\varphi(s)$ with $f(z) = z(\alpha z^2 + \beta)^s$. For standard normal Z, Stein's lemma yields $\mathbb{E}[Zf(Z)] = \mathbb{E}[f'(Z)]$. Making this substitution and simplifying yields the first recursion. For exponentially distributed Y with parameter λ and differentiable g with $\mathbb{E}[g(Y)] < \infty$, integration by parts yields

$$\mathbb{E}[g(Y)] = \frac{1}{\lambda}\mathbb{E}[g'(Y)] + g(0) = \frac{1}{\lambda^2}\mathbb{E}[g''(Y)] + \frac{1}{\lambda}g'(0) + g(0).$$

With $g(y) = (\alpha y^2 + \beta)^{s+1}$, $Y = |Z|$, and $\lambda = \sqrt{2}$, this identify yields the second recursion. □

With the ability to evaluate $\varphi(n)$ at integer n, we can find the $m \in \mathbb{N}$ at which

$$\varphi(m) \leq 1 < \varphi(m+1),$$

and this tells us $m = \lfloor \kappa/2 \rfloor$, without knowledge of κ. In other words, we can bound κ using moments of Z^2. As a first application, we have the following.

Proposition 6.2. *Monotonicity holds if*

$$\frac{\varphi(m)}{\varphi(m+1)} \geq \delta(\alpha, \beta) \tag{32}$$

or if $m \geq 2$ and

$$\min\left(\varphi(m-1), \frac{\varphi(m-2)^2}{\varphi(m-1)}\right) \geq \delta(\alpha, \beta). \tag{33}$$

Proof. Log-convexity of $\varphi(s)$ yields

$$\log 1 - \log \varphi(\kappa/2-1) \leq \log \varphi(m+1) - \log \varphi(m)\,,$$

so the ratio in (32) is a lower bound on $\varphi(\kappa/2-1)$. The log-convexity of $\varphi(s)$ on the points $(m-2, m-1, \kappa/2-1)$ implies that the left side of (33) also provides a lower bound on $\varphi(\kappa/2-1)$. □

The lower-left panel of Figure 6.2 shows the (β, α) values for which Proposition 6.2 guarantees κ monotonicity in the case of the normal distribution (or any other distribution with same first $2m+2$ moments). The simple bounds in (32)–(33) turn out to be quite powerful in the lower-right corner of the parameter region. Nevertheless, gaps remain in this corner.

We can also combine integer bounds on κ with the s-convexity bounds. As in the previous section, let ϕ_3 and ϕ_4 denote the generating functions associated with the 3-convex and 4-convex minimal distributions, respectively. If $\alpha\mu_2 + \beta > 1$, the function $\phi_3(s)$ has a minimum at

$$s_{\min,3} = \frac{\log(-\log\beta) + \log(\mu_2 - 1) - \log\log(\alpha\mu_2 + \beta)}{\log(\alpha\mu_2 + \beta) - \log\beta}\,.$$

For $\alpha\mu_2 + \beta \leq 1$ set $s_{\min,3} = \infty$. Similarly, let $s_{\min,4}$ denote the minimizer of ϕ_4 if $\alpha y_2 + \beta > 1$ and $s_{\min,4} = \infty$, otherwise.

For $\alpha y_2 + \beta > 1$ denote by $\kappa_4 > 0$ the solution of the equation $\phi_4(\kappa_4/2) = 1$, and let $m_4 = \lfloor \kappa_4/2 \rfloor$. The value of κ_4 gives useful constraints on m and κ, as shown by the following result. Denuit et al. (2000) use similar arguments to bound the Lundberg coefficient of insurance risk theory.

Lemma 6.1. *Assume $\alpha y_2 + \beta > 1$. Either $2 \leq \kappa \leq \kappa_4 < 4$, or $4 \leq \kappa_4 \leq \kappa \leq 6$, or $6 \leq \kappa \leq \kappa_4$. In particular, $m = 1$ if and only if $m_4 = 1$, and $m = 2$ if and only if $m_4 = 2$.*

Proof. For $s \in [1,2] \cup [3, \infty)$, (46) yields $\phi_4(s) \leq \varphi(s)$. In addition, we have $\varphi(1) = \phi_4(1)$ and $\varphi(2) = \phi_4(2)$ because, by construction, the first three moments of X_4 and Z^2 agree. It follows that if $\kappa/2 \in [1,2]$, then $\kappa_4/2 \in [1,2]$, and $\kappa \leq \kappa_4$. Similarly, $\varphi(3) = \phi_4(3)$ implies that if $\kappa/2 \geq 3$, then $\kappa_4 \geq \kappa$.

For $s \in [2,3]$, $\phi_4(s) \geq \varphi(s)$. In addition, $\varphi(2) = \phi_4(2)$ and $\varphi(3) = \phi_4(3)$. It follows that if $\kappa/2 \in [2,3]$, then $\kappa_4/2 \in [2,3]$, and $\kappa_4 \leq \kappa$. □

The following result provides sufficient conditions for κ monotonicity without reference to the innovations Z. The conditions are fomulated purely in terms of the 3-convex minimal and 4-convex minimal distributions that match the first two and first three moments of Z^2, respectively.

Theorem 6.2. *Suppose $\alpha y_2 + \beta > 1$. Then κ monotonicity holds under any of the following conditions:*

(i) $m_4 \neq 2$, $\kappa_4/2 - 1 \leq s_{\min,3}$ *and* $\phi_3(\kappa_4/2 - 1) \geq \delta(\alpha, \beta)$.

(ii) $m_4 \neq 2$, $\alpha\mu_2 + \beta > 1$, *and* $\phi_3(s_{\min,3}) \geq \delta(\alpha, \beta)$.

(iii) $m_4 = 2$ *and* $\phi_4(s_{\min,4}) \geq \delta(\alpha, \beta)$.

Proof. If $m_4 \neq 2$, then we know from Lemma 6.1 that $\kappa \leq \kappa_4$ and $\kappa/2 - 1 \in (0,1) \cup [2, \infty)$. By (45), $\varphi(\kappa/2-1) \geq \phi_3(\kappa/2-1) \geq \phi_3(\kappa_4/2-1)$, where the second inequality holds because ϕ_3 is decreasing on $[0, s_{\min,3}]$. Thus, $\varphi(\kappa/2-1) \geq \delta(\alpha, \beta)$, and κ monotonicity holds in case (i). In case (ii), $\alpha\mu_2 + \beta > 1$ implies that $s_{\min,3} < \infty$, and then $\varphi(\kappa/2-1) \geq \phi_3(\kappa/2-1) \geq \phi_3(s_{\min,3}) \geq \delta(\alpha, \beta)$, and κ monotonicity holds. In case (iii), Lemma 6.1 yields $m = 2$, so $\kappa/2 - 1 \in [1, 2)$, and applying (46) we get $\varphi(\kappa/2-1) \geq \phi_4(\kappa/2-1) \geq \phi_4(s_{\min,4})$. □

The region of (β, α) where κ monotonicity holds by Theorem 6.2 is shown in the upper panels of Figure 6.2 for squared innovations matching the first three moments of a squared standard normal random variable. The conditions cover almost the entire region of (α, β), except for a gap at large $\beta > 0.85$, and the region below the $\alpha y_2 + \beta = 1$ line. The region below this line has large values of κ and is covered by the ratio bounds (32)–(33).

The lower right panel of Figure 6.2 shows that combination of Proposition 6.2 and Theorem 6.2 covers nearly all of region B for normal innovations. The figure makes it tempting to conjecture that κ monotonicity holds everywhere. But numerical examples indicate that some of the gaps are genuine exceptions to monotonicity. For example, at $\alpha = 0.001$, $\beta = 0.995$, we find $\kappa = 1049.64$ and $\varphi(\kappa/2-1) = 0.9605$, but $\delta(\alpha, \beta) = 0.9638$, violating (20). In the next section, we prove the existence of such exceptions (thus excluding the possibility of numerical error) by considering limits as $\beta \to 1$.

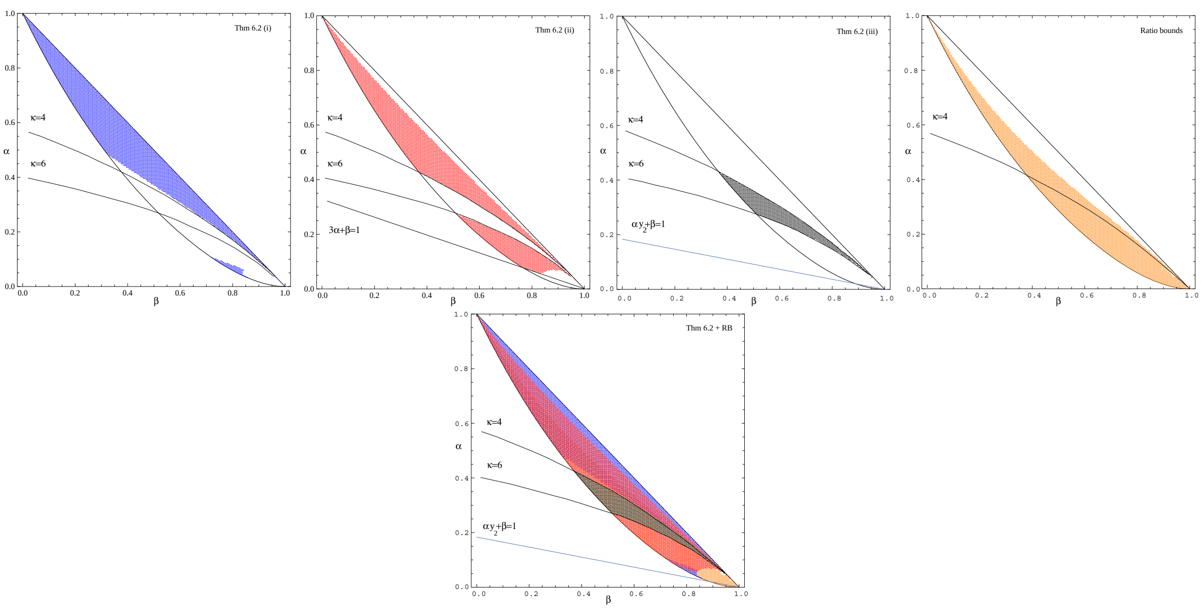

Figure 6.2: The upper panels show (β, α) points where κ monotonicity is guaranteed for $Z \sim N(0,1)$ by each case of Theorem 6.2. The lower left panel shows the region covered by Proposition 6.2, and the lower right panel shows the combined regions from Theorem 6.2 and Proposition 6.2.

7 Exceptions.

By an exception we mean a point (α, β) and an innovation distribution for which $\varphi(\kappa/2-1) < \delta(\alpha, \beta)$, implying that κ monotonicity fails. Numerical testing finds exceptions near $\beta = 1$. We investigate this region by considering limits as $\beta \uparrow 1$ with $\alpha = r(1-\beta)$, with $r \in (0,1)$.

We compare the limits of the derivatives (more precisely, difference quotients) of $\delta(\alpha, \beta)$ and $\varphi(\kappa/2-1)$ as $\beta \to 1$ along the line $\alpha = r(1-\beta)$. For δ, direct calculation yields

$$\lim_{\beta \to 1} \frac{1 - \delta(r(1-\beta), \beta)}{1-\beta} = -3 + \frac{2}{r} + r\,. \tag{34}$$

An exception is then guaranteed, for all β sufficiently close to 1, if the limit of

$$\Delta_\beta = \frac{1 - \varphi(\kappa/2-1)}{1-\beta} \tag{35}$$

is larger than the limit in (34). We consider the following class of innovations:

Sub-Gaussian innovations: There exist positive constants C, v such that

$$\mathbb{P}(|Z| > t) \le Ce^{-vt^2}\,, \text{ for all } t > 0\,.$$

This condition implies that $\psi(\theta) := \mathbb{E}[e^{\theta Z^2}]$ is finite for θ in a neighborhood of the origin. With $\bar{\theta} = \sup\{\theta \ge 0 : \psi(\theta) < \infty\}$ (which may be infinite), we add the requirement that

$$\psi(\theta) \to \infty \text{ as } \theta \uparrow \bar{\theta}. \tag{36}$$

We also require that Z satisfy (A3) for all (α, β) that satisfy (A1). Normal innovations $Z \sim N(0,1)$ fall within this class.

In a separate analysis, not reported here, we have proved the existence of exceptions for exponentially bounded and heavy-tailed innovations. Details are available from the authors.

Theorem 7.1. *For sub-Gaussian Z,*

$$\lim_{\beta\to 1, \alpha = r(1-\beta)} \Delta_\beta = 2rg_Z'(a(r))\,, \tag{37}$$

where $g_Z(c) := e^{-\frac{c}{2r}}\mathbb{E}[e^{\frac{1}{2}cZ^2}]$, and $a(r)$ is the unique strictly positive solution of $g_Z(a(r)) = 1$.

We prove this result in D but first discuss its implications. If the limit of Δ_β is larger than the limit (34), then we know that $\varphi(\kappa/2-1) < \delta(r(1-\beta), \beta)$, indicating an exception, for all β sufficiently close to 1. It thus remains to compare the limit of Δ_β in the theorem with the limit in (34).

With $Z \sim N(0,1)$, the expectation $\mathbb{E}[e^{\frac{1}{2}cZ^2}] = \frac{1}{\sqrt{1-c}}$ exists for $c < 1$ and

$$g_Z(c) = \frac{e^{-\frac{c}{2r}}}{\sqrt{1-c}}\,. \tag{38}$$

This function is convex with $g_Z(0) = 1$, $g_Z'(0) < 0$, and unique positive root $a(r)$ satisfying $a(r) = 1 - e^{-a(r)/r}$, where $g_Z'(a(r)) > 0$. Differentiating (38) gives

$$g_Z'(c) = \frac{c + r - 1}{2r(1-c)^{3/2}}e^{-\frac{c}{2r}} = \frac{c+r-1}{2r(1-c)}g_Z(c),$$

from which we obtain

$$\lim_{\beta\to 1} \Delta_\beta = 2rg_Z'(a(r)) = \frac{a(r) + r - 1}{1 - a(r)} = -1 + \frac{r}{1-a(r)}. \tag{39}$$

Exceptions occur for β near 1 if

$$-1 + \frac{r}{1-a(r)} > -3 + \frac{2}{r} + r\,. \tag{40}$$

With $x(r) = 1 - a(r)$, the fixed point equation $a(r) = 1 - e^{-a(r)/r}$ is equivalent to

$$r = \frac{x(r) - 1}{\log x(r)}, \tag{41}$$

and (40) is equivalent to

$$\frac{r^2}{x(r)} > -2r + 2 + r^2. \tag{42}$$

Equation (41) is a strictly increasing function from x to r, mapping $(0,1)$ onto $(0,1)$, so we may evaluate the limit as $r \to 0$ of the left side of (42) as

$$\lim_{r\to 0}\frac{r^2}{x(r)} = \lim_{x\to 0}\frac{(x-1)^2}{x(\log x)^2} = \infty,$$

because $\sqrt{x}\log x \to 0$ as $x \to 0$. Thus, (42) holds for all sufficiently small r. Numerically, we find that (39) is greater than (34) (so (40) holds) for $r < 0.45$. For r in this range, we will observe exceptions for all β sufficiently close to 1 and $\alpha = r(1-\beta)$.

The rather subtle structure of exceptions is surprising. The results in previous sections all point to monotonicity of κ as the "typical" case. Our bounds guarantee κ monotonicity over nearly all of the parameter space for broad classes of distributions. And yet, for some of the most important classes of innovation distributions, we find exceptions. Returning to the example of Section 4.1, this means that we cannot automatically assume that ensuring tail consistency across multiple time scales means using heavier-tailed innovations at lower frequencies, even though this is usually the case.

8 Empirical Illustration.

To illustrate our theoretical results, we apply them to several time series. We consider daily log returns on the S&P 500 stock market index (SPX), the euro/dollar exchange rate (EUR/USD), the yen/dollar exchange rate (JPY/USD), the Brazilian real/dollar exchange rate (BRL/USD), and the price of West Texas Intermediate crude oil (WTI). This set gives us an equity index, two major exchange rates, an emerging market exchange rate, and a commodity price. For each series we use 11 years of daily data, from 2010 through 2020.

We de-mean each series by subtracting its sample mean for the full time period. We then fit a GARCH(1,1) model to each de-meaned series, using the `fGARCH` package in R. We estimate parameter values using quasi-maximum likelihood estimation, which assumes normal innovations. Normal innovations facilitate comparison with results and figures in previous sections.

The parameter estimates $(\hat{\alpha}_1, \hat{\beta}_1)$ for each series are shown in Table 2. The standard errors for these estimates range from 0.006 to 0.02. The third column reports κ_1, the value of κ found by solving $\varphi(\kappa/2) = 1$, using parameter values $(\hat{\alpha}_1, \hat{\beta}_1)$ and a normal distribution for Z.

Starting from the parameters $(\hat{\alpha}_1, \hat{\beta}_1)$ for each series, we apply the temporal aggregation transformation (8)–(9) to get a sequence of parameters (α_n, β_n), $n = 1, 2, \ldots$. The sequence for each series follows a trajectory of the type illustrated in Figure 4.1, starting in region B and eventually entering region A. The $n_{\max}$ value for each series in Table 2 shows the largest n for which the trajectory remains in region B. The question of κ monotonicity is relevant in the range $n \le n_{\max}$; in region A, monotonicity is automatic.

The last four columns show the range of n values (with $n \le n_{\max}$) for which the indicated results guarantee κ monotonicity. The results collectively guarantee κ monotonicity for all series for all n. In fact, in these examples, Theorem 6.2 guarantees κ monotonicity in every case, and Proposition 6.2 covers all but one case. This pattern should not be surprising, given the near-complete coverage illustrated in Figure 6.2. Proposition 6.2 contains two sufficient conditions for κ monotonicity, (32) and (33), and we credit the proposition with ensuring monotonicity if either of these conditions is satisfied. In several cases, only one of the two conditions in Proposition 6.2 is met, but neither is uniformly more effective than the other.

We did not include Theorem 6.1 in the table because it requires stronger conditions than Theorem 6.2. We have encountered rare cases in which κ monotonicity is guaranteed by Theorem 6.1 but not by any of the results in Table 2. In our numerical search of the parameter space, we have not encountered cases in which condition (20) for κ monotonicity in Proposition 4.3 holds but for which monotonicity is not implied by any of our bounds.

As suggested by Table 2 and Figures 5.1–6.2, no single result uniformly dominates all other results in ensuring κ monotonicity. Different conditions are helpful in different parts of the (α, β) parameter space.

Table 2: Summary of κ monotonicity tests for five time series. The last four columns report the range of aggregation levels $1 \leq n \leq n_{\max}$ for which the indicated results guarantee monotonicity. Monotonicity is automatic beyond $n_{\max}$.

Market	$(\hat{\alpha}_1, \hat{\beta}_1)$	κ_1	$n_{\max}$	Thm 5.1	Unimodal bound Thm 5.3	Ratio bounds Prop. 6.2	s-cx bounds Thm 6.2
SPX	$(0.183, 0.785)$	3.9	7	all	all	all	all
EUR/USD	$(0.035, 0.962)$	7.0	67	$n \geq 39$	$n \geq 41$	all	all
JPY/USD	$(0.058, 0.936)$	5.8	32	$n \geq 12$	$n \geq 10$	all	all
BRL/USD	$(0.089, 0.898)$	5.1	17	$n \geq 3$	$n \geq 3$	all	all
WTI	$(0.116, 0.874)$	3.7	23	all	all	$n \geq 2$	all

9 Concluding Remarks.

The stationary distribution of a GARCH(1,1) process is heavy-tailed, even when the innovations driving the GARCH process are light-tailed. We have studied how the exponent κ in the power-law decay of a GARCH(1,1) process changes under a transformation of the model parameters α and β that results from temporal aggregation.

The parameter transformation we study arises in fitting GARCH models to the same time series at multiple frequencies. We showed that this parameter transformation provides a property we call forecast consistency, meaning that low-frequency and high-frequency models produce the same forecasts when both are limited to low-frequency historical data.

Our investigation of κ was motivated by a second objective we call tail consistency, requiring that models fit at different frequencies exhibit the same tail behavior in their stationary distributions. We established several results showing that κ increases under temporal aggregation of parameters, meaning that heavier-tailed innovations are required at lower frequencies to acheive tail consistency. Surprisingly, we have also proved exceptions to this pattern for several classes of innovation distributions in the region where α is close to zero and β is close to one.

These results provide guidance on the usually informal practice of fitting models at multiple frequencies. Forecast consistency argues for using the Drost and Nijman (1993) parameter transformation, rather than estimating parameters separately. Tail consistency usually argues for using heavier-tailed innovations with lower-frequency models, except for β very close to 1.

Extending these results to GARCH(p, q) models remains an open problem. The counterpart of κ for GARCH(p, q) models can be found in Basrak et al. (2002), and the temporal aggregation of parameters for these models was already derived in Drost and Nijman (1993). The challenge thus lies in understanding how the exponent in Basrak et al. (2002) changes under temporal aggregation of parameters and different innovation distributions.

A Proofs for Section 4

Proof. (Proposition 4.1.) It suffices to show that $2 < \nu_1 < \nu_2$ implies $Z^2_{\nu_2} \leq_{cx} Z^2_{\nu_1}$. Let h_ν denote the density of Z^2_ν and H_ν its cumulative distribution. By Theorem 3.A.44 of Shaked and Shanthikumar (2007), to show $Z^2_{\nu_2} \leq_{cx} Z^2_{\nu_1}$, it suffices to show that $h_{\nu_1} - h_{\nu_2}$ changes signs exactly twice on $[0, \infty)$, with sign sequence $+, -, +$.

Since $Z^2_{\nu_1}$ and $Z^2_{\nu_2}$ have the same mean, their densities must cross at least twice; this follows from combining Theorems 1.A.8 and 1.A.12 in Shaked and Shanthikumar (2007). The densities are given by $h_\nu(x) = \frac{C_\nu}{\sqrt{x}}(1 + \frac{x}{\nu-2})^{-\frac{\nu+1}{2}}$ with $C_\nu > 0$ a normalization constant. The equation $h_{\nu_1}(x) = h_{\nu_2}(x)$ thus takes the form

$$1 + \frac{x}{\nu_1 - 2} = c\left(1 + \frac{x}{\nu_2 - 2}\right)^{\frac{\nu_2+1}{\nu_1+1}},$$

for some constant $c > 0$. The left side is linear and the right side is strictly convex, so this equation can have at most two solutions in $[0, \infty)$. Thus $h_{\nu_1} - h_{\nu_2}$ changes signs exactly twice on $[0, \infty)$. We clearly have $h_{\nu_1}(x) > h_{\nu_2}(x)$ for all sufficiently large x, so the difference must have sign sequence $+, -, +$. □

B Proofs for Section 5

Proof. (Theorem 5.1.) We prove (i) last. For any $0 \le r \le s \le t$ in the domain of φ, log-convexity yields

$$\varphi(s)^{t-r} \le \varphi(r)^{t-s}\varphi(t)^{s-r}. \tag{43}$$

In case (ii), applying (43) with $r = 1$, $s = \kappa/2 - 1$, and $t = \kappa/2$, and recalling that $\varphi(1) = \alpha + \beta$ and $\varphi(\kappa/2) = 1$, yields

$$\varphi\left(\frac{\kappa}{2} - 1\right)^{(\frac{\kappa}{2}-1)} \le (\alpha + \beta)$$

and thus the upper bound in (24). The convexity of the mapping $x \mapsto x^{\kappa/2-1}$ yields the lower bound in (24) through Jensen's inequality. Monotonicity holds if the lower bound exceeds $\delta(\alpha, \beta)$, and this condition simplifies to the upper bound on κ in (ii).

For case (iii), applying (43) with $r = \kappa/2 - 1$, $s = 1$, and $t = \kappa/2$ yields

$$(\alpha + \beta) \le \varphi\left(\frac{\kappa}{2} - 1\right)^{(\frac{\kappa}{2}-1)}$$

and thus the lower bound in (25). The mapping $x \mapsto x^{\kappa/2-1}$ is concave, so Jensen's inequality yields the upper bound in (25). Monotonicity holds if the lower bound in (25) exceeds $\delta(\alpha, \beta)$, and this condition simplifies to the lower bound on κ in (iii). The trivial bound $\alpha Z^2 + \beta \ge \beta$ implies $\varphi(s) \ge \beta^s$, so κ monotonicity holds if $\beta^{\kappa/2-1} \ge \delta(\alpha, \beta)$, which yields the upper bound on κ in (iii). If (27) holds, then the lower bound exceeds the upper bound, so every κ in $(2, 4]$ satisfies at least one of the two sufficient conditions for monotonicity, and monotonicity is guaranteed.

To prove (i), suppose $\delta(\alpha, \beta) > 0$. Let Z^2 be supported on points z_1, z_2, $0 < z_1 < 1 < z_2$, with probabilities $(z_2 - 1)/(z_2 - z_1)$ and $(1 - z_1)/(z_2 - z_1)$ so that $\mathbb{E}[Z^2] = 1$. Choose z_2 large enough that $\alpha z_2 + \beta > 1/\delta(\alpha, \beta)$. By definition, $\kappa > 0$ solves

$$\frac{z_2 - 1}{z_2 - z_1}(\alpha z_1 + \beta)^{\kappa/2} + \frac{1 - z_1}{z_2 - z_1}(\alpha z_2 + \beta)^{\kappa/2} = 1. \tag{44}$$

Consider a sequence of distributions with z_2 fixed and z_1 approaching 1 from below. If κ were bounded, then as $z_1 \uparrow 1$, the second term in (44) would vanish and the first term would remain bounded away from 1; hence, $\limsup \kappa = \infty$ as $z_1 \uparrow 1$. As in (44),

$$\varphi\left(\frac{\kappa}{2} - 1\right) = \frac{z_2 - 1}{z_2 - z_1}(\alpha z_1 + \beta)^{\kappa/2-1} + \frac{1 - z_1}{z_2 - z_1}(\alpha z_2 + \beta)^{\kappa/2-1}.$$

For the last term, (44) and our choice of z_2 yield

$$\frac{1 - z_1}{z_2 - z_1}(\alpha z_2 + \beta)^{\kappa/2-1} \le \frac{1}{\alpha z_2 + \beta} < \delta(\alpha, \beta),$$

so

$$\varphi\left(\frac{\kappa}{2} - 1\right) < (\alpha z_1 + \beta)^{\kappa/2-1} + \delta(\alpha, \beta).$$

Taking any subsequence through which $z_1 \uparrow 1$ and $\kappa \to \infty$, we eventually have $\varphi(\kappa/2 - 1) < \delta(\alpha, \beta)$, where κ monotonicity fails. □

B.1 Extremal Case.

Theorem 5.2 will follow from two lemmas.

Lemma B.1. *For $s > 1$, the function $x \mapsto \varphi_{B_x}(s)$, $x \geq 1$, is continuous and strictly increasing, mapping $(1, \infty)$ onto $((\alpha + \beta)^s, \infty)$. For $s \in (0, 1)$, the function is continuous and strictly decreasing, mapping $(1, \infty)$ onto $(\beta^s, (\alpha + \beta)^s)$.*

Proof. Continuity is evident from (28). Letting $g_s(x) = (\alpha x + \beta)^s$, we have $\varphi_{B_x}(s) = g_s(0) + \frac{1}{x}[g_s(x) - g_s(0)]$. For $s > 1$, g_s is strictly convex so $\frac{1}{x}[g_s(x) - g_s(0)]$ (hence $\varphi_{B_x}(s)$) is strictly increasing in x. For $s < 1$, g_s is strictly concave so $\varphi_{B_x}(s)$ is strictly decreasing in x. The range of $\varphi_{B_x}(s)$ follows by evaluating its limits as $x \to \{1, \infty\}$. □

Lemma B.2. *Let Z be any admissible innovation with associated generating function φ. For any $s \in (1, 2)$, let $x_1 > 1$ be the unique solution to $\varphi_{B_{x_1}}(s - 1) = \varphi(s - 1)$. Then $\varphi_{B_{x_1}}(s) \leq \varphi(s)$.*

Proof. For any admissible innovation, $\beta^{s-1} < \varphi(s-1) < (\alpha+\beta)^{s-1}$, so the existence and uniqueness of x_1 follow from Lemma B.1. The last claim in the lemma follows from results of Karlin and Studden (1966). The power functions (t^{s-1}, t) and (t^{s-1}, t, t^s) form Tchebycheff systems on $[\beta, \infty)$, so $\mathbb{E}[(\alpha Z^2 + \beta)^s]$ is minimized over all distributions for Z^2 with fixed values of $\mathbb{E}[\alpha Z^2 + \beta]$ and $\mathbb{E}[(\alpha Z^2 + \beta)^{s-1}]$ by the *lower principal representation*, which is precisely the distribution of $\alpha B_{x_1} + \beta$; this is a special case of Theorem 5.1 of Karlin and Studden (1966). The theorem requires that the point $(\mathbb{E}[\alpha Z^2 + \beta], \mathbb{E}[(\alpha Z^2 + \beta)^{s-1}])$ be in the interior of the corresponding moment space. This condition holds because if Z is admissible then Z^2 is nondegenerate. □

Proof. (Theorem 5.2.) With $\kappa \in (2, 4)$ satisfying $\varphi(\kappa/2) = 1$, it follows from Lemma B.1 that there is just one x_2 at which $\varphi_{B_{x_2}}(\kappa/2) = 1$. Similarly, let x_1 be the parameter for which $\varphi_{B_{x_1}}(\kappa/2-1) = \varphi(\kappa/2-1)$. Lemma B.2 implies that $\varphi_{B_{x_1}}(\kappa/2) \leq \varphi(\kappa/2) = \varphi_{B_{x_2}}(\kappa/2)$ and thus $x_1 \leq x_2$. But then

$$\varphi(\kappa/2 - 1) = \varphi_{B_{x_1}}(\kappa/2 - 1) \geq \varphi_{B_{x_2}}(\kappa/2 - 1),$$

so if κ monotonicity holds with squared innovations B_{x_2}, it holds for all squared innovations with the same κ. □

Proof. (Corollary 5.1) For any $x > 1$ and $s \geq 0$, we have

$$(\alpha x + \beta)\varphi_{B_x}(s - 1) = \alpha\beta^{s-1}(x - 1) + \varphi_{B_x}(s).$$

Taking $s = \kappa/2$ and $x = x_2$, where, as in the proof of Theorem 5.2, $x_2 > 1$ solves $\varphi_{B_{x_2}}(\kappa/2) = 1$, yields

$$(\alpha x_2 + \beta)\varphi_{B_{x_2}}(\kappa/2 - 1) = \alpha\beta^{\kappa/2-1}(x_2 - 1) + 1.$$

Algebraic simplification now shows that the condition $\varphi_{B_{x_2}}(\kappa/2 - 1) \geq \delta(\alpha, \beta)$ is equivalent to $x_2 \leq x^*$. As $\varphi_{B_x}(\kappa/2)$ is increasing in x, the condition $x_2 \leq x^*$ is equivalent to $\varphi_{B_{x^*}}(\kappa/2) \geq 1$. The last claim in the corollary follows from (26). □

B.2 Unimodal Innovations.

Proof. (Theorem 5.3.) By Theorem V.9, p.158, in Feller (1971), any unimodal random variable Z with mode zero can be represented as $Z = VX$, with V uniform on $[0, 1]$ and X independent of V. As $\mathbb{E}[Z^2] = 1$ and $\mathbb{E}[V^2] = 1/3$, we must have $\mathbb{E}[X^2] = 3$. For $s \geq 1$, Jensen's inequality for conditional expectations yields

$$\mathbb{E}[(\alpha V^2 X^2 + \beta)^s | V] \geq (3\alpha V^2 + \beta)^s.$$

But $3V^2$ has the same distribution as U^2, so taking the expectation of both sides we get $\varphi(s) \geq \varphi_U(s)$, $s \geq 1$.

At $s = 2$, by evaluating $\mathbb{E}[U^4] = 9/5$ and $\mathbb{E}[U^2] = 1$, we get

$$\varphi_U(2) = 9\alpha^2/5 + 2\alpha\beta + \beta^2 \leq 1,$$

and therefore $\kappa_U/2 \geq 2$. We then have $\varphi(\kappa_U/2) \geq \varphi_U(\kappa_U/2) = 1$, so the strictly positive root of $\varphi(\kappa/2) = 1$ satisfies $\kappa \leq \kappa_U$. The condition $\kappa \geq 4$ then implies that $1 \leq \kappa/2 - 1 \leq \kappa_U/2 - 1$, so

$$\varphi(\kappa/2 - 1) \geq \min_{s \in [1, \kappa_U/2-1]} \varphi_U(s) \geq \delta(a, b),$$

in light of (29), and κ monotonicity follows. □

C Proof for Section 6

Proof. (Theorem 6.1.) If $Z^2 \in \mathcal{B}_s(\mu_1, \mu_2)$, then $X_3 \preceq_{3-cx} Z^2$, and $\mathbb{E}[g(X_s)] \leq \mathbb{E}[g(Z^2)]$, for all 3-convex g. With $g(x) = (\alpha x + \beta)^u$, $x \geq 0$, we have $g'''(x) = \alpha^3 u(u-1)(u-2)(\alpha x + \beta)^{u-3}$. This is non-negative for all $u \in [0, 1]$ and $u \in [2, \infty)$, which implies that g is 3-convex for this range of u, and $-g$ is 3-convex for $u \in [1, 2]$. Thus,

$$\phi_3(u) \leq \varphi(u),\ u \in [0,1] \cup [2, \infty) \quad \text{and} \quad \phi_3(u) \geq \varphi(u),\ u \in [1,2]. \tag{45}$$

Under the condition in (30) we have $\varphi(\kappa/2 - 1) \geq \delta(\alpha, \beta)$.

Similarly, the fourth derivative $g^{(4)}$ is non-negative for all $u \in [1, 2]$ and $u \geq 3$, which implies that g is 4-convex for this range of u. If $Z^2 \in \mathcal{B}_s(\mu_1, \mu_2, \mu_3)$, then $X_4 \preceq_{4-cx} Z^2$, and

$$\phi_4(u) \leq \varphi(u),\ u \in [1,2] \cup [3, \infty) \quad \text{and} \quad \phi_4(u) \geq \varphi(u),\ u \in [0,1] \cup [2,3]. \tag{46}$$

Under the condition in (31) we have $\varphi(\kappa/2 - 1) \geq \delta(\alpha, \beta)$. □

D Proof of Theorem 7.1

We evaluate the limit of $\varphi(s) = \mathbb{E}[(\alpha Z^2 + \beta)^s]$ as $s \to \infty$ and $\alpha \to 0$ with $s\alpha$ approaching a constant. Let $\bar{\theta}$ be as in (36).

Lemma D.1. *For $c/2 \in [0, \bar{\theta})$, suppose $s\alpha \to c/2$, then*

$$\lim_{\beta \to 1, \alpha = r(1-\beta)} \varphi(s) = e^{-\frac{c}{2r}} \mathbb{E}[e^{\frac{1}{2}cZ^2}] = g_Z(c).$$

Moreover, $\lim_{\beta \to 1} \alpha\kappa = a(r)$, with $a(r)$ as in Theorem 7.1.

Proof. With $\alpha = r(1-\beta)$, write $(\alpha Z^2 + \beta)^s$ as

$$(r(1-\beta)Z^2 + \beta)^s = \beta^s \left(1 + \frac{r(1-\beta)}{\beta} Z^2\right)^s.$$

With $\beta \to 1$ and $sr(1-\beta) \to c/2$, we have

$$\beta^s \to e^{-\frac{c}{2r}} \quad \text{and} \quad \left(1 + \frac{r(1-\beta)}{\beta} Z^2\right)^s \to e^{\frac{1}{2}cZ^2}. \tag{47}$$

For any $c'/2 \in (c/2, \bar{\theta})$ and all β sufficiently close to 1,

$$\left(1 + \frac{r(1-\beta)}{\beta} Z^2\right)^s \leq e^{c'Z^2/2},$$

with $\mathbb{E}[e^{\frac{1}{2}c'Z^2}] < \infty$, so by the dominated convergence theorem, we may interchange the limit and expectation as $\beta \to 1$ to get

$$\lim_{\beta\to 1} \varphi(s) = \mathbb{E}[\lim_{\beta\to 1}(\alpha Z^2 + \beta)^s] = e^{-\frac{c}{2r}}\mathbb{E}[e^{\frac{1}{2}cZ^2}] = g_Z(c).$$

The convexity of g_Z combined with $g_Z(0) = 1$, $g'_Z(0) = \mathbb{E}[Z^2 - 1/r]/2 < 0$, and (36) implies the existence and uniqueness of $a(r) > 0$ with $g_Z(a(r)) = 1$ and implies that g_Z is increasing at $a(r)$. For $\epsilon > 0$,

$$\varphi\left(\frac{a(r) \pm \epsilon}{2\alpha}\right) \to g_Z(a(r) \pm \epsilon).$$

As $g'_Z(a(r)) > 0$, for sufficiently small ϵ, $g_Z(a(r) - \epsilon) < 1 < g_Z(a(r) + \epsilon)$. It follows that $(a(r) - \epsilon)/2\alpha \leq \kappa/2 \leq (a(r) + \epsilon)/2\alpha$, for all β sufficiently close to 1, and thus that $\alpha\kappa \to a(r)$. □

Proof. (Theorem 7.1.) Because $\alpha(\kappa - 2) \to a(r)$, Lemma D.1 yields

$$\lim_{\beta\to 1} \varphi(\kappa/2 - 1) = 1. \tag{48}$$

We have

$$\begin{aligned}
\Delta_\beta = \frac{1 - \varphi(\kappa/2 - 1)}{1 - \beta} &= \frac{\mathbb{E}[(\alpha Z^2 + \beta)^{\kappa/2-1}(\alpha Z^2 + \beta - 1)]}{1 - \beta} \\
&= r\mathbb{E}[Z^2(\alpha Z^2 + \beta)^{\kappa/2-1}] - \varphi(\kappa/2 - 1)\,. \qquad (49)
\end{aligned}$$

As in (47), if $s\alpha \to c/2 \in [0, \bar{\theta})$, then

$$z^2(\alpha z^2 + \beta)^s \to z^2 e^{-\frac{c}{2r}} e^{\frac{1}{2}cz^2}.$$

Moreover, for some constant $C > 0$ and $c'/2 \in (c/2, \bar{\theta})$, we have, for all $z^2 \geq 0$,

$$z^2(\alpha z^2 + \beta)^s \leq Ce^{-\frac{c'}{2r}} e^{\frac{1}{2}c'z^2}.$$

We may therefore interchange limit and expectation to get

$$\mathbb{E}[Z^2(\alpha Z^2 + \beta)^s] \to \mathbb{E}[Z^2 e^{-\frac{c}{2r}} e^{\frac{1}{2}cZ^2}] = 2g'_Z(c) + g_Z(c)/r, \tag{50}$$

where the last expression follows from differentiating g_Z to get

$$g'_Z(c) = \frac{1}{2}e^{-\frac{c}{2r}}\mathbb{E}[(Z^2 - 1/r)e^{\frac{1}{2}cZ^2}].$$

For $s = \kappa/2 - 1$, $c = a(r)$, and the limit in (50) becomes $2g'_Z(a(r)) + 1/r$. Taking limits in (49) and recalling (48) yields

$$\lim_{\beta\to 1} \Delta_\beta = r(2g'_Z(a(r)) + 1/r) - 1 = 2rg'_Z(a(r)).$$

□

References

M. Abramowitz and I.A. Stegun. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. US Government Printing Office, Washington, 1972. 10th printing.

Carol Alexander and Emese Lazar. The continuous limit of weak garch. *Econometric Reviews*, 40(2): 197–216, 2021.

B. Basrak, R.A. Davis, and T. Mikosch. Regular variation of GARCH processes. *Stochastic Processes and Their Applications*, 99:95–116, 2002.

F. Bellini, F. Pellerey, C. Sgarra, and S.Y. Sekeh. Comparison results for GARCH processes. *Journal of Applied Probability*, 51:685–698, 2014.

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

T. Bollerslev. On the correlation structure for the generalized autoregressive conditional heteroscedastic process. *Journal of Time Series Analysis*, 9:121–131, 1988.

A. Bucher, P.N. Posch, and P. Schmidtke. Using the extremal index for value-at-risk backtesting. *Journal of Financial Econometrics*, 18(3):556–584, 2020.

A.C. Davison, S.A. Padoan, and G. Stupfler. Tail risk inference via expectiles in heavy-tailed time series. *Journal of Business & Economic Statistics*, 41(3):876–889, 2023.

M. Denuit, C. Lefévre, and M. Shaked. The s-convex orders among real random variables, with applications. *Mathematical Inequalities and Applications*, 1:585–613, 1998.

M. Denuit, C. Lefévre, and M. Shaked. On s-convex approximations. *Advances in Applied Probability*, 32:994–1010, 2000.

F.C Drost and T.E. Nijman. Temporal aggregation of GARCH processes. *Econometrica*, 61:909–927, 1993.

R. Engle. *ARCH models*. Oxford University Press, Oxford, 1995.

W. Feller. *An introduction to probability theory and its applications*. Wiley, New York, 1971. vol. II, 2nd Edition.

Christian Francq and Jean-Michel Zakoian. Estimating weak garch representations. *Econometric Theory*, 16:692–728, 2000.

Christian Francq and Jean-Michel Zakoian. *GARCH Models: Structure, Statistical Inference and Financial Applications*. John Wiley & Sons, 2019.

P. Glasserman and Q. Wu. Persistence and procyclicality in margin requirements. *Management Science*, 64:5705–5724, 2018.

J.D. Hamilton. *Time Series Analysis*. Princeton University Press, Princeton, 1994.

S. Karlin and W. Studden. *Tchebycheff systems: With applications to analysis and statistics*. Interscience Publishers, New York, 1966.

N. Mamingi. Beauty and ugliness of aggregation over time: A survey. *Review of Economics*, 68(3): 205–227, 2017.

A.J. McNeil, R. Frey, and P. Embrechts. *Quantitative Risk Management: Concepts, techniques and tools*. Princeton University Press, Princeton, 2015.

T. Mikosch and C. Stărică. Limit theory for the sample autocorrelations and extremes of a GARCH(1,1) process. *Annals of Statistics*, 28:1427–1451, 2000.

D.B. Nelson. Stationarity and persistence in GARCH(1,1) models. *Econometric Theory*, 6:318–334, 1990.

J.I. Peña, R. Rodriguez, and S. Mayoral. Tail risk of electricity futures. *Energy Economics*, 91:104886, 2020.

M. Shaked and J.G. Shanthikumar. *Stochastic Orders*. Springer, New York, 2007.

B. Su and F. Zhu. Temporal aggregation and systematic sampling for INGARCH processes. *Journal of Statistical Planning and Inference*, 219:120–133, 2022.

P. Sun and C. Zhou. Diagnosing the distribution of GARCH innovations. *Journal of Empirical Finance*, 29:287–303, 2014.

P. Teles. Testing conditional heteroscedasticity with systematic sampling of time series. *Communications in Statistics - Theory and Methods*, 52(15):5427–5450, 2023.